Large-Scale Coupled-Cluster Calculations

Dissertation zur Erlangung des Grades
„Doktor der Naturwissenschaften"
im Promotionsfach Chemie

am Fachbereich Chemie, Pharmazie und Geowissenschaften
der Johannes Gutenberg-Universität in Mainz

von

Michael Harding
geboren in Wiesbaden

Mainz, 2008

Contents

1 Introduction

Modern quantum chemistry plays a more and more decisive role in chemical research. Advances in theoretical methods as well as in computational resources extend the range of applicability continuously. Nonrelativistic quantum-chemical calculations are based on the Schrödinger equation which represents the fundamental equation for the quantum-mechanical description of atomic and molecular systems. Due to the fact there is no analytic solution to this equation for more than two particles, the main objective of quantum chemistry is to find approximate numerical solutions to this problem. These approximations have to be classified in terms of accuracy and computational effort. The corresponding applicability of quantum-chemical methods heavily depends on the required accuracy and the extent of the molecular system. Small molecules can be described by very accurate but computationally rather demanding methods, while for extended systems more pragmatic methods need to be applied.

Coupled-cluster theory [1–3] has turned out to be very accurate and reliable at the expense of being computationally rather costly. Within this framework the coupled-cluster singles and doubles model augmented by a perturbative correction for triple excitations (CCSD(T)) [4] has become the standard for accurate calculations. This method for example allows the determination of relative energies within chemical accuracy (ca. 4 kJ/mol). Computational resources limit the range of applicability for CCSD(T), so that cheaper and less accurate approximations like second-order Møller-Plesset perturbation theory (MP2) [5], Hartree-Fock (HF) [6], or density-functional theory (DFT) [7, 8] have to be used for larger molecular systems. Extending the range of applicability of highly accurate methods like CCSD(T) presents one of the challenges in quantum chemistry. Unfortunately the CCSD(T) method shows a steep operation count scaling of N^7, where N is a measure of the system size.[1] This means that doubling the size of the system would increase the overall execution time by more than two orders of magnitude. The advances in computer processor development cannot combat this steep scaling.

Figure 1.1 shows for the case of Intel processors that the number of transistors in a state-of-the-art integrated circuit (e.g., a computer processor) is roughly doubling every 24 months [9] for roughly the last four decades. Assuming that the computational performance is increasing, at best, linear[2] with the number of transistors leads to the conclusion that we would expect that it would be 14 years before the dimer of some particular molecule could be calculated in the same time that is required for the monomer today. In addition to that, limitations of other computational resources such as size and performance of fast memory, and disk space have to be taken into account.

[1] Assuming that the number of basis functions per atom is fixed.

[2] For the sake of simplicity other factors such as clock rate, cache sizes and how many add/multiply operations could be done within a clock cycle are not discussed.

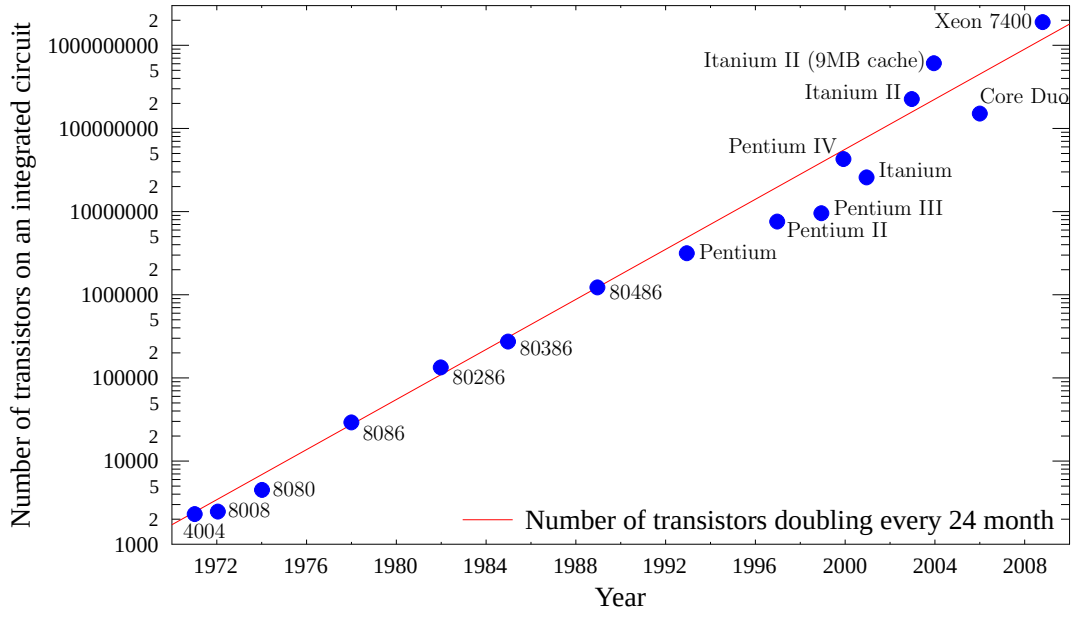

Figure 1.1: Logarithmic plot of the increase in the number of transistors within the last decades for Intel processors.

However, in principle there are three different ways out of this dilemma:

- Increase the efficiency of calculations with respect to processor usage, memory, and disk space requirements by improving the underlying mathematical structure, optimization of the computer code in terms of efficiency or use of additional physical information, e.g., the use of molecular point-group symmetry.
- Introduction of further approximations with more (e.g., Cholesky decomposition [10]) or less (e.g., local approximations [11]) controlled numerical error.
- Parallel use of several processors or computers to reduce the overall execution time.

Starting from the highly optimized program system ACES II MAB (Advanced Concepts in Electronic Structure II Mainz Austin Budapest version) [12,13] which is already optimized in terms of processor efficiency and the use of molecular point-group symmetry this work deals with the parallelization of highly accurate quantum-chemical methods and their application without introducing further approximations.
The increase of parallel computational performance by the 500 most powerful computer systems[1] ranked by their performance on the LINPACK benchmark[2] (fig. 1.2) shows approximately the same development with time for the fastest (#1), the 'slowest' (#500) and

[1] TOP500, see http://www.top500.org

[2] Performance of Various Computers Using Standard Linear Equations Software, Jack Dongarra, University of Tennessee, Knoxville TN, 37996, Computer Science Technical Report Number CS - 89 85, November 13, 2008, http://www.netlib.org/benchmark/performance.ps.

the aggregate performance (sum) which is superior to the before mentioned development of single-processor computers.

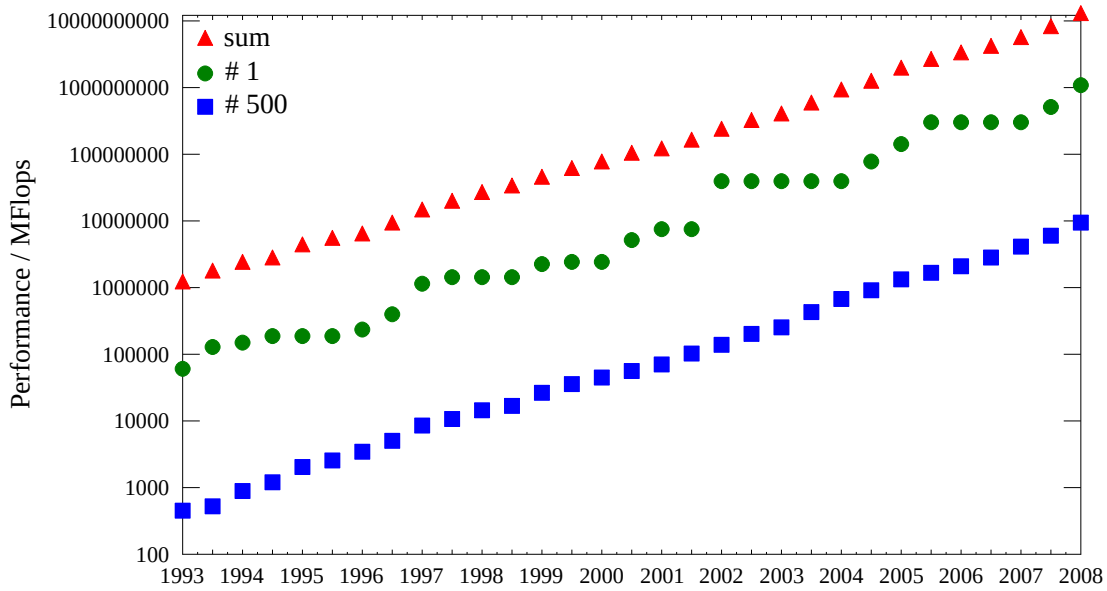

Figure 1.2: Performance of all (sum), the fastest (#1) and the 'slowest' (#500) from the list of the 500 most powerful computer systems (TOP500, see http://www.top500.org) ranked by their performance on the LINPACK benchmark.

The higher computational power can be explained with the trend of clustering larger and larger numbers of computers and the propagation of multi-core techniques, which integrate several cores (processors) in one processor enclosure.[1] To take advantage of these multi-core or multiprocessor architectures it is also necessary to execute calculations in a parallel manner. Turning back to the monomer/dimer execution time example with the current advances in computer-cluster systems one would expect the same execution time for a dimer in about 7 years. Although even that does not overcome the steep computational scaling of coupled-cluster models the advances in cluster technology together with parallel program execution vastly extends the range of applicability and is the key for numerous chemical applications in near future.

In the following chapter the underlying program system ACES II MAB which was chosen as a starting point for the parallelization and used for nearly all calculations throughout this work is briefly described. Chapter 3 deals with the design and development of the local computational resources in the last five years. Chapter 4 gives a brief sketch of the quantum-chemical methods employed for the calculation of energies and molecular proper-

[1] Lately, the doubling of the number of transistors could only be achieved by increasing the number of processor cores. The Core Duo is a dual core processor, while for example the Xeon 7400 series has 6 cores.

ties. Strategy, implementation and benchmarking of the parallelization of energy, gradients and second derivatives for coupled-cluster methods are discussed in chapter 5. The scheme employed is presented in a stepwise manner leading to an algorithm with parallelized routines for the rate-determining steps. A detailed investigation of the reduction of the overall execution time of the serial and the parallel algorithm will be carried out.
The following chapters will present typical applications for parallel highly accurate coupled-cluster calculations in the context of computational thermochemistry (chapter 6), the prediction of nuclear magnetic shielding constants (chapter 7), equilibrium geometries, and rotational spectroscopy (chapter 8).

Chapter 6 gives insight in the 'High accuracy Extrapolated Ab initio Thermochemistry' (HEAT) [14,15] protocols. Effects of increased basis-set size, higher excitations in the coupled-cluster expansion are investigated. The last two sections will show applications extending the HEAT scheme to molecules containing second-row atoms and more extended systems. In the case of the heat of formation of vinyl chloride discrepancies among experimental data of up to 17 kJ/mol will be resolved applying an extended HEAT scheme. The calculation of accurate thermochemical properties for the benzene molecule shows the limitations due to computational resources. Necessary approximations will be analyzed carefully using the available data for smaller molecules.

Quantum-chemical calculations of magnetic shielding constants have proven to be useful for the assignment and interpretation of experimental NMR spectra. In chapter 7 the findings of benchmark calculations of gas-phase ^{19}F nuclear magnetic shielding constants will be discussed. In addition, results of large-scale coupled-cluster calculations for benzene and the adamantyl cation will be presented.

In the second last chapter the determination of equilibrium geometries and other properties relevant to vibration-rotation spectroscopy is discussed. The first section raises the topic of the geometry of the hydrogen trioxy radical. The next section focuses on the determination of the empirical equilibrium geometry of the diacetylene molecule. This is followed by the application of the CCSD(T) method to the geometry and the hyperfine structure of cyanopolyynes. The last section presents large-scale calculations for the equilibrium structure of the ferrocene molecule, which vividly demonstrates the advantages associated with the developments of implementation presented in chapter 5.

Finally, chapter 9 provides a detailed summary and future prospects for the results presented in this thesis.

2 Mainz-Austin-Budapest version of the ACES II program system

ACES II (Advanced Concepts in Electronic Structure II) [12] is a program package for performing high-level quantum chemical calculations on atoms and molecules. The program suite has a rather large arsenal of high-level ab initio methods for the calculation of atomic and molecular properties. Virtually all approaches based on Møller-Plesset perturbation theory (MP) and the coupled-cluster (CC) approximation are available in ACES II. For most of these, analytic first and second-derivative approaches are available within the package.

The development of ACES II began in early 1990 in the group of Professor Rodney J. Bartlett at the Quantum Theory Project (QTP) of the University of Florida in Gainesville. During 1990 and 1991 John F. Stanton, Jürgen Gauss, and John D. Watts, supported by a few students, wrote the first version of what is now known as the ACES II program package. The integral packages (the MOLECULE package of J. Almlöf, the VPROP package of P.R. Taylor, and the integral derivative package ABACUS of T. Helgaker, P. Jørgensen J. Olsen, and H.J.Aa. Jensen) were the only parts taken from others and had been adapted for the use within ACES II.

ACES II had been originally developed for CRAY supercomputers (under the Unix-based operating system UNICOS) and, consequently, a lot of effort had been devoted to the exploitation of matrix-vector operations through optimized BLAS (Basic Linear Algebra Subprograms)-routines. However, more or less simultaneously, versions for so-called "Unix-boxes" were created. The design strategy of keeping machine-dependent code to an absolute minimum facilitated porting the program package to other computer architectures and operating systems.

A number of methodological developments have been added to the program in the last two decades: Analytic second derivatives for all coupled-cluster approaches up to full CCSDT, the calculation of nuclear magnetic resonance (NMR) chemical shifts at MP and CC levels of theory, the calculation of anharmonic force fields and vibrationally averaged properties (via numerical differentiation of analytic derivatives), relativistic corrections, corrections to the Born-Oppenheimer approximation at the CC level, nonadiabatic coupling within the equation-of-motion (EOM) framework, and several others. In addition, energies, first and second-derivative calculations for arbitrary excitation levels in configuration-interaction (CI) and CC methods are available by the tight integration of the string based many-body program MRCC of M. Kállay in ACES II MAB.

As the last merge between the original Florida version of ACES II and the version maintained in Austin and Karlsruhe and later in Mainz dates back roughly to 1995, it has been decided that both versions are now separately maintained and distributed.[1]

In 2005 the authors in Mainz, Austin and Budapest decided to provide access to the academic community. As the public release was one project within this thesis the experiences

[1] For the Florida version of ACES II, see http://www.qtp.ufl.edu/aces2,
for the Mainz-Austin-Budapest version of ACES II, see http://www.aces2.de.

will be described in the following. To distinguish between the Florida code and the local version of ACES II it was decided to rename the package to ACES II MAB (Advanced Concepts in Electronic Structure II Mainz Austin Budapest version).

First, the copyright question had to be addressed and all the contributors accepted that the package could be distributed centrally from Mainz. As a next step to keep copyright and control of the distribution process, a license agreement was created.

Realizing that the documentation of the package was outdated, a Wiki web site was created to improve on the documentation and allow all contributors to edit and extend the documentation in real time at a central place.

To obtain a license, future users only have to sign a license agreement and send it via regular mail to Mainz. After receiving the license agreement it is countersigned in Mainz and sent back together with the information to retrieve the source code.

At the beginning all support for the public release was done via email by the author of this thesis. Realizing after more than one year that questions repeat and workload increased a mailing list for the public version was established. After some time many responses came and come from the user community, while the developers actively participate in the process.

To date ACES II MAB has more than 200 licensees and is used on all continents but Africa and Antarctica. A new release including the developments of the past three years is planned for the end of 2008, while the name of the package will be changed to Cfour (Coupled-Cluster techniques for Computational Chemistry).[1]

The next chapter will discuss the design and development of the local computer environment optimized for the use of the ACES II MAB package and its recently developed parallel features (see chapter 4).

[1] http://www.cfour.de

3 A computer environment for computational chemistry

Design and development of the local computational resources in the theoretical-chemistry group at the Universität Mainz are reviewed for the last five years in the following paragraphs. First, the development was centered on the acquisition of many computers, principally to meet the demands of an increasing number of users. Later, parameters such as space, cooling and the available amount of electricity had to be taken into account. Last but not least, the required operational hours were proportional to the number of computers. New concepts like automatic configuration and installation had to be taken into account to manage local resources without additional manpower. At all stages, large coupled-cluster calculations have been carried out to benchmark promising systems with different architectures to attempt optimization of the price-performance ratio.

As the computational requirements of various theoretical models in computational chemistry differ vastly the discussion will be restricted on the needs of coupled-cluster programs, especially the ACES II MAB program (see previous chapter). Because the resource requirements for fast memory and especially disk space and performance for coupled-cluster energy and derivative calculations usually exceed those of lower level treatments like HF, MP2, or DFT, the system described could be used for those, as well.

Before 2003 mainly Alpha processor based workstations with the True64 operating system were used. Around that time, Linux based personal computers (PCs) with the i386 architecture became useful for CC calculations. For the first time they provided comparable performance and 2 GB of fast memory was available and affordable. Together with further PCs three HP Itanium II systems with 4 GB of fast memory and fast disks were added for somewhat larger calculations. Some time later, the PCs evolved from 32 to 64 bit address space with the AMD Opteron and the AMD64 processors enabling the use of more than 2 GB in a single calculation. First 2 dual Opteron systems and later 12 AMD64 single processor systems were added. Up to that time all systems were configured, installed and administered separately. The continuous changes in network and other devices during that period necessitated the most up-to-date Linux distributions (SuSE was used here), a very heterogeneous environment was established.

First, on some PC workstations and then with the equal AMD64 systems and one of the Opteron based computers first experiments with cluster configurations were studied. The cluster management packages Warewulf,[1] Clusteromatic,[2] Rocks [16–18],[3] and Clusterknoppix[4] were evaluated, and Rocks was found to provide the most convenient solution because of its flexibility and robustness. This first cluster consisted of the AMD64 machines mentioned before. About one year later, it was decided to further extend the machine park by an additional rack mounted cluster system consisting of 16 Intel Xeon based compute nodes and one dual Xeon processor based master node.

[1] http://www.perceus.org
[2] http://www.clustermatic.org
[3] http://www.rocksclusters.org
[4] http://clusterknoppix.sw.be

This time the cluster was based on a diskless installation (with local scratch disks), where all compute nodes received the operating system via network boot. The main idea was to keep the compute nodes as equal as possible. Later an extension by compute nodes based on the next generation of Intel Xeon processors was planned and implemented in the same manner. A short time after installing the extension (around autumn 2006), the decision was made to rearrange all resources for more efficient use of space, cooling and electricity. In order to further reduce the operational effort it was decided to deploy the Rocks cluster distribution (based on CentOS Linux) and to merge all available machines to one cluster.

Currently this 'historically' grown cluster consists of about 100 machines and offers three different architectures (ia32, ia64, EM64T/AMD64). The system is accessible through one frontend and one queueing system.

After this more general discussion two further issues should be discussed, namely disk performance and the type of network interconnects. Because of the high input/output (I/O) load needed in coupled-cluster calculations one single hard drive is not enough to accomplish efficient calculations. All described systems have at least two striped disks, which can write and read data in parallel. When affordable multi-processor/multi-core architectures entered the market it turned out that 6 to 8 hard disks connected to a raid-controller device represent an acceptable choice in terms of I/O performance. In general, two striped disks per core seems to be a good compromise.

As will be shown in chapter 5 there was at no time need for the implementation of a dedicated fast communication network for parallel calculations. All network interconnects are Gigabit Ethernet. Experiments with two striped network devices showed some improvement but were abandoned because of technical complications. Particular attention has to be put on the choice of network switches. Experiences with cheap network switches showed their limited bandwidths in parallel calculations. The whole cluster is connected via HP Pro Curve 2448 switches which show no bandwidth problems with the current applications.

All calculations throughout this work have been performed with the described cluster system. The measured execution times in chapters 5, 7.3 and 8.4 have been obtained on the former mentioned 16 node single core Intel Xeon (EM64T) 3.4 GHz cluster. Each node is equipped with 16 GB fast memory, 6 striped 300 GB SATA disks, and two channel bonded Gigabit Ethernet interfaces.

4 Theoretical foundations

4.1 Quantum-chemical methods

Quantum chemistry is the application of quantum mechanics to the electronic structure problem of atoms and molecules via the Schrödinger equation:

$$\hat{H}\ \Psi \quad = \quad E\ \Psi. \tag{4.1}$$

This eigenvalue equation yields the wavefunction Ψ and the energy eigenvalue E for a given Hamiltonian $\hat{H}$. The molecular Hamiltonian includes all possible interactions between electrons and nuclei, and is given by [6] (in atomic units):

$$\begin{aligned} \hat{H} \quad = \quad & \underbrace{-\sum_i \frac{1}{2}\nabla_i^2}_{\hat{T}_\mathrm{e}} \underbrace{-\sum_A \frac{1}{2M_A}\nabla_A^2}_{\hat{T}_\mathrm{n}} \\ & \underbrace{-\sum_{i,A} \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|}}_{\hat{V}_\mathrm{ne}} \underbrace{+\sum_{i<j} \frac{1}{|\mathbf{r}_i - \mathbf{r}_j|}}_{\hat{V}_\mathrm{ee}} \underbrace{+\sum_{A<B} \frac{Z_A Z_B}{|\mathbf{R}_A - \mathbf{R}_B|}}_{\hat{V}_\mathrm{nn}} . \end{aligned} \tag{4.2}$$

The first two terms describe the kinetic energy of the electrons, $\hat{T}_\mathrm{e}$, and of the nuclei, $\hat{T}_\mathrm{n}$, with ∇_i^2 and ∇_A^2 denoting the Laplacians with respect to the coordinates $\mathbf{r}_i$ of the electrons, and $\mathbf{R}_A$ of the nuclei. Coulomb interaction between charged particles are included by electron-nucleus, electron-electron and nucleus-nucleus potentials ($\hat{V}_\mathrm{ne}$, $\hat{V}_\mathrm{ee}$, $\hat{V}_\mathrm{nn}$). Nuclear charge and mass are denoted as Z_A, Z_B and M_A.

In the Born-Oppenheimer approximation [19] the total wavefunction is given as a product of nuclear and electronic wavefunction Ψ_n and Ψ_e:

$$\Psi(\{\mathbf{r}\},\{\mathbf{R}\}) \quad = \quad \Psi_\mathrm{e}(\{\mathbf{r}\},\{\mathbf{R}\})\Psi_\mathrm{n}(\{\mathbf{R}\})\ . \tag{4.3}$$

The nuclear wavefunction depends on the coordinates of the nuclei, whereas the electronic wavefunction depends on the coordinates of the electrons and parametrically on the internuclear distances. Ψ_e is the solution of the electronic Schrödinger equation:

$$\begin{aligned} \hat{H}_\mathrm{e}\Psi_\mathrm{e}(\{\mathbf{r}\},\{\mathbf{R}\}) \quad &= \quad \left(\hat{T}_\mathrm{e} + \hat{V}_\mathrm{ne} + \hat{V}_\mathrm{ee}\right) \Psi_\mathrm{e}(\{\mathbf{r}\},\{\mathbf{R}\}) \qquad (4.4) \\ &= \quad E_\mathrm{e}(\{\mathbf{R}\})\Psi_\mathrm{e}(\{\mathbf{r}\},\{\mathbf{R}\})\ . \end{aligned}$$

The eigenvalues of the electronic Schrödinger equation are the electronic energies E_e, and their geometrical dependence describes individual adiabatic potential-energy surfaces, which is the central paradigm in descriptive chemical physics. The nuclear wavefunction Ψ_n is obtained as a solution of the nuclear Schrödinger equation, $\hat{H}_\mathrm{n}\Psi_\mathrm{n} = E_\mathrm{n}\Psi_\mathrm{n}$ with $\hat{H}_\mathrm{n} = (\hat{T}_n + \hat{V}_{nn} + E_e(\{\mathbf{R}\}))$, describing rotational, translational, and vibrational dynamics of the molecule as well as chemical reactions.

Due to the fact that an exact solution of the electronic Schrödinger equation is infeasible for systems with more than one electron a number of approximate methods were developed. All methods relevant to the present work are briefly sketched in the following paragraphs.

4.1.1 Hartree-Fock theory

Hartree-Fock theory is a simple approximate method for solving the electronic Schrödinger equation [6]. The wavefunction $\Psi_{\mathrm{e}}(1,\ldots,N)$ is a single Slater determinant of N orthonormal spin orbitals φ_i.

$$\Psi_{\mathrm{e}}(1,...,N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \varphi_1(1) & \varphi_1(2) & \cdots & \varphi_1(N) \\ \varphi_2(1) & \varphi_2(2) & \cdots & \varphi_2(N) \\ \vdots & \vdots & \ddots & \vdots \\ \varphi_N(1) & \varphi_N(2) & \cdots & \varphi_N(N) \end{vmatrix} . \tag{4.5}$$

The spin orbital $\varphi_i(j)$ is a function of both the spin σ_j and spatial coordinates $\mathbf{r}_j$ of the j-th electron, i.e.

$$\varphi_i(j) = |i\rangle = \varphi_i(\mathbf{r}_j, \sigma_j) \tag{4.6}$$

and often represented as $\phi_i(\mathbf{r}_j)\psi(\sigma_j)$, where ϕ_i and ψ denote spatial orbital and spin function, respectively.
The Hartree-Fock energy E_{HF} is given by:

$$E_{\mathrm{HF}} = \langle\Psi_{\mathrm{el}}|\hat{H}_{\mathrm{el}}|\Psi_{\mathrm{el}}\rangle = \sum_i \langle i|\hat{h}|i\rangle + \frac{1}{2}\sum_{ij}\langle ij||ij\rangle. \tag{4.7}$$

The contribution from the one-electron interaction $\hat{h}$ consists of the kinetic energy and the nucleus-electron interaction which are included by the integral over spin and spatial coordinates $\tau_j \mathrel{\hat{=}} (\mathbf{r}_j, \sigma_j)$.

$$\langle i|\hat{h}|i\rangle = \int d\tau_1 \varphi_i^*(1) \underbrace{\left(\frac{1}{2}\nabla_i^2 - \sum_A \frac{Z_A}{|\mathbf{r}_i - \mathbf{R}_A|}\right)}_{\hat{h}(i)} \varphi_i(1) = h_{ii} \ , \tag{4.8}$$

The antisymmetrized integrals $\langle ij||ij\rangle$ refer to Coulomb and exchange interaction,

$$\begin{aligned} \langle ij||ij\rangle = \langle ij|ij\rangle - \langle ij|ji\rangle &= \int d\tau_1 \varphi_i^*(1) \underbrace{\int d\tau_2 \varphi_j^*(2) \frac{1}{r_{12}} \varphi_j(2)\varphi_i(1)}_{\hat{J}_j(1)\varphi_i(1)} \\ &- \int d\tau_1 \varphi_i^*(1) \underbrace{\int d\tau_2 \varphi_j^*(2) \frac{1}{r_{12}} \varphi_i(2)\varphi_j(1)}_{\hat{K}_j(1)\varphi_i(1)} \end{aligned} \tag{4.9}$$

Here, $\hat{J}_j$ and $\hat{K}_j$ denote the Coulomb and exchange operators, respectively. The interaction between an electron and the mean field of the other electrons is described by the Coulomb contribution. The exchange contribution is a consequence of the indistinguishability of the electrons. Minimizing the energy in eq. (4.7) leads to the optimal molecular orbitals (MOs). As a side condition the molecular orbitals have to be orthonormal:

$$\int d\tau_1 \varphi_i^*(1)\varphi_j(1) = \delta_{ij} \ . \tag{4.10}$$

This minimization leads to the Hartree-Fock equations

$$\hat{F}\varphi_i = \left(\hat{h} + \sum_j \left(\hat{J}_j - \hat{K}_j\right)\right)\varphi_i = \epsilon_i \varphi_i \tag{4.11}$$

with ϵ_i denoting the orbital energies. The Fock operator $\hat{F}$ is an effective one-electron operator depending on the unknown orbitals φ_i and therefore the solution results in a pseudo-eigenvalue problem, which has to be solved iteratively. This procedure is called Hartree-Fock self-consistent-field (HF-SCF) method.
Roothaan [20] and Hall [21] demonstrated that the Hartree-Fock equations (4.11) can be converted into a convenient set of algebraic matrix equations by expanding the spatial orbitals into a basis of atomic orbitals (AO) χ_ν:

$$\phi_i \quad = \quad \sum_\nu c_{\nu i}\chi_\nu \ . \tag{4.12}$$

The expansion coefficients $c_{\nu i}$ can be determined again through the HF-SCF procedure or other schemes (see subsection 4.3). With the definition of the density matrix

$$D_{\mu\nu} \quad = \quad \sum_i 2c^*_{\mu i}c_{\nu i} \tag{4.13}$$

the Hartree-Fock energy for a closed-shell molecule can be written within *restricted*-Hartree-Fock theory [6] (RHF) as

$$E_{\mathrm{HF}} \quad = \quad \sum_{\mu\nu} D_{\mu\nu}h_{\mu\nu} + \frac{1}{2}\sum_{\mu\nu\rho\sigma} D_{\mu\nu}D_{\rho\sigma}\left(\langle\mu\rho|\nu\sigma\rangle - \frac{1}{2}\langle\mu\rho|\sigma\nu\rangle\right). \tag{4.14}$$

Insertion of the expansion (4.12) into eq. (4.11) followed by projection of the Hartree-Fock equation onto $\langle\chi_\mu|$ leads to the Roothaan-Hall equations [6]:

$$\sum_\nu \underbrace{\left(h_{\mu\nu} + \sum_{\rho\sigma} D_{\rho\sigma}\left(\langle\mu\rho|\nu\sigma\rangle - \frac{1}{2}\langle\mu\rho|\sigma\nu\rangle\right)\right)}_{F_{\mu\nu}} c_{\nu i} \quad = \quad \sum_\nu \underbrace{\langle\chi_\mu|\chi_\nu\rangle}_{S_{\mu\nu}} c_{\nu i}\epsilon_i \tag{4.15}$$

with the overlap matrix elements $S_{\mu\nu}$ and the Fock matrix elements $F_{\mu\nu}$. The expansion coefficients $c_{\nu i}$ and orbital energies ϵ_i are obtained as eigenvectors and eigenvalues at the end of the iterative solution of the Roothaan-Hall equations via diagonalization of the Fock matrix; when the coefficients used to form the Fock matrix coincide with those obtained from diagonalization, the procedure has converged.

The two-electron interaction in eq. (4.9) treats the exchange interaction correctly but Hartree-Fock theory neglects the Coulomb correlation completely. This arises from the fact that the motion of an electron described by this theory depends on the mean field of the other electrons only and not on the individual positions of the other electrons. The correlation energy E_{corr} is defined as the difference of the exact non-relativistic energy E_{exact} and the Hartree-Fock energy E_{HF}:

$$E_{\mathrm{corr}} \quad = \quad E_{\mathrm{exact}} - E_{\mathrm{HF}} \ . \tag{4.16}$$

4.1.2 Møller-Plesset perturbation theory

Perturbational approaches for the description of electron correlation are based on the separation of the electronic Hamiltonian into an unperturbed $\hat{H}_0$ and a perturbed $\hat{H}'$ part:

$$\hat{H} = \hat{H}_0 + \lambda \hat{H}' \; . \tag{4.17}$$

The energy E and the wavefunction Ψ are expanded in a Taylor series, where λ is the perturbation parameter. After insertion of the expansions into the Schrödinger equation the resulting terms are sorted by the order of the parameter λ. Projection onto the eigenfunctions of the unperturbed system leads to n-th order energy expressions [6]. Møller-Plesset (MP) perturbation theory [5] defines the unperturbed Hamiltonian as a sum of Fock operators:

$$\hat{H}_0 = \sum_j \hat{F}(j) \; . \tag{4.18}$$

The perturbed part of the Hamiltonian which describes electron correlation is obtained from the difference between the true molecular electronic Hamiltonian and $\hat{H}_0$:

$$\hat{H}' = \hat{H} - \hat{H}_0 = \sum_{i<j} \frac{1}{r_{ij}} - \sum_j \hat{F}(j) \; . \tag{4.19}$$

This results in the following expressions for the second order energy [6]:

$$E^{(2)} = \frac{1}{4} \sum_{ijab} \frac{|\langle ij||ab\rangle|^2}{\epsilon_i + \epsilon_j - \epsilon_a - \epsilon_b} \tag{4.20}$$

As is the usual convention, $i, j, k, ...$ denote occupied and $a, b, c, ...$ virtual orbitals. Within the framework of the Møller-Plesset ansatz the Hartree-Fock energy (4.7) is represented by the sum of zeroth and first order energy. Electron correlation is recovered by second (MP2) and higher-order perturbation theory.

4.1.3 Configuration-interaction

Solving the Roothaan-Hall equation (4.15) within a finite AO-basis consisting of M basis functions leads to a set of M spin orbitals. The Hartree-Fock wavefunction (4.5) is constructed by the N energetically lowest spin orbitals, where N denotes the number of electrons (Aufbau principle). The HF wavefunction can be taken as a reference determinant $|\Psi_0\rangle$ for the configuration-interaction method (CI). Replacing successively the N occupied spin orbitals φ_i, φ_j, ... by virtual φ_a, φ_b, ... spin orbitals in all possible combinations leads to a set of further N-electron Slater determinants (Ψ_i^a, Ψ_{ij}^{ab}, ...), which span a complete many-electron basis for a given AO basis [22, 23]. Therefore the exact non-relativistic electronic wavefunction can be written as a linear combination of all possible excited determinants (Ψ_i^a, Ψ_{ij}^{ab}, ...) and the HF determinant. If the variation principle [6] is applied to obtain the expansion coefficients $c_{ijk...}^{abc...}$ this procedure is called full configuration-interaction method (FCI). Excited Slater determinants $\Psi_{ijk...}^{abc...}$ can be written in terms of excitation operators. Within the framework of second quantization excitations are described by quasi-particle creation $\hat{a}_a^+$, $\hat{a}_i$ and annihilation operators $\hat{a}_i^+$, $\hat{a}_a$ [24]. The FCI wavefunction is given as

$$\begin{aligned} |\Psi\rangle &= |\Psi_0\rangle + \sum_{i,a} c_i^a |\Psi_i^a\rangle + \frac{1}{4} \sum_{ij,ab} c_{ij}^{ab} |\Psi_{ij}^{ab}\rangle + \dots \\ &= |\Psi_0\rangle + \sum_{i,a} c_i^a \hat{a}_a^+ \hat{a}_i |\Psi_0\rangle + \frac{1}{4} \sum_{ij,ab} c_{ij}^{ab} \hat{a}_a^+ \hat{a}_i \hat{a}_b^+ \hat{a}_j |\Psi_0\rangle + \dots \; . \end{aligned} \tag{4.21}$$

The enormous number of possible excited Slater determinants and unknown parameters c restricts the application of FCI to small systems employing small basis sets [25]. Despite these restrictions FCI calculations play an important role for benchmarking of more approximate methods. Approximate CI methods are based on truncated CI expansions, e.g., the CISD method (configuration-interaction singles doubles) which allows only for single and double excitations [22]. These calculations are not widely used in current applications because the resulting correlation energies are not size-extensive [6].

4.1.4 Coupled-cluster theory

The following ansatz is used for the wavefunction in coupled-cluster theory (CC) [1–3, 26, 27]:

$$|\Psi_{\mathrm{CC}}\rangle \;=\; \mathrm{e}^{\hat{T}}|\Psi_0\rangle \;. \tag{4.22}$$

Here $|\Psi_0\rangle$ denotes the reference determinant, usually given by the Hartree-Fock wavefunction. The cluster operator $\hat{T}$ is defined as

$$\hat{T} \;=\; \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \cdots + \hat{T}_N = \sum_{p}^{N} \hat{T}_p \tag{4.23}$$

and consists of the sum of all single ($\hat{T}_1$), double ($\hat{T}_2$) up to N-tuple excitations ($\hat{T}_N$) of electrons from occupied to unoccupied orbitals. Each excitation is weighted by an amplitude $t_{ijk..}^{abc..}$. This leads to

$$\hat{T}_N \;=\; \frac{1}{(N!)^2} \sum_{i,j,k..}^{N} \sum_{a,b,c..}^{N} t_{ijk..}^{abc..} \hat{a}_a^+ \hat{a}_i \hat{a}_b^+ \hat{a}_j \hat{a}_c^+ \hat{a}_k \ldots \;. \tag{4.24}$$

For the determination of the amplitudes the wavefunction (4.22) is inserted in the Schrödinger equation

$$\hat{H}\; \mathrm{e}^{\hat{T}}|\Psi_0\rangle \;=\; E\; \mathrm{e}^{\hat{T}}|\Psi_0\rangle \;. \tag{4.25}$$

The correlation energy is obtained by subtraction of the Hartree-Fock energy on both sides of the equation:

$$\underbrace{\Big(\hat{H} - \langle\Psi_0|\hat{H}|\Psi_0\rangle\Big)}_{\hat{H}_{\mathrm{N}}}\, \mathrm{e}^{\hat{T}}|\Psi_0\rangle \;=\; \underbrace{(E - E_{\mathrm{HF}})}_{E_{\mathrm{corr}}}\, \mathrm{e}^{\hat{T}}|\Psi_0\rangle \;. \tag{4.26}$$

Here, the normal ordered[1] Hamiltonian $\hat{H}_{\mathrm{N}}$ is introduced, which consists of the corresponding one ($\hat{f}_{\mathrm{N}}$) and two-electron contributions ($\hat{W}_{\mathrm{N}}$)

$$\hat{f}_{\mathrm{N}} \;=\; \sum_{pq} f_{pq} \{\hat{a}_p^+ \hat{a}_q\} \tag{4.27}$$

$$\hat{W}_{\mathrm{N}} \;=\; \frac{1}{4} \sum_{pqrs} \langle pq||rs\rangle \{\hat{a}_p^+ \hat{a}_q^+ \hat{a}_s \hat{a}_r\} \;. \tag{4.28}$$

[1] A normal-ordered product of annihilation and creation operators is one in which all annihilation operators lie to the right of all creation operators. This is a useful form because a normal-ordered product gives a zero result when applied to the vacuum state.

where f_{pq} represents the matrix elements of the Fock operator,

$$f_{pq} = \langle p|\hat{h}|q\rangle + \sum_i \langle pi||qi\rangle \ . \tag{4.29}$$

Due to its complexity and the resulting computational effort the coupled-cluster problem is normally not solved in a variational manner [3]. After multiplication from the left by $\mathrm{e}^{-\hat{T}}$ eq. (4.26) is projected onto the reference determinant as well as onto all excited determinants $\Psi_{ij\cdots}^{ab\cdots}$. This yields the CC energy and the amplitudes t via:

$$\begin{aligned}
\Delta E_{\mathrm{corr}} &= \langle\Psi_0|\mathrm{e}^{-\hat{T}}\hat{H}_{\mathrm{N}}\ \mathrm{e}^{\hat{T}}|\Psi_0\rangle && (4.30)\\
0 &= \langle\Psi_i^a|\mathrm{e}^{-\hat{T}}\hat{H}_{\mathrm{N}}\ \mathrm{e}^{\hat{T}}|\Psi_0\rangle && (4.31)\\
0 &= \langle\Psi_{ij}^{ab}|\mathrm{e}^{-\hat{T}}\hat{H}_{\mathrm{N}}\ \mathrm{e}^{\hat{T}}|\Psi_0\rangle && (4.32)\\
&\ \ \ldots\\
0 &= \langle\Psi_{ij\cdots}^{ab\cdots}|\mathrm{e}^{-\hat{T}}\hat{H}_{\mathrm{N}}\ \mathrm{e}^{\hat{T}}|\Psi_0\rangle. && (4.33)
\end{aligned}$$

The resulting coupled-cluster energies correspond to the FCI result because all possible excited determinants are included [28]. The expansion of the exponential function in eq. (4.22) $\mathrm{e}^{\hat{T}} = 1+\hat{T}+\frac{1}{2!}\hat{T}^2+...$ shows that the product terms include higher excitation even in approximate coupled-cluster wavefunctions Ψ_{CC} and results in size-consistent correlation energies. This is beneficial in comparison to linear parameterization within the CI methods and therefore a systematic hierarchy of approximate coupled-cluster approaches is introduced differing by their maximum allowed excitations in the cluster operator:

$\hat{T} = \hat{T}_2$	CC doubles (CCD) [29, 30]
$\hat{T} = \hat{T}_1 + \hat{T}_2$	CC singles doubles (CCSD) [31]
$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3$	CC singles doubles triples (CCSDT) [32]
$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \hat{T}_4$	CC singles doubles triples quadruples (CCSDTQ) [33, 34]
	... .

Further reduction in complexity is obtained via the Campbell-Baker-Hausdorff expansion [35] of the similarity transformed Hamiltonian (4.30) and (4.33):

$$\mathrm{e}^{-\hat{T}}\hat{H}_{\mathrm{N}}\mathrm{e}^{\hat{T}} = \hat{H}_{\mathrm{N}} + \left[\hat{H}_{\mathrm{N}},\hat{T}\right] + \frac{1}{2!}\left[\hat{H}_{\mathrm{N}},\hat{T}\right]_2 + \frac{1}{3!}\left[\hat{H}_{\mathrm{N}},\hat{T}\right]_3 + \frac{1}{4!}\left[\hat{H}_{\mathrm{N}},\hat{T}\right]_4 \tag{4.34}$$

$$\left[\hat{H}_{\mathrm{N}},\hat{T}\right]_n = \left[\left[\hat{H}_{\mathrm{N}},\hat{T}\right]_{n-1},T\right], \text{ for n} > 1 \text{ and } \left[\hat{\mathrm{H}}_{\mathrm{N}},\hat{\mathrm{T}}\right]_1 = \left[\hat{\mathrm{H}}_{\mathrm{N}},\hat{\mathrm{T}}\right]. \tag{4.35}$$

This expansion terminates with the fifth term because the Hamiltonian consists only of one- and two-electron operators. The CC equations can be analyzed further via Wick's theorem [3], diagrammatic techniques [36] or the application of the Slater-Condon rules [6]. The exponential ansatz leads to a set of nonlinear equations which allows for an iterative solution for the amplitudes t. The resulting equations for the employed approximate CC methods are discussed in the following paragraphs.

Coupled-cluster singles doubles (CCSD)

Within the CCSD method the cluster operator consists of one and two fold excitations $\hat{T} = \hat{T}_1 + \hat{T}_2$. In order to obtain the desired amplitudes t_i^a and t_{ij}^{ab} and the energy eqs. (4.30),

(4.31) and (4.32) have to be solved. The resulting amplitude equations (4.36) and (4.37) are, c.f. [37]:

$$\begin{aligned} t_i^a D_i^a &= f_{ai} + \sum_e t_i^e F_{ae} - \sum_m t_m^a F_{mi} + \sum_{me} t_{im}^{ae} F_{me} - \sum_{nf} t_n^f \langle na||if\rangle \\ &\quad -\frac{1}{2}\sum_{mef} t_{im}^{ef}\langle ma||ef\rangle - \frac{1}{2}\sum_{men} t_{mn}^{ae}\langle nm||ei\rangle \end{aligned} \tag{4.36}$$

$$\begin{aligned} t_{ij}^{ab} D_{ij}^{ab} &= \langle ab||ij\rangle + P(ab)\sum_e t_{ij}^{ae}\left(F_{be} - \frac{1}{2}\sum_m t_m^b F_{me}\right) \\ &\quad -P(ij)\sum_m t_{im}^{ab}\left(F_{mj} + \frac{1}{2}\sum_e t_j^e F_{me}\right) + \frac{1}{2}\sum_{mn}\tau_{mn}^{ab} W_{mnij} \\ &\quad +\frac{1}{2}\sum_{ef}\tau_{ij}^{ef} W_{abef} + P(ij)P(ab)\sum_{me}\left(t_{im}^{ae} W_{mbej} - t_i^e t_m^a \langle mb||ej\rangle\right) \\ &\quad +P(ij)\sum_e t_i^e\langle ab||ej\rangle - P(ab)\sum_m t_m^a\langle mb||ij\rangle. \end{aligned} \tag{4.37}$$

The intermediates F_{pq} and W_{pqrs} for the CCSD equations are defined as follows

$$\begin{aligned} F_{ae} &= (1-\delta_{ae}) f_{ae} - \frac{1}{2}\sum_m t_m^a f_{me} + \sum_{mf} t_m^f \langle ma||fe\rangle \\ &\quad -\frac{1}{2}\sum_{mnf}\tilde{\tau}_{mn}^{af}\langle mn||ef\rangle \end{aligned} \tag{4.38}$$

$$F_{mi} = (1-\delta_{mi}) f_{mi} + \frac{1}{2}\sum_e t_i^e f_{me} + \sum_{en} t_n^e \langle mn||ie\rangle \tag{4.39}$$

$$+\frac{1}{2}\sum_{nef}\tilde{\tau}_{in}^{ef}\langle mn||ef\rangle \tag{4.40}$$

$$F_{me} = f_{me} + \sum_{nf} t_n^f \langle mn||ef\rangle \tag{4.41}$$

$$\begin{aligned} W_{mnij} &= \langle mn||ij\rangle + P(ij)\sum_e t_j^e\langle mn||ie\rangle \\ &\quad +\frac{1}{4}\sum_{ef}\tau_{ij}^{ef}\langle mn||ef\rangle \end{aligned} \tag{4.42}$$

$$W_{abef} = \langle ab||ef\rangle - P(ab)\sum_m t_m^b\langle am||ef\rangle + \frac{1}{4}\sum_{mn}\tau_{mn}^{ab}\langle mn||ef\rangle \tag{4.43}$$

$$\begin{aligned} W_{mbej} &= \langle mb||ej\rangle + \sum_f t_j^f\langle mb||ef\rangle - \sum_n t_n^b\langle mn||ej\rangle \\ &\quad -\sum_{nf}\left(\frac{1}{2}t_{jn}^{fb} + t_j^f t_n^b\right)\langle mn||ef\rangle \end{aligned} \tag{4.44}$$

with the amplitudes $\tilde{\tau}$ and τ:

$$\tilde{\tau}_{ij}^{ab} = t_{ij}^{ab} + \frac{1}{2}\left(t_i^a t_j^b - t_i^b t_j^a\right) \tag{4.45}$$

$$\tau_{ij}^{ab} = t_{ij}^{ab} + t_i^a t_j^b - t_j^a t_i^b\ . \tag{4.46}$$

The permutation operator $P(vw)$ transforms an arbitrary function $f(v,w)$ as follows:

$$P(vw)f(v,w) \quad = \quad f(v,w) - f(w,v) \ .$$

The inverse orbital-energy denominators D are defined via:

$$D_i^a \quad = \quad f_{ii} - f_{aa} \tag{4.47}$$

$$D_{ij}^{ab} \quad = \quad f_{ii} + f_{jj} - f_{aa} - f_{bb} \ . \tag{4.48}$$

The CC equations (4.36) and (4.37) are solved iteratively and contain the Fock matrix elements f_{pq} (4.29), the two-electron integrals $\langle pq||rs\rangle$ as well as the amplitudes t_i^a and t_{ij}^{ab}. As an initial guess for the t_{ij}^{ab} amplitudes results from an MP2 calculation can be used all other amplitudes are set to zero. Insertion of the initial amplitudes (or in the following iterations the amplitudes of the preceding iteration) on the right side of eqs. (4.36) and (4.37) yields after division by the corresponding weighting factor D_i^a or D_{ij}^{ab} a new set of amplitudes. This is continued until the amplitudes of two subsequent iterations are not changing within a given threshold. Insertion of the converged amplitudes into eq. (4.30) results in the CCSD energy. The overall scaling of CCSD is about $\mathcal{O}(M^6)$. With current implementations and computer resources molecules up to 30 atoms can be treated [38]. If a large enough basis set is used CCSD calculations can achieve an accuracy of about 50 kJ/mol for reaction energies [39] and 0.005 Å for bond lengths [40]. This shows that the CCSD method is a good approximation but not sufficient for very accurate predictions.

Coupled-cluster singles doubles triples (CCSDT)

Based on the CCSD equations the amplitude equations for the CCSDT model can be formulated as follows [32, 41]:

$$\begin{aligned} 0 \quad &= \quad \langle \Psi_i^a | \mathrm{e}^{-\hat{T}} \hat{H}_\mathrm{N} \mathrm{e}^{\hat{T}} | \Psi_0 \rangle \\ &= \quad T_1(\mathrm{CCSD}) + \langle \Psi_i^a | \left[\hat{H}_\mathrm{N}, \hat{T}_3 \right] | \Psi_0 \rangle \end{aligned} \tag{4.49}$$

$$\begin{aligned} 0 \quad &= \quad \langle \Psi_{ij}^{ab} | \mathrm{e}^{-\hat{T}} \hat{H}_\mathrm{N} \mathrm{e}^{\hat{T}} | \Psi_0 \rangle \\ &= \quad T_2(\mathrm{CCSD}) + \langle \Psi_{ij}^{ab} | \left[\hat{H}_\mathrm{N}, \hat{T}_3 \right] | \Psi_0 \rangle + \langle \Psi_{ij}^{ab} | \left[\left[\hat{H}_\mathrm{N}, \hat{T}_1 \right], \hat{T}_3 \right] | \Psi_0 \rangle \end{aligned} \tag{4.50}$$

$$0 \quad = \quad \langle \Psi_{ijk}^{abc} | \mathrm{e}^{-\hat{T}} \hat{H}_\mathrm{N} \mathrm{e}^{\hat{T}} | \Psi_0 \rangle. \tag{4.51}$$

$T_1(\mathrm{CCSD})$ and $T_2(\mathrm{CCSD})$ denote the contributions from the CCSD approximation, see eqs. (4.31) and (4.32). The CCSDT model gives good predictions for energies and molecular properties. Dissociation energies can be calculated with an accuracy of a few kJ/mol [42], geometries to about 0.02 Å [43]. Due to the computational scaling of $\mathcal{O}(M^8)$ the application range is limited. Therefore approximate treatments of the triples excitations are common to avoid the computationally most demanding steps, which scales with $n_{occ}^3 n_{vrt}^5$ and requires the full triples amplitudes to be stored on hard disk. Here *occ* and *vrt* denote the number of occupied and unoccupied orbitals, respectively. The most common approximation is the coupled-cluster singles and doubles scheme augmented by a perturbative treatment of triple excitations (CCSD(T)) with a scaling of $\mathcal{O}(M^7)$, which will be discussed in the next paragraph.

CCSD(T)

Within the CCSD(T) approximation the relationship between CC and perturbation theory is exploited [3]. Based on a CCSD calculation triple excitations are included via perturbation theory [4,44]. The Hamiltonian of the CC equations, $\hat{H}_{\mathrm{N}}$, can be written in a way analogous to eq. (4.17). Within this perturbation expansion triple amplitude corrections to the energy are obtained in fourth and fifth order and are calculated using the converged CCSD amplitudes. These corrections are added to the CCSD energy [3]:

$$E_{\mathrm{CCSD(T)}} = E_{\mathrm{CCSD}} + \underbrace{E^{(4)} + E^{(5)}}_{\Delta E_{\mathrm{CCSD(T)}}} \tag{4.52}$$

$$E^{(4)} = \frac{1}{36} \sum_{ijkabc} t_{ijk}^{abc*} D_{ijk}^{abc} t_{ijk}^{abc} \tag{4.53}$$

$$E^{(5)} = \frac{1}{4} \sum_{ijkabc} \langle jk||bc\rangle t_i^{a*} t_{ijk}^{abc} \; . \tag{4.54}$$

The term D_{ijk}^{abc} in eq. (4.53) consists of diagonal elements of the Fock matrix analogous to eq. (4.48). The $\hat{T}_3$ amplitudes are obtained in the following way:

$$D_{ijk}^{abc} t_{ijk}^{abc} = P(i|jk)P(a|bc) \left(\sum_d \langle bc||di\rangle t_{jk}^{ad} - \sum_l \langle la||jk\rangle t_{il}^{bc} \right) . \tag{4.55}$$

The permutation operator $P(i|jk)$ is defined via $P(i|jk)f(ijk) = f(ijk) - f(jik) - f(kij)$. The scaling behavior is reduced by one order of magnitude in comparison to the CCSDT ansatz. In addition to that the storage of the triples amplitudes as well as the iterations for the most demanding steps are avoided. The resulting energies are of similar accuracy as in the CCSDT method [24]. Due to the high accuracy of the results and less demanding computational cost CCSD(T) has become the most common CC method.

Higher-level approximate coupled-cluster methods

The accuracy of the results can be further increased systematically by including higher excitations in approximate CC methods, though this leads to an enormous increase of computational effort. The CCSDTQ method [33, 34], allowing for four-fold excitations in the cluster operator, has a computational scaling of $\mathcal{O}(M^{10})$. Therefore the range of applicability is limited to fairly small systems. The CCSDTQ model has an error of up to 1 kJ/mol in the electronic energy and is able to predict bond distances within a tenth of a picometer. A perturbative treatment of quadruple excitations can be defined in analogy to the perturbative treatment of triples. The resulting CCSDT(Q) method [45, 46] is again non iterative and scales with $\mathcal{O}(M^9)$. The lower scaling allows for an extended applicability in comparison to the full treatment of quadruple excitations.

The implementation of CC approximations including higher excitations is limited due to their complexity and number of the terms (see ref. [47]). Table 4.1 shows that the computational effort as well as the number of terms in the amplitude equations that have to be coded increases rapidly with increasing excitation level in the $\hat{T}$ operator.

Therefore implementations via computer aided techniques like automatic formula or program generation [48] as well as the so-called string-based methods [47, 49] are advantageous.

Table 4.1: Computational scaling and number of terms in approximate coupled-cluster methods.

Method	Number of terms	Computational scaling of the most demanding term
CCSD	45	$\mathcal{O}(M^6)$
CCSD(T)	50	$\mathcal{O}(M^7)$
CCSDT	99	$\mathcal{O}(M^8)$
CCSDT(Q)	107	$\mathcal{O}(M^9)$
CCSDTQ	180	$\mathcal{O}(M^{10})$
CCSDTQP	286	$\mathcal{O}(M^{12})$
CCSDTQPH	397	$\mathcal{O}(M^{14})$

4.1.5 Density-functional theory

So far only methods based on many-electron wavefunctions were considered. Density-functional theory (DFT) [8] follows a different route and replaces the many-electron wavefunction by the electron density $\rho(\{\mathbf{r}\})$ as the basic quantity. The energy of an N-electron system can be expressed as a functional of the electron density (first Hohenberg-Kohn theorem) [50].

$$E_0 = E_0[\rho_{exact}(\{\mathbf{r}\})] \tag{4.56}$$

Furthermore, the variation principle is valid for trial densities that fulfill the conditions $\int \rho_{trial}(\{\mathbf{r}\})dr = N$ and $\rho_{trial}(\{\mathbf{r}\}) \geq 0$ (second Hohenberg-Kohn theorem) [50]:

$$E[\rho_{trial}] \geq E_0^{exact} \ . \tag{4.57}$$

The real electron density minimizes the energy functional $E[\rho_{trial}]$. The exact form of the functional is unknown, the first Hohenberg-Kohn theorem only claims the existence of such a functional. $E[\rho]$ can be partitioned as follows

$$E[\rho] = T[\rho] + V_{ne}[\rho] + V_{ee}[\rho] \tag{4.58}$$

$T[\rho]$, $V_{ne}[\rho]$, and $V_{ee}[\rho]$ denote functionals of the kinetic energy, the electron-nucleus interaction and the electron-electron interaction, respectively. $V_{ne}[\rho]$ can be evaluated in a straightforward manner while the sum of $T[\rho]$ and $V_{ee}[\rho]$ defines the functional $F[\rho]$. Here, the dependence on the density is unknown.

$$F[\rho] = T[\rho] + V_{ee}[\rho] \tag{4.59}$$

$V_{ee}[\rho]$ can be subdivided into a Coulomb term $J[\rho]$ and a non-classical term, which represents all remaining interactions.
To avoid the problem of the representation of $T[\rho]$ by means of the electron density, Kohn and Sham [8] introduced a non-interacting reference system, which has the same electron density as the real system and can be represented by a Slater determinant:

$$\Psi_s = \frac{1}{\sqrt{N!}} det|\varphi_1\varphi_2 \ldots \varphi_N| \ , \tag{4.60}$$

with the orbitals φ_i. The kinetic energy $T_s[\rho]$ of the non-interacting reference system is given by

$$T_s[\rho] = \sum_i^N \langle\varphi_i| - \frac{1}{2}\nabla^2|\varphi_i\rangle. \tag{4.61}$$

Introducing the exchange-correlation functional $E_{xc}[\rho]$ the Hohenberg-Kohn functional can be written as:

$$F[\rho] = T_s[\rho] + J[\rho] + E_{xc}[\rho] \tag{4.62}$$

$$E_{xc}[\rho] = T[\rho] - T_s[\rho] + V_{ee}[\rho] - J[\rho] \tag{4.63}$$

Minimization of eq. (4.55) leads to the Kohn-Sham equations

$$\left[-\frac{1}{2}\Delta + \underbrace{\left(\int \frac{\rho(\mathbf{r_2})}{r_{12}} + V_{xc}(\mathbf{r_1}) - \sum_A \frac{Z_A}{|\mathbf{r}_1 - \mathbf{R}_A|}\right)}_{v_{eff}}\right] \varphi_i = \varepsilon_i \varphi_i, \tag{4.64}$$

which can be solved analogous to the HF equations in an iterative way. The main disadvantage is the completely undetermined exchange-correlation functional. This led to the development of a large number of approximate functionals, some of the most popular will be discussed in the following:

One possibility for the approximation of the exchange-correlation functional is the use of the corresponding expression for a homogeneous electron gas. Dirac's expression [7, 51] for the electron gas is used for the exchange term, while the correlation term is obtained from fitting an analytic form to numerical results and asymptotic limits [52]. This approach is called local-density approximation (LDA) [52].

For calculations of open-shell systems the use of spin densities is advantageous. This approach is called local spin density approximation (LSD) [53].
Since the electron or spin density is known not to be homogeneous in atoms and molecules the former approximations are improved by introducing density gradient corrections: The most popular functionals are BLYP (Slater-Dirac exchange term combined with Becke's gradient correction (B) and a correlation functional by Lee, Yang, and Parr (LYP)) [54,55] and BP86 (with the gradient-corrected correlation functional of Perdew instead of the LYP functional) [55]. In addition to gradient corrections a fraction of exact HF-exchange can be included leading to hybrid functionals. The most popular hybrid functional is B3LYP [55].

Most functionals include additional parameters which are fitted to experimental results. This practice might improve the results of calculations and add flexibility in functional development but also increases the empirical character of density-functional theory [8].

4.2 Analytic derivatives

In chemical applications the calculation of quantities directly observable experimentally is essential. The energies and wavefunctions considered to this point are not, by themselves, sufficient for this purpose. Hence, this section will focus on the calculation of properties via derivative techniques.

Most molecular properties can be expressed in terms of derivatives of the energy with respect to a certain perturbation x or combined perturbations x, y, ... [56]. Derivatives of the energy can be obtained within quantum-chemical methods via numerical or analytical differentiation. On the one hand finite-difference schemes are simple to implement, because only energies depending on the perturbation have to be calculated. On the other hand the limited numerical accuracy and the possibly large number of perturbations restrict the range of applicability. In addition to that it is nearly impossible to include complex wavefunction parameters in such a scheme, which is, for example, necessary for the calculation of magnetic properties [56].

The application of analytic-derivative techniques overcomes the problems mentioned here, but requires more effort in derivation and implementation of the associated equations [57]. The conceptual ideas for HF and CC theory will be presented in the following paragraphs.

In general the electronic energy E and the corresponding wavefunction parameters c (e.g., MO coefficients, CI coefficients or CC amplitudes) depend on the perturbation x. Via the chain-rule:

$$E = E(x, c(x)) \quad \Rightarrow \quad \left(\frac{\mathrm{d}E}{\mathrm{d}x}\right) = \left(\frac{\partial E}{\partial x}\right) + \left(\frac{\partial E}{\partial c(x)}\right)\left(\frac{\partial c(x)}{\partial x}\right) \tag{4.65}$$

is obtained. For the first term derivatives of the integrals are needed because the energy depends on the perturbation x via $\hat{H}$ and χ_ν. In the second term derivatives of the wavefunction parameters are found.

4.2.1 Hartree-Fock theory

Within the framework of variational methods like the HF-SCF or the CI method the derivatives of the perturbed wavefunction parameters vanish, as the variation principle requests minimization of the energy with respect to c.

$$\left(\frac{\partial E}{\partial c(x)}\right)_{x=0} = 0\ . \tag{4.66}$$

Orthonormality of the molecular orbitals (see eqs. (4.10) and (4.12)) is ensured by setting up the following side condition:

$$\left(\sum_{\mu\nu} c^*_{\mu i} S_{\mu\nu} c_{\nu j} - \delta_{ij}\right) = 0. \tag{4.67}$$

The HF energy gradient then is obtained for the case of RHF by differentiating eq. (4.14) with respect to the perturbation x:

$$\begin{aligned}\frac{\mathrm{d}E}{\mathrm{d}x} &= \sum_{\mu\nu} D_{\mu\nu}\frac{\partial h_{\mu\nu}}{\partial x} + \frac{1}{2}\sum_{\mu\nu\sigma\rho} D_{\mu\nu}D_{\sigma\rho}\left(\frac{\partial\langle\mu\sigma|\nu\rho\rangle}{\partial x} - \frac{1}{2}\frac{\partial\langle\mu\sigma|\rho\nu\rangle}{\partial x}\right) \\ &\quad - \sum_{\mu\nu} W_{\mu\nu}\frac{\partial S_{\mu\nu}}{\partial x}\ .\end{aligned} \tag{4.68}$$

where W denotes the energy-weighted density matrix.

$$W_{\mu\nu} = 2\sum_i c^*_{\mu i}\epsilon_i c_{\nu i} \tag{4.69}$$

The following expression is obtained for second derivatives at the HF-SCF level [58, 59] by straightforward differentiation of eq. (4.60):

$$\begin{aligned}\frac{\mathrm{d}^2E}{\mathrm{d}x\mathrm{d}y} &= \sum_{\mu\nu} D_{\mu\nu}\frac{\partial^2 h_{\mu\nu}}{\partial x\partial y} + \frac{1}{2}\sum_{\mu\nu\sigma\rho} D_{\mu\nu}D_{\sigma\rho}\left(\frac{\partial^2\langle\mu\sigma|\nu\rho\rangle}{\partial x\partial y} - \frac{1}{2}\frac{\partial^2\langle\mu\sigma|\rho\nu\rangle}{\partial x\partial y}\right)\\ &\quad - \sum_{\mu\nu} W_{\mu\nu}\frac{\partial^2 S_{\mu\nu}}{\partial x\partial y} + \sum_{\mu\nu}\frac{\partial D_{\mu\nu}}{\partial y}\left(\frac{\partial h_{\mu\nu}}{\partial x} + \sum_{\sigma\rho} D_{\sigma\rho}\left(\frac{\partial\langle\mu\sigma|\nu\rho\rangle}{\partial x} - \frac{1}{2}\frac{\partial\langle\mu\sigma|\rho\nu\rangle}{\partial x}\right)\right)\\ &\quad - \sum_{\mu\nu}\frac{\partial W_{\mu\nu}}{\partial y}\frac{\partial S_{\mu\nu}}{\partial x}\end{aligned} \tag{4.70}$$

with

$$\frac{\partial D_{\mu\nu}}{\partial x} = 2\sum_i\left(\frac{\partial c^*_{\mu i}}{\partial x}c_{\nu i} + c^*_{\mu i}\frac{\partial c_{\nu i}}{\partial x}\right) \tag{4.71}$$

and

$$\frac{\partial W_{\mu\nu}}{\partial x} = 2\sum_i\left(\frac{\partial c^*_{\mu i}}{\partial x}\varepsilon_i c_{\nu i} + c^*_{\mu i}\varepsilon_i\frac{\partial c_{\nu i}}{\partial x}\right) + \sum_{ij} c^*_{\mu i}\frac{\partial\varepsilon_{ij}}{\partial x}c_{\nu j}\ . \tag{4.72}$$

Eq. (4.70) contains first derivatives of the MO coefficients, which are determined by differentiating eq. (4.15) or Brillouin's theorem ($f_{ai} = 0$) and parameterized in the following manner in terms of the unperturbed MO coefficients:

$$\frac{\partial c_{\mu i}}{\partial x} = \sum_p c_{\mu p}U^x_{pi}\ . \tag{4.73}$$

Only the coefficients U^x_{ai} need to be determined through the coupled-perturbed Hartree-Fock (CPHF) equations [58, 59] which takes for "real" perturbations the following form:

$$\sum_e\sum_m(4\langle ae|im\rangle - \langle ae|mi\rangle - \langle am|ei\rangle + \delta_{ae}\delta_{im}(\varepsilon_a - \varepsilon_i))U^x_{em} = B^x_{ai} \tag{4.74}$$

with

$$\begin{aligned}B^x_{ai} &= -\sum_{\mu\nu} c^*_{\mu a}c_{\nu i}\left(\frac{\partial h_{\mu\nu}}{\partial x} + \sum_{\sigma\rho} D_{\sigma\rho}\left(\frac{\partial\langle\mu\sigma|\nu\rho\rangle}{\partial x} - \frac{1}{2}\frac{\partial\langle\mu\sigma|\rho\nu\rangle}{\partial x}\right)\right)\\ &\quad + \sum_{mn} S^x_{mn}(2\langle am|in\rangle - \langle am|ni\rangle) + S^x_{ai}\varepsilon_i\end{aligned} \tag{4.75}$$

and

$$S^x_{pq} = \sum_{\mu\nu} c^*_{\mu p}\frac{\partial S_{\mu\nu}}{\partial x}c_{\nu q} \tag{4.76}$$

while the U^x_{ij} coefficients (which are not unique due to the invariance of the HF-SCF energy with respect to unitary rotation of the occupied orbitals) are commonly set to $-\frac{1}{2}S^x_{ij}$ in order to fulfill the differentiated orthonormality condition. The CPHF equations are linear and can be solved using iterative techniques [59].

4.2.2 Coupled-cluster theory

For non-variational approaches like CC methods eq. (4.66) is not valid. Nevertheless, elimination of the derivatives of the wavefunction parameters from the gradient expression is possible. For this purpose an energy functional $\tilde{E}$ is set up containing the equations that determine t as additional constraints multiplied by Lagrange multipliers λ_i. Within CC theory these constraints correspond to the amplitude equations (4.33). The derivative of the energy functional with respect to the perturbation yields an expression that does not contain derivatives of the wavefunction with respect to the perturbation but requires the additional solution of the conditional equation for λ. The energy functional for CC theory is defined by:

$$\tilde{E} = \langle\Psi_0|\mathrm{e}^{-\hat{T}}\hat{H}\mathrm{e}^{\hat{T}}|\Psi_0\rangle + \sum_p \lambda_p\langle\Phi_p|\mathrm{e}^{-\hat{T}}\hat{H}\mathrm{e}^{\hat{T}}|\Psi_0\rangle \tag{4.77}$$

$$= \langle\Psi_0|(1+\hat{\Lambda})\mathrm{e}^{-\hat{T}}\hat{H}\mathrm{e}^{\hat{T}}|\Psi_0\rangle\ , \tag{4.78}$$

with the amplitude equations (4.33) as side condition. The abbreviated form in eq. (4.78) contains a deexcitation operator $\hat{\Lambda}$ [60] which is defined by analogy to the cluster operator:

$$\langle\Psi_0|\hat{\Lambda} = \sum_p \lambda_p\langle\Psi_p| \tag{4.79}$$

$$\hat{\Lambda} = \underbrace{\sum_{i,a}\lambda_a^i\hat{a}_i^+\hat{a}_a}_{\hat{\Lambda}_1} + \underbrace{\frac{1}{4}\sum_{ij,ab}\lambda_{ab}^{ij}\hat{a}_i^+\hat{a}_a\hat{a}_j^+\hat{a}_b}_{\hat{\Lambda}_2} + \ldots\ . \tag{4.80}$$

No derivatives of $\hat{T}$ and $\hat{\Lambda}$ are required for energy gradients [61]:

$$\frac{\mathrm{d}\tilde{E}}{\mathrm{d}x} = \langle\Psi_0|(1+\hat{\Lambda})\mathrm{e}^{-\hat{T}}\frac{\mathrm{d}\hat{H}}{\mathrm{d}x}\mathrm{e}^{\hat{T}}|\Psi_0\rangle\ . \tag{4.81}$$

In addition to that the following set of linear equations has to be solved in order to make $\tilde{E}$ stationary with respect to both t and λ amplitudes:

$$0 = \langle\Psi_0|(1+\hat{\Lambda})\left(\mathrm{e}^{-\hat{T}}\hat{H}\mathrm{e}^{\hat{T}} - E_{\mathrm{CC}}\right)|\Psi_p\rangle\ . \tag{4.82}$$

Furthermore, the CC energy depends on the MO coefficients which are determined at the HF level. This leads to an extended energy functional of the from:

$$\tilde{E} = \langle\Psi_0|(1+\hat{\Lambda})\mathrm{e}^{-\hat{T}}\hat{H}\mathrm{e}^{\hat{T}}|\Psi_0\rangle + \sum_{ia} Z_{ia}f_{ai} + \sum_{pq} I_{pq}\left(\sum_{\mu\nu} c_{\mu p}S_{\mu\nu}c_{\nu q}^* - \delta_{pq}\right)\ . \tag{4.83}$$

Here Z_{ia} denotes the Lagrange multiplier, which ensure Brillouin's theorem ($f_{ai} = 0$) [6] as side condition. I_{pq} is an element of an effective energy-weighted density matrix and is used as multiplier to conserve the orthonormality of the orbitals. Differentiation of the extended energy functional including orbital relaxation yields in the AO basis:

$$\frac{\mathrm{d}E}{\mathrm{d}x} = \sum_{\mu\nu} D_{\mu\nu}\frac{\partial h_{\mu\nu}}{\partial x} + \sum_{\mu\nu\rho\sigma}\Gamma_{\mu\nu\rho\sigma}\frac{\partial\langle\mu\nu|\sigma\rho\rangle}{\partial x} + \sum_{\mu\nu} I_{\mu\nu}\frac{\partial S_{\mu\nu}}{\partial x}. \tag{4.84}$$

$D_{\mu\nu}$, $\Gamma_{\mu\nu\rho\sigma}$, and $I_{\mu\nu}$ denote an effective one-particle matrix, the two-particle density matrix and the energy-weighted density matrix, respectively [62]. The explicit expressions depend on the applied approximate CC method (see e.g. references [60, 63]).

Analytic second derivatives can be determined by further differentiation of the gradient expressions employing the (2n+1) and the (2n+2) rules [24, 56, 64–66].

$$\begin{aligned}\frac{\mathrm{d}^2\tilde{E}}{\mathrm{d}x\mathrm{d}y} &= \langle\Psi_0|(1+\hat{\Lambda})\mathrm{e}^{-\hat{T}}\frac{\mathrm{d}^2\hat{H}}{\partial x\partial y}\mathrm{e}^{\hat{T}}|\Psi_0\rangle \\ &+\langle\Psi_0|(1+\hat{\Lambda})[\mathrm{e}^{-\hat{T}}\frac{\mathrm{d}\hat{H}}{\partial x}\mathrm{e}^{\hat{T}},\frac{\mathrm{d}\hat{T}}{\mathrm{d}y}]|\Psi_0\rangle \\ &+\langle\Psi_0|(1+\hat{\Lambda})[\mathrm{e}^{-\hat{T}}\frac{\mathrm{d}\hat{H}}{\partial y}\mathrm{e}^{\hat{T}},\frac{\mathrm{d}\hat{T}}{\mathrm{d}x}]|\Psi_0\rangle \\ &+\langle\Psi_0|(1+\hat{\Lambda})[[\mathrm{e}^{-\hat{T}}\hat{H}\mathrm{e}^{\hat{T}},\frac{\mathrm{d}\hat{T}}{\mathrm{d}x}],\frac{\mathrm{d}\hat{T}}{\mathrm{d}y}]|\Psi_0\rangle\ . \end{aligned} \tag{4.85}$$

The expression given in eq. (4.78) contains only derivatives of the cluster operator The required perturbed amplitudes are obtained by differentiation of the first order CC equations with respect to x or y:

$$\langle\Psi_p|\mathrm{e}^{-\hat{T}}\frac{\partial\hat{H}}{\partial x}\mathrm{e}^{\hat{T}}|\Psi_0\rangle + \langle\Psi_p|[\mathrm{e}^{-\hat{T}}\hat{H}\mathrm{e}^{\hat{T}},\frac{\mathrm{d}\hat{T}}{\mathrm{d}x}]|\Psi_0\rangle = 0. \tag{4.86}$$

For some cases it is advantageous to rearrange eq. (4.78) [65], this expression can also be obtained by straightforward differentiation of gradient:

$$\begin{aligned}\frac{\mathrm{d}^2\tilde{E}}{\mathrm{d}x\mathrm{d}y} &= \langle\Psi_0|(1+\hat{\Lambda})\mathrm{e}^{-\hat{T}}\frac{\mathrm{d}^2\hat{H}}{\partial x\partial y}\mathrm{e}^{\hat{T}}|\Psi_0\rangle \\ &+\langle\Psi_0|(1+\hat{\Lambda})[\mathrm{e}^{-\hat{T}}\frac{\mathrm{d}\hat{H}}{\partial x}\mathrm{e}^{\hat{T}},\frac{\mathrm{d}\hat{T}}{\mathrm{d}y}]|\Psi_0\rangle \\ &+\langle\Psi_0|\frac{\mathrm{d}\hat{\Lambda}}{\mathrm{d}x}\mathrm{e}^{-\hat{T}}\frac{\mathrm{d}\hat{H}}{\partial x}\mathrm{e}^{\hat{T}}|\Psi_0\rangle\ . \end{aligned} \tag{4.87}$$

In eq. (4.80) derivatives of the Lagrange multipliers appear but derivative amplitudes are required only for one of the two perturbations. When dealing with two different classes of perturbations like in the case of NMR shieldings only 6 perturbed equations have to be solved, while the use of eq. (4.78) requires the solution of the perturbed amplitude equations for the $3N_n$ components of the nuclear magnetic moments and the 3 components of the magnetic field.

In this work the third and fourth order derivatives with respect to the nuclear coordinates are obtained by numerical differentiation of analytic second derivatives [67].

4.3 Basis sets

According to eq. (4.12) all quantum-chemical methods discussed so far require the use of a one-electron basis composed of a finite number of atomic orbitals, centered at the atomic nucleus within the molecule. Nowadays, linear combinations of Gaussian orbitals are chosen because it is easier to calculate two-electron and other integrals with Gaussian basis functions than with, for example, Slater orbitals. The basis sets employed in this work will be described briefly in the following.

4.3.1 Correlation-consistent basis sets

The basis sets developed by Dunning and coworkers [68–73] are designed to provide systematic convergence to the complete basis set (CBS) limit and can be used together with extrapolation techniques [74,75]. For first- and second-row atoms, the basis sets are cc-pVXZ where X=D, T, Q, 5, 6. The 'cc-p', stands for 'correlation consistent polarized' and the 'V' indicates that they are valence only basis sets.[1] They include successively larger sets of polarization (correlating) functions (d, f, g, etc.) as can be found in table 4.2.

The cc-pVXZ basis sets can be augmented with core functions [76–78] for improving the description when correlating all electrons (ae). The resulting polarized core-valence basis sets are termed cc-pCVXZ where X=D, T, Q, 5, 6.

A variation of the cc-pCVXZ basis sets is provided by the cc-pwCVXZ sets where X=T, Q, 5 [77]. For these basis sets the core functions are optimized with respect to both core and valence correlation energy and thus provide an improved description of the valence-correlation energy in comparison to the cc-pCVXZ sets. The corresponding contraction schemes are given in the tables 4.3, 4.4, and 4.5.

For the calculation of excited states and electric properties the polarized consistent valence or (weighted) core valence basis sets can be further augmented by additional sets of diffuse functions for each angular momentum [72, 79–81]. This is indicated by 'aug-' in front of the 'cc-p'.

Table 4.2: Contraction scheme of correlation-consistent polarized valence basis sets of Dunning and co-workers for hydrogen.

Basis set	Contraction for H	aug
cc-pVDZ	4s1p $\longrightarrow$ 2s1p	+ 1s1p
cc-pVTZ	5s2p1d $\longrightarrow$ 3s2p1d	+ 1s1p1d
cc-pVQZ	6s3p2d1f $\longrightarrow$ 4s3p2d1f	+ 1s1p1d1f
cc-pV5Z	8s4p3d2f1g $\longrightarrow$ 5s4p3d2f1g	+ 1s1p1d1f1g
cc-pV6Z	10s5p4d3f2g1h $\longrightarrow$ 6s5p4d3f2g1h	+ 1s1p1d1f1g1h

[1] It is quite common in applications of correlated methods to use the frozen-core (fc) approximation, in which the lowest-lying molecular orbitals are excluded from the correlation treatment. The frozen core for atoms lithium to neon typically consists of the 1s atomic orbital. A justification for this approximation is that the inner-shell electrons of an atom are less sensitive to their environment than are the valence electrons. Thus the error introduced by freezing the core orbitals should be nearly constant for molecules containing the same types of atoms.

Table 4.3: Contraction scheme of correlation-consistent polarized valence basis sets of Dunning and co-workers for first-row atoms.

Basis set	Contraction for C, N, O, F	aug	C
cc-pVDZ	9s4p1d ⟶ 3s2p1d	+ 1s1p1d	+1s1p
cc-pVTZ	10s5p2d1f ⟶ 4s3p2d1f	+ 1s1p1d1f	+2s2p1d
cc-pVQZ	12s6p3d2f1g ⟶ 5s4p3d2f1g	+ 1s1p1d1f1g	+3s3p2d1f
cc-pV5Z	14s8p4d3f2g1h ⟶ 6s5p4d3f2g1h	+ 1s1p1d1f1g1h	+4s4p3d2f1g
cc-pV6Z	16s10p5d4f3g2h1i ⟶ 7s6p5d4f3g2h1i	+ 1s1p1d1f1g1h1i	+5s5p4d3f2g1h

Table 4.4: Contraction scheme of correlation-consistent polarized valence basis sets of Dunning and co-workers for second-row atoms.

Basis set	Contraction for Cl and Si	aug	C
cc-pVDZ	12s8p1d ⟶ 4s2p1d	+ 1s1p1d	+1s1p
cc-pVTZ	15s9p2d1f ⟶ 5s4p2d1f	+ 1s1p1d1f	+2s2p1d
cc-pVQZ	16s11p3d2f1g ⟶ 6s5p3d2f1g	+ 1s1p1d1f1g	+3s3p2d1f
cc-pV5Z	20s12p4d3f2g1h ⟶ 7s6p4d3f2g1h	+ 1s1p1d1f1g1h	+3d2f1g[a]

[a] 20s12p are recontracted to 11s10p.

Table 4.5: Contraction scheme of correlation-consistent weighted core-polarized valence basis sets of Dunning and co-workers for carbon and iron.

Basis set	cc-pwCVTZ	cc-pwCQZ
C	12s7p3d1f ⟶ 6s5p3d1f	15s9p5d3f1g ⟶ 8s7p5d3f1g
Fe	22s18p10d3f2g ⟶ 9s8p6d3f2g	25s21p14d4f3g2h ⟶ 10s9p7d4f3g2h

4.3.2 Basis sets for the calculation of nuclear magnetic shielding constants

Previous experience (see e.g., [82, 83]) indicates that the basis sets developed by Schäfer et al. [84] augmented with polarization functions [82] are well suited for the calculation of NMR chemical shifts. A series of different sets derived from the original work of Schäfer et al. [84], a "double-zeta polarization (dzp)" type of basis, larger contracted sets denoted as tzp, tz2p, qz2p, and pz3d2f as well as the large uncontracted 13s9p4d3f/8s3p2d and 15s11p4d3f/10s4p3d sets is employed. The exponents for the polarization functions have been chosen as described in Ref. [82] in the case of dzp, tzp, tz2p, and qz2p and have been taken from Ref. [68] in all other cases.

Table 4.6: Contraction schemes of the basis sets used in the NMR chemical shift calculations.

Basis set	Contraction for C,N,O,F	Contraction for H
dzp	8s4p1d ⟶ 4s2p1d	4s1p ⟶ 2s1p
tzp	9s5p1d ⟶ 5s3p1d	5s1p ⟶ 3s1p
tz2p	9s5p2d ⟶ 5s3p2d	5s2p ⟶ 3s2p
qz2p	11s7p2d ⟶ 6s4p2d	6s2p ⟶ 3s2p
pz3d2f	13s8p3d2f ⟶ 8s5p3d2f	8s3p2d ⟶ 5s3p2d
13s9p4d3f	13s9p4d3f	8s3p2d
15s11p4d3f	15s11p4d3f	10s4p3d

4.3.3 Atomic natural orbital basis sets

In atomic natural orbital (ANO) basis sets the contraction coefficients are defined by the natural orbitals obtained from atomic calculations using large primitive basis sets. In this thesis, a smaller (ANO1) and a larger (ANO2) contraction of the ANO basis set of Taylor and Almlöf [85] is used for the calculation of some of the spectroscopic properties. These basis sets are always used together with the frozen-core approximation.

Table 4.7: Contraction schemes employed for the ANO basis sets of Taylor and Almlöf.

Basis set	Contraction for C,N,O,F	Contraction for H
ANO1	13s8p6d4f ⟶ 4s3p2d1f	8s6p4d ⟶ 4s2p1d
ANO2	13s8p6d4f2g ⟶ 5s4p3d2f1g	8s6p4d3f ⟶ 4s3p2d1f

4.3.4 Split-valence basis sets

This type of basis defines different contractions for core atomic orbitals and valence orbitals. In this work the valence quadruple-zeta plus polarization basis set def2-QZVPP of Weigend and Ahlrichs [86, 87] is employed for near basis-set limit DFT benchmark calculations.

Table 4.8: Contraction scheme of the def2-QZVPP basis sets of Weigend and Ahlrichs.

Basis set	def2-QZVPP
H	7s3p2d1f ⟶ 4s3p2d1f
C	15s8p3d2f1g ⟶ 7s4p3d2f1g
Si	20s14p4d2f1g ⟶ 9s6p4d2f1g
Fe	24s18p10d4f2g ⟶ 11s6p5d4f2g

5 Parallel coupled-cluster calculations

Application of CC methods to larger chemical problems is limited by the rapidly increasing computational effort with growing number of electrons and basis functions.
A detailed analysis reveals that for CC methods like coupled-cluster singles and doubles (CCSD) [31] and coupled-cluster singles and double scheme augmented by a perturbative treatment of triple excitations (CCSD(T)) [4] the limiting factor is CPU time and not storage requirements. If N is chosen as a measure of the system size, storage requirements for two-electron integrals, coupled-cluster amplitudes and intermediates scale as N^4. In contrast to that, the operation count scales as N^6 for CCSD and N^7 for CCSD(T). Nowadays, efficient implementations allow calculations at the CCSD(T) level of theory with up to 800 basis functions [88–91].
While for almost all methods numerous efficient parallel algorithms have been developed, the number of parallel implementations of CC methods has only increased in recent years [92–100]. However, no CC code capable of calculating general second-order molecular properties at the CC level using analytical derivatives has been so far adapted for parallel architectures. The main reason for this is that the mathematical structure of the CC equations makes an efficient fully-parallel implementation or re-implementation to be very demanding and time consuming. In this work an alternative approach, namely the adaptation of an efficient serial algorithm to parallel environments is demonstrated.
This ansatz is based only on the message passing interface (MPI) [101] and works without an additional layer of complexity provided by specialized libraries like Global Arrays (GA) [102, 103], Distributed Data Interface (DDI) [91, 104], or Array Files (AF) [105, 106] which are used in common implementations. All non-parallel steps run redundantly on every available processor at the same time.
The employed strategy is presented in a stepwise manner leading to an algorithm with parallelized routines for the time-determining steps in the CCSD and CCSD(T) energy, gradient, and analytical second-derivative calculations. A detailed investigation of the time-determining steps in CCSD and CCSD(T) calculations and the reduction of the overall execution time in the parallel algorithm is carried out.

5.1 Parallelization strategy for coupled-cluster energies and derivatives

A common approach for the parallelization of CC algorithms is to minimize storage requirements while aiming at constant (but in practice high) total network communication by distributing parts of integrals, amplitudes and intermediates and communicating the pieces as needed by other processors. While this approach guarantees scaling of the algorithm in the limit of a large number of processors [96, 107], high-speed and expensive network connections are required. Furthermore, the structure of such an algorithm as well as the communication overhead, arising through dead times in which a node awaits data, may shift the crossover point with respect to efficient serial algorithms to a large numbers of nodes.
Following a different route, a replicated storage scheme in order to minimize communica-

tion is applied. Most of the quantities needed in the CC iterations are stored completely on every node in order to avoid communication of intermediate quantities. In contrast to the algorithm outlined by Olson et al. [91], the work presented here is tailored to cluster architectures with moderate hardware specifications, assuming relatively slow interconnect structures like Gigabit-Ethernet. This ensures applicability for high speed interconnects as well. Furthermore, it is assumed that memory is available to store the full set of T_1 and T_2 amplitudes locally on every node in fast memory and on hard disk. The storage of the amplitudes rarely becomes a bottleneck: If one assumes a molecule with 20 occupied and 600 virtual spin orbitals, which would correspond to a basis set of better than quadruple-zeta quality, then the number of T_2 amplitudes, which scales as occ^2vrt^2, would be of the order of a few hundred millions, which roughly corresponds to 1.5 GB of memory, if no symmetry is used.

In the actual algorithm, parts of intermediates or integrals will be contracted with the amplitudes on different nodes to give parts of the resulting quantity. In a final step, the amplitudes are updated and broadcast to all nodes. While allowing for distributed contractions during the CC iterations at minimal communication, this strategy has two drawbacks. Primarily, it does not allow for optimal scaling in the limit of a large number of processors as the distribution costs scale in the worst case linearly with the size of the distributed entity and the number of processors. The exact scaling behavior for the communication depends on the employed communication hardware and the algorithm used. Furthermore, it does not reduce the storage requirements for the replicated quantities like the T_2 amplitudes or the MO integrals excluding the four-virtual index integrals. The latter are treated in partial AO algorithms which eliminate the need for a full transformation of the two-electron integrals and only require storage of the AO integrals. Therefore, it is straightforward to calculate and store AO integrals in a distributed manner which are usually the largest quantity in terms of disk space in modern CC algorithms. Then the needed MO integrals are calculated once in a semi-parallel way and then are fully stored on each node (see 5.1.2 and 5.1.3). Together with the increased availability of large and cheap direct attached disk space the distributed storage of AO integrals makes it obsolete to recalculate or approximate these in every new step. At the same time, the efficiency of this algorithm is improving for increasing number of basis functions and/or correlated electrons: The time-determining steps can be more efficiently parallelized (see 5.2). However, calculations that are not feasible due to memory or disk space limitations (for the MO integrals) will also not be feasible when multiple processors are used.

For methods like CCSD or CCSD(T), communication costs associated with the replication of T_2-like quantities are usually at least two orders of magnitude smaller than the CPU time required for their computation. For CCSD(T), the scaling of CPU time is N^7 (occ^3vrt^4) while storage and communication costs grow as occ^2vrt^2 per compute node. Thus, the distribution of the time-determining steps to a number of processors in the way described above leads to a major reduction of overall execution time (walltime), especially when large calculations are considered. A detailed discussion and examples for the different aspects of this parallelization strategy will be given in the following subsections.

In subsection 5.1.1 the parallelization of the CCSD(T) perturbative triples for energies, gradients, and second derivatives starting from an implementation [65, 108, 109] that proves to be an ideal structure for the adaptation to parallel architectures will be described.

An analysis of the time-determining steps in CCSD energy, gradient and second derivative calculations is carried out in subsection 5.1.2 and the modification of the AO-based calculation of the leading term (the so-called particle-particle-ladder term that includes contraction over two virtual indices) is described.

In subsection 5.1.3 further issues for the optimization of the parallel code are described concerning the evaluation of two-electron integrals, the Hartee-Fock self-consistent-field (HF-SCF) procedure and the integral transformation. For the benchmark calculations correlation-consistent and correlation-consistent core valence Dunning basis sets [68, 76, 80] have been used throughout.

5.1.1 Parallel algorithm for the perturbative triples contributions to CCSD(T) energies, gradients, and second derivatives

The first step in the parallelization of the CCSD(T) scheme is to realize that almost any large CCSD(T) calculation is dominated by the evaluation of the perturbative triples contribution. In table 5.1 the timings for several standard serial CCSD(T) calculations are summarized. While the time-determining step for CCSD scales as occ^2vrt^4, the computational bottleneck of the perturbative triples correction scales as occ^3vrt^4. Thus, in comparison to CCSD the time spent for the (T) correction more rapidly increases with the number of electrons and this renders the computation of the perturbative triples correction the time-determining step in CCSD(T) calculations.

Table 5.1: Timings for the perturbative triples step in CCSD(T) energies relative to the total walltime of the CC part.

Molecule	Basis set	Number of electrons	Number of basis functions	Fraction of the perturbative triples step in a CCSD(T) calculation (%)
H_2O	cc-pCVTZ	10	115	13
H_2O	cc-pCVQZ	10	144	13
H_2O	cc-pCV5Z	10	218	13
Cl_2	cc-pCVTZ	34	118	52
Cl_2	cc-pCVQZ	34	218	52
Benzene	cc-pCVDZ	42	138	51
Benzene	cc-pCVTZ	42	354	57
Hexachlorobenzene	cc-pVDZ	168	192	60

The parallelization of the perturbative triples correction starts from the energy expressions as given in section 4.1.4

The basic scheme utilized in the serial ACES II MAB algorithm for the formation of the fourth and fifth-order T_3 amplitudes is an outer loop over an index triple i, j, k of the t_{ijk}^{abc} amplitudes. For energy calculations, for example, blocks of a, b, c index triples are calculated within the loop one at a time and immediately used to form the $E^{(4)}$ and $E^{(5)}$ energy contributions. In this way, storage of the full triples amplitudes is circumvented, as has also been reported on many occasions in the literature [92, 110, 111].

The same scheme is used in the parallel algorithm (compare fig. 5.1): Assuming that

the T_1 and T_2 amplitudes and the corresponding integrals are fully or at least partially locally available on all nodes, each node can independently form a, b, c energy contributions. Only a single number per node, namely the summed up energy contributions, has to be communicated. In a final step, the energy contributions from the i, j, k blocks are summed up to give the total energy correction. In this way, the parallelization of the (T) energy contributions can be achieved in a straightforward fashion.

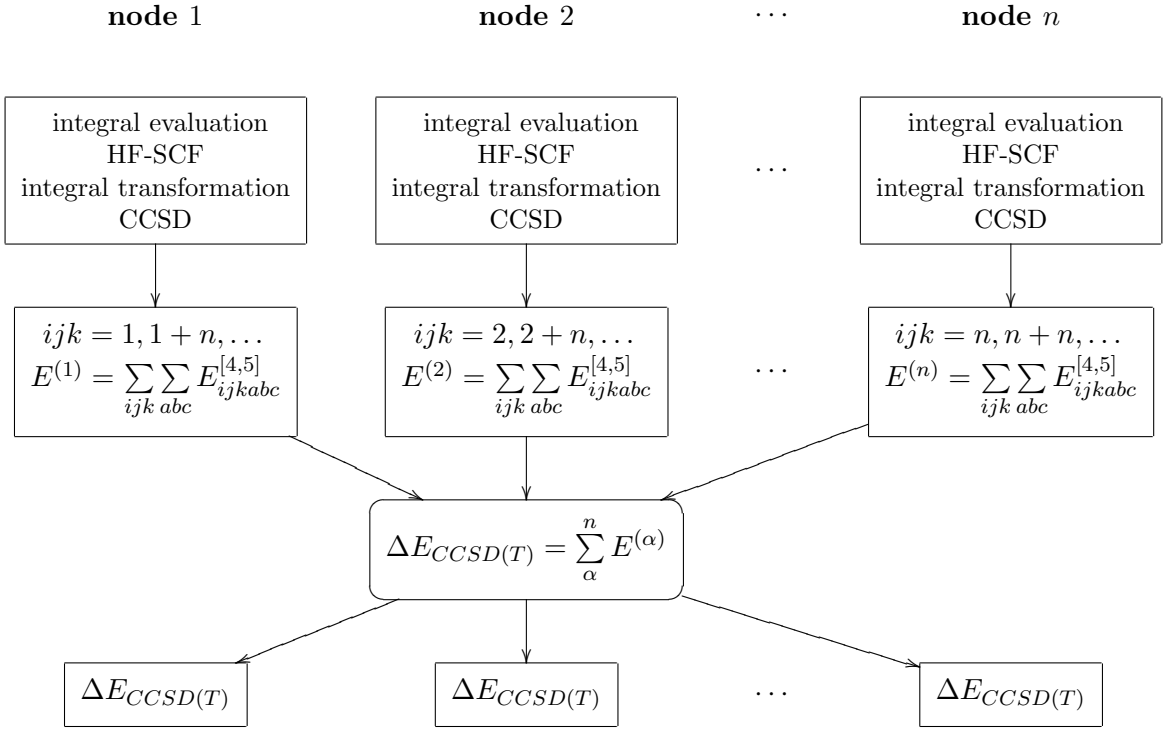

Figure 5.1: Graphical representation of the parallel calculation of the perturbative triples on n compute nodes.

For the first and second CC derivatives, Watts et al. [112] as well as Gauss and Stanton [65, 113] describe a CCSD(T) algorithm based on the idea of Lee and Rendell [114] to use perturbed canonical orbitals in order to avoid the re-computation of the triples amplitudes. The independent calculation of a, b, c blocks without storage of the full triple amplitudes can also be applied to derivative calculations. Here, an outer loop of an index triple i, j, k is also present and after the formation of an a, b, c block of T_3 amplitudes within the loop not only the energy increment, but also the contributions to the unperturbed effective one-particle density matrices, the two-particle density matrices and the contributions to the inhomogeneous terms of the Λ equations have to be calculated from the triple amplitudes. Furthermore, the same loop structure can be used afterwards to construct the perturbed triple amplitudes $\frac{\partial t^{abc}_{ijk}}{\partial x}$ and $\frac{\partial \tilde{t}^{abc}_{ijk}}{\partial x}$,[1] the corresponding contributions to the perturbed density matrices and terms of the perturbed Λ equation for each perturbation [65, 113]. Thus, the perturbative triple amplitudes, the perturbed amplitudes and all quantities to which they contribute can be calculated. In parallel, again, a, b, c blocks of amplitudes can be calculated and contracted independently into intermediate quantities that are held on each node and only need to

[1] The quantity $\tilde{t}^{abc}_{ijk}$ arises as a disconnected term in the fifth-order energy correction of the perturbative triples correction.

be communicated once, namely after all i, j, k index triples are processed. However, while for energy calculations only one double-precision quantity needs to be communicated, for gradients and second derivatives the corresponding contributions to the two-particle density matrices must be exchanged and summed as well.

To illustrate the efficiency of this scheme, energy calculations, geometry optimizations, and calculations of NMR chemical shifts have been carried out for the benzene molecule using Dunning's correlation consistent core/valence basis sets. All calculations for the benzene molecule where carried out at the all-electron CCSD(T)/cc-pVQZ geometry (r_{CH}=1.0800 Å, r_{CC}=1.3911 Å), which was taken from ref. [115]. The timings for the perturbative triples part of the algorithm are displayed in table 5.2. Due to machine load and other influences the timings can be assumed to be accurate to a few seconds walltime.

Table 5.2: Timings for the parallel perturbative triples step in CCSD(T) energies, one step of the geometry optimization and the calculation of NMR chemical shifts as analytical second derivatives for the benzene molecule (walltime in seconds). The number of basis functions is given in parentheses.

	number of nodes	1	2	4	8	16
cc-pCVDZ (138)	Energy	75	40	19	9	5
	Geometry	251	126	63	31	16
	NMR shieldings	2936	1452	727	363	192
cc-pCVTZ (342)	Energy	2163	1081	540	269	146
	Geometry	6594	3241	1619	809	426
	NMR shieldings	80779	40285	20171	10527	5171
cc-pCVQZ (684)	Energy	29019	14514	7225	3592	1758
	Geometry	82171	40999	20425	10207	5489
cc-pCV5Z (1200)	Energy	238882	119469	59764	29895	15184

The timings for the perturbative energy correction in the CCSD(T) algorithm scale almost perfectly up to 16 processors, even for the smallest basis set. It should be noted that the communication time required for the distribution of intermediate quantities calculated in the a, b, c loop of the perturbative triples is typically on the order of a few minutes at most, using Gigabit Ethernet interconnection. This is the case even for the largest examples and the largest numbers of nodes tested. In contrast to this, the time required for the parallel computation of the triples quantities themselves is usually of the order of hours for these examples.

By using this straightforward scheme for the adaptation of the serial code to cluster architectures, the overall time for the most CPU-time intensive steps in CCSD(T) calculations can be scaled down efficiently. However, the effort for the underlying CCSD calculation that precedes the calculation of the perturbative triples has not been discussed so far, but now becomes the dominant step in the overall execution time. The next subsection focuses on this issue.

5.1.2 Analysis and parallelization of time-determining steps in the CCSD energy, gradient, and second-derivative calculations

From the previous considerations it becomes obvious that the parallelization of the (T) step in large-scale CCSD(T) calculations allows significant reduction of the overall computation time up to a certain point. Increasing the number of nodes, however, does not lead to a further gain in execution time, if the effort for the underlying serial CCSD calculation exceeds the time for the parallel calculation of the perturbative triples. Thus, the next meaningful step in the parallelization of the CCSD(T) method is to identify and parallelize bottlenecks in the CCSD algorithm. The time-determining steps in a CCSD energy calculation are the so-called particle-particle ladder terms that scale as occ^2vrt^4 [116] (compare eq. (4.37)[1]):

$$t_{ij}^{ab} D_{ij}^{ab} \leftarrow \frac{1}{2} \sum_{ef} \tau_{ij}^{ef} \langle ab| |ef\rangle \tag{5.1}$$

where the intermediate

$$\tau_{ij}^{ab} = t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a \tag{5.2}$$

is used. Here and in the following paragraphs $A \leftarrow B$ denotes that B contributes to the determination of A.
It should be noted that for CC energy and derivative calculations terms including $\langle ab| |cd\rangle$ integrals or corresponding integral derivatives can, in general, be identified as the computationally dominating contributions. For large basis sets the quartic dependence on the number of virtual indices will usually render this contraction expensive in terms of computational time.

One problem of the formulation in eq. (5.1) is that the molecular-orbital integrals always represent a storage bottleneck due to their lack of sparsity. As a consequence, the common practice in modern CC programs is an AO integral-based algorithm in which the corresponding amplitudes are first partially transformed to the AO basis in a N^5 procedure

$$\tau_{ij}^{\mu\nu} = \sum_{ef} c_{\mu e} c_{\nu f} \tau_{ij}^{ef} \tag{5.3}$$

and then contracted with the AO integrals:

$$Z_{ij}^{\mu\nu} = \sum_{\sigma\rho} \langle \mu\nu|\sigma\rho\rangle \, \tau_{ij}^{\sigma\rho}. \tag{5.4}$$

Afterwards, the resulting intermediate will be transformed back and processed in the MO basis [117–123] as follows:

$$Z_{ij}^{ab} = \sum_{\mu\nu} c_{\mu a} c_{\nu b} Z_{ij}^{\mu\nu}. \tag{5.5}$$

This has the advantage that only the more sparse AO integrals need to be stored and processed.

While the operation count of the time-determining step scales as occ^2ao^4 in this scheme instead of occ^2vrt^4, in practice the smaller number of non-zero AO integrals leads to a significantly reduced I/O and operation count. Thus, in almost all relevant cases the AO based

[1] In practice the factorization of the CC equations leads to additional terms that are included in this contraction.

algorithm outperforms the straightforward MO based scheme. Realizing that parallelizing this single contribution will lead to a major reduction of overall time in each CCSD iteration, the AO based term has been chosen as a starting point for improving on the parallel CCSD(T) code.

The basic loop structure for which the power of multiple processors can be used effectively is an outer loop over batches of AO integrals that are read from disk and contracted with all matching T_2 amplitudes in the AO basis (c.f. eq. (5.4)). After the contraction has been carried out, the resulting Z_{ij}^{ab} intermediate is communicated and the CCSD iteration is continued.
It should be noted that the communication required after the corresponding intermediate has been formed scales at most as occ^2vrt^2 per compute node. It is expected that this step can be parallelized efficiently without running into communication bottlenecks.
Table 5.3 shows the timings of the CCSD iterations for the benzene molecule, where the AO ladder term eq. (4.38) has been parallelized in the described fashion.

Table 5.3: Time per CCSD iteration in comparison to the time spent on the particle-particle ladder term for the benzene molecule (walltime in seconds). The number of basis functions is given in parentheses.

	Number of nodes	1	2	4	8	16
cc-pCVDZ (138)	Time per CCSD iteration	5.2	3.9	3.0	2.5	2.4
	Time for AO ladder term	3.4	2.0	1.1	0.6	0.5
cc-pCVTZ (342)	Time per CCSD iteration	135	79	52	39	33
	Time for AO ladder term	113	57	31	16	9
cc-pCVQZ (684)	Time per CCSD iteration	1902	1027	584	365	257
	Time for AO ladder term	1761	885	445	226	115

As can be seen in the first column, the calculation of the AO-based particle-particle ladder term dominates the time for one iteration, even for the smallest examples by more than 60 percent. Furthermore, the parallelization of this term in batches of AO integrals results in an almost perfect reduction of the walltime for this contribution up to 16 processors and, thus, to a significant reduction of the overall time per iteration, especially for larger examples.

For the calculation of analytic gradients terms analogous to those in energy calculations appear in the CCSD Λ equations [116]:

$$\lambda_{ab}^{ij} D_{ij}^{ab} \leftarrow \frac{1}{2} \sum_{ef} \lambda_{ef}^{ij} \langle ef | | ab \rangle \,. \tag{5.6}$$

Within the ACES II MAB program package this term is calculated employing the same AO integral-based scheme as for the CCSD equations. Thus, the time-determining N^6 step in the gradient calculations can be parallelized in the same way as the corresponding term in the energy calculation.

For second-derivative calculations, the contributions that have to be considered occur in

the equations for the perturbed amplitudes and perturbed Λ :

$$\frac{\partial t_{ij}^{ab}}{\partial \chi} D_{ij}^{ab} \leftarrow \frac{1}{2} \sum_{ef} \frac{\partial t_{ij}^{ef}}{\partial \chi} \langle ef | | ab \rangle \tag{5.7}$$

$$\frac{\partial \lambda_{ab}^{ij}}{\partial \chi} D_{ij}^{ab} \leftarrow \frac{1}{2} \sum_{ef} \frac{\partial \lambda_{ef}^{ij}}{\partial \chi} \langle ef | | ab \rangle \,. \tag{5.8}$$

By parallelizing these contributions in the AO based scheme, the overall computational cost of the most time-consuming steps in the CCSD gradient and analytical second-derivative calculations can be reduced as well. Due to the dominance of these steps compared to the overall time per CCSD iteration this simple strategy improves the overall CCSD time significantly if multiple processors are used.
Table 5.4 summarizes the timings for the corresponding modules that include the steps described above for calculations of NMR chemical shifts for the benzene molecule.

Table 5.4: Timings for the solution of the lambda equations and the solution of the perturbed amplitude and lambda equations for the benzene molecule (walltime in seconds).

	Number of nodes	1	2	4	8	16
Lambda equations	cc-pCVDZ	65	48	40	35	34
	cc-pCVTZ	1819	1147	828	653	603
	cc-pCVQZ	25166	14808	9510	6887	5854
Perturbed amplitude and lambda equations	cc-pCVDZ	607	459	390	345	331
	cc-pCVTZ	18526	12190	9255	7972	6022

From the last columns of tables 5.4 and 5.2 it becomes clear that for this example the overall time required for the CCSD and CCSD(T) derivative equations is of the same order of magnitude as the time for the evaluation of the perturbative triples part, if 16 processors are used. Thus, the CCSD part of the calculation will still dominate the overall time for CCSD(T) derivative calculations if more than 16 processors are used. This is mainly due to contributions that have not been parallelized. These are of lower scaling or related with the transformation of integral derivatives.

So far only the particle-particle ladder terms for the CCSD, Λ, perturbed amplitudes and perturbed Λ equations are parallelized. To further improve the algorithm, one could utilize parallel matrix multiplication routines for CCSD contributions that scale as occ^3vrt^3. This pushes the crossover point from which the preceding CCSD calculation will dominate over the CCSD(T) calculation time to larger number of processors for larger examples. For larger molecules with medium-sized basis sets this crossover point should shift to 32 or even 64 processors so that large-scale cluster architectures could be used to carry out calculations that would take months on a single processor within days.

5.1.3 Further optimization issues

The scheme for parallelizing the AO-ladder terms described in subsection 5.1.2 requires only equally distributed integrals to be present on the different compute nodes. As a consequence,

the evaluation of the integrals is carried out in parallel and, in turn, used for a parallel HF-SCF procedure and integral transformation. Each node calculates and stores only a part of the integrals, and thus during the HF-SCF procedure only an incomplete Fock-matrix is built on each node, which is then exchanged between the compute nodes. The total Fock matrix is simply the sum of these incomplete matrices. The rest of the algorithm is unaltered. For the integral transformation all AO integrals are read in only once and communicated in the form of an intermediate array. After transformation of the MO integrals with two and three virtual indices each node stores all calculated MO integrals locally. The timings (in seconds) for parallel integral evaluation, HF-SCF and integral transformation for the benzene molecule are shown in table 5.5. This simple scheme does not only result in a reduced storage requirement per node but also in a reduction of the overall time for the integral evaluation, transformation and HF-SCF execution time. It should be noted, that for some cases, even super linear scaling of the evaluation of the HF-SCF energy can be observed. This is due cache effects resulting from the different memory hierarchies of a modern computer: In parallel computing, not only the numbers of processors change, so does the size of the accumulated caches or input/output (I/O) buffers, more or even all data can fit into caches and the memory access time reduce dramatically, which causes the extra speedup in addition to that from the actual computation [124].

Table 5.5: Timings for the treatment of two-electron integrals, the HF-SCF and the integral transformation for the benzene molecule (walltime in seconds). The number of basis functions is given in parentheses.

	Number of nodes	1	2	4	8	16
cc-pCVDZ (138)	Integral evaluation	8.5	4.6	2.2	1.1	0.7
	HF-SCF	1.8	1.1	0.6	0.4	0.3
	Integral transformation	1.8	1.4	1.3	1.1	1.3
cc-pCVTZ (342)	Integral evaluation	340	167	90	45	24
	HF-SCF	57	30	16	9	5
	Integral transformation	61	37	29	23	21
cc-pCVQZ (684)	Integral evaluation	19379	10012	4906	2518	1505
	HF-SCF	1009	453	225	122	67
	Integral transformation	1096	601	399	267	233

An important issue in parallel implementations is to avoid load-balancing problems. In the work presented here, HF-SCF, the integral transformation, and any CCSD-like equations are automatically balanced by the local calculation and storage of equally sized amounts of two-electron integrals at the beginning of the calculation. The remaining steps in the calculation of the perturbative triples are balanced on average by the large number of these contributions. This applies for the calculation of energies, gradients, and any second-order properties. In practice, even multiprocessor systems do not show balancing problems since the load is kept equal on every processor. For the actual implementation we assume that dedicated nodes are available. However, load-balancing problems will arise if heterogeneous resources are used or if compute nodes have different loads due to other calculations. This issue will have to be addressed in further developments of the current algorithm.

In addition, for the special case of analytic second derivatives, one could also use a coarse-grained parallelization scheme [125] on top of the one suggested here. Namely, one could distribute the different perturbations that are calculated independently, for example B_x, B_y and B_z for NMR chemical shifts or geometric perturbations in the case of harmonic frequencies, to different processors. While this has not been carried out in the approach presented here, it is a straightforward addition to any code that could further help to improve the scaling of the algorithm with the number of processors.

5.2 Results and discussion

Nowadays more than 500 GB of disk space and 16 GB of RAM are readily available even on single cluster nodes within medium sized computer clusters, so it is not foreseeable that memory or storage will present a bottleneck for large-scale calculations. As has been noted before, only serial calculation times of the order of month or years will render large-scale CCSD(T) calculations unfeasible. While a parallel implementation cannot combat the steep scaling of high-level CC methods, the power of parallel computer architectures expands the range of applicability dramatically.

In table 5.6 the results of some representative benchmark calculations for energies and gradients are summarized. The timings for CCSD(T) energy calculations for benzene and cyclohexene and the timings of one step of the geometry optimization of the adamantyl cation ($C_{10}H_{15}^+$) are given.

Table 5.6: Overall timings for the CCSD(T) energy calculations of benzene and cyclohexene and for one step of the geometry optimization of the adamantyl cation.

Molecule	Comput. symmetry	Basis set	Number of basis functions	Number of nodes	Execution time [h]
Benzene [a]	D_{2h}	cc-pV5Z	876	16	4
Benzene [a]	D_{2h}	cc-pV6Z	1386	16	21
Cyclohexene[a,b]	C_2	aug-cc-pVQZ	940	16	40
Adamantyl cation	C_s	cc-pVTZ	510	9	90

[a] Energies (frozen core) for benzene cc-pV5Z, cc-pV6Z and cyclohexene are -231.8916163, -231.8987752 and -234.2797258 Hartree; [b] carried out at the (fc)MP2/cc-pVTZ geometry.

For the high-symmetry case benzene in a hextuple-zeta basis set the serial energy calculation would take about two weeks and is reduced to less than a day using 16 processors. From table 5.6 it also becomes obvious that the number of basis functions is not the only factor when considering the size of a system, but also symmetry and the distribution of orbitals among the irreducible representations, as well as the ratio of occupied to virtual orbitals.

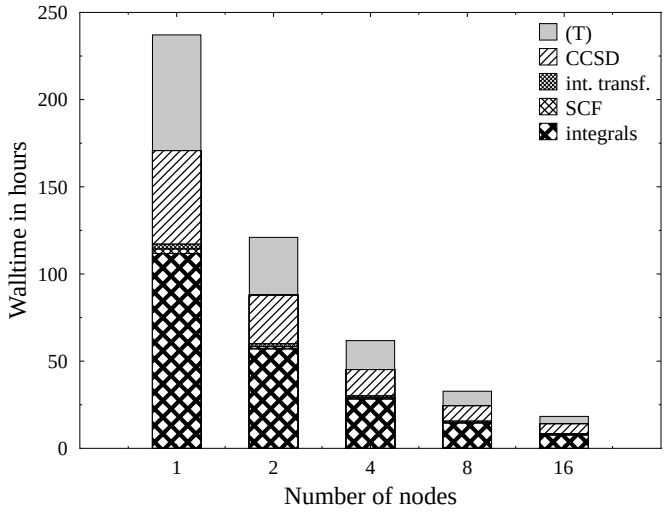

Figure 5.2: Composition of the overall time for one step in the CCSD(T) energy calculation of the benzene molecule at the cc-pCV5Z level of theory (1200 basis functions)

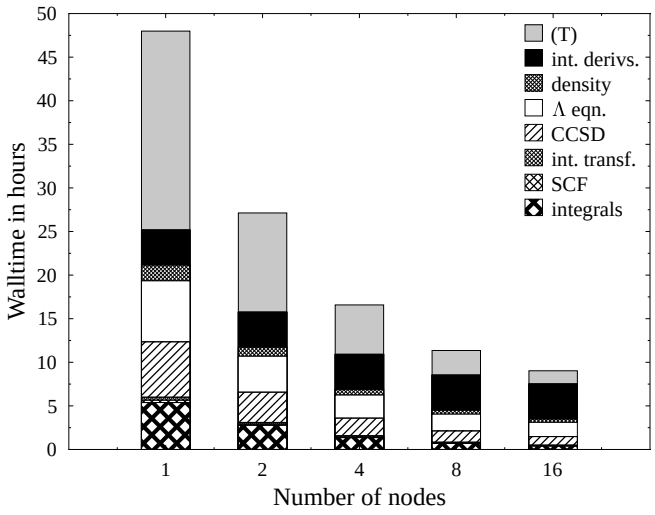

Figure 5.3: Composition of the overall walltime for one step in the geometry optimization of the benzene molecule at the CCSD(T)/cc-pCVQZ level of theory (684 basis functions)

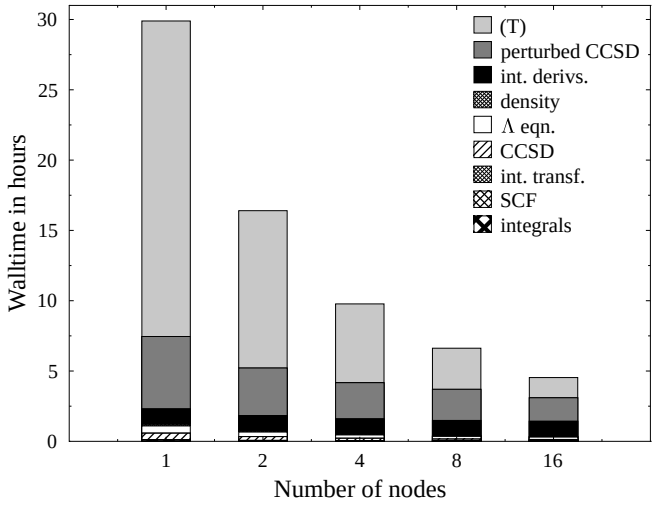

Figure 5.4: Composition of the overall walltime for the calculation of the NMR chemical shifts of the benzene molecule at the CCSD(T)/cc-pCVTZ level of theory (342 basis functions)

Figures 5.2, 5.3, and 5.4 give a more detailed view on the different steps required for three smaller CCSD(T) calculations, namely for the energy, the gradient for one step of a geometry optimization and the calculation of NMR chemical shifts for the benzene molecule.

For the gradient calculation, which takes about 47 hours on a single processor, it can be seen that the calculation of the perturbative triples contribution takes only about half of the overall time of the optimization step. Using the current algorithm it is possible to scale down this contribution and also the calculation of the two-electron integrals, the CCSD and Λ equations as well as the integral transformation and the contributions to the density matrices. Using 16 processors the overall time is reduced to less than ten hours. At this point steps that have not been considered for parallelization in the algorithm dominate the overall time, so that the usage of a larger numbers of nodes would not yield significant further speedups.

The calculation of the NMR chemical shifts with a larger basis set shows a different profile. Here the perturbative triples contributions to the second derivatives clearly dominate compared to the other steps, such as the SCF or CCSD calculations. Using 16 processors the overall time of 30 hours can be reduced to less than 5 hours. After this point, the remaining steps in the perturbed CCSD equations that have not been parallelized dominate the overall calculation time.

Figures 5.5 and 5.6 give a detailed insight for the scaling of CCSD(T) energy and derivative calculations wall time. The speedup is defined by: $S_p = \frac{t_1}{t_p}$, where p is the number of processors, t_1 is the execution time of the sequential algorithm and t_p is the execution time of the parallel algorithm using p processors.

As has been discussed in the previous subsections, the scaling of computing time with the number of nodes is improving for increasing system size as the importance of the parallelized, time-determining steps is even larger.

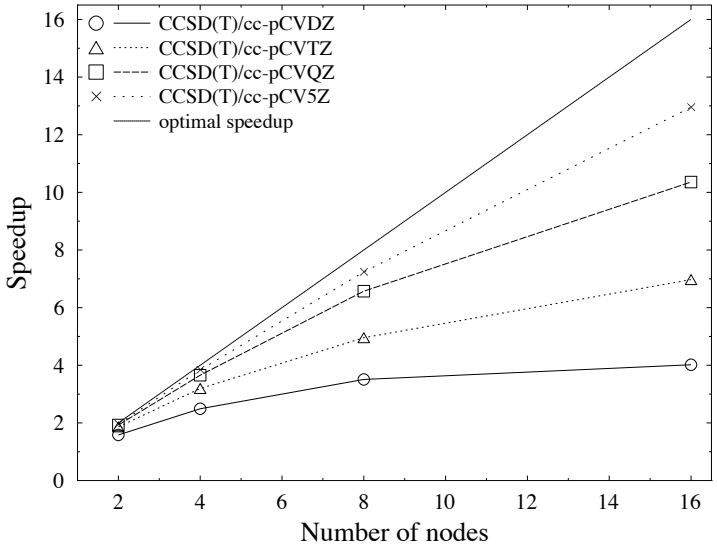

Figure 5.5: Parallel scaling of CCSD(T) energy calculations for the benzene molecule using different basis sets

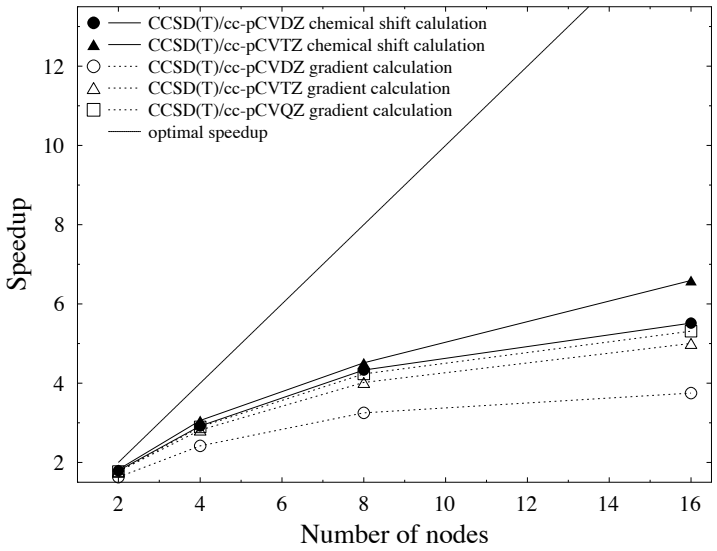

Figure 5.6: Parallel scaling of CCSD(T) first and second analytical energy derivative calculations for the benzene molecule using different basis sets

5.3 Conclusions

A detailed analysis of the time-determining steps in CC energy, gradient, and second derivative calculations shows that for almost all practical applications only a few terms completely dominate the overall computation time. This motivates a straightforward strategy for the parallelization of CCSD and CCSD(T) energies, gradients and second derivatives that has been outlined and implemented in this thesis. Starting from the highly efficient serial implementation of the ACES II MAB computer code an adaptation for affordable workstation clusters has been obtained by parallelizing the most time-consuming steps of the algorithm. This also includes the calculation of parallel CCSD energies, gradients, and second derivatives using unrestricted (UHF) and restricted open shell (ROHF) Hartree-Fock references, parallel UHF-CCSD(T) energies, gradients, and second derivatives, parallel ROHF-CCSD(T) energies, as well as parallel equation-of-motion CCSD energies and gradients for closed and open shell references. Furthermore virtually the same scheme as for the perturbative triples in CCSD(T) has been applied to the calculations of CCSDT(Q) energies [126] using a pilot implementation in ACES II MAB. Based on the experiences of the presented study the MRCC program of M. Kállay [94, 127] has been modified for an OpenMP[1] as well as a MPI parallel implementation for parallel generalized coupled-cluster energy, gradient and second-derivative calculations [128]

The central aspect of the implementation presented here are the replication of the cluster amplitudes and the distributed evaluation, storage, and access of the two-electron integrals to arrive at an algorithm for which sufficient local memory and disk space are necessary but which is not dependent on sophisticated high-speed network connections.
Benchmark calculations for systems with up to 1300 basis functions show that the resulting algorithm for energies, gradients and second derivatives at the CCSD and CCSD(T) level of theory exhibits good scaling with the number of processors as long as the terms that are the time-determining steps in the serial calculation still dominate the overall time in the parallel computation. It is important to note that the communication steps within the algorithm are at no point a limiting factor in the current implementation, even if 16 or more processors are used. Nevertheless, at larger numbers of nodes the algorithm will break down, as steps in the CCSD algorithm that have not been parallelized prevent a better scaling of the overall execution time, especially for small systems and large number of nodes. The current limitation of the parallel implementation becomes obvious for more than 16 processors. However, an analysis of the algorithm leads us to the conclusion that the scaling behavior is much better for larger examples, where the time-determining steps that have been parallelized dominate the overall execution time more strongly.
If a very rough estimate is allowed at this point - implementations of this kind would open the field of application for the CC hierarchy of high accuracy *ab initio* methods to systems of about 30 atoms in a triple-zeta basis or about 15 atoms in a quadruple-zeta basis.
Typical applications will be presented in the next chapters like high accuracy calculations for structures and energies, vibrational frequencies or properties related to NMR spectroscopy and state-of-the-art spectroscopy.

[1] http://www.openmp.org

6 Computational thermochemistry

Significant progress has been achieved recently in the field of theoretical high-accuracy thermochemistry [14, 15, 129–132]. The calculation of thermodynamic properties (heats of formation, total atomization energies, ionization potentials, and bond energies) with better than chemical accuracy (1 kcal/mol $\sim$ 4 kJ/mol) has become possible due to

- methodological developments enabling the routine inclusion of higher excitations in electron-correlated quantum-chemical calculations [127],
- extrapolation techniques for estimating the basis-set limit [74, 75] in quantum-chemical calculations,
- the possibility to routinely account for anharmonic effects in the calculation of zero-point vibrational energy corrections using second-order vibrational perturbation theory (VPT2) (see for example [133–136]),
- the availability of reliable experimental data, obtained within the active thermochemical tables (ATcT) of Ruscic and co-workers [137], for the calibration of the theoretical schemes.

In this way, protocols such as HEAT [14, 15], W3 [130] and W4 [131], as well as the focal-point analysis [132] have been developed and are capable of providing thermochemical energies (i.e., atomization energies or heats of formation) with an accuracy of 1 kJ/mol or even better.

6.1 High accuracy extrapolated ab initio thermochemistry

According to the thermochemical **H**igh accuracy **E**xtrapolated **A**b initio **T**hermochemistry (HEAT) [14,15] protocols, the total energy of a molecule can be obtained within the following additivity assumption

$$\begin{aligned} E &= E^{\infty}_{\mathrm{HF}} + \Delta E^{\infty}_{\mathrm{CCSD(T)}} + \Delta E_{\mathrm{CCSDT}} + \Delta E_{\mathrm{HLC}} \\ &\quad + \Delta E_{\mathrm{REL}} + \Delta E_{\mathrm{SO}} + \Delta E_{\mathrm{DBOC}} + \Delta E_{\mathrm{ZPE}} \end{aligned} \tag{6.1}$$

with E^{∞}_{HF} as the basis-set limit value for the Hartree-Fock energy, $\Delta E^{\infty}_{\mathrm{CCSD(T)}}$ the basis-set limit for the correlation energy obtained at the CCSD(T) level [4], $\Delta E_{\mathrm{CCSDT}}$ and ΔE_{HLC} accounting for additional correlation effects which are not covered by the CCSD(T) treatment. ΔE_{ZPE} denotes the contributions to the vibrational zero-point energies, ΔE_{REL} accounts for relativistic corrections, and ΔE_{DBOC} is the diagonal Born-Oppenheimer correction (DBOC) necessary due to the use of the Born-Oppenheimer approximation.

The following paragraphs describe in detail how the various terms in eq. (6.1) are obtained.

6.1.1 Molecular geometries

All calculations have been carried out with geometries optimized at the CCSD(T) level using the cc-pVQZ basis set. The geometries are obtained in calculations with all electrons correlated and are listed in ref. [14].

6.1.2 HF and CCSD(T) energy

The sum of E^{∞}_{HF} and $\Delta E^{\infty}_{\mathrm{CCSD(T)}}$ provides an estimate for the non-relativistic electronic energy within the Born-Oppenheimer approximation using the CCSD(T) method for the treatment of electron correlation. Following common practice, these two energy contributions have been obtained separately using basis-set extrapolation techniques [75] based on energies obtained with hierarchical series of basis sets. For the HF energy, calculations are carried out for this purpose with the augmented correlation-consistent polarized core-valence basis-sets aug-cc-pCVXZ. The corresponding basis-set limit is then obtained with the following empirical extrapolation formula advocated by Feller [74]

$$E^{X}_{\mathrm{HF}} = E^{\infty}_{\mathrm{HF}} + a\exp(-bX), \tag{6.2}$$

where E^{X}_{HF} is the HF-SCF energy obtained with the aug-cc-pCVXZ basis-set. The parameters a and b as well as the extrapolated HF-SCF energy E^{∞}_{HF} are determined uniquely from three energies.

For the correlation energy, a formula derived from the atomic partial-wave expansion is used for the extrapolation to the basis-set limit [75],

$$\Delta E^{X}_{\mathrm{CCSD(T)}} = \Delta E^{\infty}_{\mathrm{CCSD(T)}} + \frac{a}{X^3}. \tag{6.3}$$

In eq. (6.3), $\Delta E^{X}_{\mathrm{CCSD(T)}}$ is the CCSD(T) correlation energy obtained with the aug-cc-pCVXZ basis set. The parameter a and the estimated basis-set limit $\Delta E^{\infty}_{\mathrm{CCSD(T)}}$ are determined from two calculations. Different from the Wn schemes [129–131], the correlation energy is not divided in a valence and a core-correlation energy part and instead both contributions are treated up to the CCSD(T) level together. For the open-shell systems, all calculations have been performed using an unrestricted Hartree-Fock (UHF) reference.

6.1.3 Higher-level correlation effects

It should not be forgotten that the (T) correction is based on perturbation-theory arguments [4, 138]. Thus, when aiming at high accuracy in cases where triples corrections are large, it appears necessary to investigate correlation effects beyond CCSD(T) and, for example, to ask to what extent CCSD(T) results differ from those obtained from a full treatment of triple excitations with the CCSDT model. Unfortunately, it is usually not possible to perform the required CCSDT calculations with the same large basis sets as the CCSD(T) computations, and this is the major reason for separating the correlation energy in a CCSD(T) contribution and those due to treatments beyond CCSD(T). This separation is well justified, as all contributions beyond CCSD(T) are expected to be rather small and the largest contribution to the correlation energy is covered at the CCSD(T) level. It thus seems to be

sufficient to estimate these higher-order contributions using smaller basis sets and by taking into account valence correlation only. The difference between CCSDT and CCSD(T) is estimated by extrapolating the corresponding energy difference

$$\Delta E_{\mathrm{CCSDT}} = E^{\mathrm{TQ}}_{\mathrm{CCSDT}}(\mathrm{fc}) - E^{\mathrm{TQ}}_{\mathrm{CCSD(T)}}(\mathrm{fc}) \tag{6.4}$$

where TQ denotes that the corresponding contribution has been obtained by extrapolating the frozen-core CCSDT and CCSD(T) energies obtained with cc-pVTZ and cc-pVQZ basis sets. fc indicates in eq. (6.4) that the calculations have been performed in the frozen-core approximation. Open-shell systems are treated using again a UHF reference function.

Recent investigations [14,15,130,131,139] have convincingly demonstrated that the CCSDT level is not sufficient for reaching kJ/mol accuracy. Due to the implementation of general CC models [47, 127, 140], CC calculations beyond CCSDT are nowadays routinely possible though still very demanding. Inclusion of quadruple excitations are feasible due to the fact that even small basis-set calculations (at the cc-pVDZ level, for example) provide realistic estimates of its importance. In addition, the availability of the CCSDT(Q) scheme [126, 141] in which quadruple excitations are treated in a perturbative manner, similar to that of triple excitations in CCSD(T), increases the range of application for this kind of studies. The higher-order correlation contributions beyond CCSDT can be estimated via

$$\Delta E_{\mathrm{HLC}} = E^{\mathrm{cc-pVDZ}}_{\mathrm{CCSDT(Q)}}(\mathrm{fc}) - E^{\mathrm{cc-pVDZ}}_{\mathrm{CCSDT}}(\mathrm{fc}), \tag{6.5}$$

$$\Delta E_{\mathrm{HLC}} = E^{\mathrm{cc-pVDZ}}_{\mathrm{CCSDTQ}}(\mathrm{fc}) - E^{\mathrm{cc-pVDZ}}_{\mathrm{CCSDT}}(\mathrm{fc}), \tag{6.6}$$

or

$$\Delta E_{\mathrm{HLC}} = E^{\mathrm{cc-pVDZ}}_{\mathrm{CCSDTQP}}(\mathrm{fc}) - E^{\mathrm{cc-pVDZ}}_{\mathrm{CCSDT}}(\mathrm{fc}). \tag{6.7}$$

6.1.4 Zero-point vibrational energies

Zero-point vibrational energy (ZPE) contributions are obtained via a second-order perturbation theory (VPT2) treatment [142].

$$\Delta E_{\mathrm{ZPE}} = \sum_i \frac{\omega_i}{2} + G_0 + \frac{1}{4}\sum_{i<j} x_{ij}. \tag{6.8}$$

with the required force fields computed at the all-electron CCSD(T)/cc-pVQZ level. In eq. (6.8), the sums run over all normal modes i with ω_i denoting the harmonic vibrational frequencies and x_{ij} the corresponding anharmonicity constants. Expressions for the latter can be found, for example, in ref. [142]. To avoid resonance contributions to ΔE_{ZPE}, it is necessary to include the so-called G_0 contribution, see refs. [15, 143–146].

6.1.5 Diagonal Born-Oppenheimer correction

As recent studies [147, 148] indicated that electron-correlation effects have only a modest impact on the diagonal Born-Oppenheimer (DBO) correction, this correction, which accounts for errors due to the Born-Oppenheimer approximation, is usually computed at the HF-SCF

level [149]. Calculations within the HEAT schemes are performed using the HF-SCF/aug-cc-pVTZ and in order to ensure that electron-correlation effects are marginal the CCSD/aug-cc-pCVQZ level of theory [147] is applied, thereby using an UHF reference for open-shell systems.

6.1.6 Relativistic effects

Scalar-relativistic effects on the total energy (ΔE_{REL}) are included by computing the corresponding corrections in a perturbative manner with the mass-velocity and Darwin one-electron terms as perturbation [150]. As discussed, for example, by Davidson et al. [151] this is a reasonable approximation as long as molecules with light elements only are considered, as it is the case in the present study. Calculations have been performed at the CCSD(T) level using the aug-cc-pCVTZ basis and an UHF ansatz for the open-shell systems. This treatment was extended to include the corresponding two-electron Darwin term [152,153] referred to as MVD2 in the following and is used in the improved HEAT schemes [15,148].

6.1.7 Overview and status

Table 6.1 gives an overview of the different HEAT approaches. In ref. [15] the following naming scheme has been defined:

HEAT**IJK**-HLC, where the cardinal numbers **IJK** denote the aug-cc-pCVXZ (X=**I,J,K**) basis used in the HF-SCF energies. The aug-cc-pCVXZ (X=**J, K**) basis sets are used in the extrapolation of the CCSD(T) correlation energy. HLC indicates the treatment for correlation treatment beyond CCSDT employing the cc-pVDZ basis and the frozen-core approximation. Using this naming scheme HEAT-345(Q) denotes that HF-SCF calculations are performed with aug-cc-pCVTZ, aug-cc-pCVQZ, and aug-cc-pCV5Z. The latter two basis sets are used in the all-electron CCSD(T) calculations, while higher level correlation effects are estimated at the CCSDT(Q) level of theory. All other terms of eq. (6.1) are treated like presented in table 6.1.

Table 6.1: Summary of the various contributions to the total energy in different HEAT schemes.

Term of eq. (6.1)	HEAT	HEAT345-(Q)	HEAT345-Q	HEAT345-QP
E_{HF}^{∞}	HF-SCF/aug-cc-pCVXZ (X=3, 4, 5)			
$\Delta E_{\mathrm{CCSD(T)}}^{\infty}$	CCSD(T)/aug-cc-pCVXZ (X=4, 5)			
$\Delta E_{\mathrm{CCSDT}}$	(fc)CCSDT/cc-pCVXZ (X=3, 4)			
ΔE_{HLC}	(fc)CCSDTQ/ cc-pVDZ[a]	(fc)CCSDT(Q)/ cc-pVDZ	(fc)CCSDTQ/ cc-pVDZ	(fc)CCSDTQP/ cc-pVDZ
ΔE_{Rel}	MVD1	MVD2 CCSD(T)/aug-cc-pCVTZ		
ΔE_{SO}	(fc)RECP-SO-CISD/cc-pVDZ			
ΔE_{DBOC}	HF-SCF/aug-cc-pVTZ	(ae)CCSD/aug-cc-pCVQZ		
ΔE_{ZPE}	(ae)CCSD(T)/cc-pVQZ[b]	(ae)CCSD(T)/cc-pVQZ		

[a]ROHF reference; [b]the G_0 contribution is neglected.

6.2 Improvements and overview for the HEAT schemes

Table 6.2 shows the results of the various contributions presented in table 6.1. Values are given as difference to the total atomization energies (TAEs) from the ATcT. The second column gives the results of the original HEAT protocol. The following four columns show the effect of adding individually the contributions arising from the G_0 contribution (Δ_{G_0}), the change from ROHF to UHF reference in the estimate for the quadruple excitations (Δ_{RU}), the two-electron Darwin contribution (Δ_{D2}) and the electron-correlation effects on the DBOC (ΔE_{DBOC}^{CCSD}). HEAT345-Q indicates the inclusion of all of these contributions (compare table 6.1).

Table 6.2: Effects of adding individually the contributions arising from the G_0 term (Δ_{G_0}), the change from ROHF to UHF reference in the estimate for the quadruple excitations (Δ_{RU}), the two-electron Darwin contribution (Δ_{D2}) and the electron-correlation effects on the DBOC (ΔE_{DBOC}^{CCSD}). Errors (in kJ/mol) in atomization energies, compared with ATcT values based on the Core (Argonne) Thermochemical Network, C(A)TN, ver. 1.064.

	ATcT[a]	HEAT	$+\Delta_{G_0}$[b]	$+\Delta_{RU}$[b]	$+\Delta_{D2}$[b]	$+\Delta E_{DBOC}^{CCSD}$[c]	HEAT345-Q
N_2	941.14 ± 0.03	-0.07	-0.07	-0.07	0.01	-0.12	-0.05
H_2	432.07 ± 0.00	0.39	0.29	0.39	0.40	0.23	0.14
F_2	154.51 ± 0.13	-0.07	-0.06	-0.06	-0.04	-0.08	-0.04
CO	1072.13 ± 0.09	-0.32	-0.32	-0.31	-0.28	-0.36	-0.32
O_2	493.69 ± 0.00	-0.34	-0.34	-0.66	-0.32	-0.38	-0.69
C_2H_2	1626.16 ± 0.24	-0.21	-0.37	-0.21	-0.23	-0.38	-0.55
CCH	1075.05 ± 0.25	0.15	0.01	0.16	0.12	0.02	-0.15
CF	545.61 ± 0.63	-0.12	-0.12	-0.12	-0.10	-0.14	-0.13
CH_2	752.70 ± 0.26	-0.31	-0.19	-0.31	-0.33	-0.37	-0.27
CH	334.66 ± 0.23	-0.14	-0.16	-0.14	-0.13	-0.23	-0.26
CH_3	1209.63 ± 0.13	0.30	-0.03	0.30	0.29	0.17	-0.16
CN	745.52 ± 0.25	-1.26	-1.26	-1.18	-1.23	-1.29	-1.20
CO_2	1598.27 ± 0.09	-1.08	-0.24	-1.08	-1.06	-1.14	-0.27
H_2O_2	1055.23 ± 0.07	-0.43	-0.46	-0.43	-0.40	-0.59	-0.59
H_2O	917.83 ± 0.03	0.42	0.36	0.42	0.43	0.30	0.25
HCN	1268.26 ± 0.17	-0.04	0.00	-0.04	-0.02	-0.15	-0.09
HCO	1132.87 ± 0.19	-0.37	-0.21	-0.41	-0.37	-0.49	-0.34
HF	565.97 ± 0.01	0.46	0.41	0.46	0.46	0.40	0.36
HNO	823.65 ± 0.12	-0.71	-0.77	-0.71	-0.67	-0.88	-0.90
HO_2	694.46 ± 0.22	0.10	0.23	0.08	0.13	0.04	0.18
NH_2	713.32 ± 0.19	0.63	0.59	0.63	0.64	0.53	0.50
NH_3	1157.25 ± 0.04	0.28	0.29	0.28	0.30	0.14	0.16
NH	328.43 ± 0.29	-0.60	-0.63	-0.61	-0.59	-0.65	-0.68
NO	626.85 ± 0.08	-0.63	-0.64	-0.73	-0.59	-0.68	-0.73
OF	213.47 ± 0.47	0.09	0.09	0.07	0.12	0.07	0.08
OH	425.62 ± 0.03	0.52	0.48	0.52	0.53	0.46	0.43
$\bar{\Delta}$[d]		-0.13	-0.12	-0.14	-0.11	-0.21	-0.20
max		-1.26	-1.26	-1.18	-1.23	-1.29	-1.20
Δ_{std}^{d}		0.48	0.42	0.49	0.47	0.47	0.42

[a] data taken from ref. [148]; [b] contribution taken from refs. [15, 148]; [c] contribution taken from refs. [148, 154]; [d] mean deviation ($\bar{\Delta}$) and standard deviation (Δ_{std}) with respect to the ATcT data.

The original HEAT protocol performs well in terms of accuracy and gives differences

of below 1 kJ/mol for all considered molecules beside CN and CO_2. The effects from the change in the reference function of the the quadruples contribution, the two-electron Darwin contribution, and the correlated DBOC treatment of DBOC are rather small. The G_0 contribution shows only a significant improvement for the CO_2 molecule and seems to be needed for achieving an accuracy of better than 1 kJ/mol in this case. Statistical analysis of the considered summation schemes reveals similar quality for all of them. The mean deviation and the standard deviation are nearly constant with about 0.2 and 0.5 kJ/mol respectively, while the CN molecule remains in all cases the source for the maximum deviation (-1.2 kJ/mol). The effects of different treatment of the HLC contribution are shown in table 6.3.

Table 6.3: Errors (in kJ/mol) in atomization energies calculated with the HEAT345-(Q), HEAT345-Q, and HEAT345-QP protocols, compared with ATcT values based on the Core (Argonne) Thermochemical Network, C(A)TN, ver. 1.064.

	HEAT345-(Q)	HEAT345-Q	HEAT345-QP
N_2	0.60	-0.05	0.42
H_2	0.14	0.14	0.14
F_2	0.40	-0.04	0.12
CO	0.09	-0.32	-0.19
O_2	-0.17	-0.69	-0.26
C_2H_2	-0.22	-0.55	-0.23
CCH	0.02	-0.15	0.16
CF	0.11	-0.13	-0.20
CH_2	-0.27	-0.27	-0.26
CH	-0.27	-0.26	-0.25
CH_3	-0.16	-0.16	-0.14
CN	0.37	-1.20	-0.72
CO_2	0.64	-0.27	-0.11
H_2O_2	-0.19	-0.59	-0.39
H_2O	0.36	0.25	0.29
HCN	0.49	-0.09	0.33
HCO	0.18	-0.34	-0.19
HF	0.42	0.36	0.37
HNO	-0.34	-0.90	-0.55
HO_2	0.64	0.18	0.39
NH_2	0.51	0.50	0.52
NH_3	0.22	0.16	0.21
NH	-0.69	-0.68	-0.67
NO	-0.12	-0.73	-0.39
OF	0.70	0.08	0.22
OH	0.43	0.43	0.44
$\bar{\Delta}^a$	0.15	-0.20	-0.04
max	0.70	-1.20	-0.72
Δ^a_{std}	0.37	0.42	0.36

[a] Mean deviation ($\bar{\Delta}$) and standard deviation (Δ_{std}) with respect to the ATcT data.

Interestingly the computationally less demanding and more approximate HEAT345-(Q) approach outperforms the HEAT345-Q scheme and seems to be closer to the more complete HEAT345-QP protocol [15]. In order to investigate the remaining discrepancies to the ATcT values all contributions of the HEAT schemes will be discussed in the following paragraphs. Particular attention will be paid on the basis-set convergence and core-correlation effects.

6.2.1 Basis-set convergence of HF-SCF and CCSD(T)

Starting from the HEAT345-QP scheme the HF-SCF contribution is varied applying the aug-cc-pCVXZ (X=3, 4, 5, and 6) basis set where 345 and 456 denote a three-point extrapolation using the corresponding cardinal numbers according to eq. (6.2). The results are shown in table 6.4.

Table 6.4: Variation of the HF-SCF contribution within the HEAT345-QP scheme by using either the aug-cc-pCVXZ (X=3, 4, 5, and 6) basis set or in the case of 345 and 456 the three-point extrapolation according to eq. (6.2). Errors (in kJ/mol) in atomization energies, compared with ATcT values based on the Core (Argonne) Thermochemical Network, C(A)TN, ver. 1.064.

	3	4	5	345	6	456
N_2	-1.71	0.47	0.55	0.42	0.55	0.55
H_2	-0.49	0.01	0.12	0.14	0.14	0.15
F_2	0.83	0.06	0.04	0.12	0.03	0.02
CO	-2.67	0.15	0.09	-0.19	0.08	0.07
O_2	-2.01	0.18	0.05	-0.26	0.05	0.05
C_2H_2	-2.54	-0.53	-0.23	-0.23	-0.19	-0.19
CCH	-2.26	-0.08	0.24	0.16	0.27	0.17
CF	-1.12	-0.20	-0.12	-0.20	-0.13	-0.13
CH_2	-1.46	-0.53	-0.30	-0.26	-0.27	-0.26
CH	-0.66	-0.31	-0.24	-0.25	-0.23	-0.23
CH_3	-1.70	-0.50	-0.19	-0.14	-0.15	-0.14
CN	-3.19	-0.75	-0.62	-0.72	-0.62	-0.62
CO_2	-2.94	0.84	0.47	-0.11	0.43	0.42
H_2O_2	-0.61	0.02	-0.16	-0.39	-0.17	-0.16
H_2O	-0.61	0.47	0.45	0.29	0.46	0.47
HCN	-1.68	0.37	0.45	0.33	0.45	0.45
HCO	-2.31	0.08	0.06	-0.19	0.06	0.05
HF	0.29	0.63	0.52	0.37	0.52	0.52
HNO	-1.94	0.00	-0.21	-0.55	-0.22	-0.22
HO_2	-0.40	0.77	0.63	0.39	0.63	0.63
NH_2	-0.33	0.52	0.59	0.52	0.61	0.61
NH_3	-1.24	0.09	0.26	0.21	0.30	0.30
NH	-1.04	-0.63	-0.62	-0.67	-0.61	-0.61
NO	-2.34	0.02	-0.10	-0.39	-0.11	-0.11
OF	0.40	0.33	0.27	0.22	0.26	0.26
OH	0.13	0.58	0.53	0.44	0.54	0.54
$\bar{\Delta}^a$	-1.29	0.08	0.10	-0.04	0.10	0.10
max	-3.19	0.84	0.63	-0.72	0.63	0.63
Δ^a_{std}	1.09	0.44	0.36	0.36	0.36	0.36

[a] Mean deviation ($\bar{\Delta}$) and standard deviation (Δ_{std}) with respect to the ATcT data.

The triple-zeta basis gives poor results for the HF-SCF contribution. The use of quadruple-zeta and pentuple-zeta basis sets seem to be enough for kJ/mol accuracy. The aug-cc-pCV5Z set gives converged results as can be seen with the identical results for mean and standard deviation for aug-cc-pCV5Z, aug-cc-pCV6Z, and the extrapolated results from aug-cc-pCVXZ with X=4, 5, and 6. Furthermore it can be noted that the values obtained by extrapolation for X=3, 4, and 5 deteriorate the HF-SCF/aug-cc-pCV5Z results. Comparison between aug-

cc-pCV6Z and extrapolated estimates with X=4, 5, and 6 show rather small differences. It remains questionable if the empirical extrapolation of HF-SCF energies is useful in the context of computational thermochemistry. Usually the effort for those calculations is negligible compared to all other steps involved and a calculation with an aug-cc-pCV6Z basis should be possible for all cases. For the sake of consistency with the definitions for the HEAT schemes made before the 456 extrapolated HF-SCF energies will be used in all following steps.

To further investigate the basis-set convergence of the CCSD(T) correlation energy contribution the basis is varied employing the aug-cc-pCVXZ (X=3, 4, 5, and 6) basis set, where 45 and 56 denote two-point extrapolation of the correlation energies using eq. (6.3).

Table 6.5: Variation of the CCSD(T) correlation energy contribution by employing the aug-cc-pCVXZ (X=3, 4, 5, and 6) basis set, with 45 and 56 denoting two-point extrapolation of the correlation energies using eq. (6.3). Errors (in kJ/mol) in atomization energies, compared with ATcT values based on the Core (Argonne) Thermochemical Network, C(A)TN, ver. 1.064.

	3	4	5	45	6	56
N_2	-17.08	-10.23	-1.66	0.55	0.46	-0.22
H_2	-3.25	-1.18	-0.54	0.15	-0.29	0.04
F_2	-3.06	-1.50	-0.15	0.02	0.38	-0.03
CO	-6.71	-5.23	1.29	0.07	2.60	-0.30
O_2	-6.36	-4.21	-0.30	0.05	0.83	0.03
C_2H_2	0.89	-1.64	6.56	-0.19	8.06	-0.98
CCH	1.16	-1.98	5.40	0.17	6.26	-0.56
CF	-2.32	-1.51	1.47	-0.13	1.87	-0.21
CH_2	6.40	0.84	4.01	-0.26	4.05	-0.57
CH	4.06	-0.19	1.60	-0.23	1.60	-0.35
CH_3	2.71	0.31	4.88	-0.14	5.11	-0.56
CN	-8.20	-7.24	0.17	-0.62	1.41	-1.23
CO_2	-9.25	-6.70	3.28	0.42	6.19	-0.25
H_2O_2	-13.87	-5.35	0.38	-0.16	1.59	-0.81
H_2O	-9.12	-2.64	1.03	0.47	1.64	-0.08
HCN	-5.85	-5.83	2.07	0.45	4.62	-0.34
HCO	-4.05	-3.65	3.17	0.05	4.48	-0.36
HF	-5.14	-0.50	0.76	0.52	1.17	0.30
HNO	-12.51	-7.32	-0.35	-0.22	1.19	-0.62
HO_2	-9.92	-4.10	0.65	0.63	1.80	0.27
NH_2	-5.59	-2.49	1.52	0.61	2.09	0.17
NH_3	-9.38	-3.57	1.96	0.30	2.70	-0.36
NH	-3.08	-2.28	-0.24	-0.61	0.11	-0.82
NO	-11.99	-7.51	-1.49	-0.11	0.35	-0.37
OF	-5.32	-2.32	-0.16	0.26	0.59	0.22
OH	-4.22	-1.05	0.78	0.54	1.13	0.30
$\bar{\Delta}^a$	-5.42	-3.43	1.39	0.10	2.38	-0.30
max	-17.08	-10.23	6.56	0.63	8.06	-1.23
Δ^a_{std}	5.60	2.80	2.08	0.36	2.18	0.40

[a] Mean deviation ($\bar{\Delta}$) and standard deviation (Δ_{std}) with respect to the ATcT data.

In contrast to the convergence behavior at the HF-SCF level CCSD(T) correlation energies obtained without extrapolation are not even close to 1 kJ/mol accuracy. Particularly, the consideration of the maximum errors leads to the conclusion that extrapolation of correlation

energies is definitively needed for chemical and sub-chemical accuracy. The 45 extrapolated energies used in the original HEAT protocol correspond to a "Pauling point". For the succeeding cardinal numbers the estimates become worse. The large differences between the 45 and 56 extrapolated estimates show clearly that even a sextuple-zeta basis set is not sufficient for achieving 1 kJ/mol accuracy.

6.2.2 Basis-set convergence of higher-level correlation effects

Basis-set convergence of the valence CCSDT contribution

Starting with the 56 extrapolated values from the last section the basis-set convergence of the ΔE_{CCSDT} valence correlation energy contribution is investigated by variation of the cc-pVXZ basis with X=3, 4, and 5. 34 and 45 denote two-point extrapolation of the energy using eq. (6.3).

Table 6.6: Basis-set convergence of the ΔE_{CCSDT} contribution by variation of the cc-pVXZ basis with X=3, 4, and 5. 34 and 45 denote two-point extrapolation of the energy using eq. (6.3). Errors (in kJ/mol) in atomization energies, compared with ATcT values based on the Core (Argonne) Thermochemical Network, C(A)TN, ver. 1.064.

	3	4	34	5	45
N_2	0.50	0.09	-0.22	-0.10	-0.29
H_2	0.04	0.04	0.04	0.04	0.04
F_2	0.32	0.12	-0.03	0.00	-0.12
CO	0.44	0.01	-0.30	-0.17	-0.37
O_2	0.48	0.22	0.03	0.07	-0.09
C_2H_2	-0.24	-0.67	-0.98	-0.84	-1.03
CCH	0.20	-0.24	-0.56	-0.43	-0.63
CF	0.22	-0.03	-0.21	-0.13	-0.24
CH_2	-0.49	-0.53	-0.57	-0.54	-0.56
CH	-0.31	-0.33	-0.35	-0.34	-0.36
CH_3	-0.40	-0.49	-0.56	-0.53	-0.56
CN	-0.40	-0.88	-1.23	-1.12	-1.38
CO_2	0.95	0.26	-0.25	-0.03	-0.34
H_2O_2	-0.05	-0.49	-0.81	-0.70	-0.91
H_2O	0.33	0.09	-0.08	-0.02	-0.13
HCN	0.42	-0.02	-0.34	-0.21	-0.40
HCO	0.40	-0.04	-0.36	-0.23	-0.44
HF	0.58	0.42	0.30	0.35	0.27
HNO	0.15	-0.29	-0.62	-0.50	-0.72
HO_2	0.80	0.49	0.27	0.35	0.20
NH_2	0.34	0.24	0.17	0.20	0.17
NH_3	-0.02	-0.22	-0.36	-0.30	-0.39
NH	-0.78	-0.80	-0.82	-0.81	-0.81
NO	0.28	-0.10	-0.37	-0.27	-0.46
OF	0.54	0.35	0.22	0.26	0.16
OH	0.44	0.36	0.30	0.32	0.28
$\bar{\Delta}^a$	0.18	-0.09	-0.30	-0.22	-0.35
max	0.95	-0.88	-1.23	-1.12	-1.38
Δ^a_{std}	0.42	0.38	0.40	0.39	0.42

[a] Mean deviation ($\bar{\Delta}$) and standard deviation (Δ_{std}) with respect to the ATcT data.

The mean valence correlation energy additivity error for the ΔE_{CCSDT} contribution is 0.05 kJ/mol when comparing HEAT345-QP and HEAT345-QP including ΔE_{CCSDT} obtained with 45 extrapolation of cc-pVQZ and cc-pV5Z. The largest discrepancy with 0.15 kJ/mol is observed for the CN molecule. The very small changes when going from the extrapolated triple and quadruple zeta results to quintuple zeta or the extrapolated 45 results show that the 34 extrapolated value used in the original HEAT scheme can be regarded as converged.

Basis-set convergence of the valence quadruple contributions

In order to investigate the basis-set convergence of the $\Delta E_{CCSDT(Q)}$ and the ΔE_{CCSDTQ} valence correlation energy contributions is the cc-pVXZ basis is varied with X=2, 3, and 4. 34 denotes a two-point extrapolation of the $\Delta E_{CCSDT(Q)}$ energy using eq. (6.3). All estimates are based on the 45 extrapolated values from the last section replacing the HLC contribution by the $\Delta E_{CCSDT(Q)}$ or ΔE_{CCSDTQ} contribution.

Table 6.7: Basis-set convergence of the $\Delta E_{CCSDT(Q)}$ and ΔE_{CCSDTQ} contributions by variation of the cc-pVXZ basis with X=3, and 4. 34 denotes a two-point extrapolation of the energy using eq. (6.3). Errors (in kJ/mol) in atomization energies, compared with ATcT values based on the Core (Argonne) Thermochemical Network, C(A)TN, ver. 1.064.

	2	3	4	34	2	3
N_2	-0.11	0.14	0.42	0.63	-0.76	-0.48
H_2	0.04	0.04	0.04	0.04	0.04	0.04
F_2	0.16	0.08	0.37	0.59	-0.29	-0.39
CO	-0.09	-0.02	0.19	0.33	-0.50	-0.41
O_2	0.00	-0.12	0.14	0.34	-0.52	-0.65
C_2H_2	-1.03	-0.63	-0.40	-0.22	-1.35	-0.90
CCH	-0.77	-0.28	-0.04	0.14	-0.94	-0.48
CF	0.06	-0.02	0.11	0.21	-0.18	–
CH_2	-0.57	-0.60	-0.58	-0.56	-0.56	-0.60
CH	-0.38	-0.37	-0.36	-0.34	-0.36	-0.35
CH_3	-0.57	-0.61	-0.57	-0.54	-0.58	-0.61
CN	-0.28	0.54	0.87	1.11	-1.86	-1.28
CO_2	0.41	0.21	0.54	0.79	-0.50	-0.64
H_2O_2	-0.71	-0.96	–	–	-1.12	–
H_2O	-0.07	-0.36	-0.27	-0.21	-0.17	-0.45
HCN	-0.24	0.11	0.36	0.55	-0.81	-0.42
HCO	-0.07	-0.01	–	–	-0.59	-0.52
HF	0.33	0.00	0.04	0.07	0.26	-0.07
HNO	-0.51	-0.34	-0.04	0.18	-1.07	-0.88
HO_2	0.46	0.38	–	–	0.00	-0.18
NH_2	0.15	0.07	0.13	0.17	0.14	0.07
NH_3	-0.38	-0.48	-0.38	-0.32	-0.44	-0.51
NH	-0.83	-0.88	-0.86	-0.84	-0.82	-0.88
NO	-0.18	-0.05	0.23	0.43	-0.80	-0.68
OF	0.63	0.61	0.85	1.03	0.02	-0.15
OH	0.27	0.13	0.17	0.20	0.27	0.10
$\bar{\Delta}^a$	-0.16	-0.13	0.04	0.16	-0.52	-0.44
max	-1.03	-0.96	0.87	1.11	-1.86	-1.28
Δ^a_{std}	0.43	0.40	0.44	0.50	0.51	0.33

[a] Mean deviation ($\bar{\Delta}$) and standard deviation (Δ_{std}) with respect to the ATcT data.

For the $\Delta E_{CCSDT(Q)}$ contribution extrapolation is needed. 34 seems to be close to convergence. When going from cc-pVDZ to cc-pVTZ a rather large change for the maximum error and the standard deviation with respect to the ATcT data is observed.

6.2.3 Core-correlation effects

Core-correlation effects may play some role even with the high-level correlation treatment. To investigate these effects calculations with the core-polarized valence basis sets were carried out. The estimates start from the HEAT456-QP scheme replacing the CCSDT contribution by their all-electron counterpart employing the correlation-consistent sets with additional core functions.

Core-correlation effects for the full triples contribution

Table 6.8 shows estimates that are based on the HEAT456-QP schemes. The basis-set convergence of the ΔE_{CCSDT} core and valence correlation energy contributions is investigated by variation of the cc-pVXZ basis with X=3, and 4, while in the all-electron calculation the core-polarized cc-pCVXZ are employed. 34 and C34 denote two-point extrapolation of the energy using eq. (6.3).

Table 6.8: Basis-set convergence of the ΔE_{CCSDT} core and valence correlation energy contributions by variation of the cc-pVXZ basis with X=3, and 4, while in the all-electron calculation the core-polarized cc-pCVXZ are employed. 34 and C34 denote two-point extrapolation of the energy using eq. (6.3). Errors (in kJ/mol) in atomization energies, compared with ATcT values based on the Core (Argonne) Thermochemical Network, C(A)TN, ver. 1.064.

	3	C3	4	C4	34	C34
N_2	0.50	0.57	0.09	0.24	-0.22	0.00
H_2	0.04	0.04	0.04	0.04	0.04	0.04
F_2	0.32	0.43	0.12	0.25	-0.03	0.12
CO	0.44	0.42	0.01	0.08	-0.30	-0.17
O_2	0.48	0.56	0.22	0.34	0.03	0.17
C_2H_2	-0.24	-0.21	-0.67	-0.55	-0.98	-0.80
CCH	0.20	0.30	-0.24	-0.04	-0.56	-0.29
CF	0.22	0.19	-0.03	-0.01	-0.21	-0.16
CH_2	-0.49	-0.50	-0.53	-0.52	-0.57	-0.54
CH	-0.31	-0.28	-0.33	-0.29	-0.35	-0.29
CH_3	-0.40	-0.40	-0.49	-0.46	-0.56	-0.51
CN	-0.40	-0.21	-0.88	-0.59	-1.23	-0.87
CO_2	0.95	0.96	0.26	0.39	-0.25	-0.03
H_2O_2	-0.05	-0.03	-0.49	-0.42	-0.81	-0.71
H_2O	0.33	0.31	0.09	0.10	-0.08	-0.06
HCN	0.42	0.44	-0.02	0.10	-0.34	-0.15
HCO	0.40	0.35	-0.04	0.00	-0.36	-0.26
HF	0.58	0.55	0.42	0.40	0.30	0.28
HNO	0.15	0.23	-0.29	-0.13	-0.62	-0.39
HO_2	0.80	0.84	0.49	0.57	0.27	0.37
NH_2	0.34	0.34	0.24	0.27	0.17	0.22
NH_3	-0.02	0.00	-0.22	-0.16	-0.36	-0.28
NH	-0.78	-0.78	-0.80	-0.79	-0.82	-0.79
NO	0.28	0.32	-0.10	0.01	-0.37	-0.21
OF	0.54	0.58	0.35	0.41	0.22	0.29
OH	0.44	0.43	0.36	0.36	0.30	0.30
$\bar{\Delta}^a$	0.18	0.21	-0.09	-0.02	-0.30	-0.18
max	0.95	0.96	-0.88	-0.79	-1.23	-0.87
Δ^a_{std}	0.42	0.41	0.38	0.36	0.40	0.36

[a] Mean deviation ($\bar{\Delta}$) and standard deviation (Δ_{std}) with respect to the ATcT data.

The effects of including core correlation to the ΔE_{CCSDT} contribution are more pronounced than those observed for the larger basis set in the valence energy (compare table 6.6).

Core-correlation effects for the quadruples contributions

In the following table the estimates are based on the HEAT456-(Q) and HEAT456-Q schemes. In addition, the core-contribution effects from the full triples contribution are included with the 34 extrapolated values from the previous section. The basis-set convergence of the $\Delta E_{CCSDT(Q)}$ and the ΔE_{CCSDTQ} valence correlation energy contributions is investigated by variation of the cc-pVXZ basis with X=2, and 3, while in the all-electron calculation the core-polarized cc-pCVXZ are employed.

Table 6.9: Investigation of the $\Delta E_{CCSDT(Q)}$ and ΔE_{CCSDTQ} core and valence correlation energy contributions by variation of the cc-pVXZ basis with X=2, and 3, while in the all-electron calculation the core-polarized cc-pCVXZ are employed. Errors (in kJ/mol) in atomization energies, compared with ATcT values based on the Core (Argonne) Thermochemical Network, C(A)TN, ver. 1.064.

	2	C2	3	C3	2	C2
N_2	0.18	0.22	0.43	0.51	-0.47	-0.45
H_2	0.04	0.04	0.04	0.04	0.04	0.04
F_2	0.39	0.41	0.32	0.34	-0.05	-0.05
CO	0.11	0.14	0.18	0.28	-0.30	-0.29
O_2	0.27	0.29	0.14	0.19	-0.25	0.25
C_2H_2	-0.80	-0.80	-0.41	-0.35	-1.13	-1.13
CCH	-0.43	-0.40	0.06	0.14	-0.60	-0.59
CF	0.15	0.17	0.06	0.14	-0.09	-0.08
CH_2	-0.56	-0.57	-0.59	-0.60	-0.55	-0.56
CH	-0.31	-0.31	-0.30	-0.31	-0.29	-0.29
CH_3	-0.52	-0.53	-0.56	-0.57	-0.53	-0.54
CN	0.22	0.32	1.04	1.22	-1.35	-1.32
CO_2	0.72	0.76	0.52	0.70	-0.19	-0.17
H_2O_2	-0.51	-0.50	-0.75	-0.70	-0.91	-0.92
H_2O	0.01	0.01	-0.28	-0.26	-0.09	-0.10
HCN	0.01	0.03	0.36	0.43	-0.56	-0.56
HCO	0.11	0.14	0.17	0.27	-0.41	-0.41
HF	0.34	0.34	0.01	0.03	0.28	0.27
HNO	-0.18	-0.15	-0.01	0.07	-0.74	-0.73
HO_2	0.63	0.65	0.55	0.62	0.17	0.17
NH_2	0.21	0.21	0.13	0.14	0.20	0.20
NH_3	-0.26	-0.26	-0.36	-0.36	-0.32	-0.33
NH	-0.82	-0.81	-0.86	-0.86	-0.80	-0.80
NO	0.07	0.10	0.20	0.29	-0.55	-0.54
OF	0.77	0.80	0.74	0.83	0.15	0.16
OH	0.30	0.30	0.15	0.16	0.29	0.29
$\bar{\Delta}^a$	0.00	0.02	0.04	0.12	-0.35	-0.33
max	-0.82	-0.81	1.04	1.22	-1.35	-1.32
Δ^a_{std}	0.43	0.44	0.46	0.49	0.43	0.44

[a] Mean deviation ($\bar{\Delta}$) and standard deviation (Δ_{std}) with respect to the ATcT data.

For both CCSDT(Q) and CCSDTQ the changes employing core-polarization functions together with the inclusion of the core electrons in the correlation treatment are rather small when regarding the double-zeta sets. For the perturbative treatment of quadruple excitations calculations employing triple-zeta basis sets were possible. The differences are on the same order as for the 34 extrapolated valence energy contributions. The contribution of core correlation might be considered as converged because the difference between the mean deviations is only about 0.1 kJ/mol when going from cc-pCVDZ to cc-pCVTZ, while the valence-only calculations yield a difference of 0.04 kJ/mol when going from the double-zeta to the triple-zeta set.

6.2.4 Current best estimates

The results from the previous paragraphs allow the discussion of the most complete HEAT scheme to date (HEAT456-QP), and in addition the construction of even more complete schemes. First, the most advanced scheme that could be applied to all molecules will be presented: In addition to the HEAT456-QP scheme the ΔE_{CCSDT} contribution is treated employing the 45 extrapolated values. The differences of 34 extrapolated ΔE_{CCSDT} contributions is used in order to correct for core-correlation effects. The core-polarized basis sets are used for the all-electron calculations, while the polarized valence sets are employed for the frozen-core calculations. Although the use of different basis sets for valence and all-electron correlation might present an approximation, the consequences will be rather small. The ΔE_{HLC} contribution is estimated by the sum of $\Delta E_{CCSDT(Q)}$ at the cc-pCVTZ level, the difference ΔE_{CCSDTQ} and $\Delta E_{CCSDT(Q)}$ at the cc-pCVDZ level and the valence pentuples contribution.

The next higher level could only be applied to a subset of the molecules considered here. Up to the CCSDT level this scheme is identical to the scheme described before. The ΔE_{HLC} contribution is given by the sum of $\Delta E_{CCSDT(Q)}$ obtained from 34 extrapolation, $\Delta E_{CCSDT(Q)}$ core correction defined via the difference in the $\Delta E_{CCSDT(Q)}$ contributions obtained from CCSDT(Q)/cc-pCVTZ and (fc)CCSDT(Q)/cc-pVTZ calculations, the difference of ΔE_{CCSDTQ} and $\Delta E_{CCSDT(Q)}$ at the cc-pVTZ level, a core correction for the full CCSDTQ contribution obtained in a similar manner as for the perturbative quadruples employing the cc-pCVDZ and cc-pVDZ sets, and again the valence pentuples contribution.

The last column of table 6.10 presents the scheme described before with additional consideration of hextuples (in the frozen-core approximation) obtained with the cc-pVDZ basis set.

Table 6.10: Errors (in kJ/mol) in atomization energies of HEAT345-(Q), HEAT456-QP and the best estimates, compared with ATcT values based on the Core (Argonne) Thermochemical Network, C(A)TN, ver. 1.064.

	HEAT345-(Q)	HEAT456-QP	HEAT456-QP[a] +improvements[a]	HEAT456-QP +improvements[b]	HEAT456-QPH +improvements[c]
N_2	0.60	-0.22	0.24	0.75	0.79
H_2	0.14	0.04	0.04	0.04	0.04
F_2	0.40	-0.03	-0.04	0.43	–
CO	0.09	-0.30	-0.09	0.29	0.30
O_2	-0.17	0.03	-0.05	0.40	0.45
C_2H_2	-0.22	-0.98	-0.41	0.05	–
CCH	0.02	-0.56	0.19	0.57	0.59
CF	0.11	-0.21	-0.20	–	–
CH_2	-0.27	-0.57	-0.58	-0.55	-0.55
CH	-0.27	-0.35	-0.29	-0.26	-0.26
CH_3	-0.16	-0.56	-0.56	-0.49	-0.49
CN	0.37	-1.23	-0.10	0.23	0.26
CO_2	0.64	-0.25	-0.16	0.48	0.48
H_2O_2	-0.19	-0.81	-1.01	–	–
H_2O	0.36	-0.08	-0.39	-0.23	-0.22
HCN	0.49	-0.34	0.19	0.67	–
HCO	0.18	-0.36	-0.19	–	–
HF	0.42	0.30	-0.06	0.01	0.02
HNO	-0.34	-0.62	-0.27	0.27	0.27
HO_2	0.64	0.27	0.27	–	–
NH_2	0.51	0.17	0.15	0.24	0.24
NH_3	0.22	-0.36	-0.40	-0.21	-0.21
NH	-0.69	-0.82	-0.83	-0.81	-0.81
NO	-0.12	-0.37	-0.09	0.38	0.41
OF	0.70	0.22	0.26	0.54	–
OH	0.43	0.30	0.15	0.19	0.20
$\bar{\Delta}^d$	0.15	-0.30	-0.16	0.13	0.08
max	0.70	-1.23	-1.01	-0.81	-0.81
Δ^d_{std}	0.37	0.40	0.33	0.41	0.44

[a] HEAT456-QP+T45+TC34+(Q)C3+QC2+P (see text);
[b] HEAT456-QP+T45+TC34+(Q)34+((Q)C3-(Q)3)+(Q3-(Q)3)+(QC2-(Q)C2)-(Q2-(Q)2)+P (see text);
[c] values of [b] including the contribution from (fc)CCSDTQPH/cc-pVDZ;
[d] mean deviation ($\bar{\Delta}$) and standard deviation (Δ_{std}) with respect to the ATcT data.

It is clear from the error analysis that the theoretically best HEAT approaches do not necessarily have the best statistical properties. For example, the largest error for all methods using a particular treatment for the ΔE_{HLC} is larger when the 456-based extrapolations are used instead of the simpler 345-based extrapolations. However, all mean deviations with respect to the ATcT data are about 0.5 kJ/mol, and a detailed analysis of them is not straightforward. Efforts to distinguish between the quality of the various HEAT approaches are simply not meaningful on the basis of such a statistical comparison. Therefore, a few selected molecules are chosen for a detailed investigation of deficiencies that exist for the most complete approach presented here. It is self-evident that none of the individual contributions in the HEAT protocol associated with the non-relativistic and scalar-relativistic energies is experimentally determinable. One can only base the level of confidence in a particular

correlation energy on the convergence of the energy contributions; that is, molecules with a low HLC probably have more accurate correlation energies than those with larger ΔE_{HLC} values. A similar faith must be associated with the DBOC contribution but fortunately this is usually of a magnitude such that a 10-20% error in the calculation would amount to no more than ca. 0.05 kJ/mol in the TAE. However, the remaining contributions, the spin-orbit stabilization energy and the zero-point vibrational energy can be determined in the laboratory, subject to certain assumptions. And for a few of the simpler systems in the test suite, the corresponding data are available. Let us first investigate the simple hydrides which have extremely tight ATcT error bars (HF, OH and H_2O) of ≤ 0.1 kJ/mol. For all of these systems, the error in the HEAT-based SO and ZPE corrections can be quantified, and improved TAEs can be calculated using the remaining contributions.

HF

Hydrogen fluoride represents a somewhat puzzling case. The straightforward HEAT-456QP atomization energy is 0.31 kJ/mol above the current ATcT value. Use of the exact zero-point correction of 2050.8 cm^{-1} [155] (which is 14.3 cm^{-1} below the HEAT value) acts to further increase this discrepancy – the molecule is now predicted to be overbound by 0.48 kJ/mol. As the HLC for HF is less than 1 kJ/mol, experience suggests that basis-set effects would have only a small effect and actually make the agreement worse. However, the HEAT-based spin-orbit stabilization of the fluorine atom undershoots the experimental value of 134.7 cm^{-1} [156] by 8.7 cm^{-1}, and application of the appropriate adjustment brings the theoretical value closer; exact ZPE and spin-orbit contributions lead to a TAE that is 0.37 kJ/mol above the ATcT value. The HLC for HF is predicted surprisingly poorly with a small basis set and in fact decreases with basis-set size. In general, the HLC determined in a small basis set will undershoot the correct result (the same is true for individual contributions to the CCSD(T) contributions; it is just the extrapolations that tend to overshoot the correct result), meaning that the contribution to the TAE obtained with HEAT-based values of ΔE_{HLC} will be slightly too small. The cc-pVDZ basis set overshoots the extrapolated 34 result for the CCSDT(Q) contribution to the TAE by roughly 0.3 kJ/mol. This accounts, nearly quantitatively, for the remaining error in the TAE for hydrogen fluoride.

OH radical

The theoretically most complete atomization energy is calculated to be 425.82 kJ/mol, which is 0.20 kJ/mol above the quite precise ATcT value of 425.62 $\pm$ 0.03 kJ/mol.

It turns out that largest in magnitude amongst these is the error due to the HEAT-based SO contribution, which corresponds to the center-of-gravity between the non-relativistic energy (which is equal to the eigenvalue of the Born-Oppenheimer electronic Hamiltonian) and the true ground-state electronic energy, a quantity that we refer to here as the spin-orbit stabilization energy (SOSE). The exact SOSE for OH is -38.2 cm^{-1} (-0.46 kJ/mol) [148,157], which is 0.32 kJ/mol smaller in magnitude than the HEAT result of -65.2 cm^{-1}. Hence, the error in the HEAT-based SOSE for the atomization energy of OH is $0.32 + 0.11 = 0.43$ kJ/mol, which decreases the theoretical TAE to 425.38 KJ/mol, which is 0.24 kJ/mol below the ATcT value. With a residual error of -0.24 kJ/mol after correcting the SOSE, the error just changed the sign for the OH radical.

For OH, the experimental value of the ZPE is 1850.69 cm^{-1} (22.139 kJ/mol) [155], which is 0.11 kJ/mol below the HEAT value. Thus, the theoretical atomization energy is increased by this amount, which results in a value of 425.51 kJ/mol, which is still 0.1 kJ/mol below the ATcT value.

H_2O

The water molecule is another interesting case, where the complexities associated with treating the spin-orbit interaction in hydroxyl are not present, water being a closed-shell singlet molecule. The HEAT456-QP atomization energy of 917.76 kJ/mol is but 0.07 kJ/mol below the ATcT estimate of 917.84 ± 0.03 kJ/mol. However, this agreement is somewhat fortuitous and degrades when proper account is taken for the spin-orbit and ZPE contributions. For the former, the error of 0.11 kJ/mol in the HEAT value for the oxygen atom spin-orbit stabilization energy [156] must be added to the HEAT-456QP result, raising it to 917.88 kJ/mol, this value being in better agreement with ATcT. The situation is further improved when one includes the ZPE error; the semi-experimental value of 55.76 kJ/mol [158] is 0.03 kJ/mol below the HEAT value, lowering the theoretical value to 917.85 kJ/mol. However, again there is something of a Pauling point even at this level. The ΔE_{HLC} contribution to the TAE of H_2O is overestimated with the cc-pVDZ basis set by about 0.14 kJ/mol (compare table 6.7). This, together with basis-set effects of the $\Delta E^{\infty}_{CCSD(T)}$ and E^{∞}_{HF} corrections, clearly shows that one must be very careful in analyzing theoretical thermochemical results at such advanced levels of theory. Perfect, or nearly perfect, agreement is always fortuitous and it is totally unclear whether improving a particular contribution is really beneficial.

6.2.5 Discussion

The previous section makes clear that the overall error in HEAT-based schemes for thermochemistry is so small that even more sophisticated treatments of individual contributions and/or consideration of effects such as the angular-momentum coupling in the hydroxyl radical change results in subtle ways that do not necessarily improve results. Indeed, it seems clear that easily the most practical and useful general HEAT method is probably HEAT345-(Q), which is also the least expensive approach. At this point, it is appropriate to go through each contribution to the HEAT energy and discuss some general trends.

The CCSD(T)/cc-pVQZ geometries used in the HEAT protocol are not the true equilibrium geometries. An intrinsic error of ca. 0.005 Å has to be considered. Obviously, any error in geometry will result in an atomization energy that is a bit too small. It is always further from the bottom of the well to the asymptote than from some point above the minimum. However, the fact that the energy is stationary in this vicinity means that the resulting error will be too small. A typical force constant is 0.5 hartree $bohr^{-2}$, which is equivalent to about 4000 kJ mol^{-1} $Å^{-2}$. Hence, an error of 0.005 Å in a bond distance translates to 0.05 kJ/mol. Therefore it is unlikely that such effects will be important for diatomic molecules, but certainly need to be considered for larger systems.

E^{∞}_{HF} is apparently quite well converged; differences between TAE contributions calculated from aug-cc-pCV5Z, aug-cc-pCV6Z, and HEAT-456 HF-SCF energies are so small as to suggest that E^{∞}_{HF} from HEAT-456 is probably in error by less than 0.05 kJ/mol in all cases. The same is not true for the HEAT-345 E^{∞}_{HF} results.

$\Delta E^{\infty}_{CCSD(T)}$ is of course one of the largest sources of error and the principal term of importance for the correlation energy. Here, the error is apparently systematic. While CCSD(T) energies calculated in all of the basis sets used here lead to TAE contributions that are too small (correlation effects always give a positive contribution to the TAE), the partial-wave extrapolations tend to overshoot the results. The finite basis set calculations converge (slowly) from below; the extrapolations appear to converge from above. The ΔE_{CCSDT} and ΔE_{HLC} contributions have been analyzed in detail by Karton et al. [159], and it seems that the strategy for ΔE_{CCSDT} that is used in the HEAT approach which is also based on extrapolation – also gives a slightly too positive contribution to the TAE, as well. However, the magnitude of the difference is smaller than that of $\Delta E^{\infty}_{CCSD(T)}$, and typically on the order of 0.1 kJ/mol.

The overestimation of the correlation contribution to the TAE at the CCSDT level is, in HEAT-based thermochemistry, offset to a large degree by a compensating error in the ΔE_{HLC} term. Boese et al. [130] were the first to investigate the basis-set effects in CCSDTQ and related methods vis-a-vis the TAE. One can do calculations with larger and larger basis sets, and thereby approach the exact $\Delta E^{\infty}_{CCSD(T)}$ and ΔE_{CCSDT} contributions, albeit with considerable computational demands, and then attempt to also account quantitatively for the ΔE_{HLC} using large basis sets. There is no doubt that this is the only way to establish definitive results, and to make predictions of fundamental thermochemical parameters when extraordinarily high precision is required. However, it is very demanding, and it seems that such an approach will always encounter severe restrictions with regard to its scope of application. The other approach, which is clearly more practical, is to use the empirical cancellation between the overshooting of the $\Delta E^{\infty}_{CCSD(T)}$ and ΔE_{CCSDT} contributions and the undershooting of the ΔE_{HLC} contribution. Each has its merits, and it is always satisfying to understand why a method performs "better than it ought to", as does HEAT345-(Q).

Most of the remaining contributions have rather smaller error bars. First, the ΔE_{REL} contribution is based on first-order perturbation theory, and contains both one- and two-electron Darwin and mass-velocity contributions. In ref. [148] it is found that the larger basis set tends to affect TAE contributions by at most 0.07 kJ/mol (for, not surprisingly, F_2). The spin-orbit contribution ΔE_{SO} is a somewhat larger source of error, as the calculations for the atoms C, O, and F have errors (with respect to the exact spectroscopic stabilizations) of -0.02, 0.11 and 0.10 kJ/mol, with the HEAT-based stabilization energies larger in magnitude for O and F. Hence, this error leads to an error of the associated TAE contribution for closed-shell molecules by a magnitude that is easily calculated from stoichiometric considerations. The DBOC contribution changes little when one goes from the HF-based treatment of this correction that was used in ref. [14] and ref. [15] to the CCSD result used here. It is expected to be accurate to well within 0.1 kJ/mol.

Finally, the zero-point vibrational correction can be determined essentially exactly from spectroscopic data on diatomics and is also occasionally well-known for polyatomic molecules such as, for example, water. Results reported earlier in this work, together with the empirical knowledge that quantum-chemical calculations of vibrational energy levels tend to approach the correct answer from above, suggests that the calculated ΔE_{ZPE} contribution to the atomization energies is too large in magnitude; a more accurate treatment of this contribution would tend to increase atomization energies.

A summary of the effectively systematic errors encountered in the HEAT scheme is shown

in Table 6.11, where the tendency towards error cancellation is evident, as is the cause of the seemingly anomalous behavior of HF. It should be emphasized that the error bars for HEAT (about 0.5 kJ/mol mean deviation with respect to the ATcT data) apply only to molecules containing a very few first-row atoms.

Table 6.11: Guide to typical sizes and signs of errors in the contributions to HEAT-based atomization energies for molecules with two to four first-row atoms.

Contribution to TAE	Sign of Typical Error in HEAT-based theory	Greater than 0.2 kJ/mol ?
E_{HF}^{∞}	Essentially exact	No
$\Delta E_{CCSD(T)}^{\infty}$	+	Can be
ΔE_{CCSDT}	+	No
ΔE_{HLC}	-	Can be
	$+^a$	Can be
ΔE_{REL}	varies	No
ΔE_{SO}	+ for hydrocarbons	Can be
	- otherwise	Can be
ΔE_{DBOC}	Unknown	No
ΔE_{ZPE}	-	Can be

[a]For molecules containing OH and FH bonds.

Calculating atomization energies is very difficult, and the magnitude of the error ultimately will become proportional to the size of the system. For example, it is entirely unrealistic to expect a HEAT-based calculation of the total atomization energy of benzene (see section 6.4) to have this sort of accuracy. It is a necessary consequence of size extensivity that the characteristic relative error in atomization energies will remain constant, but the absolute error is clearly going to grow linearly with the size of the system. "1 kJ/mol accuracy" applies only within the small molecule (two or three first-row atoms, at most) approximation. Hence, the achievement of chemical accuracy is really a relative goal.

6.2.6 Conclusions

The results of this work serve to make a general statement about very high-accuracy theoretical thermochemistry. The accuracies that can now be obtained - via improvements in quantum chemistry made in the last decade (general coupled-cluster approaches, basis-set extrapolation techniques, and the development of large basis sets necessary to adequately use the extrapolations) - are at the point where it is extremely difficult to judge the quality of comparably accurate strategies. As a result, a critical analysis of the calculated numbers beyond the relatively qualitative discussion epitomized by the content of table 6.11 is not straightforward in general, but must take the form of the brief discussions of OH, HF, and H_2O earlier in this chapter. For other systems, where the uncertainty in the ATcT data is much greater, it really is not possible to make any sort of detailed analysis at all.

The time-honored idea of "Pauling points" is valuable here, even at what might seem to be an absurdly high level of theory. The in a relative sense rather simple HEAT345-(Q) method is in fact statistically the most accurate of the various approaches, despite being the cheapest. While the argument can be made that there are cases in which HEAT345-(Q) might be prone to larger errors, the conclusion that this simplest HEAT method is the most

pragmatic and offers the best compromise between accuracy and cost cannot be refuted. To improve systematically upon HEAT345-(Q), the results of this work suggests that:

- More accurate equilibrium geometries are needed based on CCSD(T) calculations employing core-polarized quadruple-zeta basis sets or better based on extrapolation schemes similar to the HEAT protocol [160].
- The CCSD(T) energy has to be calculated with at least a septuple-zeta basis again using extrapolation techniques, alternatively CCSD(T)-F12/aug-cc-pCV6Z [161] calculations might as well resolve the unsatisfactory situation.
- The HLC contributions have to be treated with large basis sets and extrapolation techniques. Even on this levels of theory core-correlation effects cannot be neglected.
- Even on the HEAT345-(Q) level more accurate SOSEs are needed. This might be accomplished with the recent progress made by treating the SO interactions on the CC level [162, 163].
- The ZPE contributions can be improved using extrapolation techniques as can be seen in ref. [164] together with variational treatments for the determination of anharmonicities like the discrete variable representation approaches [165, 166].

The lesson learned in this work and refs. [131] and [159] is that all of these improvements must be made to effect an overall improvement in the quality of the results. The field of high-accuracy thermochemistry has clearly reached an impasse, but quite a satisfactory one since the level of accuracy achievable now is sufficient for almost all needs.

Beside improvements of the accuracy the HEAT methods themself provide an interesting starting point for a possibly more accurate description of the thermochemistry for molecules containing second row atoms, larger molecules, potential-energy surfaces, transition states and molecular properties.[1]

The next two sections of this thesis will address the application to molecules containing second-row atoms and the total atomization energy of benzene.

[1] First attempts for geometries, and harmonic frequencies were already made as can be seen in refs. [160,164].

6.3 High accuracy extrapolated ab initio thermochemistry of vinyl chloride

A recent benchmark study of reaction energies using local coupled-cluster techniques revealed several discrepancies between the computed results and the available experimental data [167, 168]. The problematic reactions involve vinyl chloride (CH_2CHCl) and isocyanic acid (HNCO). The discrepancies were of the order of 25 kJ/mol [167] and thus larger than the typical error of the CCSD(T) model used in the study. While a thorough theoretical investigation of the thermochemistry of HNCO has been presented in ref. [145] this section will focus on vinyl chloride. To be more specific, the following two reactions are investigated

$$C_2H_4 + Cl_2 \longrightarrow CH_2CHCl + HCl \text{ (R1)}$$
$$C_2H_2 + HCl \longrightarrow CH_2CHCl \text{ (R2)}$$

as well as the heat of formation of vinyl chloride.
The reported experimental standard heats of formation at a temperature of 298.15 K for CH_2CHCl span a range from -21 to -38.1 kJ/mol [169–177]. For the first reaction the values, when computed from the available data in the literature, again for a temperature of 298.15 K, vary between -106.7 and -123.8 kJ/mol and for the second they range from -96.3 to -113.4 kJ/mol. The only direct measurement reports a value of -100.66 kJ/mol for the second reaction [171].

Considering the recent achievements in theoretical high-accuracy thermochemistry [14, 15, 130, 131], it should be possible to provide values with a remaining uncertainty of about 1-2 kJ/mol for these reaction energies and for the heat of formation of vinyl chloride. The current unsatisfactory situation concerning these enthalpies will be resolved by performing a corresponding computational investigation using a slightly modified HEAT protocol [14, 15].

6.3.1 Differences to the original HEAT protocol

The procedure applied is closest to what has been termed the HEAT345-(Q) protocol. However, due to the fact that the systems considered here contain chlorine, i.e., a second-row element, a few modifications are necessary. Those are (1) use of all-electron CCSD(T)/cc-pCVQZ geometries instead of CCSD(T)/cc-pVQZ geometries, thus recognizing the increased importance of a proper treatment of core-correlation effects in the present case; (2) for the same reason, the ZPE corrections have been obtained using the cc-pCVXZ basis-sets. Furthermore, different to the original HEAT protocol, the ZPE term is partitioned into harmonic ($\Delta E_{ZPE}^{anharm.}$) and an anharmonic ($\Delta E_{ZPE}^{harm.}$) contributions which are obtained separately with different basis-sets (X=Q for the harmonic force field and X=T for the anharmonic force field) in order to keep computational cost at an acceptable level; (3) additional consideration of temperature corrections.

The DBO corrections have been obtained at the HF-SCF level. Spin-orbit effects were included in first order based on experimental data [156,178,179] and account for the difference between the non-relativistic (weighted average) of the states and the true ground state.

6.3.2 Temperature effects

As the final goal is a comparison with experimental thermochemical energies determined at a temperature of 298.15 K, it is essential to also include temperature contributions in this

investigation. For this purpose, it is assumed that the rotational degrees of freedom can be treated classically and that the harmonic approximation is sufficient for the treatment of vibrational degrees of freedom. The temperature correction ΔE_T to the enthalpies of formation and the reaction energies is then given as

$$\Delta E_T = \Delta E_T^{trans} + \Delta E_T^{rot} + \Delta E_T^{v} \tag{6.9}$$

where ΔE_T^{trans} and ΔE_T^{rot} are the classical corrections of $\pm\frac{1}{2}RT$ for each translational and rotational degree of freedom gained or lost in the reaction, and R is the universal gas constant. ΔE_T^{v} is the vibrational temperature term,

$$\Delta E_T^{v} = RT \sum_i \frac{\frac{hc\nu_i}{k_B T}}{\exp\left(\frac{hc\nu_i}{k_B T}\right) - 1} \tag{6.10}$$

with ν_i denoting the harmonic vibrational frequencies in cm^{-1} and k_B the Boltzmann constant. Possible numerical problems caused by low-frequency modes do not appear with the considered molecules because the lowest frequency is about 400 cm^{-1}. Finally, to obtain enthalpies instead of energies, the pressure-volume work term ΔpV has to be added, which is equal to $+RT$ for each mole of gas produced and $-RT$ for each mole consumed, thereby assuming ideal gas behavior.

6.3.3 Results and discussion

The computed total energies for the species investigated in the present work are summarized in table 6.12 together with the individual contributions from eq. (6.1) and required for the application of the HEAT scheme. Table 6.13 then reports the atomization energies for H_2, HCl, Cl_2, C_2H_2, C_2H_4, and CH_2CHCl, as they are obtained from the data given in table 6.12. Again, the individual contributions are repeated, and, in addition, the theoretical atomization energies (for all species except CH_2CHCl) are compared with those derived within the active thermochemical table (ATcT) of Ruscic and coworkers [137]. This comparison once more confirms the high accuracy that can be achieved with state-of-the-art thermochemical protocols such as HEAT, as the remaining discrepancies between theory and experiment (i.e., the ATcT values) are in all cases, but C_2H_4 below 0.5 kJ/mol.

Higher-level correlation energy contributions (HLC) are, as it is well known [14, 15, 130, 131], again important. They amount to 3 - 5 kJ/mol (sum of $\Delta E_{\mathrm{CCSDT}}$ and $\Delta E_{\mathrm{CCSDT(Q)}}$) and affect the computed heats of formation by a few kJ/mol. The effect is particularly pronounced for the systems with multiple bonds, but also non-negligible for Cl_2. Again, as it has been observed before [15], the CCSDT and CCSDT(Q) contribution to the heats of formation, but not for the total energies, have opposite signs and partially cancel, so that the total "beyond CCSD(T)" correction amounts to less than 1 kJ/mol.

Anharmonic contributions to the ZPE are of the same size as the HLC and thus important to consider. These corrections are particularly important for the larger systems, i.e., C_2H_4 and CH_2CHCl, probably due to the increased number of normal modes.

Table 6.12: Contributions to the HEAT345(Q) total energies (in atomic units) for the nine species studied in this work.

	E_{HF}	$\Delta E_{CCSD(T)}$	ΔE_{CCSDT}	ΔE_{HLC}	$\Delta E_{ZPE}^{harm.}$	$\Delta E_{ZPE}^{anharm.}$	ΔE_{rel}	ΔE_{DBOC}	ΔE_{SO}	Total
H	-0.500022	0.000000	0.000000	0.000000	0.000000	0.000000	-0.000007	0.000272	0.000000	-0.499756
H_2	-1.133661	-0.040912	0.000000	0.000000	0.010032	-0.000141	-0.000010	0.000460	0.000000	-1.164232
C	-37.693774	-0.151042	-0.000466	-0.000021	0.000000	0.000000	-0.015090	0.001660	-0.000135	-37.858869
Cl	-459.489895	-0.665245	-0.000767	-0.000144	0.000000	0.000000	-1.404007	0.005940	-0.001340	-461.555458
HCl	-460.112797	-0.713899	-0.000540	-0.000250	0.006850	-0.000061	-1.403590	0.006143	0.000000	-462.218145
Cl_2	-919.010698	-1.395150	-0.000836	-0.000703	0.001276	-0.000003	-2.807653	0.011876	0.000000	-923.201891
C_2H_2	-76.855711	-0.480387	0.000223	-0.001018	0.026531	-0.000241	-0.029756	0.003673	0.000000	-77.336687
C_2H_4	-78.070890	-0.518008	-0.000167	-0.000622	0.050966	-0.000801	-0.029685	0.004224	0.000000	-78.564983
C_2H_3Cl	-537.013383	-1.199564	-0.000297	-0.001083	0.042668	-0.000543	-1.433286	0.009898	0.000000	-539.595589

Table 6.13: Contributions to atomization enthalpies (in kJ/mol) for the molecules in this set. ATcT values were taken from ref. [131].

	E_{HF}	$\Delta E_{CCSD(T)}$	ΔE_{CCSDT}	ΔE_{HLC}	$\Delta E_{ZPE}^{harm.}$	$\Delta E_{ZPE}^{anharm.}$	ΔE_{rel}	ΔE_{DBOC}	ΔE_{SO}	total	ATcT	ATcT-total
H_2	350.81	107.41	0.00	0.00	-26.34	0.37	-0.01	0.22	0.00	432.47	432.06 ± 0.00	-0.41
HCl	322.62	127.74	-0.60	0.28	-17.99	0.16	-1.11	0.18	-3.52	427.77	427.65 ± 0.00	-0.13
Cl_2	81.15	169.76	-1.83	1.09	-3.35	0.01	-0.95	0.01	-7.04	238.85	239.24 ± 0.00	0.39
C_2H_2	1229.05	468.14	-3.03	2.56	-69.66	0.63	-1.15	0.50	-0.71	1626.15	1625.98 ± 0.07	-0.35
C_2H_4	1793.89	566.91	-2.01	1.52	-133.81	2.10	-1.37	0.48	-0.71	2227.00	2225.88 ± 0.06	-1.12
C_2H_3Cl	1669.49	609.73	-3.68	2.35	-112.03	1.42	-2.42	0.47	-4.23	2161.11		

Relativistic corrections to the total energies are rather large (> 1 hartree) as soon as the species contains chlorine, but these effects tend to cancel when thermochemical energies are computed. The relativistic effects on the heats of formation are at a few kJ/mol, rather small and of similar magnitude as the HLC and the anharmonic corrections to the ZPE. Consideration of spin-orbit corrections is essential, as they amount about 3.5 kJ/mol per chlorine. For carbon these corrections are with 0.35 kJ/mol significantly smaller, but also not negligible when aiming at sub-kJ/mol accuracy.

The theoretical prediction for the atomization energies of H_2, HCl, Cl_2, C_2H_2, and C_2H_4 agree well with previous theoretical high-accuracy values from HEAT or Wn studies [14, 15, 130, 131, 180]. For CH_2CHCl, the predicted atomization energy is 2161.1 kJ/mol.

Table 6.14: Calculated standard enthalpies of formation at standard conditions (in kJ/mol) in comparison with experimental values.

	H_2	HCl	Cl_2	Acetylene	Ethene	Vinyl chloride
$\Delta_f H^0(298.15K)$	-0.19	-92.19	0.51	228.42	52.42	-22.15
Experimental (NIST)	0.0	-92.31	0.0	226.7	52.47	-21 to - 38.1

In table 6.14, the calculated standard enthalpies of formation, $\Delta_f H^0(298.15K)$, are compared with available experimental values. The theoretical values have been obtained from the atomization energies reported in table 6.13 using the NIST-CODATA standard enthalpies of formation for the atoms (hydrogen: 217.998 ± 0.006 kJ/mol, carbon: 716.68 ± 0.45 kJ/mol, chlorine: 121.301 ± 0.008 kJ/mol [181]) as well as computed temperature corrections. Good agreement is found for all considered species with the deviations between theory and experiment in the range of 0.19 to -1.62 kJ/mol. The largest discrepancy is seen for acetylene with about 2 kJ/mol, which is possibly related to the large error bar on the experimental value of the standard enthalpy of formation for the carbon atom. For CH_2CHCl, it is assumed that the remaining error in the theoretical prediction of $\Delta_f H^0(298.15K)$ is of similar magnitude to that for the other investigated species. A conservative estimate therefore might be ± 2 kJ/mol. On the other hand, the experimental value for $\Delta_f H^0(298.15K)$ of CH_2CHCl settles in the range of -21.0 to -38.1 kJ/mol. The prediction of -22.2 ± 2 kJ/mol supports the values of the upper half of that range and necessarily rejects the others. Table 6.15 provides a detailed comparison with the experimental values.

Table 6.15: Various experimental results for the standard enthalpy of formation (in kJ/mol) of vinyl chloride.

Reference	Method	$\Delta_f H^0(298.15K)$
Lacher, Emery, et al., 1956 [169, 170]	calorimetry hydrogenation	37.2 ± 0.8
Lacher, Gottlieb, et al., 1962 [171]	calorimetry	33.8 ± 1.2
Cox and Pilcher, 1970 [172]	reanalyzation of data from [169, 170]	38.1 ± 0.84
Cox and Pilcher, 1970 [172]	reanalyzation of data from [171]	35.3 ± 1.4
Alfassi, Golden, et al., 1973 [173]	heat of equilibrium	21
Levanova, Treger, et al., 1976 [174]	heat of equilibrium	29
Ritter, 1991 [175]	thermochemical network	21.7
Gurvich et al., 1991 [176]	reanalyzation	23.0 ± 2.1
Manion, 2002 [177]	reanalyzation	22.0 ± 3

Only the values from refs. [173, 175, 176] and [177] are consistent with the computations. The value from ref. [173] has been determined via the heat of equilibrium, while that from ref. [175] has been obtained using a thermochemical network and the values from [176, 177] were found by reinvestigation of the experimental literature. The other values, obtained by either calorimetry or via the heat of equilibrium are all too high in absolute terms and clearly outside the error bar estimate for the calculated value. In table 6.16, the evaluated reaction enthalpies for $C_2H_4 + Cl_2 \longrightarrow CH_2CHCl + HCl$ and $C_2H_2 + HCl \longrightarrow CH_2CHCl$ are compared with the values directly obtained in experiment or derived from the experimental heat of formation of CH_2CHCl. Consistent with the findings for the heat of formation, good agreement is again only seen with the values derived from those given in refs. [173, 175, 176] and [177]. The other data [169–172, 174] should again be considered unreliable.

Table 6.16: Contributions to the reaction enthalpies (in kJ/mol) for the reactions $C_2H_4 + Cl_2 \longrightarrow CH_2CHCl + HCl$ (R1) and $C_2H_2 + HCl \longrightarrow CH_2CHCl$ (R2).

	R1	R2
E_{HF}^{∞}	-117.08	-117.82
$\Delta E_{\mathrm{CCSD(T)}}^{\infty}$	-0.80	-13.86
$\Delta E_{\mathrm{CCSDT}}$	0.44	0.05
ΔE_{HLC}	-0.02	0.49
$\Delta E_{\mathrm{ZPE}}^{\mathrm{harmonic}}$	-7.15	24.38
$\Delta E_{\mathrm{ZPE}}^{\mathrm{anharmonic}}$	0.52	-0.63
ΔE_{REL}	1.21	0.16
ΔE_{SO}	0.00	0.00
ΔE_{DBOC}	-0.16	0.22
$\Delta E_{\mathrm{T}}(298.15K)$	0.79	-6.92
$\Delta_r H(0K)$	-123.03	-107.01
$\Delta_r H(298.15K)$	-122.25	-114.10
$\Delta_r H^{0\mathrm{exp.}}(298.15K)$ [169, 170]	-107.60	-97.20
$\Delta_r H^{0\mathrm{exp.}}(298.15K)$ [172]	-106.70	-96.30
$\Delta_r H^{0\mathrm{exp.}}(298.15K)$ [171]	-111.00	-100.66
$\Delta_r H^{0\mathrm{exp.}}(298.15K)$ [172]	-109.50	-99.10
$\Delta_r H^{0\mathrm{exp.}}(298.15K)$ [173]	-123.80	-113.40
$\Delta_r H^{0\mathrm{exp.}}(298.15K)$ [174]	-115.80	-105.40
$\Delta_r H^{0\mathrm{exp.}}(298.15K)$ [175]	-123.09	-112.69
$\Delta_r H^{0\mathrm{exp.}}(298.15K)$ [176]	-121.80	-111.40
$\Delta_r H^{0\mathrm{exp.}}(298.15K)$ [177]	-122.80	-112.40

The individual contributions to the two reaction energies are listed in table 6.16. It is interesting to note that the characteristics of both reactions are rather different. For the first reaction, electron-correlation effects are small (-0.38 kJ/mol), while for the second reaction the latter are with -13.32 kJ/mol more sizable. For the second reaction, the ZPE effects and the temperature effects are larger, thus rendering an accurate prediction of the reaction enthalpy for this reaction is more challenging than for the first. However, the best estimate for those two reactions are -122.2 and 113.8 kJ/mol with again a conservative error estimate of ± 2 kJ/mol.

A very recent threshold photoelectron photoion coincidence (TREPICO) study [182] confirmed again the experimental values of refs. [173, 175, 176] and [177] and is with 22.6 ± 3.2 kJ/mol in excellent agreement with the heat of formation at 298.15 K computed in this study.

6.4 High accuracy extrapolated ab initio thermochemistry of benzene

The question arises as to how accurate the HEAT protocol performs for even larger systems than vinyl chloride. The total energies estimated by the HEAT formalism are size-extensive. The relative error in the TAE will remain constant but the absolute error is clearly going to grow linearly with the size of the system. To investigate this behavior, the benzene molecule appears to be a good choice. Because of its high symmetry even 6 heavy atoms can be treated with a slightly modified HEAT approach. In a recent publication Parthiban and Martin [183] obtain rather good agreement with experiment while neglecting completely higher-level correlation effects and DBOC. For an extended system, such as benzene, only the HEAT345-(Q) scheme can be considered with available computational resources. All contributions except $\Delta E^{\infty}_{CCSD(T)}$, ΔE_{CCSDT}, and ΔE_{ZPE} can be evaluated as defined for the HEAT345-(Q) scheme. $\Delta E_{CCSD(T)}$ could not be obtained using the augmented core-polarized basis sets due to convergence problems with the CCSD equations. The CCSD/aug-cc-pCV5Z calculation (1566 basis functions) diverged. This problem could not be resolved using an aug-cc-pV5Z basis (1092 basis functions), while skipping the augmentation functions on the hydrogen atoms. Further attempts employing various algorithms for convergence acceleration and the use of the MOLPRO program package [93] instead of ACES II MAB did not lead to any improvement. Therefore the augmentation functions were skipped and the cc-pCVQZ (684 basis functions) and cc-pCV5Z (1200 basis functions) basis sets are used in this study. The ΔE_{CCSDT} contribution is estimated with the cc-pVTZ basis set only due to memory limitations of the available computer hardware. The contribution from the perturbative quadruples employing the cc-pVDZ basis set (47 billion quadruple excitations) could be evaluated within 32 hours on the 12 node AMD64 cluster (see chapter 3). Again due to convergence problems with the Dunning basis sets, only the harmonic contribution to the ZPE could be evaluated employing the cc-pVQZ basis set. The anharmonic contribution was evaluated using the cc-pVTZ basis set. The results of this more approximate HEAT scheme are presented in the following table.

Table 6.17: Contributions to atomization enthalpy (in kJ/mol) for the benzene molecule.

E^{∞}_{HF}	+	4294.7
$\Delta E^{\infty}_{\mathrm{CCSD(T)}}$	+	1439.8
$\Delta E_{\mathrm{CCSDT}}$	-	8.9
ΔE_{HLC}	+	6.8
ΔE_{REL}	-	4.0
ΔE_{SO}	-	2.3
ΔE_{DBOC}	+	1.0
ΔE_{ZPE}	-	260.8
TAE		5466.0
exp.		5465.2 ± 1.8

The experimental TAE of 5465.2 ± 1.8 kJ/mol is derived from the $\Delta_f H^0_{gas}$ value of Prosen et al. [184], using the same heats of formation for the hydrogen and carbon atoms as in ref. [15] (hydrogen 216.07 kJ/mol; carbon 711.58 kJ/mol) and the $H_{289} - H_0$ correction given in ref. [183]. The excellent agreement with the calculated values of 5466.0 kJ/mol

appears fortuitous and deserves further analysis. The results of table 6.17 and ref. [183] lead to the conclusion that in the case of the benzene molecule the contributions from ΔE_{CCSDT}, $\Delta E_{CCSDT(Q)}$, and ΔE_{DBOC} almost cancel. This explains the accidental excellent agreement with experiment obtained in ref. [183]. The TAE given there is 5466.8 kJ/mol. As already stated one should not expect the same accuracy for rather large molecules, such as benzene, as for the small molecules in the HEAT set. The HF contribution should have no large error. When going from the quadruple-zeta to the quintuple-zeta basis set the value changes only by 0.61 kJ/mol, while the extrapolation adds another 0.11 kJ/mol. As expected these differences are more pronounced for the benzene molecule than for the smaller molecules considered earlier in this chapter. Table 6.18 shows the individual effects of the approximations to $\Delta E^{\infty}_{\mathrm{CCSD(T)}}$, $\Delta E_{\mathrm{CCSDT}}$, and ΔE_{ZPE} for the closed-shell systems in the HEAT set as well as the performance of the whole procedure in comparison to the ATcT data.

Table 6.18: Individual effects of the approximations to $\Delta E^{\infty}_{\mathrm{CCSD(T)}}$, $\Delta E_{\mathrm{CCSDT}}$, and ΔE_{ZPE} for the closed-shell systems in the HEAT set as well as the performance of the whole procedure HEAT'345-(Q) in comparison to the ATcT data.

	HEAT345-(Q)	$\Delta_{(T)}{}^a$	$\Delta_T{}^b$	$\Delta_{ZPE}{}^c$	HEAT'345-(Q)^d
N_2	0.60	0.48	1.31	0.60	1.19
H_2	0.14	0.15	0.14	0.14	0.15
F_2	0.40	0.83	0.74	0.40	1.17
CO	0.09	-0.01	0.84	0.09	0.74
C_2H_2	-0.22	-0.02	0.52	0.15	1.09
CO_2	0.64	0.80	1.85	0.62	1.99
H_2O_2	-0.19	1.25	0.58	-0.20	2.01
H_2O	0.36	1.38	0.78	0.36	1.80
HCN	0.49	2.65	1.12	0.52	3.31
HF	0.42	1.26	0.70	0.42	1.54
HNO	-0.34	0.15	0.43	-0.33	0.93
NH_3	0.22	1.18	0.56	0.25	1.55
$\bar{\Delta}^e$	0.22	0.84	0.80	0.25	1.46
max	0.64	2.65	1.85	0.62	3.31
Δ^e_{std}	0.33	0.77	0.45	0.30	0.79

[a]CCSD(T)/aug-cc-pCV45Z-CCSD(T)/cc-pCV45Z; [b]CCSDT/cc-pV34Z-CCSDT/cc-pVTZ; [c]sum of harmonic contribution using the cc-pVQZ basis set and the anharmonic contribution from the corresponding calculation with the cc-pVTZ basis set; [d]total approximate HEAT scheme; [e] mean deviation ($\bar{\Delta}$) and standard deviation (Δ_{std}) with respect to the ATcT data.

The approximation made for the CCSD(T) correlation energy has the largest influence on the investigated set of molecules. For 5 of the 12 molecules the "1 kJ/mol accuracy" is lost. Interestingly, the result for the molecule most similar to benzene, namely acetylene, is not affected too much by skipping the augmentation functions in the basis set. A systematic positive shift of about 0.8 kJ/mol is observed for the applied approximation to the ΔE_{CCSDT} contribution. The use of different basis sets for the harmonic and anharmonic contribution to the zero-point energy seem to have modest effects on the computed TAE. Further investigations for the ΔE_{ZPE} contribution is carried out using different quadruple-zeta and triple-zeta quality basis sets. The results are shown in table 6.19.

Table 6.19: Zero-point energy contributions for the benzene molecule calculated with and without the frozen-core approximation applying triple and quadruple-zeta quality basis sets (in kJ/mol).

Method	$\Delta E_{ZPE}^{harmonic}$	$\Delta E_{ZPE}^{anharmonic}$	ΔE_{ZPE}
(fc)CCSD(T)/ANO2	263.0	–	260.0 [a]
(fc)CCSD(T)/ANO1	262.9	-3.0	259.9
(fc)CCSD(T)/cc-pVQZ	262.2	–[b]	260.5 [a]
(fc)CCSD(T)/cc-pVTZ	261.8	-1.7	260.1
(ae)CCSD(T)/cc-pVQZ	263.3	–[b]	260.8 [a]
(ae)CCSD(T)/cc-pVTZ	264.2	-2.5	261.7

[a]Sum of harmonic contribution using the quadruple-zeta basis set and the anharmonic contribution from the corresponding calculation with the triple-zeta quality basis set.
[b]The perturbed lambda equations could not be converged to sufficiently tight convergence criteria.

The quadruple-zeta results for the harmonic contribution, $\Delta E_{ZPE}^{harmonic}$, agree within 1 kJ/mol, while with the triple-zeta basis sets the ZPE is underestimated in the frozen-core approximation and overestimated by the all-electron calculations. The differences in the anharmonic contribution, $\Delta E_{ZPE}^{anharmonic}$, are slightly more than 1 kJ/mol. Hopefully, the final result of an (fc)CCSD(T)/ANO2 calculation will resolve this remaining uncertainty. For the sake of consistency with the definition of the HEAT schemes the results form the all-electron calculations were taken for the current estimate (compare table 6.17). A conservative error estimate for this contribution in comparison to the full HEAT345-(Q) scheme might be 1 kJ/mol, which is in excellent agreement with the semi-empirical best estimate for the ZPE of 259.7 kJ/mol from ref. [183].

Taking into account the first-order spin-orbit contributions derived from the hyperfine structure of the constituent atoms yields a ΔE_{SO} of -2.13 kJ/mol [183], which is only 0.2 kJ/mol higher than the value obtained within the HEAT345-(Q) approach. For extended organic molecules it is important to account for correlation effects on the DBO correction [154]. Ref. [154] gives a value of +0.6 kJ/mol for the contribution to the TAE obtained at the MP2 level of theory which is only 0.4 kJ/mol lower than the HF-SCF estimate. Consideration of the higher-level treatment of the DBO correction and the experimental spin-orbit corrections leads to 5466.8 kJ/mol as best estimate for the TAE of benzene, which is accidentally the same value obtained in ref. [183].

Assessment of the error in the final result remains difficult. The calculated TAE is clearly within the experimental error bar. Statistical analysis from table 6.18 would suggest 1.5 kJ/mol and 0.8 kJ/mol for mean and standard deviation, respectively. A conservative error estimate is to multiply the corresponding error of acetylene by 3 which leads to a value of 3.3 kJ/mol. As can be observed in table 6.18 the final error will be positive. This leads to the final estimate of 5466,0 kJ/mol with an uncertainty of 3.3 kJ/mol which might be written as 5467.7 $\pm$ 1.7 kJ/mol. Consideration of the introduced approximations and the size of the molecule leads to the conclusion that even for a molecule such as benzene the TAE can be obtained with better than chemical accuracy, although it remains open to future work if sub-kJ/mol accuracy can be achieved in this case by *ab initio* thermochemistry schemes.

7 Accurate prediction of nuclear magnetic shielding constants

The nuclear magnetic shielding tensor $\underline{\underline{\sigma}}$ can be calculated via the second derivatives of the energy with respect to an external magnetic field $\mathbf{B}$ and the magnetic moment of the k-th nucleus $\mathbf{m}_k$:

$$\underline{\underline{\sigma}}_k = \left(\frac{\partial^2 E}{\partial \mathbf{m}_k \partial \mathbf{B}} \right)_{\mathbf{m}_k=0, \mathbf{B}=0} \quad . \tag{7.1}$$

Experimentally, isotropic relative NMR chemical shifts δ are determined which are defined by the following relation:

$$\delta = \frac{1}{3} Tr(\underline{\underline{\sigma}}_{ref}) - \frac{1}{3} Tr(\underline{\underline{\sigma}}_k) \; . \tag{7.2}$$

As a consequence for quantum-chemical calculations the shielding tensors for the molecule of interest as well as for the reference compound (e.g., tetramethylsilane (TMS) in the case of ^{1}H and ^{13}C shieldings) have to be determined with the same method and basis set.

The use of finite basis sets for the calculation of magnetic properties causes a gauge-origin dependence of the computed values for magnetic properties [56]. This simply means that the results of the calculations depend on an arbitrarily chosen gauge-origin $\mathbf{R}_0$ for the vector potential of the external magnetic field. One possibility to solve this problem is to multiply the AOs with a phase factor which depends on the magnetic field. This results in perturbation-dependent basis functions $\chi_\nu(\mathbf{B})$,

$$|\chi_\nu(\mathbf{B})\rangle = exp\left(-\frac{ie}{2\hbar} (\mathbf{B} \times [\mathbf{R}_\nu - \mathbf{R}_0]\mathbf{r} \right) |\chi_\nu(\mathbf{0})\rangle \quad , \tag{7.3}$$

with $\mathbf{R}_\nu$ as the center of the basis function χ_ν. These AOs are usually called "Gauge-Including Atomic Orbitals" (GIAOs) [185–190].

In the following sections benchmark calculations for gas-phase ^{19}F nuclear magnetic shielding constants as well as for two larger molecules, namely benzene and the adamantyl cation will be presented.

7.1 Quantitative prediction of gas-phase ^{19}F nuclear magnetic shielding constants

^{19}F NMR spectroscopy has become more and more common besides the well established ^{1}H and ^{13}C NMR techniques [191]. Fluorine-19, at 100 % natural abundance, is a 1/2 spin nucleus and therefore the effort required to obtain ^{19}F NMR spectra is virtually the same as that involved in getting proton NMR data [191]. Because a fluorine nucleus is on average surrounded by nine electrons rather than a single electron as is the case with hydrogen, the range of fluorine chemical shifts and the sensitivity of fluorine chemical shifts to the details of the local environment are much higher. ^{19}F NMR spectroscopy therefore is one of the most important analytic methods in fluorine chemistry.

Quantum-chemical investigations have evolved to be an important tool for the interpretation of experimentally obtained NMR spectra [192, 193], and many examples for ^{19}F NMR chemical shifts calculations can be found in the literature [194–202]. Concerning the accuracy of the quantum-chemical determination of NMR chemical shifts, a recent benchmark investigation of gas-phase ^{13}C nuclear magnetic shielding constants [83] has demonstrated that near quantitative accuracy can be achieved if state-of-the-art techniques are employed: (1) calculations are performed at accurate equilibrium geometries, (2) gauge-including atomic orbitals (GIAOs) [185–190] are used, (3) large (uncontracted) basis sets are employed, (4) electron correlation is adequately treated by employing the CCSD(T) model, and (5) vibrational zero-point corrections are treated by second-order vibrational perturbation theory (VPT2) [142]. However, requirements for electron-correlated calculations on fluorine compounds are higher than for carbon because of the larger number of electrons surrounding the nucleus and, in particular, because of the free electron pairs. This can be seen in the difficulties in obtaining reasonable results for F_2 [203] and oxygen-fluorine compounds such as OF_2. Therefore, the importance of electron-correlation contributions to nuclear magnetic shielding constants is expected to be larger than for ^{13}C shieldings [83]. Due to the quite large range of fluorine shielding constants of about 800 ppm theoretical predictions with a deviation in 5 ppm of the experimental value can be regarded as excellent. The promising results of the ^{13}C NMR benchmark investigation [83] and the availability of a set of gas-phase ^{19}F NMR data [204,205] motivated us for a similar study as in ref. [83] for ^{19}F NMR chemical shifts. However, as temperature effects are rather large for F_2 [203] those will be, unlike in ref. [83], included in the present investigation.

7.1.1 Computational details

^{19}F nuclear magnetic shielding constants have been computed using a variety of theoretical approaches for a set of 28 molecules which is shown in fig. 7.1. It has been chosen for two reasons: First, for many of the molecules in this set experimental gas-phase ^{19}F NMR data are available for comparison [204, 205]. Second, the set includes molecules with fluorine bonded to hydrogen, carbon, oxygen, nitrogen, and fluorine, as well as different neighboring groups.

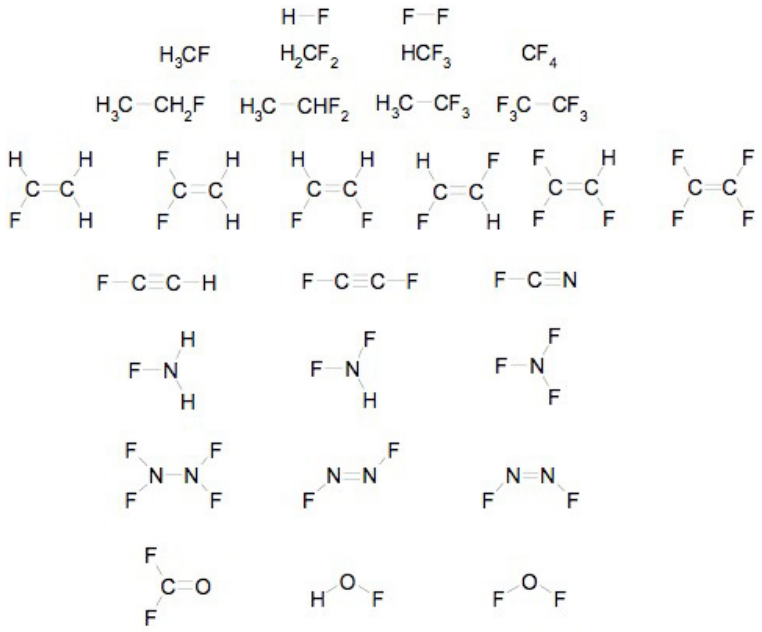

Figure 7.1: Molecules considered in the investigation of ^{19}F nuclear magnetic shielding constants.

For the investigation of electron-correlation effects, shielding constants have been calculated at the following levels of theory: (1) HF-SCF [190], (2) MP2 [206], (3) CCSD [207], and (4) CCSD(T) [208]. For the two diatomic molecules HF and F_2, additional calculations at the CCSDT [209] and CCSDTQ [66] level have been carried out. In addition, shielding constants have been computed within the framework of DFT using the BP86 functional [55, 210] and the B3LYP hybrid functional [211] as implemented in the TURBOMOLE package [212].

For all employed computational levels the shielding tensor has been evaluated via the corresponding second derivatives of the energy with respect to external magnetic field and nuclear magnetic moments (compare eq. (7.1)). GIAOs have been employed to provide rapid basis-set convergence and to ensure gauge-origin independence. Previous experience (see e.g., [83]) indicates that the basis sets developed by Schäfer et al. [84] augmented with polarization functions [82] are well suited for the calculation of chemical shifts. Basis-set convergence is monitored by performing shielding calculations with a series of different sets derived from the original work of Schäfer et al. in ref. [84]. From these, a "triple-zeta polarization (tzp)" type of basis, larger contracted sets denoted as tz2p, qz2p, and pz3d2f as well as the large uncontracted 13s9p4d3f/8s3p2d and 15s11p4d3f/10s4p3d sets are employed.

Since chemical shieldings are often very sensitive to the geometry chosen for the calculation and highly accurate experimental equilibrium geometries are not available for all molecules considered here, theoretical geometries obtained at the CCSD(T) level using the cc-pVTZ basis set correlating all electrons are used.

The importance of zero-point vibrational corrections is investigated by computing the corresponding contributions via VPT2 [142]. For this purpose, the shielding tensor $\underline{\underline{\sigma}}$ is expanded around the equilibrium geometry r_e in a Taylor series up to second order in terms of the normal coordinates Q_r. The expectation value

$$\underline{\underline{\sigma_0}} = \underline{\underline{\sigma}}(r_e) + \sum_r \left(\frac{\partial \underline{\underline{\sigma}}}{\partial Q_r} \right)_{Q_r=0} \langle Q_r \rangle + \frac{1}{2} \sum_{r,s} \left(\frac{\partial^2 \underline{\underline{\sigma}}}{\partial Q_r Q_s} \right)_{Q_r=0, Q_s=0} \langle Q_r Q_s \rangle \tag{7.4}$$

then yields the vibrationally averaged shielding tensor σ_0 if the proper expressions for $\langle Q_r \rangle$ and $\langle Q_r Q_s \rangle$ are used:

$$\langle Q_r \rangle = -\frac{\hbar}{4\omega_r^2} \sum_s \frac{k_{rss}}{\omega_s} \tag{7.5}$$

and

$$\langle Q_r Q_s \rangle = \delta_{rs} \frac{\hbar}{2\omega_r} \tag{7.6}$$

from ref. [142] with ω_r as the harmonic frequency of the r-th normal mode and k_{rst} as the corresponding cubic force constants. Eq. (7.6) corresponds to the usual harmonic-oscillator (zeroth-order) expression for the square displacement of Q_r, while eq. (7.5) is the corresponding first-order expression for the expectation value of Q_r within a perturbative treatment of anharmonic effects.

Temperature effects are taken into account by replacing the expectation values for Q_r and Q_r^2 by the corresponding expressions for finite temperatures T:

$$\langle Q_r \rangle = -\frac{\hbar}{4\omega_r^2} \left[\sum_s \frac{k_{rss}}{\omega_s} \; coth\left(\frac{\hbar\omega_s}{2k_B T}\right) - 2k_B T \sum_\alpha \frac{a_r^{(\alpha\alpha)}}{I_{\alpha\alpha}} \right] \tag{7.7}$$

and

$$\langle Q_r Q_s \rangle = \delta_{rs} \frac{\hbar}{2\omega_r} \; coth\left(\frac{\hbar\omega_r}{2k_B T}\right), \tag{7.8}$$

where k_B is the Boltzmann constant. $I_{\alpha\alpha}$ denotes the moment of inertia of the molecule with respect to the α-th principal axis, and $a_r^{(\alpha\alpha)}$ is the first coefficient for the expansion of $I_{\alpha\alpha}$ in terms of the normal coordinate Q_r. The additional *coth* terms in eq. (7.7) and (7.8) are due to the population of excited vibrational levels, while the second term in eq. (7.7) accounts for the population of higher rotational states. This term usually provides the main contribution and is here treated using a classical approximation [213].

The required cubic force fields and property derivatives are computed by numerical differentiation of analytic second derivatives with respect to the normal coordinates Q_r (for further details, see ref. [214, 215]). Vibrational corrections are computed at HF-SCF level using a doubly polarized triple-zeta (tz2p) basis as well as at the MP2 level using the cc-pVTZ basis for the force fields and the qz2p basis for the shielding calculations.

A rigorous comparison of calculated nuclear magnetic shieldings with experimental data necessitates a suitable computational approach. Therefore, the initial discussion will focus on the dependence of computed shieldings on molecular geometry (following subsection), electron-correlation (subsection 7.1.3) and basis set effects (subsection 7.1.4), as well as the effect of vibrational averaging and corrections for finite temperature (subsection 7.1.5). The last subsection is then devoted to a comparison of the different theoretical results with the experimental gas-phase data [204, 205].

7.1.2 Geometry dependence

Table 7.1 illustrates the dependence of the computed ^{19}F shielding constants on the choice of equilibrium geometries. CCSD(T)/cc-pVQZ geometries have been shown to be very reliable

for first-row compounds [216, 217] with errors (in comparison to accurate experimental r_e geometries) of 0.002 to 0.003 Å in the bond distances and a few tenths of degrees in the bond angles. Errors in CCSD(T)/cc-pVTZ geometries are only marginally larger, and therefore those geometries have been adopted as the standard choice in the current investigation. Compared to CCSD(T) geometries, the quality of the MP2/cc-pVTZ structures is substantially lower, with errors up to 0.006 to 0.007 Å in the bond distances. It is nevertheless useful to include in the current study the MP2/cc-pVTZ geometries, since the use of MP2 geometries often represents a cost-effective compromise.

Table 7.1: Dependence of computed ^{19}F nuclear magnetic shielding constants (in ppm) on the employed equilibrium geometry. All calculations have been performed at the CCSD(T) level using the qz2p basis set

.

Geometry	MP2/cc-pVTZ	CCSD(T)/cc-pVTZ	CCSD(T)/cc-pVQZ
HF	418.0	418.4	418.8
CFH_3	480.7	479.7	479.5
CH_2F_2	351.9	351.4	351.3
CHF_3	287.0	286.9	287.2
CF_4	269.6	269.8	270.7
CFCH	425.9	425.7	426.6
CFCF	478.4	477.6	478.5
F_2CO	232.9	233.9	235.2
NH_2F	400.5	394.8	395.4
NHF_2	224.7	219.7	220.7
NF_3	87.3	81.5	87.7
FN=NF(cis)	100.8	96.6	97.9
FN=NF(trans)	138.4	133.6	134.9
HOF	204.6	185.8	188.3
OF_2	-2.7	-18.9	-13.5
F_2	-140.7	-189.0	-182.1

Differences in the results for the MP2/cc-pVTZ and CCSD(T)/cc-pVTZ geometries are less than 1 ppm for all molecules containing only fluorine-carbon or fluorine-hydrogen bonds. For molecules where fluorine is bonded to nitrogen the change is about 6 ppm, while MP2 cannot describe geometries of molecules containing oxygen-fluorine and fluorine-fluorine bonds accurately enough. The corresponding differences in the computed shifts are larger than 16 ppm.

Changes when going from CCSD(T)/cc-pVTZ to CCSD(T)/cc-pVQZ geometries are of the same size (i.e., about 0.4 to 0.9 ppm) for fluorine bonded to carbon and hydrogen. The corresponding change for F_2 amounts to +6.9 ppm, which is much smaller than the difference of -48.3 ppm between MP2 and CCSD(T). When discussing the geometry dependence of computed shieldings, it is also important to realize that the bond distance variation is only one of two relevant parameters. The change in the computed shieldings also depends on the sensitivity of the shielding to the geometry change which can be, for example, monitored via the corresponding geometrical derivatives of the shieldings [218] and seems to be more pronounced for ^{19}F than ^{13}C.

7.1.3 Electron correlation

Electron-correlation effects have been investigated in shielding calculations at the HF-SCF, MP2, CCSD, and CCSD(T) level, as well as at two DFT levels employing the CCSD(T)/cc-pVTZ optimized geometries and the qz2p basis set. Table 7.2 summarizes the corresponding nuclear magnetic shielding constants. As CCSD(T) calculations have been shown to provide near-quantitative accuracy for energies and other properties [24], the CCSD(T)/qz2p values will be used as reference for comparison.

Table 7.2: Comparison of different quantum-chemical methods in the calculation of ^{19}F nuclear magnetic shielding constants (in ppm). All calculations have been performed with the qz2p basis set using CCSD(T)/cc-pVTZ optimized geometries.

	HF-SCF	MP2	CCSD	CCSD(T)	BP86	B3LYP
HF	413.7	423.8	418.1	418.4	409.9	410.4
CH_3F	486.6	486.6	481.2	479.7	460.9	465.2
CH_2F_2	367.2	353.9	355.9	351.4	317.4	326.3
CHF_3	303.4	287.7	291.8	286.9	249.0	258.9
CF_4	281.6	270.8	273.6	269.8	234.0	242.3
CH_3CH_2F	439.2	430.4	429.1	425.7	397.5	404.3
CH_3CHF_2	341.1	323.4	327.5	322.2	284.9	295.5
CH_3CF_3	290.8	273.0	278.1	273.0	234.2	245.4
CF_3CF_3	312.1	296.8	300.7	296.0	257.9	268.6
CH_2=CHF	342.8	325.8	333.7	327.3	262.9	282.3
CFH=CFH(cis)	384.5	376.1	378.0	373.5	321.2	335.4
CFH=CFH(trans)	413.9	404.4	404.4	399.9	350.1	365.1
CH_2=CF_2	302.7	289.7	296.8	291.7	237.2	252.2
CF2=CHF(E)	350.0	337.8	341.5	336.8	287.4	301.1
CF2=CHF(Z)	319.8	308.7	314.2	309.5	257.0	270.5
CF_2=CHF	431.5	423.2	420.6	416.5	372.9	385.2
CF_2=CF_2	357.3	344.5	347.3	342.5	294.7	307.5
CFCH	428.4	431.8	428.1	425.7	390.6	400.4
CFCF	479.4	486.2	478.6	477.6	452.6	458.4
F_2CO	238.1	237.6	237.9	233.9	183.0	194.5
CF_3CN	281.6	264.9	270.4	265.0	221.4	233.8
NH_2F	438.8	407.8	400.5	394.8	367.2	379.9
NHF_2	254.6	225.9	228.7	219.7	169.8	183.4
NF_3	111.0	81.7	92.7	81.5	15.7	30.1
NF_2NF_2	186.3	159.5	167.4	158.1	102.5	114.3
FN=NF(cis)	96.2	99.2	105.8	96.6	9.9	23.0
FN=NF(trans)	147.3	137.3	143.6	133.6	51.9	68.9
HOF	285.6	190.0	197.1	185.8	147.9	171.4
OF_2	20.6	-8.3	-5.6	-18.9	-86.8	-71.3
F_2	-177.3	-171.3	-175.2	-189.0	-269.4	-258.7
$\bar{\Delta}^a$	-18.1	-3.8	-5.9		46.7	34.7
Δ^a_{std}	18.9	4.4	3.5		18.2	15.6

[a] Mean deviation ($\bar{\Delta}$) and standard deviation (Δ_{std}) with respect to the CCSD(T) data.

Fig. 7.2 shows the difference of the results obtained at the HF-SCF, MP2, and CCSD level from the CCSD(T) results. The difference between CCSD(T) and DFT results is displayed in fig. 7.3.

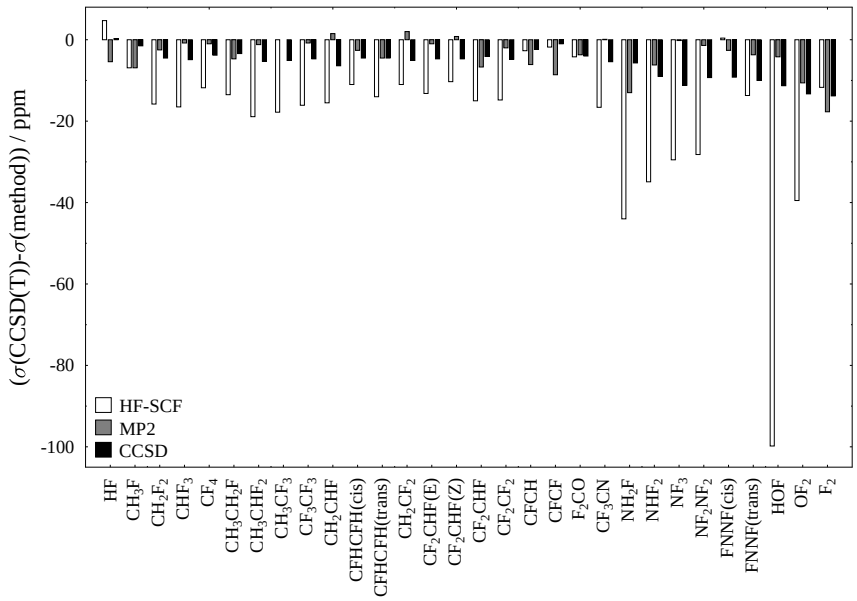

Figure 7.2: Deviation of ^{19}F nuclear magnetic shielding constants calculated using HF-SCF, MP2 and DFT methods from the CCSD(T) reference data. All results have been obtained for CCSD(T)/cc-pVTZ optimized geometries using the qz2p basis set.

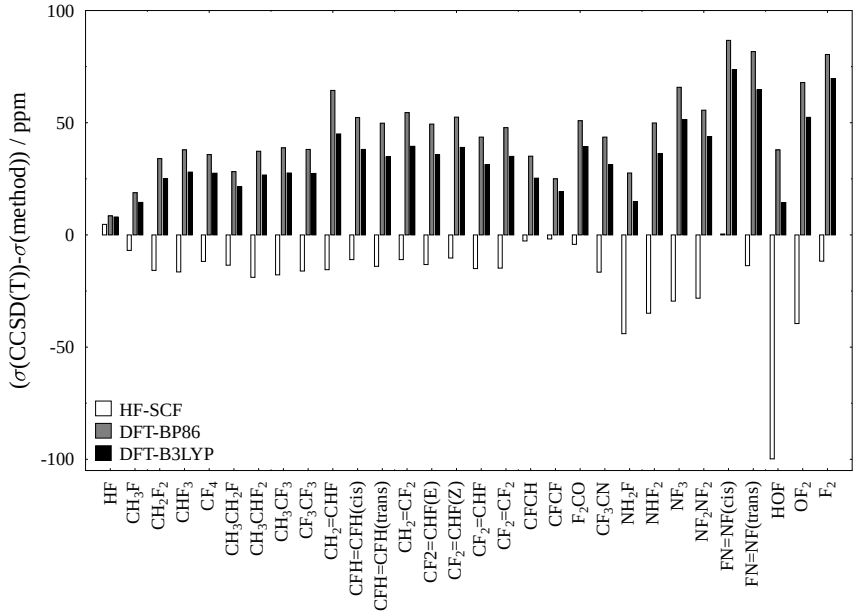

Figure 7.3: Deviation of ^{19}F nuclear magnetic shielding constants calculated using HF-SCF and DFT methods from the CCSD(T) reference data. All results have been obtained for CCSD(T)/cc-pVTZ optimized geometries using the qz2p basis set.

The differences between CCSD(T) and HF-SCF (as a measure for electron correlation)

amount to as much as 100 ppm for the molecules considered here (compare fig. 7.2). The correlation contributions are smaller (less than 20 ppm) for fluorine involved exclusively in bonds to hydrogen, carbon, or fluorine, while they are considerably larger for fluorine bonded to nitrogen. Very large electron-correlation effects of up to 100 ppm are observed in systems with oxygen-fluorine bonds (HOF, OF_2).

MP2 recovers most of the electron-correlation contributions, but also tends to slightly overestimate the latter. CCSD calculations, on the other hand, yield much smaller corrections than MP2 and CCSD(T) in line with the observation that the CCSD approximation generally tends to underestimate electron-correlation effects [24].

It is difficult to assess the ultimate accuracy of the CCSD(T) reference data. Calculations of chemical shieldings at the full CCSDT level suggest that the CCSD(T) model tends to overestimate triple excitation contributions [209]. The accuracy of CCSD(T) calculations is therefore expected to be high, probably comparable to or even better than that provided by CCSDT. However, this conclusion needs to be checked by calculations with explicit consideration of higher excitations.

Recently, the implementation of second derivatives for general coupled-cluster methods has been reported [66], and preliminary results have been presented for the nuclear magnetic shieldings of CO and H_2O up to the full configuration-interaction (FCI) limit within the frozen-core approximation. Using Dunning's cc-pVDZ basis set and the frozen-core approximation the changes between the CCSD(T) and the CCSDTQ level are less than 0.3 ppm. The CCSDTQ results turned out to be more or less equal to the FCI results. The lack of results for ^{19}F shielding constants, the frozen-core approximation and especially the use of the (for the calculation of shieldings not well suited) cc-pVDZ basis set motivated us to further investigate the effect of higher excitations.

Table 7.3 presents all-electron calculations for the ^{19}F shieldings of HF and F_2 up to the CCSDTQ level employing the tzp basis set.

Table 7.3: Comparison of the electron-correlation treatment of different quantum-chemical methods in the calculation of ^{19}F nuclear magnetic shielding constants for HF and F_2 (in ppm). All calculations have been performed with the tzp basis set using CCSD(T)/cc-pVTZ optimized geometries.

			HF	F_2
σ(HF-SCF)			414.02	-161.75
σ(MP2)	-	σ(HF-SCF)	+9.53	+9.01
σ(CCSD	-	σ(MP2)	-5.59	-7.13
σ(CCSD(T))	-	σ(CCSD)	+0.06	-13.36
σ(CCSDT)	-	σ(CCSD(T))	-0.08	-3.13
σ(CCSDTQ)	-	σ(CCSDT)	+0.01	-1.57

The convergence behavior with respect to the applied excitation level in the coupled-cluster model is very different for the two considered molecules. While the HF molecule is treated very accurately at the CCSD level and the remaining error in the treatment of electron-correlation effects is smaller than 0.1 ppm, the shielding for F_2 still shows a large change of -1.57 ppm when going from the CCSDT to the CCSDTQ level. Unfortunately, both cases considered here are not comparable to the other molecules in the benchmark set. For both molecules, HF and F_2, MP2 gives results that are worse than the corresponding

CCSD values, which is different from most of the other molecules. Therefore, the MP2 behavior cannot be used as an indicator for the need of a higher-level correlation treatment. Nevertheless, it is noteworthy that electron-correlation effects beyond CCSD(T) can amount to 5 ppm. A further increase in computational resources might enable a more detailed study of these effects, in particular, as similar effects for HOF and OF_2 are expected.

Fig. 7.3 and table 7.2 show that DFT systematically overestimates ^{19}F shielding constants. The deviations from CCSD(T) are uniformly positive (between +8 and +87 ppm). For most of the systems considered here, the error (with respect to the CCSD(T) reference) is even larger than the corresponding deviation obtained at the HF-SCF level. Hence, unlike for many other applications, DFT does not provide a systematic improvement compared to HF-SCF. It should be noted that this finding is much more pronounced for ^{19}F shieldings than for ^{13}C shieldings (compare with ref. [83]).

7.1.4 Basis-set convergence

Table 7.4 reports the results of the basis-set investigation using basis sets ranging from tzp up to large uncontracted sets. To illustrate the convergence behavior, the estimated "basis-set error" is defined as the deviation from the results obtained with the large 13s9p4d3f basis. In this way, a basis-set error of 5 to 15 ppm can be attributed to the tzp set. For the qz2p set, significantly smaller deviations (about 1 to 5 ppm) are observed, while for the larger pz3d2f set the basis-set error is less than 2 ppm. It thus seems that the 13s9p4d3f results can be considered close to the basis-set limit.

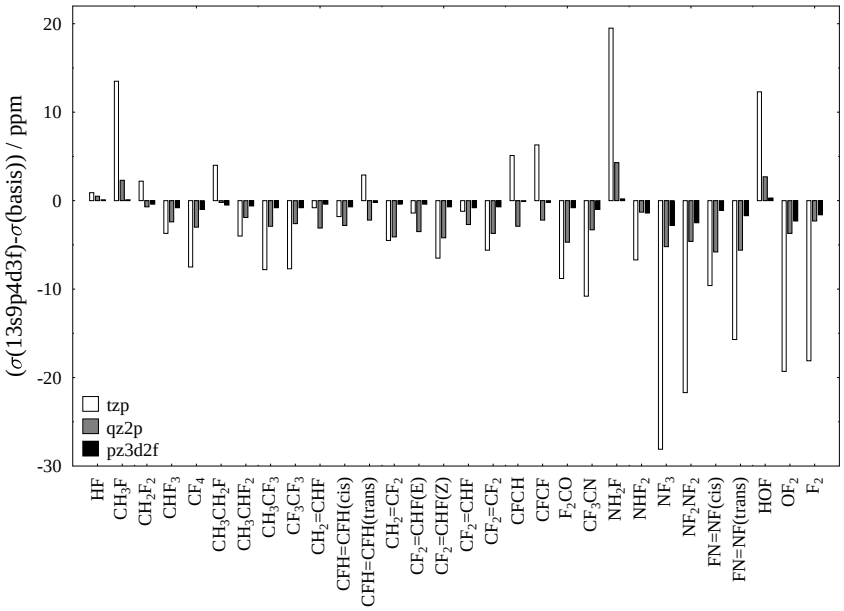

Figure 7.4: Basis-set convergence in the calculation of ^{19}F nuclear magnetic shielding constants. All results have been obtained at the CCSD(T) level employing CCSD(T)/cc-pVTZ optimized geometries.

The remaining basis-set error in the 13s9p4d3f calculations is estimated to be about 0.5 ppm on the basis of a few additional calculations with the larger 15s11p4d3f set (compare table 5). A more detailed analysis of the data presented in table 7.4 and fig. 7.4 reveals that the sign of the "basis-set error" is more or less always the same. Deviations from the 13s9p4d3f results are mostly positive and improvement of the basis leads to a decrease in the computed shieldings. This systematic behavior explains why accurate relative chemical shifts are often obtained with modest basis sets, while the accurate prediction of nuclear magnetic shielding constants requires very large basis sets [203]. The error of the tzp basis shows a different sign for eight molecules in the benchmark set, therefore at least an additional polarization function (tz2p) or the qz2p basis set is recommended for the calculation of ^{19}F shieldings.

Table 7.4: Basis-set convergence of quantum-chemical calculations of ^{19}F nuclear magnetic shielding constants (in ppm). All calculations have been performed at the CCSD(T) level using CCSD(T)/cc-pVTZ optimized geometries.

	tzp	qz2p	pz3d2f	13s9p4d3f	15s11p4d3f
HF	418.0	418.4	418.8	418.9	419.1
CH_3F	468.5	479.7	481.9	482.0	482.0
CH_2F_2	348.5	351.4	351.1	350.7	
CHF_3	288.2	286.9	285.3	284.5	
CF_4	274.3	269.8	267.8	266.8	
CH_3CH_2F	421.6	425.8	426.1	425.6	
CH_3CHF_2	324.3	322.2	320.9	320.3	
CH_3CF_3	277.9	273.0	270.9	270.1	
CF_3CF_3	301.1	296.0	294.2	293.4	
CH_2=CHF	325.0	327.3	324.6	324.2	
CFH=CFH(cis)	372.5	373.5	371.4	370.7	
CFH=CFH(trans)	394.8	399.9	397.9	397.7	
CH_2=CF_2	292.1	291.7	288.0	287.6	
CF_2=CHF(E)	334.7	336.8	333.7	333.3	
CF_2=CHF(Z)	311.8	309.5	306.0	305.3	
CF_2=CHF	415.0	416.5	414.6	413.8	
CF_2=CF_2	344.4	342.5	339.5	338.8	
CFCH	417.7	425.7	422.9	422.8	422.8
CFCF	469.1	477.6	475.6	475.4	475.4
F_2CO	238.0	233.9	230.0	229.2	
CF_3CN	272.5	265.0	262.7	261.7	
NH_2F	379.6	394.8	398.9	399.1	
NHF_2	225.1	219.7	219.8	218.4	
NF_3	104.4	81.5	79.1	76.3	
NF_2NF_2	175.2	158.1	156.0	153.5	
FN=NF(cis)	100.4	96.6	91.9	90.8	
FN=NF(trans)	143.7	133.6	129.7	128.0	
HOF	176.2	185.8	188.2	188.5	188.4
OF_2	-3.3	-18.9	-20.3	-22.6	-25.0
F_2	-173.2	-189.0	-189.7	-191.3	-193.2
$\bar{\Delta}^a$	-4.2	-2.4	-0.8		
Δ^a_{std}	10.4	2.4	0.8		

[a] Mean deviation ($\bar{\Delta}$) and standard deviation (Δ_{std}) with respect to the 13s9p4d3f data.

7.1.5 Vibrational corrections and temperature effects

Table 7.5 summarizes the computed zero-point vibrational corrections to the ^{19}F shieldings. It appears that for ^{19}F shieldings these corrections lie in a range of -5 to -21 ppm and therefore are far from negligible when considering absolute shieldings. As the corrections are highly systematic (i.e., in all considered cases negative) the corresponding effects on relative chemical shifts are smaller. However, even then vibrational effects are not negligible for molecules containing nitrogen-fluorine or oxygen-fluorine bonds.

Comparison of the vibrational corrections computed at HF-SCF and MP2 level show that these corrections are underestimated at the former level, which is contrary to previous results for ^{13}C shieldings [83].

Table 7.5: Computed zero-point vibrational (ZPVC) and temperature corrections (TC) for T = 300 K for ^{19}F nuclear magnetic shielding constants (in ppm).

Chemical shifts:	HF-SCF/tz2p		MP2/qz2p	
Force field:	HF-SCF/tz2p		MP2/cc-pVTZ	
	ZPVC	TC	ZPVC	TC
HF	-8.11	-0.31	-8.22	-0.34
CH_3F	-7.94	-0.19	-8.81	-0.28
CH_2F_2	-8.34	-0.21	-9.67	-0.31
CHF_3	-7.53	-0.24	-8.91	-0.37
CF_4	-5.76	-0.35	-6.83	-0.51
CH_3CH_2F	-9.99	-0.50	-11.88	-0.66
CH_3CHF_2	-8.88	-0.68	-10.83	-0.84
CH_3CF_3	-7.02	-0.86	-8.66	-1.06
CF_3CF_3	-6.22	-1.16	-7.51	-1.33
CH_2=CHF	-9.47	-0.25	-9.95	-0.40
CFH=CFH(cis)	-7.65	-0.45	-8.63	-0.63
CFH=CFH(trans)	-6.68	-0.41	-7.83	-0.62
CH_2=CF_2	-6.89	-0.28	-7.79	-0.43
CF_2=CHF(E)	-5.61	-0.38	-6.48	-0.59
CF_2=CHF(Z)	-6.14	-0.40	-6.74	-0.56
CF_2CHF	-5.95	-0.54	-7.24	-0.78
CF_2CF_2	-5.37	-0.56	-6.39	-0.86
CFCH	-6.41	-0.31	-7.59	-1.02
CFCF	-4.54	-0.52	-5.91	-2.06
F_2CO	-5.68	-0.29	-6.12	-0.41
CF_3CN	-6.47	-0.62	-7.13	-0.86
NH_2F	-7.50	-0.38	-12.22	-0.85
NHF_2	-10.37	-0.59	-15.37	-1.15
NF_3	-12.10	-0.63	-17.66	-1.38
NF_2NF_2	-11.35	-1.14	-17.03	-2.02
FN=NF(cis)	-12.63	-0.88	-16.14	-1.64
FN=NF(trans)	-11.24	-0.98	-14.43	-1.70
HOF	-6.34	-0.55	-17.79	-1.77
OF_2	-12.28	-1.17	-21.19	-2.49
F_2	-13.87	-1.57	-20.77	-2.86

The vibrational corrections given here for the HF-SCF level are in good agreement with the values reported in ref. [200]. The observed differences for fluoromethane, difluoromethane,

trifluoromethane, tetrafluoromethane, fluoroethane, 1,1,1-trifluoroethane, hexafluoroethane, fluoroethene, trans-1,2-difluoroethene, and 1,1-difluoroethene are below 1 ppm. For 1,1-difluoroethane, cis-1,2-difluorethene, and trifluoroethene larger deviations of about 2 ppm are found. The latter differences might be explained with the slightly different perturbational approach used in ref. [200] for the description of zero-point vibrational effects based on a Taylor expansion around an effective geometry. It should also be noted that a smaller basis set has been used in ref. [200].

The MP2 level is clearly a reasonable compromise, since the corresponding CCSD(T) calculations are very demanding and the differences in vibrational corrections computed at MP2 and CCSD(T) are generally small [83], at least for molecules without nitrogen-fluorine, oxygen-fluorine, or fluorine-fluorine-bonds. Furthermore, one has to consider errors due to the perturbational treatment of the vibrational corrections which, however, amount to a few percent only and thus are probably negligible.[1]

Temperature corrections are typically one order of magnitude smaller than the corresponding vibrational corrections [203, 219]. This is confirmed by the calculations presented in table 7.5. All temperature corrections are negative and vary from -0.2 to -2.9 ppm. HF-SCF and MP2 give similar results, while results obtained with HF-SCF are always somewhat smaller than the corresponding MP2 values. Temperature corrections for molecules containing nitrogen-fluorine, oxygen-fluorine, or fluorine-fluorine bonds are expected to be underestimated, which is supported by the comparison of the temperature correction of F_2 of -4.64 ppm from ref. [203] with the value of -2.86 ppm obtained in the present study. Temperature corrections are small in general and they might be neglected especially for relative ^{19}F shieldings. However, they can be obtained without additional computational cost if vibrational corrections are calculated within a perturbative scheme.

7.1.6 Comparison with experimental gas-phase data

After calibration of the applied computational methods a detailed comparison between theory and experiment is presented in the following subsection. Even though relative shifts are more relevant from a chemical point of view, absolute shieldings will be considered first. Table 7.6 presents the best theoretical estimates in comparison with the experimental data. The calculated estimates are obtained by augmenting CCSD(T)/13s9p4d3f results (see table 7.4) with the zero-point vibrational and temperature corrections obtained at the MP2 level (see table 7.5). The corresponding experimental values are derived from the gas-phase ^{19}F NMR shifts reported in ref. [204, 205] using $\sigma_{T=300K}(\mathrm{HF}) = 409.6 \pm 1.0$ ppm [203].

[1] Numerical calculations of vibrational averaged shieldings on a large set of diatomic and triatomic molecules indicate that the error of the perturbational treatment is well below a few % of the corresponding correction. (M.E. Harding, J. Vázquez, G. Diezemann, and J. Gauss, to be published).

Table 7.6: Comparison of ^{19}F nuclear magnetic shielding constants (in ppm) obtained from theory and experiment for $T = 300$ K

	Theory[a]	Experiment[b]		Theory[a]	Experiment[b]
HF	410.3	409.6	CF_2CHF	405.8	403.8
CH_3F	472.9	470.6	CF_2CF_2	331.6	330.4
CF_2H_2	340.7	338.7	CFCH	414.2	—
CF_3H	275.2	273.7	CFCF	467.4	—
CF_4	259.5	258.6	F_2CO	222.7	221.2
CH_3CH_2F	413.1	—	CF_3CN	253.7	251.7
CH_3CHF_2	308.6	306.0	NH_2F	386.0	—
CH_3CF_3	260.4	258.3	NHF_2	201.9	—
CF_3CF_3	284.6	282.6	NF_3	57.3	49.9
CH_2CHF	313.9	—	NF_2NF_2	134.5	—
CHF=CHF (cis)	361.4	—	FNNF (cis)	73.0	—
CHF=CHF (trans)	389.3	—	FNNF (trans)	111.9	—
CH_2CF_2	277.4	278.6	HOF	168.9	—
CF_2CHF(E)	326.2	324.9	OF_2	-46.3	—
CF_2CHF(Z)	298.0	297.1	F_2	-214.9	-233.2

[a] CCSD(T)/13s9p4d3f calculations augmented by vibrational and temperature corrections
[b] obtained from experimental relative shifts [204, 205] using $\sigma_{T=300\,K}(\mathrm{HF}) = 409.6 \pm 1.0$ ppm [203].

The deviations of the theoretical estimates from the experimental reference data are shown in fig. 7.5.

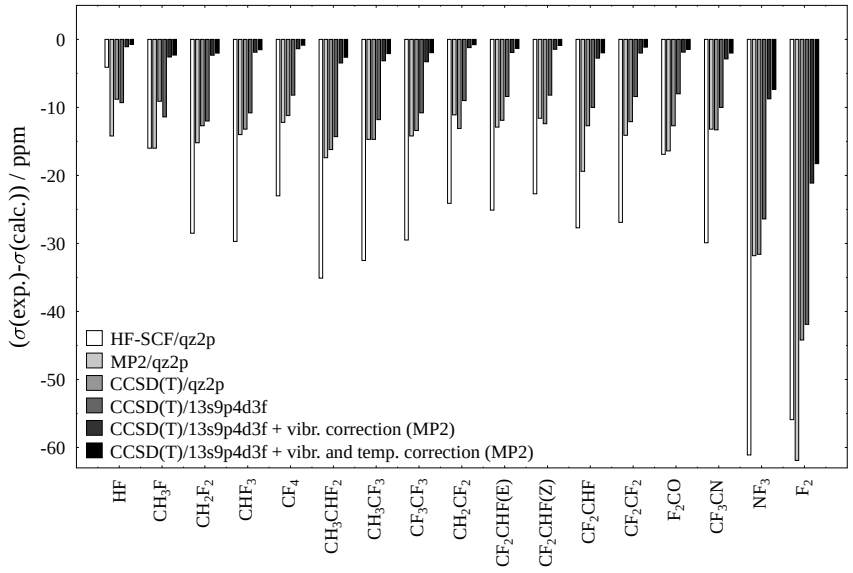

Figure 7.5: Comparison of experimental and calculated ^{19}F nuclear magnetic shielding constants.

To demonstrate the importance of the various contributions, the deviations for the HF-SCF/qz2p, MP2/qz2p, CCSD(T)/qz2p, as well as the CCSD(T)/13s9p4d3f results with and without consideration of vibrational and temperature effects are shown.

With the F_2 molecule being a well-known problem in electronic-structure theory [24], F_2 is excluded from the analysis and will be discussed separately in the following paragraph. Table 7.6 and fig. 7.5 clearly reveal the high accuracy achieved in the present calculations.

Table 7.7: Statistical analysis of the errors (in ppm) in the calculated ^{19}F nuclear magnetic shielding constants in comparison with experiment excluding the F_2 molecule. Numbers in parentheses show the statistical analysis including F_2.

	Mean deviation	Standard deviation
HF-SCF/qz2p	-27.1 (-28.7)	11.8 (13.4)
MP2/qz2p	-15.5 (-18.3)	4.8 (12.2)
CCSD/qz2p	-18.3 (-20.6)	7.3 (11.9)
CCSD(T)/qz2p	-13.7 (-15.5)	5.1 (8.9)
CCSD(T)/pz3d2f	-11.8 (-13.7)	5.0 (9.1)
CCSD(T)/13s9p4d3f	-11.1 (-12.9)	4.4 (8.9)
CCSD(T)/13s9p4d3f + vibr. corrections (HF)	-2.1 (-2.5)	1.4 (1.3)
CCSD(T)/13s9p4d3f + vibr. and temp. corrections (HF)	-3.5 (-4.8)	2.9 (6.2)
CCSD(T)/13s9p4d3f + vibr. corrections (MP2)	-2.6 (-3.7)	1.8 (4.8)
CCSD(T)/13s9p4d3f + vibr. and temp. corrections (MP2)	-1.9 (-2.9)	1.6 (4.2)
BP86/qz2p	27.4 (27.9)	11.3 (11.1)
B3LYP/qz2p	16.8 (17.3)	7.8 (7.8)
HF-SCF/qz2p + vibr. and temp. corrections (HF)	-19.5 (-20.7)	10.8 (11.6)
MP2/qz2p + vibr. and temp. corrections (HF)	-7.9 (-10.2)	3.6 (10.8)
MP2/qz2p + vibr. and temp. corrections (MP2)	-7.1 (-9.1)	2.8 (8.7)
BP86/qz2p + vibr. and temp. corrections (HF)	34.5 (34.4)	10.8 (11.1)
BP86/qz2p + vibr. and temp. corrections (MP2)	35.8 (37.1)	11.2 (12.0)
B3LYP/qz2p + vibr. and temp. corrections (HF)	23.9 (24.8)	7.3 (8.0)
B3LYP/qz2p + vibr. and temp. corrections (MP2)	25.2 (26.5)	7.7 (9.0)

A statistical analysis of the data (see table 7.7) shows that the mean deviation from experiment is only -1.9 ppm for the best theoretical estimates; the standard deviation is 1.6 ppm. The analysis also demonstrates the importance of the individual contributions: inclusion of electron correlation changes the mean deviation from -27.1 ppm (HF-SCF/qz2p) to -13.7 ppm (CCSD(T)/qz2p) while the standard deviation is reduced from 11.8 to 5.1 ppm. A further decrease in the mean deviation (in absolute terms) is achieved in the CCSD(T) calculations by using larger basis sets (-13.7 ppm for CCSD(T)/qz2p, -11.8 ppm for CCSD(T)/pz3d2f, and -11.1 ppm for CCSD(T)/13s9p4d3f) as well as by including vibrational corrections (-11.1 ppm for CCSD(T)/13s9p4d3f without vibrational corrections and -2.6 ppm for the same computational level with consideration of zero-point vibrational effects at the MP2 level). Even the inclusion of the rather small temperature corrections calculated at the MP2 level further improves the agreement between theory and experiment by 0.7 ppm and 0.2 ppm for mean and standard deviation, respectively. However, consideration of the vibrational and temperature corrections as obtained in HF-SCF calculations leads to an inconsistent picture. Consideration of vibrational effects again significantly improves the statistical measures, while the inclusion of the temperature effects slightly detoriates them. The addition

of vibrational and temperature corrections to the HF-SCF and DFT results has no significant effect on mean and standard deviation. In the case of HF-SCF the mean deviation is reduced by 10 ppm but the standard deviation is only lowered by about 1 ppm. For both functionals employed at the DFT level the mean deviation is increased by 7 to 9 ppm, the standard deviation remains nearly unaltered. At least for the functionals employed, the DFT results cannot be improved by considering a theoretically more complete approach. Because of these findings and the fair quality of the results for absolute and relative ^{19}F shieldings it cannot be recommended to compute ^{19}F NMR chemical shifts at the DFT level using these two functionals.

MP2, in combination with vibrational and temperature corrections obtained at the HF-SCF or MP2 level, yields mean deviations of 7.9 and 7.1 ppm, respectively. The standard deviation is of the order of 3 ppm. For molecules with hydrogen-fluorine or carbon-fluorine bonds a standard deviation of 2.2 ppm is found. The combination of MP2/qz2p shieldings with vibrational and temperature effects computed at the HF-SCF/tz2p level thus provides a good compromise between computational cost and accuracy.

In order to assess the accuracy achieved, a few additional points should be mentioned. First, it is essential to realize that the uncertainty in the experimental ^{19}F NMR scale amounts to 1.0 ppm because of the uncertainty in the experimental ^{19}F shielding constant of HF [203]. However, this error is entirely systematic and for all absolute shieldings the same. Thus, a change in σ(HF) would affect the mean deviation, but not the standard deviation. Second, it appears that nearly all computed values are too low in comparison with experiment (negative mean deviation). On the basis of the current results, it can be assumed that remaining electron-correlation and basis-set effects are small, i.e., at most 2 to 3 ppm. A smaller uncertainty can be attributed to vibrational effects. More difficult to assess are remaining geometry effects, but the error here is (except for a few critical cases such as, for example, F_2) small and amounts to no more than a few tenths of ppm. Overall, it should be noted that the residual total error in the calculations presented here is at most 3 to 4 ppm and, thus, the achieved agreement between theory and experiment is more than satisfactory.
Relative chemical shifts are chemically more important, but more difficult to analyze in a benchmark study. The reason is that conclusions are affected by the arbitrary choice of the reference compound. It is not necessary to choose here as reference the compound for which $\delta = 0$ ppm, as relative shifts can be obtained in a more general manner via $\delta = \sigma_{ref} - \sigma + \delta_{ref}$. δ_{ref} corresponds to the (experimental) relative shift for the reference compound chosen to convert the absolute shieldings to the relative scale. The subtraction of two shielding constants allows for error cancellations but at the same time the use of relative shifts eliminates all errors due to uncertainties in the experimental absolute shielding scale and thus seems to be the preferred choice. The effect of error cancellation is seen in fig. 7.6 where relative shifts with respect to the HF molecule are displayed.

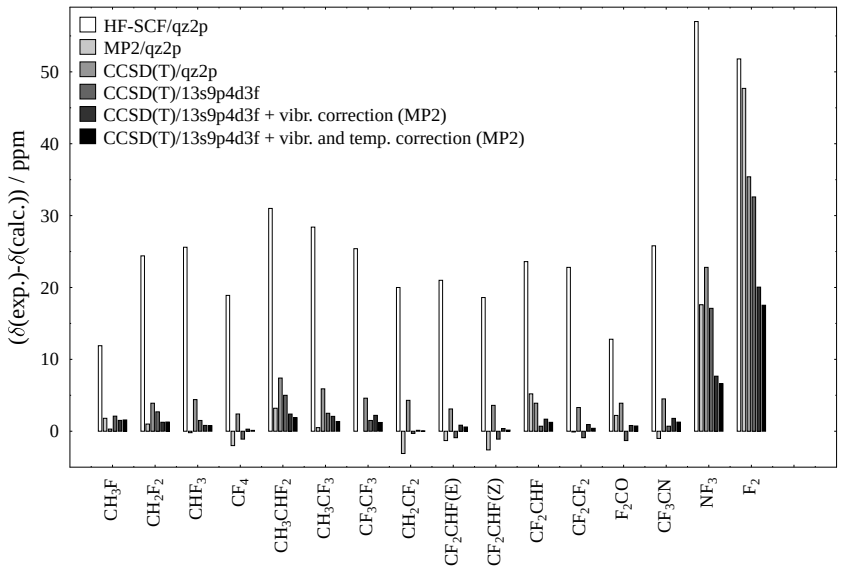

Figure 7.6: Comparison of experimental and calculated relative ^{19}F NMR chemical shifts (δ, vs the HF molecule as reference).

The mean deviation from experiment is reduced, but the standard deviation remains the same as for the absolute shieldings. The mean deviations are now -24.5 (HF-SCF/qz2p), -1.4 (MP2/qz2p), -5.2 (CCSD(T)/qz2p), -1.9 (CCSD(T)/13s9p4d3f), -1.6 ppm (CCSD(T)/ 13s9p4d3f plus vibrational corrections at the MP2 level), and -1.3 ppm when considering temperature effects as well. Interestingly, the DFT results are not significantly better for relative than for absolute shieldings. The corresponding mean deviations are, with -29.5 (BP86) and -18.7 ppm (B3LYP), higher than for all of the conventional quantum-chemical methods.

Basis-set enlargement mainly affects the mean deviation of absolute shieldings (compare table 7.7). It thus appears that reliable predictions of relative shifts are usually possible at lower theoretical levels. This finding, which is consistent with previous experience [83, 203, 220], is reflected by the often successful use of HF-SCF calculations for predicting NMR chemical shifts (see, for example, refs. [220]) as well as the excellent results obtained with MP2 and CCSD(T) with basis sets of relatively modest size (see, for example, ref. [193]).

As has already been seen in previous subsections and noted in ref. [203], F_2 is a difficult case and requires a separate discussion. The experimental result is -233.2 ppm [204], while so far the theoretical estimate is only -214.9 ppm. Table 7.7 reveals that the F_2 molecule as a single case out of 17 has dramatic impact on the statistical analysis and therefore has been taken out of the analysis in the previous subsection. Consideration of zero-point vibrational and temperature corrections at the CCSD(T) level (force field computed with the cc-pVTZ basis, shieldings with the qz2p basis) changes the prediction to -233.0 ppm. The inclusion of full triples and quadruples in the coupled-cluster treatment lowers the value again by 4.7 ppm (assuming additivity of these contributions). Employing the large 15s11p4d3f uncontracted basis set lowers the value by another 2.1 ppm (compare table 7.4). Furthermore, geometry effects of +6.9 ppm (see table 7.1), which cancel the latter contributions should not be

ignored. This yields a total estimate of -222.2 which still differs by -10 ppm from experiment. A quantitative description of the absolute shielding constants of F_2 seems impossible using the approaches considered in this study due to the high sensitivity of the calculated shielding constants on the used geometry, the slow basis set convergence, as well as the rather large vibrational and temperature corrections.

It is noteworthy that both employed DFT functionals describe the F_2 molecule with about the same (poor) quality as all other molecules. This can be seen in the nearly unchanged mean and standard deviation when including or excluding the computed values for F_2

7.1.7 Conclusions

The accurate prediction of ^{19}F nuclear magnetic shielding constants and relative NMR chemical shifts is a challenging task. Nevertheless, this investigation demonstrates that near-quantitative agreement with experiment (about 2 ppm deviation) can be achieved if (1) electron correlation is adequately treated by employing the CCSD(T) model, (2) large (uncontracted) basis sets are used, (3) GIAOs are used to ensure gauge-origin independence, (4) calculations are performed at accurate equilibrium geometries (obtained from CCSD(T)/cc-pVTZ calculations correlating all electrons), (5) vibrational averaging and temperature corrections via VPT2 are included. For the CCSD(T)/13s9p4d3f calculations corrected for vibrational effects, mean and standard deviation from experiment are -1.9 and 1.6 ppm, respectively. Lower-level methods are much less satisfactory. If only molecules with hydrogen-fluorine or carbon-fluorine bonds are considered MP2/qz2p shieldings corrected for vibrational and temperature effects at the HF/tz2p level might offer a good price-performance ratio. The use of DFT with the two popular functionals BP86 and B3LYP, however, cannot be recommended, as results are not even better than the corresponding HF-SCF values. Remaining problem cases are molecules with fluorine-nitrogen and fluorine-oxygen bonds, for which generally a (slightly) larger deviation from experiment is seen. As has been shown for the F_2 molecule, the best currently available theoretical treatments are not sufficient for those cases. It seems that for those systems, coupled-cluster calculations beyond CCSD(T) are needed together with highly accurate equilibrium geometries and reliable treatments of vibrational and temperature effects.

7.2 Benchmark calculation for the ^{13}C NMR chemical shifts of benzene

A previous benchmark study of ^{13}C NMR chemical shifts demonstrated that methods like CCSD(T) can achieve an accuracy of 2-4 ppm [83]. Due to limitations of the serial implementation for the calculation of shieldings and of the computational resources the largest cases considered in ref. [83] were tetrafluoromethane and acetone. One example for which accurate benchmark results are of immediate interest is benzene as computational studies, especially applying density-functional theory, on all kinds of substituted benzene species are abundant in the literature [221–223]. Thus, the parallel algorithm for the calculation of second-order properties has been applied to perform CCSD(T) calculations of NMR chemical shifts of benzene using various basis sets in order to estimate the basis-set limit. Here, the new algorithm allows for very large basis sets even for a system with 12 atoms and 42 correlated electrons.

Table 7.8: ^{13}C NMR chemical shifts (in ppm) for the benzene molecule using various basis sets, the experimental values have been taken from ref. [224] using the absolute shifts of carbon monoxide ($\sigma_{T=300K} = 0.9 \pm 0.9$ ppm) [203], for a detailed description of the scheme for the computation of the zero-point vibrational correction see previous chapter.

Basis	Number of basis functions	$\sigma(^{13}C)$
cc-pCVDZ	138	82.07
cc-pCVTZ	342	66.61
tz2p	198	68.42
qz2p	228	64.95
pz3d2f	474	63.22
vib corr.[a]		-3.43
Total		59.79
Experiment		57.1 [224]

[a] The cubic force field was calculated at the MP2/cc-pVTZ level and the NMR shieldings for the displacements at the MP2/qz2p level.

The results including NMR chemical shifts and also zero-point vibrational corrections are given in table 7.8. An analysis of the basis set convergence leads to the conclusion that the Dunning basis sets which have been optimized for energies from post HF correlation methods and that are fairly diffuse are not very suitable for the calculation of NMR chemical shifts that probe the electron density closer to the nucleus. Even the Dunning core-valence basis sets that are augmented with tight functions do not perform as well as the corresponding Ahlrichs basis sets [82,84] that have been optimized at the HF level of theory and are significantly more compact. Furthermore, comparison to experiment once again demonstrates the importance of the inclusion of zero-point vibrational corrections if one aims at quantitative accuracy in the prediction of NMR chemical shifts.

7.3 NMR chemical shifts of the 1-adamantyl cation

Quantum chemistry provides important tools for the assignment and interpretation of NMR spectra and has shown to be useful in many cases (see e.g., [225–233]). Adamantane has been used extensively for systematic studies of its various cations [234]. From those the bridgehead 1-adamantyl cation, which has been prepared and characterized under long-lived stable conditions in superacidic solution [235], is of particular interest. Surprisingly, the usual quantum-chemical procedures, among them also rather elaborate MP2 calculations, are not able to reproduce the experimental data [235] for the 1-adamantyl cation (fig. 7.7) in a satisfactory manner [236]. As shown in table 7.9, all applied methods predict the correct shielding pattern, but fairly large discrepancies between all theoretical methods and experiment persist. The ^{13}C NMR chemical shifts for the C_1, i.e., the carbocationic center C^+, is reproduced quite well by the BP86 and the MP2 calculations, while for the shieldings of the other carbons larger differences are noted. For C_1 and C_3 B3LYP in combination with a large split-valence basis gives a deviation of more than 14 ppm to the experimental data. The differences for C_2 and C_4 are with about 6 to 7 ppm much smaller. The values obtained at the HF-SCF level show deviations of different sign but similar magnitude in comparison to the B3LYP calculations. The shift for C_4 is accidentally in good agreement with experiment, while the result for the carbon C_1 overshoots the measurement by 32 ppm. Ref. [237] gives similar results for the carbocationic center employing HF-SCF and B3LYP calculations together with much smaller Pople-type split-valence basis sets. A different picture can be drawn for the ^{1}H NMR chemical shifts which show excellent agreement between experiment and all quantum-chemical methods except at the HF-SCF level.

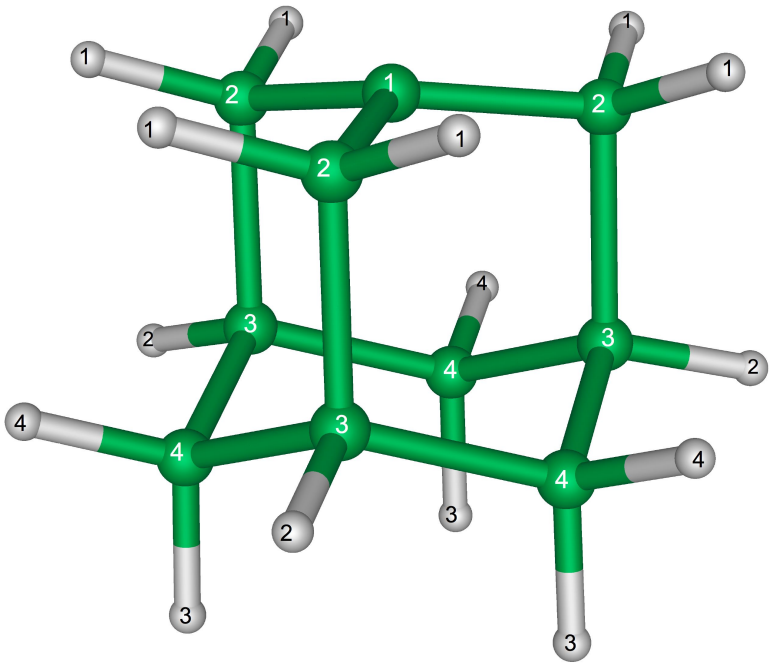

Figure 7.7: The 1-adamantyl cation with the employed numbering scheme.

Table 7.9: NMR chemical shifts (in ppm) of the 1-adamantyl cation computed with various methods. Atom numbering according to fig. 7.7.

Geometry Chemical shifts	BP86/def2-QZVPP BP86/def2-QZVPP	B3LYP/def2-QZVPP B3LYP/def2-QZVPP	HF-SCF/tz2p HF-SCF/tz2p	MP2/cc-pVTZ MP2/qz2p	experiment [235]
C_1	299.9	314.8	332.4	299.7	300
C_2	75.1	73.2	60.9	72.4	65.7
C_3	112.9	104.5	63.9	96.7	86.8
C_4	44.3	42.4	32.2	40.5	34.5
H_1	3.89	3.83	3.75	3.96	4.19
H_2	5.60	5.17	3.74	5.23	5.19
H_3	2.32	2.15	1.81	2.14	2.42
H_4	2.14	2.01	1.74	2.17	2.30

To resolve the unsatisfactory situation a more detailed study is carried out in the following. Since chemical shieldings are often very sensitive to the geometry chosen for the calculation, a (ae)CCSD(T)/cc-pVTZ geometry optimization was carried out.[1] Table 7.10 shows the CCSD(T) results for the geometrical parameters in comparison with HF-SCF and MP2 values obtained with the same basis set.

Table 7.10: Selected geometrical parameters of the 1-adamantyl cation. Atom numbering according to fig. 7.7. Bond lengths in Å, angles in degree.

Parameter	HF-SCF/cc-pVTZ	MP2/cc-pVTZ	CCSD(T)/cc-pVTZ
R(C_1-C_2)	1.457	1.438	1.449
R(C_2-C_3)	1.599	1.611	1.610
R(C_3-C_4)	1.531	1.516	1.524
R(C_2-H_1)	1.078	1.082	1.085
R(C_3-H_2)	1.080	1.087	1.088
R(C_4-H_3)	1.083	1.088	1.089
R(C_4-H_4)	1.083	1.087	1.090
$\sphericalangle$ ($C_1C_2C_3$)	98.4	97.9	98.2
$\sphericalangle$ ($H_1C_2H_1$)	111.1	112.4	112.2
$\sphericalangle$ ($C_2C_3C_4$)	108.7	108.2	108.3
$\sphericalangle$ ($C_3C_4C_3$)	109.7	109.7	109.7
$\sphericalangle$ ($H_3C_4H_4$)	106.9	107.1	107.3
$\sphericalangle$ ($C_2C_3H_2$)	106.0	104.9	105.3

In general, it is observed that for most of the structural parameters MP2 gives larger correlation effects than CCSD(T), defined here as the difference between the individual parameter obtained at the HF-SCF level and with the correlated method. The largest differences with about 0.01 Å are seen for the C_1-C_2 and C_3-C_4 distances and should have a significant influence on the shielding of the carbocationic center. The changes of the carbon-hydrogen bond lengths and the bond angles are negligible.

For the CCSD(T)/cc-pVTZ geometry NMR chemical shift calculations at the MP2 and

[1] Employing the parallel algorithm a geometry optimization cycle within computational C_s symmetry took about 90 hours on 9 compute nodes.

CCSD(T) levels were carried out, the results are presented in table 7.11. For center C_1 a large shift of about 8 ppm is observed at the MP2 level, while the other ^{13}C and ^{1}H shieldings remain nearly unchanged. Comparison to the shieldings calculations at the CCSD(T)/qz2p level shows, as expected, improved agreement with the experimental values. To be more specific the C_1 is shifted 1.7 ppm upfield all other shieldings are about 4 to 5 ppm closer to experiment. Since all presented shieldings are reported with TMS as reference, the question might arise if the cc-pVTZ basis set provides enough flexibility for correlating all electrons in the geometry optimization of the reference compound containing with silicon a second-row atom. Therefore, the effect of dropping seven core electrons and employing a cc-pCVTZ was tested for the chemical shieldings of TMS at the CCSD(T)/qz2p level. Both tests show negligible differences in the shieldings, being +0.04 ppm and -0.68 ppm for ^{13}C, and 0.00 and -0.13 ppm for ^{1}H for the cc-pCVTZ and the partially dropped core, respectively. The qz2p basis set seems to be close enough to the basis-set limit, as it can be concluded from the consideration of the additional values for the dzp and the tzp basis in table 7.11. The changes due to the use of larger basis sets are expected to be on the order of 1-2 ppm for the ^{13}C centers.

Table 7.11: NMR chemical shifts of the 1-adamantyl cation computed at the CCSD(T) and MP2 level using the CCSD(T)/cc-pVTZ geometry. Atom numbering according to fig. 7.7.

Method	CCSD(T)	CCSD(T)	CCSD(T)	experiment	MP2
Basis set	dzp	tzp	qz2p	[235]	qz2p
C_1	290.6	301.0	306.8	300	308.5
C_2	65.1	67.9	69.9	65.7	73.3
C_3	85.6	88.9	91.8	86.8	96.3
C_4	35.6	37.5	36.1	34.5	40.7
H_1	3.73	3.84	3.90	4.19	3.97
H_2	4.72	4.84	4.95	5.19	5.11
H_3	1.95	2.12	2.13	2.42	2.13
H_4	1.93	2.08	2.10	2.30	2.12

Although the CCSD(T)/qz2p computations show acceptable results, additional effects as zero-point vibrational and temperature corrections should be considered. Because zero-point vibrational effects usually are important [83], those were determined applying VPT2 (see section 7.1.1). Unfortunately, high-level treatments of vibrational corrections for a molecule of the size of the adamantyl cation are too demanding at the present time. The use of HF-SCF/tz2p for the required harmonic and anharmonic force fields and the NMR chemical shifts certainly represents a compromise but should give an appropriate description of the zero-point vibrational effects. In addition to that temperature corrections at the same level of theory were evaluated as this could be done without additional computational cost (compare section 7.1.1). Table 7.12 shows the equilibrium CCSD(T)/qz2p shieldings, δ_e, at the CCSD(T)/cc-pVTZ geometry in the first column, in the second column, δ_0, with zero-point vibrational corrections added, and the third column presents the final theoretical value, $\delta_{T=220K}$, including as well temperature corrections.

Table 7.12: Final analysis of the theoretically evaluated chemical shifts of the 1-adamantyl cation. Atom numbering according to fig. 7.7.

	δ_e	δ_0	$\delta_{T=220K}$	experiment (T=220K), [235]
C_1	306.8	300.9	300.3	300
C_2	69.9	70.1	69.9	65.7
C_3	91.8	93.2	92.9	86.8
C_4	36.1	36.8	36.7	34.5
H_1	3.90	3.92	3.87	4.19
H_2	4.95	5.07	5.06	5.19
H_3	2.13	2.21	2.21	2.42
H_4	2.06	2.16	2.16	2.30

The inclusion of zero-point vibrational effects most notably improves the value for the carbocationic center by 6 ppm, the shielding for the C_3 is slightly deteriorated, and all others remain nearly unchanged. The temperature corrections are one magnitude smaller than the vibrational effects and do not improve significantly the theoretical prediction.

The final values are in good agreement with the measurements. Only the calculated shieldings for C_2 and C_3 show slightly larger differences to experiment. The remaining discrepancies might be explained by the former mentioned 1-2 ppm uncertainty due to basis-set incompleteness in the calculation of NMR chemical shifts, the low-level treatment of vibrational effects, and the neglect of solvent effects in super-acid solutions. Furthermore, there is the possibility of a rather small systematic shift because the measurements were obtained relative to an external capillary. However, a more elaborate treatment for the geometry is not expected to change the computed shieldings [83].

8 Calculation of equilibrium geometries and spectroscopic properties

The knowledge of molecular structures is of particular importance for the understanding of the chemical and physical properties of molecules. Therefore, the accurate determination of molecular geometries is one of the major applications and challenges of quantum chemistry [24, 238]. Equilibrium geometries (r_e) that are defined via minima of the Born-Oppenheimer potential-energy surface can be determined in quantum chemical calculations using the gradient techniques described in chapter 4.

The molecular structure obtained by rotational spectroscopy is not necessarily the same as that evaluated in quantum-chemical calculations. The former is usually an effective (r_0) structure, that is obtained by a fit of the experimental rotational constants B_0^{exp} to the structural parameters. The often used r_s structures are obtained analyzing experimental rotational constants of a series of isotopically substituted species such that vibrational effects on the structural parameters are minimized [239, 240].

Determination of empirical equilibrium structures from rotational spectroscopy

The key to the empirical or semi-experimental equilibrium structure r_e^{emp} is the availability of accurate anharmonic force fields from quantum-chemical computations which enables the determination of individual zero-point vibrational contributions. The rotational constants B_υ are expanded in the vibrational quantum numbers υ:

$$B_\upsilon = B_\mathrm{e} - \sum_r \alpha_r \left(\upsilon_r + \frac{1}{2}\right) + \sum_{r \geq s} \gamma_{rs} \left(\upsilon_r + \frac{1}{2}\right)\left(\upsilon_s + \frac{1}{2}\right) + \dots \; . \tag{8.1}$$

B_e is the equilibrium rotational constant, the sums with α_r and γ_{rs} over the normal coordinates r and s account for the vibrational corrections to the rotational constant. The corrections due to the γ_{rs} terms are usually neglected because of their small magnitude. The α_r are determined via second-order vibrational perturbation theory (VPT2) [142]. This leads to the following expression for the zero-point vibrational contributions to the rotational constants,

$$\Delta B_e^{calc} - \Delta B_{\upsilon=0}^{calc} = \Delta B_0^{calc} = \frac{1}{2} \sum_r \alpha_r \tag{8.2}$$

for each isotopologue. In eq. (8.2), the vibrational correction ΔB_0^{calc} is given as the sum of the calculated rotation-vibration interaction constants α_r. With those corrections ΔB_0^{calc}, it is possible to obtain (empirical) equilibrium values for the rotational constants according to

$$B_e^{emp} = B_0^{exp} + \Delta B_0^{calc}. \tag{8.3}$$

The equilibrium structure is then obtained by a least-squares fit of the structural parameters to the equilibrium moments of inertia I_e (see, e.g., ref. [67]).

$$I_e = \frac{h}{8\pi^2 B_e} \tag{8.4}$$

Beside rotational constants and their vibrational corrections so-called hyperfine parameters can be evaluated via quantum-chemical calculations. Two of these hyperfine parameters, namely the quadrupole coupling and the nuclear spin rotation constants, will be described in the following. The quadrupole coupling χ_{ij} is defined as,

$$\chi_{ij} = -\frac{eQ_n q_{ij}^n}{\hbar} \tag{8.5}$$

where e is the elementary charge, Q_n is the quadrupole constant of the corresponding nucleus n, which can be taken from the literature [241] and q_{ij}^n is the electric-field gradient. The electric-field gradient is determined as a first-order property. Zero-point vibrational effects can be calculated using essentially the same VPT2 scheme as discussed in section 7.1.1 for the shielding tensor. The shielding tensors in eqs. (7.4) and (7.5) are simply replaced by the electric-field gradient tensors.

Coupling between the nuclear spin and the rotation of a molecule cause additional splittings in the rotational spectra [242]. The corresponding molecular parameters are the so-called nuclear spin-rotation tensors $\underline{\underline{\mathcal{M}}}_n$ which are given as the second derivative of the molecular energy with respect to nuclear spin $\mathbf{I}_n$ of the n-th nucleus and the rotational angular momentum $\mathbf{J}$:

$$\underline{\underline{\mathcal{M}}}_n = -\frac{\partial^2 E}{\partial \mathbf{I}_n \partial \mathbf{J}} \ . \tag{8.6}$$

The quantum-chemical calculation of $\underline{\underline{\mathcal{M}}}_n$ suffers from a slow basis-set convergence and an unphysical origin dependence concerning the involved electronic angular momentum. These problems can be overcome by using perturbation-dependent basis functions (often referred to as rotational London orbitals [243]):

$$\chi_\mu(\mathbf{r}, \mathbf{J}) = \exp(i(\mathbf{I}^{-1}\mathbf{J}) \times \mathbf{R}_\mu) \cdot \mathbf{r}) \ \chi_\mu(\mathbf{r}) \ . \tag{8.7}$$

For linear molecules only the xx component of the nuclear spin-rotation tensor, C, will be discussed.

8.1 The geometry of the hydrogen trioxy radical

The hydrogen trioxy radical HO_3 is the parent compound for a class of species that may be formed from the reaction of alkoxy radicals with molecular oxygen and is present in the atmosphere [244–246]. Indeed, on the basis of recent spectroscopic studies, which indicated that HO_3 has an HO-OO bond energy of up to 25 kJ mol^{-1}, Murray et al. [247] suggested that as much as two-thirds of OH might be tied up as HO_3 in the coldest regions of the atmosphere. The energetics of HO_3 will not be discussed here except it should be noticed that, in general quantum-chemical calculations predict significantly lower bond energies for HO_3 [248–252].

Through the use of Fourier-transform microwave (FTMW) spectroscopy, Suma et al. [253] succeeded in obtaining the pure rotational spectra of HO_3 and DO_3. The rotational constants for HO_3 and DO_3 are entirely consistent with the rotational structure seen earlier in the ($v_1 = 0 \rightarrow 2$) transition by Murray et al. [247].

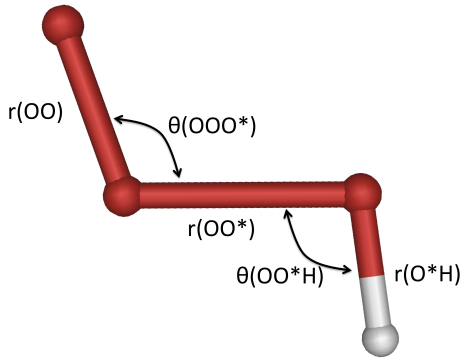

Figure 8.1: Trans-HO_3 with the employed naming scheme for the structural parameters.

Given the planarity of HO_3, the four independent rotational constants from the two isotopic species are insufficient to determine the structure. This requires specification of five parameters - two oxygen-oxygen distances, the OH distance, and the OOO and OO*H angles (compare fig. 8.1). It is clear from the work of Suma et al. [253] that the trans, rather than the cis, isomer is seen. The procedure applied in the microwave study was to fix the terminal oxygen-oxygen distance, thereby assuming a "perturbed" O_2, as well as the OH distance, to values obtained in their multi-reference configuration-interaction (MRCI) calculations: r(OO) = 1.225 Å and r(O*H) = 0.972 Å. This resulted in three parameters to fit to the four independent rotational constants and yields: r(OO*) = 1.688 Å; θ(OOO*) = 111.0°; and θ(OO*H) = 90.0°, the first and third of these being most important in the present context.

The equilibrium structure of trans-HO_3 obtained with the extrapolation scheme by Heckert et al. [254] which attempts to approximate a complete basis set (CBS) at the CCSD(T) level [255] can be found in table 8.1, together with the r_0 structure from ref. [253]. There is a large discrepancy for the central oxygen-oxygen distance as well as for θ(OO*H). The too short prediction of r(OO*) is in line with lower-level and more standard calculations: central oxygen-oxygen distances obtained in geometry optimizations at other levels of coupled-cluster theory are also given in table 8.1. The two most demanding calculations reported here, the CCSD(T)/CBS and the full CCSDT [256] geometry optimizations, were each carried out with

an unrestricted Hartree-Fock reference wavefunction. At lower levels of theory, unrestricted and restricted reference functions were used with only slight differences in the resulting equilibrium structures as seen in table 8.1.

Table 8.1: Structural parameters for trans-OOO*H. (Bond lengths in Å, angles in degrees.)

	r(OO)	r(OO*)	r(O*H)	θ(OOO*)	θ(OO*H)
FTMW[a,b]	-	1.688	-	111.0	90.0
MRCI/aug-cc-pVTZ[b]	1.225	1.677	0.972	110.2	95.9
UHF-(fc)CCSD(T)/CBS[c]	1.226	1.582	0.969	109.7	97.5
UHF-(fc)CCSDT/cc-pVTZ	1.234	1.596	0.969	109.8	96.7
UHF-(fc)CCSD(T)/cc-pVTZ	1.233	1.590	0.969	109.8	96.8
ROHF-(fc)CCSD(T)/cc-pVTZ	1.233	1.587	0.969	109.7	96.8
UHF-(fc)CCSD(T)/cc-pVQZ	1.227	1.589	0.968	109.8	97.1
ROHF-(fc)CCSD(T)/cc-pVQZ	1.228	1.587	0.968	109.8	97.1
UHF-(fc)CCSD(T)/aug-cc-pVTZ	1.230	1.607	0.972	109.7	97.1
ROHF-(fc)CCSD(T)/aug-cc-pVTZ	1.230	1.605	0.972	109.7	97.1
UHF-(fc)CCSD(T)/ANO1	1.231	1.608	0.971	109.7	96.8
ROHF-(fc)CCSD(T)/ANO1	1.231	1.605	0.971	109.7	96.9
UHF-(fc)CCSD(T)/ANO2	1.228	1.589	0.969	109.7	97.2
ROHF-(fc)CCSD(T)/ANO2	1.229	1.586	0.969	109.7	97.3

[a] Effective r_0 structure based on constrained parameters;
[b] ref. [253];
[c] ref. [255].

At the (fc)CCSD(T)/cc-pVDZ level the central oxygen-oxygen bond length is unusually long with r_e(OO*) approximately 1.75 Å. However, the bond length contracts as the basis is improved. Differences between cc-pVTZ, cc-pVQZ and a large ANO basis set are quite small and in good agreement with the extrapolated complete basis set result. Consideration of triple excitation effects is important. With the cc-pVTZ basis set, triple excitation effects lengthen r_e(OO*) to 1.590 Å from r_e(OO*) = 1.520 Å found at the UHF-(fc)CCSD level. A full optimization at the (fc)CCSDT/cc-pVTZ level indicates that the difference between the full CCSDT basis-set limit for the critical r_e(OO*) parameter and that corresponding to CCSD(T) is on the order of 0.01 Å, with the former being longer.

The equilibrium OO*H bond angle is insensitive to theory. There is no chance that the *ca.* 5 degree discrepancy is due to a shortcoming of the theoretical methods used. It is quite clear that the full configuration-interaction complete basis-set value of θ_e(OO*H) is within two degrees of 97°, which is more than five degrees larger than the value from the r_0 structure. Both angles in the effective structure differ significantly from the r_e structure.

An anharmonic force field has been calculated with (fc)CCSD(T)/aug-cc-pVTZ and the VPT2 treatment has been applied to estimate the vibrational contribution to the observed rotational constants. This led to the finding that r_0(OO*) should exceed r_e(OO*) by 0.03 Å. However, the presence of a very anharmonic low-frequency mode [1] points to potential problems of using VPT2 in this case, and underscores the importance of large-amplitude motion in this molecule.

A calculation of the equilibrium geometry of the related FO_2 molecule at the (fc)CCSD(T)/cc-pVQZ level gives an FO distance of 1.633 Å, in relatively good agreement

[1] At the UHF (fc)CCSD(T)/aug-cc-pVTZ level a harmonic frequency of 125 cm^{-1} and an anharmonic frequency of 101 cm^{-1} is obtained.

with the r_0 distance of 1.649 Å reported by Hirota and collaborators [257]. A similar VPT2 calculation for FO_2 leads to the expectation that the r_0 and r_e FO distances differ by only about 0.01 Å for FO_2, so the electronic structure calculation appears to predict r_e(FO) quite accurately for this molecule. Given the similarity in electronic structure, it would be hard to explain why a theoretical method that works so well for FO_2 should fail for HO_3.
Based on these findings it is assumed that the equilibrium structure of trans-HO_3 exhibits a central oxygen-oxygen bond length in the range 1.58-1.62 Å and an equilibrium bond angle of 95-99° which is in contradiction to the estimates of 1.688 Å and 90° given by Suma et al.

8.2 The empirical equilibrium structure of diacetylene

Diacetylene, H−C≡C−C≡C−H, the simplest polyacetylene, is of importance in organic chemical synthesis [258] and is also an abundant astronomical species. In space, it was detected through infrared spectroscopy in the atmosphere of Saturn's moon Titan [259] and in the protoplanetary nebula CRL618 [260]. In the laboratory, it has been the subject of numerous spectroscopic studies in the infrared and also in the microwave regime (see refs. [261,262] and references therein). Spectroscopic data were also used to evaluate its structural parameters using different approaches. An advanced experimental near-equilibrium $(r_m^{\rho})_{corr}$ structure [1] was reported by Tay et al. [261] making use of experimental rotation-vibration data for a total of nine different isotopic species. In addition, various quantum-chemical calculations were performed with an empirically corrected frozen-core CCSD(T)/cc-pVQZ calculation [264] representing the highest theoretical level employed to date.

The equilibrium structure of diacetylene is of interest, as it represents one of the simplest systems with conjugated carbon-carbon multiple bonds. The question is how conjugation affects the structure and in particular shortens the central carbon-carbon single bond. It is well known that low-level quantum-chemical calculations are not reliable in this respect, with Hartree-Fock calculations typically underestimating the conjugation and density-functional theory overestimating it [265]. High-level calculations, preferably at the coupled-cluster (CC) level, are needed for reliable predictions. To determine the accuracy of the CC calculations, it is essential to provide an reliable equilibrium structure based on experimental data.

In a recent Fourier transform microwave spectroscopic study, the rotational spectra of four new isotopologues of monodeutero diacetylene were characterized [262] raising the number of investigated isotopic species of diacetylene to a total of thirteen. Experimental B_0 values of all thirteen isotopologues studied through rotation-vibration [261,266] and rotational spectroscopy [262] are collected in table 8.2. The least-squares fits for the structural parameters have been performed including the data for all thirteen isotopologues. The fits have been performed with respect to the moments of inertia using the same weight for all considered isotopic species (further details are given in ref. [267]).

Table 8.2 summarizes in addition to the experimental rotational constants B_0 also the corresponding calculated vibrational corrections as obtained at the (fc)MP2/cc-pVTZ, (fc)CCSD(T)/cc-pVTZ, (fc)CCSD(T)/cc-pVQZ, CCSD(T)/cc-pCVQZ, and (fc)CCSD(T)/ANO2 levels of theory. Analysis of all these calculations allows one to draw the following conclusions (numbers given in the following with respect to the parent isotopic species): (a) MP2 calculations overestimate the vibrational corrections in comparison with CCSD(T) computations (e.g., 2.1 MHz for (fc)MP2/cc-pVTZ and 0.91 MHz for (fc)CCSD(T)/cc-pVTZ); (b) the use of quadruple-zeta basis sets is recommended, as the smaller cc-pVTZ and cc-pCVTZ basis sets yield vibrational corrections with rather large errors (e.g. 0.91 MHz for (fc)CCSD(T)/cc-pVTZ in comparison with 1.92 MHz for (fc)CCSD(T)/cc-pVQZ and 2.22 MHz for CCSD(T)/cc-pCVTZ in comparison with 1.84 MHz for CCSD(T)/cc-pCVQZ); (c) the use of the ANO2 set yields similar results as the (fc)MP2/cc-pVTZ calculation.

[1] The $(r_m^{\rho})_{corr}$ structure is calculated from scaled moments of inertia I_m^{ρ} by the method of least squares. The moments of inertia, I_m^{ρ} are scaled by the ratio (ρ) of the substitution moment of inertia, I_s, and the moment of inertia of the chosen parent isotopomer, I_0: $I_m^{\rho} = (2\rho - 1)I_0$. For molecules containing hydrogen atoms a simple empirical correction [263] is applied which accounts for the elongation of bonds to hydrogen when deuterium substitutions are used.

This analysis shows also that the computed vibrational corrections (at the CCSD(T)/cc-pCVQZ and (fc)CCSD(T)/cc-pVQZ levels) should have an accuracy of about 0.2 to 0.3 MHz. This accuracy is sufficient for obtaining an accurate equilibrium structure of diacetylene from the empirical B_e rotational constants and, in turn, it is more than adequate for a theoretical prediction of the ground-state rotational constants B_0. For the latter, the errors in the theoretical determination of high-level B_e values due to remaining basis-set errors and still missing electron-correlation contributions certainly are larger than the zero-point vibrational correction ΔB_0^{calc} [268]. Table 8.3 gathers the new extensive set of structural parameters for diacetylene and reports the different empirical structures of diacetylene as obtained from using the vibrational corrections given in table 8.2 and the pure theoretical structures based on high-level CC calculations using a variety of basis sets. In addition, structural parameters from the literature are given [261, 262, 264].

Comparing the five different empirical r_e^{emp} structures, values of 1.062 Å, 1.208 Å, and 1.373 Å are obtained for the C−H distance, the C≡C triple bond, and C–C single bond, respectively, irrespective of the vibrational corrections used. The differences in the vibrational corrections affects the bond lengths only in the fourth decimal. Nevertheless, the changes when going from the cc-pVTZ to the cc-pVQZ or cc-pCVQZ corrections are not entirely negligible and the results obtained with the (fc)CCSD(T)/cc-pVQZ, CCSD(T)/cc-pCVQZ, and CCSD(T)/ANO2 corrections are considered the most reliable. The statistical uncertainties in all of the fits are small being 0.0004 Å (CC single bond), 0.0003 Å (CC triple bond), and 0.0001 Å (CH bond) and hence in all cases below 0.001 Å.

Comparison of the derived empirical r_e structure with the pure theoretical geometries reveals good agreement with the results obtained at the highest applied quantum-chemical level (CCSD(T)/cc-pCV5Z). With 0.001 Å and less the discrepancies are in the expected range. A closer analysis of the extensive set of theoretical data in table 8.3 furthermore demonstrates the importance of core correlation, as the (fc)CCSD(T) calculations yield too long bond distances (by about 0.0015 Å for the CH distances and 0.003 Å for the CC distances).

Table 8.2: Experimental rotational constants B_0 and computed zero-point vibrational corrections ΔB_0 (all values in MHz) for the various isotopologues of diacetylene.

		ΔB_0				
Isotopic Species		(fc)MP2/cc-pVTZ	(fc)CCSD(T)/cc-pVTZ	(fc)CCSD(T)/cc-pVQZ	CCSD(T)/cc-pCVQZ	(fc)CCSD(T)/ANO2
HCCCCH	4389.3019(39)[a]	2.098	0.912	1.922	1.838	2.083
DCCCCD	3809.2433(66)[b]	−0.207	−1.110	−0.203	−0.237	−0.005
$H^{13}C^{13}C^{13}C^{13}CH$	4098.8959(36)[b]	2.145	1.064	1.970	1.890	2.108
$H^{13}C^{13}CCCH$	4243.7325(111)[b]	2.130	0.997	1.953	1.871	2.103
$H^{13}CCCCH$	4258.5465(105)[b]	2.112	0.953	1.912	1.829	2.063
$HC^{13}CCCH$	4371.6291(45)[b]	2.113	0.955	1.962	1.879	2.121
$H^{13}C^{13}C^{13}CCH$	4224.7392(99)[b]	2.138	1.032	1.987	1.905	2.134
$H^{13}C^{13}CC^{13}CH$	4115.0556(42)[b]	2.133	1.027	1.935	1.855	2.075
HCCCCD	4084.45342(7)[c]	0.815	−0.219	0.745	0.688	0.927
$H^{13}CCCCD$	3964.11797(17)[c]	0.871	−0.141	0.774	0.718	0.946
$HC^{13}CCCD$	4066.49893(16)[c]	0.830	−0.178	0.782	0.725	0.962
$HCC^{13}CCD$	4071.64202(16)[c]	0.851	−0.160	0.800	0.743	0.980
$HCCC^{13}CD$	3977.69016(15)[c]	0.870	−0.144	0.775	0.718	0.947

[a] Ref. [266]; [b] ref. [261]; [c] ref. [262].

For the sake of completeness, table 8.3 also gives parameters from other equilibrium or near-equilibrium structure determinations. Reasonable agreement is found between the best empirical equilibrium structures r_e^{emp} ((fc)CCSD(T)/cc-pVQZ, CCSD(T)/cc-pCVQZ), and (fc)CCSD(T)/ANO2) and the near-equilibrium $(r_m^{\rho})_{corr}$ structure of Tay et al. [261] with the largest discrepancy being in the central C−C bond. Somewhat larger discrepancies are found in comparison to the r_e structure from ref. [262]. However, the latter structure was only based on a rough empirical correction for the difference between the r_s and r_e structure (without determining the vibrational corrections explicitly) and hence is not comparable from a rigorous point of view. Good agreement is found for the structure given by Botschwina and Puzzarini [264] where a theoretical structure obtained at the CCSD(T)/cc-pVQZ level of theory was empirically corrected through comparison to an empirical equilibrium structure of HC_3N. Finally, a comparison of the r_e^{emp} structures and the r_0 structure reveals the importance of including vibrational corrections in structural determinations. In the latter structure the CH distances are significantly shorter (by about 0.005 Å) whereas the CC single bond is longer (by about 0.002 Å) than in the r_e^{emp} structures. A similar conclusion also holds for the comparison with the r_s structure reported in ref. [262] for HCCCCD.

Comparison of the r_e^{emp} parameters of diacetylene with those of similar molecules reveals that the C–H bond lengths found in diacetylene (1.6016 Å, this work), fluoroacetylene (1.6014 Å, ref. [264]), methyldiacetylene (1.6013 Å, ref. [271]), acetylene (1.0618 Å, ref. [269]), HC_3N (1.0623 Å, ref. [270]) and methylacetylene (1.061 Å, ref. [272]) are practically identical. Clearly, conjugation effects in diacetylene manifest themselves in the C≡C bond length: whereas in acetylene a value of 1.2029(1) Å is found [269], the C≡C triple bond is significantly longer in diacetylene, namely 1.2084 Å. The latter value is in good agreement with those found in similar conjugated systems. Recent examples are CC distances in the ethynyl group of the substituted diacetylenes HC_4F and CH_3C_4H (1.2080 Å [264] and 1.2085 Å [271], respectively) and also the C≡C distances in branched species such as *cis*-hex-3-ene-1,5-diyne and (*Z*)-pent-2-en-4-ynenitrile (1.208(3) Å [265] and 1.207(3) Å [273]). Given the prototypical C=C equilibrium bond length in ethylene of 1.3305 Å [274]) the formal C−C single bond length in diacetylene of 1.3727 Å found here is very short (and comparable to monofluorodiacetylene (1.3729 Å) and methyldiacetylene (1.3734 Å) indicative of some double-bond character. Furthermore, conjugation i.e. π electron delocalization is more pronounced in diacetylene (and HC_4F/CH_3C_4H) than in HC_3N where the C≡C bond is found to be shorter (intermediate between acetylene and diacetylene) and the C−C bond longer compared to diacetylene (see table 7.3).

Several infrared studies of the vibrational fundamentals, overtones and hot bands of the main isotopic species of diacetylene are found in the literature (see ref. [261] and references therein). Table 8.4 presents harmonic and fundamental frequencies of diacetylene as obtained via VPT2. The required harmonic and anharmonic force fields were obtained by CCSD(T) calculations using the cc-pVTZ, cc-pVQZ, and ANO2 basis sets in the frozen-core approximation and the cc-pCVQZ basis correlating all electrons. The accuracy of the reported fundamental frequencies obtained using Dunning's quadruple-zeta basis sets is probably only about 2-5 wavenumbers, as it turned out impossible to converge the required CC second-derivative calculations at the displaced points with very tight convergence thresholds. In addition, the corresponding (fc)MP2/cc-pVTZ results are reported. Comparison of the theoretical fundamental frequencies and the experimental values collected in ref. [275] reveals that

Table 8.3: Equilibrium structures of diacetylene and related molecules (Å).

Method	r_{C-H}	$r_{C\equiv C}$	r_{C-C}
(fc)MP2/cc-pVTZ	1.0620	1.2194	1.3687
(fc)CCSD(T)/cc-pVTZ	1.0638	1.2149	1.3789
(fc)CCSD(T)/cc-pVQZ	1.0633	1.2119	1.3769
(fc)CCSD(T)/ANO2	1.0631	1.2118	1.3766
(fc)CCSD(T)/cc-pVQZ+corr (ref. [264])[a]	1.0615	1.2087	1.3720
(fc)CCSD(T)/cc-pV5Z	1.0630	1.2111	1.3764
(fc)CCSD(T)/cc-pV6Z	1.0630	1.2109	1.3762
CCSD(T)/cc-pCVTZ	1.0633	1.2121	1.3770
CCSD(T)/cc-pwCVTZ	1.0630	1.2111	1.3763
CCSD(T)/cc-pCVQZ	1.0621	1.2091	1.3742
CCSD(T)/cc-pwCVQZ	1.0620	1.2089	1.3741
(fc)CCSD(T)/cc-pwCVQZ	1.0632	1.2115	1.3766
CCSD(T)/cc-pCV5Z	1.0617	1.2083	1.3737
r_0	1.0561	1.2079	1.3752
$(r_m^\rho)_{corr}$ (ref. [261])	1.0613(1)	1.2096(1)	1.3708(2)
r_e (ref. [262])	1.0609	1.2104	1.3709
r_e^{emp} ((fc)MP2/cc-pVTZ)	1.0623	1.2077	1.3736
r_e^{emp}((fc)CCSD(T)/cc-pVTZ)	1.0620	1.2083	1.3732
r_e^{emp}((fc)CCSD(T)/cc-pVQZ)	1.0616	1.2084	1.3727
r_e^{emp}(CCSD(T)/cc-pCVQZ)	1.0615	1.2085	1.3727
r_e^{emp}((fc)CCSD(T)/ANO2)	1.0614	1.2085	1.3726
r_e^{emp} acetylene (ref. [269])	1.0618	1.2029	—
r_e^{emp} HC_3N (ref. [270])	1.0623	1.2059	1.3761
r_e^{emp} HC_4F (ref. [264])[b]	1.0614	1.2080	1.3731
r_e^{emp} H_3C-C_4-H (ref. [271])	1.6013(3)	1.2085(6)/1.2091(16)	1.3734(14)

[a] Empirically corrected *ab initio* structure, see text;
[b] $r_{C\equiv C}$ given refers to the ethynyl group.

the results obtained with the cc-pVXZ (X=T, Q) sets turn out less reliable, while better agreement with experiment is seen for the other two basis sets. For both, the cc-pCVQZ and the ANO2 set, the computed frequencies match the experimental values within about 10 cm^{-1}. This finding again documents the suitability of the ANO basis sets for frequency calculations, while the Dunning basis sets (at least when using the valence sets) show some deficiencies [276–280].

Table 8.4: Vibrational fundamentals of diacetylene (in cm^{-1}).

	Experiment[b]	(fc)MP2/cc-pVTZ		(fc)CCSD(T)/cc-pVTZ		(fc)CCSD(T)/cc-pVQZ		CCSD(T)/cc-pCVQZ		(fc)CCSD(T)/ANO2	
		harm.	anharm.	harm.	anharm.	harm.	anharm.	harm.	anharm.	harm.	anharm.
$\nu_1(\Sigma_g^+)$	3332	3486	3361	3458	3329	3457	3281	3463	3333	3462	3330
$\nu_2(\Sigma_g^+)$	2189	2194	2147	2233	2188	2235	2184	2243	2197	2237	2190
$\nu_3(\Sigma_g^+)$	872	897	867	887	849	892	855	894	865	891	861
$\nu_4(\Sigma_u^+)$	3334	3482	3369	3454	3339	3458	3288	3465	3338	3461	3329
$\nu_5(\Sigma_u^+)$	2022	2008	1970	2051	2014	2057	2016	2064	2031	2056	2020
$\nu_6(\Pi_g)$	626	620	625	623	643	632	616	636	635	639	627
$\nu_7(\Pi_g)$	483	487	452	474	419	481	484	484	490	482	479
$\nu_8(\Pi_u)$	628	629	602	633	614	634	616	640	635	640	627
$\nu_9(\Pi_u)$	220	230	217	227	216	220	225	221	223	220	219

The accuracy of the reported fundamental frequencies obtained using Dunning's quadruple-zeta basis sets is probably only a couple of wavenumbers, see text.

[b] Ref. [275] and references therein.

8.3 Geometry and hyperfine structure cyanopolyynes

8.3.1 Geometry and hyperfine structure of deuterated cyanoacetylene

Cyanoacetylene, HC_3N, has been detected in interstellar space as early as 1971 [281]. It is so abundant in massive star-forming regions [282] and circumstellar shells of late-type stars such as CRL 618 [283] that even excited vibrational states of ^{13}C species have been observed. The data obtained for the protoplanetary nebula CRL 618 were used to derive a physical model for this short-lived stage in the evolution of stars [283]. The detection of DC_3N in the cold, dark cloud TMC-1 dates back more than a quarter of a century [284].

There have been numerous reports on investigations of the rotational spectra of HC_3N [285, 286], its ^{13}C [287] and ^{15}N [287, 288] substituted species, including doubly substituted ones [287]. Mallinson and DeZafra [289] reported several transitions of DC_3N in the ground and low-lying excited vibrational states up to 212 GHz. The ^{14}N and D hyperfine structure of the $J = 1 - 0$ transition had been studied previously with an FTMW [290] and a beam maser spectrometer [291], respectively.

High-level quantum-chemical calculations have been performed to test the accuracy of such calculations and to guide the experiments within a collaboration with the spectroscopy group at Cologne (Prof. Dr. S. Schlemmer, H. Spahn, Dr. H.S.P. Müller, Dr. Thomas F. Giesen, and Dr. Jens-Uwe Grabow) in the case of the nuclear spin-rotation values which are difficult to determine in the case of the DC_3N molecule.

Calculations for the equilibrium structure as well as the nuclear quadrupole and spin-rotation coupling parameters for ^{14}N and D have been performed at the CCSD(T) level. In the calculations of the spin-rotation tensors, perturbation-dependent basis functions as described in ref. [243] have been used to ensure fast basis-set convergence. All hyperfine-structure calculations have been performed at the CCSD(T)/cc-pVQZ geometry. The calculated hyperfine parameters usually refer to the minimum of the potential-energy surface, i. e. they are equilibrium values, whereas the experimental values usually have been obtained for the ground vibrational state. Therefore, zero-point vibrational corrections to the quadrupole-coupling constants have been computed for HC_3N and DC_3N using VPT2 as described in section 7.1.1 for shieldings, c.f. the beginning of this chapter. These calculations were performed at the MP2/cc-pVQZ level. Vibrational corrections to the nuclear spin-rotation values are not yet available for linear molecules.

The results of the calculations are summarized in table 8.5. Rather large changes occur when going from the cc-pVDZ to the cc-pVTZ basis set, while the corresponding changes from the cc-pVTZ to the cc-pVQZ are much smaller. Changes due to the use of a quintuple zeta set are even smaller. Therefore, the values obtained at the CCSD(T)/cc-pVQZ level can be considered as essentially converged with respect to basis-set size. Core-correlating polarization functions seem to be only necessary in the electric-field gradient calculations (needed for the quadrupole-coupling constants) where they amount to about 3 % even at the quadruple-zeta level. On the other hand, the effects of the core-polarization basis functions are rather small for the spin-rotation constants. For C(N) a large change is only seen at the double-zeta level, but the differences become smaller with increasing basis-set size. As there are no core-correlating basis functions at H (or D) it is not surprising that their spin-rotation

Table 8.5: DC_3N bond lengths (Å) and ^{14}N and D nuclear quadrupole (MHz) and spin-rotation coupling parameters (kHz) calculated at the CCSD(T) level employing cc-pVXZ (X = D, T, and Q) and cc-pCVXZ (X = D, T, Q, and 5) basis sets in comparison with experimental results.

Basis set	r(D−C)	r(C≡C)	r(C−C)	r(C≡N)	$eQq(^{14}N)$	eQq(D)	$C(^{14}N)$	C(D)
cc-pVDZ	1.0793	1.2300	1.3968	1.1815	−3.553	0.254	0.820	−0.075
cc-pVTZ	1.0592	1.2074	1.3749	1.1633	−4.350	0.226	0.897	−0.073
cc-pVQZ	1.0622	1.2064	1.3761	1.1606	−4.456	0.220	0.911	−0.072
cc-pCVDZ	1.0784	1.2278	1.3956	1.1799	−3.743	0.254	0.840	−0.075
cc-pCVTZ	1.0642	1.2093	1.3809	1.1637	−4.244	0.225	0.908	−0.073
cc-pCVQZ	1.0630	1.2062	1.3782	1.1607	−4.334	0.220	0.916	−0.072
cc-pCVQZ+vib[b]	–	–	–	–	−4.298	0.207	–	–
cc-pCV5Z	–	–	–	–	−4.348	0.218	–	–
cc-pCV5Z+vib[b]	–	–	–	–	−4.312	0.205	–	–
exp.[c]	1.0623	1.2059	1.3761	1.1606	−4.318	0.203	1.058	−0.097

[a] The calculations for the nuclear quadrupole and spin-rotation parameters have been performed at the CCSD(T)/cc-pVQZ geometry. All electrons were correlated.

[b] Including vibrational corrections calculated at the MP2/cc-pVQZ level.

[c] Mixed experimental/*ab initio* structural parameters for HC_3N from ref. [292]; hyperfine structural parameters from [293].

constants more or less remain unchanged when going from the cc-pVXZ to the cc-pCVXZ basis sets.[1]

Hyperfine parameters for DC_3N had been experimentally determined previously [290,291]. The agreement between computed and experimental hyperfine parameters is very good to excellent as can be seen in table 8.5. The experimental nuclear spin-rotation values seem to be slightly larger in magnitude than the theoretical values, however, the differences are not significant. The eQq(N) value calculated with a quadruple-zeta basis set including core-correlating basis functions agrees well with the experimental value; inclusion of vibrational corrections (computed at the MP2/cc-pVQZ level) further improves the agreement (the theoretical vibrationally averaged values are −4.298 MHz for eQq(N) of DC_3N) and −4.294 MHz for eQq(N) of HC_3N). The remaining discrepancies of about 20 to 30 kHz are mostly due to basis-set effects, as calculations at the CCSD(T)/cc-pCV5Z level (with MP2/cc-pVQZ vibrational corrections included) provide improved values (−4.312 and −4.308 MHz for DC_3N and HC_3N, respectively) in comparison with experiment. Possible short-comings of the calculated vibrational corrections cannot be ruled out entirely. The vibrational corrections reduce the small differences between the calculated equilibrium quadrupole parameters and the experimental ground state values.

The calculated equilibrium eQq(D) value differs by almost 10 % from the experimental value. Vibrational effects are here more important, as they amount to about 0.013 MHz. The corresponding calculation at the CCSD(T)/cc-pCV5Z level (with vibrational corrections included) leads with a final theoretical value of 0.205 MHz again to excellent agreement between theory and experiment.

[1] The spin-rotation constants $C(^{13}C)$ for the isotopologues containing ^{13}C are given for completeness reasons as 1.683, 1.235, and 1.285 kHz for $DCC^{13}CN$, $DC^{13}CCN$, and $D^{13}CCCN$, respectively, computed at the CCSD(T)/cc-pCVQZ level.

8.3.2 Geometry and hyperfine structure cyanobutadiyne and cyanohexatriyne

Based on the experiences with the cyanoacetylene similar calculations were performed for the sucessive cyanopolyynes, cyanobutadiyne (HC_5N) and cyanohexatriyne (HC_7N), taking advantage of the presented parallelization scheme. The structural parameters of HC_5N and HC_7N calculated at the (ae)CCSD(T)/cc-pVQZ level in comparison with selected values from the literature are given in table 8.6. As has been seen before in the case of DC_3N

Table 8.6: Bond lengths for HC_5N and HC_7N calculated at the CCSD(T)/cc-pVQZ level (this work) in comparison with selected values from the literature (Å).

	HC_5N			HC_7N	
	this work	$r_{emp}{}^{a,b}$	$r_0^{exp\ b}$	this work	$r_0^{exp\ c}$
r(H−C)	1.0622	1.60220(3)	1.0571(2)	1.0621	1.0570(4)
r(C≡C)	1.2097	1.20948(5)	1.2086(3)	1.2107	1.2101(6)
r(C−C)	1.3695	1.36519(12)	1.3618(7)	1.3629	1.3610(9)
r(C≡C)	–	–	–	1.2169	1.2141(19)
r(C−C)	–	–	–	1.3577	1.3616(27)
r(C≡C)	1.2128	1.21284(17)	1.2196(10)	1.2148	1.2149(20)
r(C−C)	1.3706	1.37079(12)	1.3659(7)	1.3694	1.3657(9)
r(C≡N)	1.1621	1.16209(5)	1.1608(3)	1.1625	1.1611(6)

[a] Vibrational corrections obtained from a (fc)CCSD(T)/cc-pVQZ force field [294]; [b] ref. [294]; [c] ref. [295].

the experimental carbon-carbon and carbon-nitrogen distances are well reproduced by the CCSD(T)/cc-pVQZ calculations. In addition it is noteworthy that the results for the HC_5N molecule are nearly identical to those of the empirical structure from ref. [294]. The calculated

Table 8.7: ^{14}N and D nuclear quadrupole (MHz) and spin-rotation coupling parameters (kHz) calculated at the CCSD(T) level using CCSD(T)/cc-pVQZ geometries in comparison with experimental results.

	Basis set	$eQq(^{14}N)$	eQq(D)	$C(^{15}N)$	$C(^{14}N)$	$C(^{1}H)$
HC_5N	cc-pVQZ	-4.436	0.220	-0.395	0.289	-0.148
	cc-pCVQZ	-4.315	0.220	-0.397	0.290	-0.148
	exp.[a]	-4.242(30)	–	–	–	–
HC_7N	cc-pVQZ	-4.436	0.221	-0.062	0.122	-0.063
	cc-pCVQZ	-4.315	0.221	–	–	–
	exp.[b]	-4.29(16)	–	–	–	–

[a] Ref. [296]; [b] ref. [295].

$eQq(^{14}N)$ values are in good agreement for both considered molecules. The calculations employing the core-polarized valence basis sets (534 and 702 basis functions for HC_5N and HC_7N, respectively) give again better estimates than those without special core-polarization functions. The analysis of the experimental spin-rotation constants by the spectroscopy group at Cologne is still in progress, so that a comparision with experiment is currently not possible. Similar agreement as in the case of DC_3N is expected.

8.4 The equilibrium structure of ferrocene

Within the last 25 years many attempts were made to determine the molecular structure of ferrocene by applying various quantum-chemical methods [297–301]. A recent study [302] presented first calculations employing analytical CCSD(T) gradients to this problem using, however, a relatively small basis set (373 basis functions) and the frozen-core approximation. Up to now, quantum-chemical models have been unable to determine the equilibrium metal-ligand distance, a quantity that is not directly accessible to experiment but often used for benchmark studies in the framework of density-functional theory. The structural parameters of ferrocene in its eclipsed (equilibrium, compare fig. 8.2) and staggered (saddle point, compare fig. 8.2) conformations have been determined in the present work using analytic CCSD(T) gradients correlating all 96 electrons with a full triple-zeta quality basis set. Using the cc-pVTZ basis set (508 basis functions) one geometry cycle takes about 2.3 days when performing the calculation on 15 nodes. With the cc-pwCVTZ basis set (672 basis functions) a geometry cycle takes about 8.8 days using 14 nodes. The results in comparison with experimental data [300, 303–307], a previous coupled-cluster study [302], and DFT calculations are presented in tables 8.8 and 8.9.

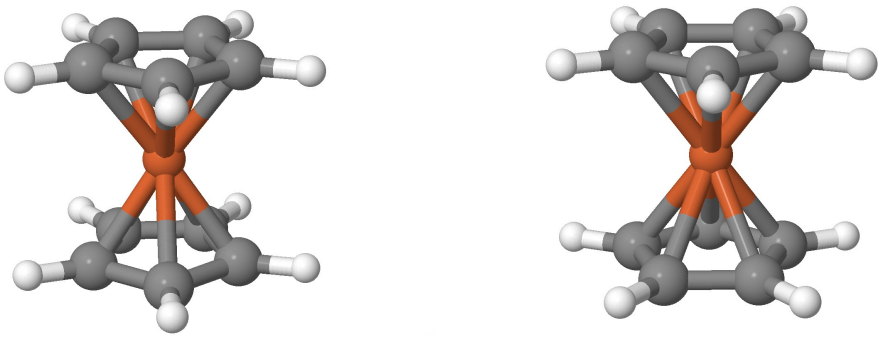

Figure 8.2: The ferrocene molecule in its eclipsed (left) and staggered (right) conformation.

Table 8.8: Structure parameters of the eclipsed conformation of ferrocene. Bond lengths given in Å, angles in degrees.

	Theory					Experiment		
Method	B3LYP def2-QZVPP	BP86 def2-QZVPP	CCSD(T)[a,b] TZ2P+f	CCSD(T) cc-pVTZ	CCSD(T) cc-pwCVTZ	GED	XRD	ND
Fe-C_5	1.691	1.650	1.655	1.639	1.648	1.661 [c,i]		
Fe-C	2.079	2.051	2.056	2.039	2.047	2.054 [d]	2.059 [f]	2.031 [g,i]
C-C	1.423	1.433	1.433	1.426	1.427	1.435 [d]	1.431 [f]	1.415 [h,i]
C-H	1.077	1.085	1.077	1.075	1.079	1.080 [d]		1.073 [h,i]
∢ C_5-H	0.82	0.66	1.03	0.45	0.52	3.7 [e]		1.7 [g,i]

[a] Only 66 valence electrons were correlated; [b] ref. [302]; [c] ref. [303]; [d] ref. [300]; [e] ref. [304]; [f] ref. [305]; [g] ref. [306]; [h] ref. [307]; [i] parameters not corrected for thermal motion.

The experimental values, as acquired by the techniques of electron diffraction (GED) [300,303,304], X-ray diffraction (XRD) [306,307] and neutron diffraction (ND) [305] show no outstanding agreement with the computed geometry parameters. GED, XRD and ND results are vibrationally averaged distances, and therefore different from what is being calculated. Thus, a comparison with experiment is not straightforward. B3LYP apparently overestimates the metal-ligand (Fe-C_5) and the iron-carbon (Fe-C) distance, while BP86 gives better results. The values obtained at the CCSD(T) level show a relatively pronounced basis-set dependence. For both conformers the (fc)CCSD(T)/TZ2P+f calculations predict the longest bond distances, while the CCSD(T)/cc-pVTZ computation predicts probably too short bond lengths. The addition of the weighted core-polarization functions (CCSD(T)/cc-pwCVTZ) show an elongation of the metal-ligand and the iron-carbon distance. Carbon-carbon and carbon-hydrogen distances seem to be close to the final values for the presented CC calculations. The observed change in both conformations between the cc-pVTZ and the cc-pwCVTZ is only 0.001 Å for the carbon-carbon distances and 0.003 Å for the hydrogen-carbon distances. The out of plane angle $\sphericalangle$ C_5-H between the plane of the 5 cyclopentadienyl carbons and the individual hydrogens might be overestimated in the experimental measurements as all theoretical models predict values about 1°.

Table 8.9: Structure parameters of the staggered conformation of ferrocene. Bond lengths given in Å, angles in degrees.

Method	B3LYP def2-QZVPP	BP86 def2-QZVPP	CCSD(T)[a,b] TZ2P+f	CCSD(T) cc-pVTZ	CCSD(T) cc-pwCVTZ
Fe-C_5	1.695	1.654	1.659	1.642	1.652
Fe-C	2.082	2.055	2.058	2.041	2.050
C-C	1.422	1.433	1.432	1.425	1.426
C-H	1.077	1.085	1.077	1.075	1.078
$\sphericalangle$ C_5-H	0.96	0.99	1.34	0.67	0.61

[a] Only the 66 valence electrons were correlated; [b] ref. [302].

CCSD(T)/cc-pwCVQZ (1289 basis functions) calculations for both conformations are under way and will hopefully finally resolve the question regarding the equilibrium structure of ferrocene.

9 Summary

Coupled-cluster theory provides one of the most successful concepts in electronic-structure theory. This work covers the parallelization of coupled-cluster energies, gradients, and second derivatives and its application to selected large-scale chemical problems, beside the more practical aspects such as the publication and support of the quantum-chemistry package ACES II MAB and the design and development of a computational environment optimized for coupled-cluster calculations. The main objective of this thesis was to extend the range of applicability of coupled-cluster models to larger molecular systems and their properties and therefore to bring large-scale coupled-cluster calculations into day-to-day routine of computational chemistry. The applications cover the areas of benchmark and application of computational thermochemistry, nuclear magnetic shielding constants, and spectroscopic properties. The following paragraphs provide a detailed summary and future prospects for the results presented in this thesis.

Parallel coupled-cluster calculations
A straightforward strategy for the parallelization of CCSD and CCSD(T) energies, gradients and second derivatives has been outlined and implemented. Starting from the highly efficient serial implementation of the ACES II MAB computer code an adaptation for affordable workstation clusters has been obtained by parallelizing the most time-consuming steps of the algorithms. This includes the calculation of parallel CCSD energies, gradients, and second derivatives using unrestricted (UHF) and restricted open shell (ROHF) Hartree-Fock references, parallel UHF-CCSD(T) energies, gradients, and second derivatives, parallel ROHF-CCSD(T) energies, as well as parallel equation-of-motion CCSD energies and gradients for closed- and open-shell references. Furthermore virtually the same scheme as for the perturbative triples in CCSD(T) has been applied to the calculations of CCSDT(Q) energies using a pilot implementation in ACES II MAB. Based on the experiences of the presented study the MRCC program has been modified for an OpenMP as well as a MPI parallel implementation for parallel generalized coupled-cluster energy, gradient and second-derivative calculations.

The central aspects of this implementation are the replication of the cluster amplitudes and the distributed evaluation, storage, and access of the two-electron integrals to arrive at an algorithm for which sufficient local memory and disk space are necessary but which is not dependent on sophisticated high-speed network connections.

Benchmark calculations for systems with up to 1300 basis functions and the applications within this thesis have shown that the resulting algorithm for energies, gradients and second derivatives at the CCSD and CCSD(T) level of theory exhibits good scaling with the number of processors and substantially extends the range of applicability.

Computational thermochemistry
The results of this work serve to make a general statement about very high-accuracy theoretical thermochemistry. The accuracies that can now be obtained are at the point where it is extremely difficult to judge the quality of comparably accurate strategies. The time-honored idea of "Pauling points" is valuable here, even at what might seem to be an absurdly high

level of theory. The HEAT345-(Q) method is in fact statistically the most accurate of the various approaches, despite being the cheapest. The conclusion that this simplest HEAT method is the most pragmatic and offers the best compromise between accuracy and cost cannot be realistically refuted. To improve systematically upon HEAT345-(Q), the results of this work suggests that:

- More accurate equilibrium geometries are needed based on CCSD(T) calculations employing core-polarized quadruple-zeta basis sets or extrapolation schemes similar to the HEAT protocol.
- The CCSD(T) energy has to be calculated with at least a septuple-zeta basis again using extrapolation or R12/F12 techniques.
- The HLC contributions have to be treated with large basis sets and extrapolation techniques, while core-correlation effects cannot be neglected.
- More accurate quantum-chemical spin-orbit stabilization energies are needed.
- The ZPE contributions can be improved using extrapolation techniques together with variational treatments for the determination of anharmonicities.

All of these improvements must be made to effect an overall improvement in the quality of the results. The field of high-accuracy thermochemistry has clearly reached an impasse, but quite a satisfactory one since the level of accuracy achievable now is sufficient for almost all needs.

The HEAT methods themself offer an interesting starting point for a possibly more accurate description of the thermochemistry for molecules containing second-row atoms, larger molecules, potential-energy surfaces, transition states and molecular properties.

As described for vinyl chloride the HEAT scheme has been generalized for molecules containing second-row atoms. The standard heat of formation of vinyl chloride is predicted to be -20.8 ± 2 kJ/mol. This prediction allowed the different experimental reported values to be discriminated.

In the case of the benzene molecule it has been shown that even for molecules of this size chemical accuracy can be achieved, while with increasing size the absolute error in the computed thermochemical properties will increase linearly with the system size. The final estimate of an more approximate HEAT scheme for the TAE is with 5467.7 ± 1.7 kJ/mol in good agreement with experiment. It is expected that open-shell systems, such as the phenyl radical, and closed-shell systems up to the size of naphthalene can be treated with similar approximate HEAT schemes yielding as well chemical accuracy for the calculated total atomization energies. Larger systems will become more challenging in terms of accuracy and computational and might be out of reach for the next decade.

Accurate prediction of nuclear magnetic shielding constants

The accurate prediction of ^{19}F nuclear magnetic shielding constants and relative NMR chemical shifts is a challenging task. Nevertheless, near-quantitative agreement with experiment (about 2 ppm deviation) can be achieved if:

- The CCSD(T) model is used to account for electron correlation.
- Large (uncontracted) basis sets are used.
- GIAOs are used to ensure gauge-origin independence,
- The calculations are performed at accurate equilibrium geometries.
- Vibrational averaging and temperature corrections via VPT2 are included.

For the CCSD(T)/13s9p4d3f calculations corrected for vibrational effects, mean and standard deviation from experiment are -1.9 and 1.6 ppm, respectively. Lower-level methods are much less satisfactory. If only molecules with hydrogen-fluorine or carbon-fluorine bonds are considered MP2/qz2p shieldings corrected for vibrational and temperature effects at the HF/tz2p level might offer a good price-performance ratio. The use of DFT with the two popular functionals BP86 and B3LYP, however, cannot be recommended, as results are not even better than the corresponding HF-SCF values. Remaining problem cases are molecules with fluorine-nitrogen and fluorine-oxygen bonds, for which generally a (slightly) larger deviation from experiment is seen. As has been shown for the F_2 molecule, the best currently available theoretical treatments are not sufficient for those cases. It seems that for those systems, coupled-cluster calculations beyond CCSD(T) are needed together with highly accurate equilibrium geometries and reliable treatments of vibrational and temperature effects.

Applying a very similar level of theory for the calculation of the ^{13}C NMR chemical shifts of benzene led again to quantitative agreement with experimental gas-phase data.

The study of the bridgehead 1-adamantyl cation resolved earlier discrepancies of lower-level theoretical treatment. Interestingly, the correlation effects on the geometry turned out to be the key to the accurate description of the NMR shieldings of this molecule. The NMR shieldings obtained at geometries from HF, DFT, and MP2 calculations show large deviations to the experimental values. Consideration of electron-correlation effects at the CCSD(T) level employing a quadruple-zeta basis set, GIAOs for the calculation of the NMR shieldings as well as vibrational corrections and temperature effects yield good agreement with the experimental data.

The results for benzene and the adamantyl cation show current possibilities in highly accurate determination of nuclear magnetic shielding constants as an example for second-order properties. With the next decade, the advances in computational resources together with the availability of parallel coupled-cluster programs will enable the accurate quantum-chemical description of systems constituted of 10-20 heavy atoms with similar accuracy as presented for the adamantyl cation.

Accurate equilibrium geometries and spectroscopic properties

The equilibrium structure of trans-HO_3 exhibits according to high-level coupled-cluster calculations a central oxygen-oxygen bond length in the range 1.58-1.62 Å and an equilibrium bond angle of 95-99° which is in contradiction to the experimental estimates of 1.688 Å and 90°.

The equilibrium structure of diacetylene has been determined based on the combination of experimental rotational constants B_0 of thirteen isotopic species and zero-point vibrational

corrections ΔB_0 calculated at various quantum-chemical levels. The empirical equilibrium structures obtained agree to within 10^{-3} Å irrespective of the theoretical level employed. The new recommended equilibrium structure of diacetylene is $r_{C-H} = 1.0615$ Å, $r_{C\equiv C} = 1.2085$ Å, and $r_{C-C} = 1.3727$ Å. The structural parameters at the highest level of theory applied (CCSD(T)/cc-pCV5Z) are $r_{C-H} = 1.0617$ Å, $r_{C\equiv C} = 1.2083$ Å, $r_{C-C} = 1.3737$ Å. Evaluation of harmonic and anharmonic force fields calculated at the CCSD(T)/cc-pCVQZ and (fc)CCSD(T)/ANO2 levels yielded harmonic and anharmonic vibrational frequencies being in good agreement with experiment, while corresponding (fc)MP2 and (fc)CCSD(T) calculations with the cc-pVXZ (X=T, Q) sets are less satisfactory.

High-level quantum-chemical calculations at the CCSD(T) level on the hyperfine structure parameters of the cyanopolyynes were found to be in excellent agreement with experiment. Thus, such calculations can be used to predict these properties for unknown species or for species for which these properties are difficult to determine experimentally. Furthermore, it could be shown that small differences between experimental quadrupole-coupling parameters and those calculated with large basis sets can largely be accounted for by including vibrational corrections to the computed values.

Within the last 25 years many attempts were made to determine the molecular structure of ferrocene by applying various quantum-chemical methods. This work presented the theoretically most accurate description to date and reveals the need for a calculation using a larger quadruple-zeta basis set. Core-correlation cannot be neglected when aiming at an accuracy of better than 0.05 Å. CCSD(T)/cc-pwCVQZ (1289 basis functions) calculations for both conformations are under way and will hopefully finally resolve the question regarding the equilibrium structure of ferrocene.

In summary, it can be stated that the implemented parallel coupled-cluster algorithms substantially extend the range of applicability, while advances in computer resources within the next two decades will enable calculations for systems up to three times larger than those considered here.

A Technical details

All presented execution time measurements were obtained on a 16 node single core 3.4 GHz Intel Xeon (EM64T) cluster with 2 MByte L2 Cache and 16 GB DDR-333 RAM each node. For the network communication a channel bonded Gigabit Ethernet was used. Channel bonding was set up using the two already built-in network interfaces of the compute nodes by using the standard Linux kernel drivers. This resulted in about 50 % more network throughput in comparison to one single Gigabit Ethernet connection per node. For the parallel implementation the message passing interface (MPI) [101] is used. The results presented here are obtained by using LAM/MPI [308, 309]. All communication in the parallel implementation is done by the MPI_ALLREDUCE subroutine.

All calculations except those at the DFT and the CCSDT(Q), CCSDTQ, CCSDTQP, and CCSDTQPH levels have been performed with the Mainz-Austin-Budapest version of the ACES II (ACES II MAB) program package [12], some of them were carried out using the recently developed parallel version [310]. The DFT calculations have been performed with the TURBOMOLE suite of programs [212,311] using tight convergence criteria (typically $< 10^{-8}$ a.u.) and very accurate settings (grid 6) for the numerical integration grid. The CCSDTQ calculations have been carried out using the string-based many-body code MRCC [94,127,128] interfaced to ACES II MAB.

The RHF-CCSD(T)/aug-cc-pCV5Z energy calculations in chapter 6.3 were performed using the MOLPRO package [93].

Bibliography

[1] J. Čížek, J. Chem. Phys. **45**, 4256 (1966).

[2] J. Gauss, in *Encyclopedia of Computational Chemistry*, Editor P. v. R. Schleyer et al., Wiley, New York, 1998, p. 615.

[3] T. D. Crawford and H. F. Schaefer III, Rev. Comp. Chem. **14**, 33 (2000).

[4] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, Chem. Phys. Lett. **157**, 479 (1989).

[5] C. Møller and M. S. Plesset, Phys. Rev. **46**, 618 (1934).

[6] A. Szabo and N. S. Ostlund, *Modern Quantum Chemistry*, Dover Publications Mineola, New York, 1996.

[7] J. C. Slater, Phys. Rev. **81**, 385 (1951).

[8] R. G. Parr and W. Yang, *Density-Functional Theory of Atoms and Molecules*, Oxford Science Publications, 1994.

[9] G. E. Moore, Electronics **35**, 114 (1965).

[10] H. Koch and A. S. de Merás, J. Chem. Phys. **113**, 508 (2000).

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

[12] J. F. Stanton, J. Gauss, J. D. Watts, W. J. Lauderdale, and R. J. Bartlett, Int. J. Quantum Chem. Symp. **26**, 879 (1992).

[13] J. F. Stanton, J. Gauss, J. D. Watts, P. G. Szalay, R. J. Bartlett, with contributions from, A. A. Auer, D. E. Bernholdt, O. Christiansen, M. E. Harding, M. Heckert, O. Heun, C. Huber, D. Jonsson, J. Jusèlius, W. J. Lauderdale, T. Metzroth, C. Michauk, D. R. Price, K. Ruud, F. Schiffmann, A. Tajti, M. E. Varner, J. Vázquez, and including the integral packages: MOLECULE (J. Almlöf and P. R. Taylor), PROPS (P. R. Taylor), ABACUS (T. Helgaker, H. J. Aa. Jensen, P. Jørgensen and J. Olsen), ACES II Mainz-Austin-Budapest version, 2005-2008, see http://www.aces2.de.

[14] A. Tajti, P. G. Szalay, A. G. Császár, M. Kállay, J. Gauss, E. F. Valeev, B. A. Flowers, J. Vázquez, and J. F. Stanton, J. Chem. Phys. **121**, 11599 (2004).

[15] Y. J. Bomble, J. Vázquez, M. Kállay, C. Michauk, P. G. Szalay, A. G. Császár, J. Gauss, and J. F. Stanton, J. Chem. Phys. **125**, 064108 (2006).

[16] M. J. Katz, P. M. Papadopoulos, and G. Bruno, IEEE International Conference on Cluster Computing, 2002. Proceedings. , 47 (2002).

[17] P. Papadopoulos, M. Katz, and G. Bruno, Concurrency and Computation: Practice & Experience (2003).

[18] P. Papadopoulos, C. Papadopoulos, M. Katz, W. Link, and G. Bruno, Int. J. High Performance Computing Applications (2004).

[19] M. Born and J. R. Oppenheimer, Ann. Physik **389**, 457 (1927).

[20] C. C. Roothaan, Rev. Mod. Phys. **23**, 69 (1951).

[21] G. G. Hall, Proc. Roy. Soc. **A205**, 541 (1951).

[22] I. Shavitt, in *Modern Theoretical Chemistry*, volume 3, page 189, Editor H. F. Schaefer III, Plenum Press, New York, 1977.

[23] I. Shavitt and H. F. Schaefer III, *Advances in Quantum Chemistry: The Configuration Interaction Method*, Academic Press, New York, 1999.

[24] T. Helgaker, P. Jørgensen, and J. Olsen, *Molecular Electronic-Structure Theory*, Wiley, Chichester, 2000.

[25] J. M. Junquera-Hernández, J. Sánchez-Marín, G. Bendazzoli, and S. Evangelisti, J. Chem. Phys. **120**, 8405 (2004).

[26] J. Čížek, Adv. Chem. Phys. **14**, 35 (1969).

[27] J. Čížek and J. Paldus, Int. J. Quantum. Chem. **5**, 359 (1971).

[28] H. J. Monkhorst, Int. J. Quantum. Chem. Symp. **11**, 421 (1977).

[29] R. J. Bartlett and G. D. Purvis, Int. J. Quantum. Chem. **14**, 561 (1978).

[30] J. A. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkley, Int. J. Quantum Chem. **14**, 545 (1978).

[31] G. D. Purvis and R. J. Bartlett, J. Chem. Phys. **76**, 1910 (1982).

[32] J. Noga and R. J. Bartlett, J. Chem. Phys. **89**, 3401 (1988).

[33] N. Oliphant and L. Adamowicz, J. Chem. Phys. **96**, 3739 (1992).

[34] S. A. Kucharski and R. J. Bartlett, J. Chem. Phys. **97**, 4282 (1992).

[35] S. Hassani, *Mathematical Physics*, Springer, 1999.

[36] S. A. Kucharski and R. J. Bartlett, Adv. Quantum. Chem. **18**, 281 (1986).

[37] J. F. Stanton, J.Gauss, J. D. Watts, and R. J.Bartlett, J. Chem. Phys. **94**, 4334 (1991).

[38] H. Koch, A. S. de Meras, T. Helgaker, and O. Christiansen, J. Chem. Phys. **104**, 4157 (1996).

[39] K. L. Bak, P. Jørgensen, J. Olsen, T. Helgaker, and W. Klopper, J. Chem. Phys. **112**, 9229 (2000).

[40] F. Pawlowski, A. Halkier, P. Jørgensen, K. L. Bak, T. Helgaker, and W. Klopper, J. Chem. Phys. **118**, 2539 (2002).

[41] G. E. Scuseria and H. F. Schaefer III, J. Chem. Phys. **89**, 382 (1988).

[42] J. A. Sordo, J. Chem. Phys. **114**, 1974 (2001).

[43] J. Gauss and J. F. Stanton, J. Chem. Phys. **116**, 1773 (2002).

[44] R. J. Bartlett, J. D. Watts, S. A. Kucharski, and J. Noga, Chem. Phys. Letters **165**, 513 (1990).

[45] Y. J. Bomble, M. Kállay, J. Gauss, and J. F. Stanton, J. Chem. Phys. **123**, 054101 (2005).

[46] M. Kállay and J. Gauss, J. Chem. Phys. **123**, 214105 (2005).

[47] M. Kállay and P. R. Surján, J. Chem. Phys. **115**, 2945 (2001).

[48] S. Hirata, J. Phys. Chem. A **107**, 9887 (2003).

[49] J. Olsen, J. Chem. Phys. **113**, 7140 (2000).

[50] P. Hohenberg and W. Kohn, Phys. Rev. **136**, B864 (1964).

[51] P. A. M. Dirac, Proc. Royal Soc. (London) A **123**, 714 (1929).

[52] W. Kohn and L. J. Sham, Phys. Rev. **140**, A1133 (1965).

[53] A. K. Rajagopal, Adv. Chem. Phys. **41**, 59 (1980).

[54] R. Kolle and O. Salvetti, Chim. Acta **37**, 329 (1975).

[55] A. D. Becke, Phys. Rev. A **38**, 3098 (1988).

[56] J. Gauss, in *Modern Methods and Algorithms of Quantum Chemistry*, Editor J. G. Gotendorst, Neumann Institute for Computing, Jülich, 2002.

[57] H. B. Schlegel, in *Modern Electronic Structure Theory*, Editor D. R. Yarkony, World Scientific, Singapore, 1995, p. 459.

[58] J. Geratt and I. M. Mills, J. Chem. Phys. **49**, 1719 (1968).

[59] J. Pople, R. Krishnan, H. B. Schlegel, and J. S. Binkley, Int. J. Quantum Chem. Symp. **13**, 225 (1979).

[60] A. C. Scheiner, G. E. Scuseria, J. E. Rice, T. J. Lee, and H. F. Schaefer III, J. Chem. Phys. **87**, 5361 (1987).

[61] L. Adamowicz, W. D. Laidig, and R. J. Bartlett, Int. J. Quant. Chem. Symp. **18**, 245 (1984).

[62] J. E. Rice and R. D. Amos, Chem. Phys. Lett. **122**, 585 (1985).

[63] J. Gauss and J. F. Stanton, Phys. Chem. Chem. Phys. **2**, 2047 (2000).

[64] H. Koch, H. J. Aa. Jensen, T. Helgaker, P. Jørgensen, G. E. Scuseria, and H. F. Schaefer III, J. Chem. Phys. **92**, 4924 (1990).

[65] J. Gauss and J. F. Stanton, Chem. Phys. Letters **276**, 70 (1997).

[66] M. Kállay and J. Gauss, J. Chem. Phys. **120**, 6841 (2004).

[67] J. F. Stanton and J. Gauss, Int. Rev. Phys. Chem. **19**, 61 (2000).

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

[69] D. E. Woon and T. H. Dunning, Jr., J. Chem. Phys. **98**, 1358 (1993).

[70] A. Wilson, T. van Mourik, and T. H. Dunning, Jr., J. Mol. Struct. THEOCHEM **388**, 339 (1996).

[71] A. K. Wilson, D. E. Woon, K. A. Peterson, and T. H. Dunning, Jr., J. Chem. Phys. **110**, 76677 (1999).

[72] T. van Mourik and T. H. Dunning, Jr., Int. J. Quantum Chem. **76**, 205 (2000).

[73] N. B. Balabanov and K. A. Peterson, J. Chem. Phys. **123**, 064107 (2005).

[74] D. Feller, J. Chem. Phys. **96**, 6104 (1992).

[75] T. Helgaker, W. Klopper, H. Koch, and J. Noga, J. Chem. Phys. **106**, 9639 (1997).

[76] D. E. Woon and T. H. Dunning, Jr., J. Chem. Phys. **103**, 4572 (1995).

[77] K. A. Peterson and T. H. Dunning, Jr., J. Chem. Phys. **117**, 10548 (2002).

[78] N. J. DeYonker, K. A. Peterson, and A. Wilson, J. Phys. Chem. A **111**, 11383 (2007).

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

[80] D. E. Woon and T. H. Dunning, Jr., J. Chem. Phys. **100**, 2975 (1994).

[81] T. van Mourik, Mol. Phys. **96**, 529 (1999).

[82] J. Gauss, J. Chem. Phys. **99**, 3629 (1993).

[83] A. A. Auer, J. Gauss, and J. F. Stanton, J. Chem. Phys. **118**, 10407 (2003).

[84] A. Schäfer, H. Horn, and R. Ahlrichs, J. Chem. Phys. **97**, 2571 (1992).

[85] J. Almlöf and P. R. Taylor, J. Chem. Phys. **86**, 4070 (1987).

[86] F. Weigend and R. Ahlrichs, Phys. Chem. Chem. Phys. .

[87] F. Weigend, F. Furche, and R. Ahlrichs, J. Chem. Phys. **119**, 12753 (2003).

[88] M. Kállay and J. Gauss, J. Chem. Phys **123**, 214105 (2005).

[89] P. Botschwina, Theor. Chem. Acc. **114**, 350 (2005).

[90] J. G. Hill, J. A. Platts, and H. J. Werner, Phys. Chem. Chem. Phys. **8**, 4072 (2006).

[91] R. M. Olson, J. L. Bentz, R. A. Kendall, M. W. Schmidt, and M. S. Gordon, J. Chem. Theory Comput. **3**, 1312 (2007).

[92] J. D. Watts, Parallel Computing **26**, 857 (2000).

[93] H.-J. Werner, P. J. Knowles, R. Lindh, F. R. Manby, M. Schütz, P. Celani, T. Korona, G. Rauhut, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, C. Hampel, G. Hetzer, A. W. Lloyd, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, P. Palmieri, R. Pitzer, U. Schumann, H. Stoll, A. J. Stone, R. Tarroni, and T. Thorsteinsson, Molpro, version 2006.1, a package of ab initio programs, 2006, see http://www.molpro.net.

[94] M. Kállay, MRCC, 2007, see http://www.mrcc.hu.

[95] A. Köhn and C. Hättig, J. Chem. Phys. **118**, 7751 (2003).

[96] S. Hirata, J. Phys. Chem. A **107**, 9887 (2003).

[97] E. Aprá, T. L. Windus, T. P. Straatsma, E. J. Bylaska, W. de Jong, S. Hirata, M. Valiev, M. Hackler, L. Pollack, K. Kowalski, R. Harrison, M. Dupuis, D. M. A. Smith, J. Nieplocha, V. Tipparaju, M. Krishnan, A. A. Auer, E. Brown, G. Cisneros, G. Fann, H. Fruchtl, J. Garza, K. Hirao, R. Kendall, J. Nichols, K. Tsemekhman, K. Wolinski, J. Anchell, D. Bernholdt, P. Borowski, T. Clark, D. Clerc, H. Dachsel, M. Deegan, K. Dyall, D. Elwood, E. Glendening, M. Gutowski, A. Hess, J. Jaffe, B. Johnson, J. Ju, R. Kobayashi, R. Kutteh, Z. Lin, R. Littlefield, X. Long, B. Meng, T. Nakajima, S. Niu, M. Rosing, G. Sandrone, M. Stave, H. Taylor, G. Thomas, J. van Lenthe, A. Wong, and Z. Zhang, "NWChem, A Computational Chemistry Package for Parallel Computers, Version 4.7" , Pacific Northwest National Laboratory, Richland, Washington 99352 (2005).

[98] M. W. Schmidt, K. K. Baldridge, J. A. Boatz, S. T. Elbert, M. S. Gordon, J. H. Jensen, S. Koseki, N. Matsunaga, K. A. N. S. Su, T. L. Windus, M. Dupuis, and J. A. Montgomery, J. Comput. Chem. **14**, 1347 (1993).

[99] PQS version 3.2; Parallel Quantum Solutions, 2005, see http://www.pqs-chem.com.

[100] J. Noga, W. Klopper, T. Helgaker, and P. Valiron, DIRCCR12-OS, 2004, see http://www-laog.obs.ujf-grenoble.fr/~valiron/ccr12.

[101] The MPI Forum, MPI: a message passing interface. In Proceedings of the 1993 ACM/IEEE conference on Supercomputing, ACM Press: Portland, OR, United States, 1993.

[102] J. Nieplocha, R. J. Harrison, and R. Littlefield, Proc. Supercomputing 1994; IEEE Computer Society Press: Washington, D.C. , 340 (1994).

[103] J. Nieplocha, B. Palmer, V. Tipparaju, K. Manojkumar, H. Trease, and E. Aprá, Int. J. High Perform, Comput. Appl. **20**, 203 (2006).

[104] R. M. Olson, M. W. Schmidt, M. S. Gordon, and A. P. Rendell. Enabling the Efficient Use of SMP Clusters: The GAMESS/DDI Approach. In Supercomputing, 2003 ACM/IEEE Conference, Phoenix, AZ, 2003, p 41.

[105] A. R. Ford, T. Janowski, and P. Pulay, J. Comput. Chem. **28**, 1215 (2007).

[106] T. Janowski, A. R. Ford, and P. Pulay, J. Chem. Theory Comput. **3**, 1368 (2007).

[107] G. Baumgartner, A. A. Auer, D. E. Bernholdt, A. Bibireata, V. Choppella, D. Cociorva, X. Gao, R. Harrison, S. Hirata, S. Krishanmoorthy, S. Krishnan, C.-C. Lam, M. Nooijen, R. Pitzer, J. Ramanujam, P. Sadayappan, and A. Sibiryakov, Proceedings of the IEEE **93**, 276 (2005).

[108] J. D. Watts, J. Gauss, and R. J. Bartlett, Chem. Phys. Letters **200**, 1 (1992).

[109] P. G. Szalay, J. Gauss, and J. F. Stanton, Theor. Chem. Acc. **100**, 5 (1998).

[110] A. P. Rendell, M. F. Guest, and R. A. Kendall, J. Comput. Chem. **14**, 1429 (1993).

[111] A. A. Auer, G. Baumgartner, D. E. Bernholdt, A. Bibireata, V. Choppella, D. Cociorva, X. Gao, R. Harrison, S. Krishanmoorthy, S. Krishnan, Q. Lu, C.-C. Lam, M. Nooijen, R. Pitzer, J. Ramanujam, P. Sadayappan, and A. Sibiryakov, Mol. Phys. **104**, 211 (2006).

[112] J. D. Watts, J. Gauss, and R. J. Bartlett, J. Chem. Phys. **98**, 8718 (1993).

[113] J. Gauss and J. F. Stanton, J. Chem. Phys. **104**, 2574 (1996).

[114] A. P. Rendell and T. J. Lee, J. Chem. Phys. **94**, 6219 (1991).

[115] J. Gauss and J. F. Stanton, J. Phys. Chem. A **104**, 2865 (2000).

[116] J. Gauss, J. F. Stanton, and R. J. Bartlett, J. Chem. Phys. **95**, 2623 (1991).

[117] W. Meyer, J. Chem. Phys. **64**, 1975 (1975).

[118] R. Ahlrichs and C. Zirz, Proceedings of the workshop "Molecular Physics and Quantum Chemistry" (Wollongong, 1980) .

[119] R. Ahlrichs and C. Zirz, Theor. Chim. Acta **36**, 275 (1976).

[120] J. A. Pople, J. S. Binkley, and R. Seeger, Int. J. Quantum Chem. Symp. **10**, 1 (1976).

[121] C. Hampel, K. Peterson, and H.-J. Werner, Chem. Phys. Lett. **192**, 1 (1992).

[122] H. Koch, O. Christiansen, R. Kobayashi, P. Jørgensen, and T. Helgaker, Chem. Phys. Lett. **228**, 233 (1994).

[123] J. Gauss and J. F. Stanton, J. Chem. Phys. **103**, 3561 (1995).

[124] E. Fossgård and K. Ruud, J. Comput. Chem. **27**, 326 (2006).

[125] D. Price, J. Vázquez, P. G. Szalay, M. E. Harding, J. Gauss, and J. F. Stanton, to be published, 2008.

[126] Y. J. Bomble, J. F. Stanton, M. Kállay, and J. Gauss, J. Chem. Phys. **123**, 054101 (2005).

[127] M. Kállay and P. R. Surján, J. Chem. Phys. **113**, 1359 (2000).

[128] M. Kállay and M. E. Harding, parallel version of the string-based general coupled-cluster program MRCC, 2006, see http://www.mrcc.hu.

[129] J. M. L. Martin and G. de Oliveira, J. Chem. Phys. **111**, 1843 (1999).

[130] A. D. Boese, M. Oren, O. Atasoylu, J. M. L. Martin, M. Kállay, and J. Gauss, J. Chem. Phys. **120**, 4129 (2004).

[131] A. Karton, E. Rabinovich, J. M. L. Martin, and B. Ruscic, J. Chem. Phys. **125**, 144108 (2006).

[132] A. G. Császár, W. D. Allen, and H. F. Schaefer III, J. Chem. Phys. **108**, 9751 (1998).

[133] D. A. Clabo, W. D. Allen, R. B. Remington, Y. Yamaguchi, and H. F. Schaefer III, Chem. Phys. **123**, 187 (1988).

[134] W. Schneider and W. Thiel, Chem. Phys. Lett. **157**, 367 (1989).

[135] J. F. Stanton and J. Gauss, Int. Rev. Phys. Chem. **19**, 61 (2000).

[136] T. A. Ruden, P. R. Taylor, and T. Helgaker, J. Chem. Phys. **119**, 1951 (2003).

[137] B. Ruscic, R. E. Pinzon, M. L. Morton, G. von Laszevski, S. J. Bittner, S. G. Nijsure, K. A. Amin, M. Minkoff, and A. F. Wagner, J. Phys. Chem. A **108**, 9979 (2004).

[138] J. F. Stanton, Chem. Phys. Lett. **281**, 130 (1997).

[139] T. A. Ruden, T. Helgaker, P. Jørgensen, and J. Olsen, Chem. Phys. Lett. **371**, 62 (2003).

[140] J. Olsen, J. Chem. Phys. **113**, 7140 (2000).

[141] M. Kállay and J. Gauss, J. Chem. Phys. **123**, 214105 (2005).

[142] I. M. Mills, in Modern Spectroscopy: Modern Research, edited by K. N. Rao and C. W. Matthews, Academic Press, New York, 1972, p. 115.

[143] V. Barone and C. Minichino, J. Mol. Struct.: THEOCHEM **330**, 365 (1995).

[144] V. Barone, J. Chem. Phys. **120**, 3059 (2004).

[145] M. S. Schuurman, S. R. Muir, W. D. Allen, and H. F. Schaefer III, J. Chem. Phys. **120**, 11586 (2004).

[146] J. Vázquez and J. F. Stanton, Mol. Phys. **104**, 377 (2006).

[147] J. Gauss, A. Tajti, M. Kállay, J. F. Stanton, and P. G. Szalay, J. Chem. Phys. **125**, 144111 (2006).

[148] M. E. Harding, J. Vázquez, B. Ruscic, A. K. Wilson, J. Gauss, and J. F. Stanton, J. Chem. Phys. **128**, 114111 (2008).

[149] N. C. Handy, Y. Yamaguchi, and H. F. Schaefer III, J. Chem. Phys. **84**, 4481 (1986).

[150] R. D. Cowan and M. Griffin, J. Opt. Soc. Am. **66**, 1010 (1976).

[151] R. L. Martin, J. Phys. Chem. **87**, 750 (1983).

[152] W. Klopper, J. Comput. Chem. **18**, 20 (1997).

[153] C. Michauk and J. Gauss, J. Chem. Phys. **127**, 044106 (2007).

[154] A. Tajti, P. G. Szalay, and J. Gauss, J. Chem. Phys. **127**, 014102 (2007).

[155] K. P. Huber and G. Herzberg, *Molecular Spectra and Molecular Structure IV: Constants of Diatomic Molecules*, Van Nostrand, New York, 1979.

[156] C. E. Moore, *NSRDS-NBS-35*, U. S. National Bureau Standards: Gaithersburg, MD, 1970.

[157] B. Ruscic, A. F. Wagner, L. B. Harding, R. L. Asher, D. Feller, D. A. Dixon, K. A. Peterson, Y. Song, X. Qian, C.-Y. Ng, J. Liu, W. Chen, and D. W. Schwenke, J. Chem Phys. A. **106**, 2727.

[158] A. G. Császár and I. M. Mills, Spectrochim. Acta **A53**, 1101 (1997).

[159] A. Karton, P. R. Taylor, and J. M. L. Martin, J. Chem. Phys. **127**, 064104 (2007).

[160] M. Heckert, M. Kállay, D. P. Tew, W. Klopper, and J. Gauss, J. Chem. Phys **125**, 044108 (2006).

[161] W. Klopper, F. R. Manby, S. Ten-no, and E. F. Valeev, Int. Rev. Phys. Chem. **25**, 427 (2006).

[162] F. Wang, J. Gauss, and C. van Wüllen, J. Chem. Phys **129**, 064113 (2008).

[163] K. Klein and J. Gauss, submitted to J. Chem. Phys. .

[164] D. P. Tew, W. Klopper, M. Heckert, and J. Gauss, J. Phys. Chem. A **111**, 11242 (2007).

[165] J. C. Light, I. P. Hamilton, and J. V. Lill, J. Chem. Phys. **82**, 1400 (1985).

[166] M. Harding, J. Vázquez, J. F. Stanton, and J. Gauss, to be published. (2008).

[167] K. Pflüger, H.-J. Werner, A. Schäfer, and M. Schütz, to be published (2008).

[168] H.-J. Werner and K. Pflüger, in *Ann. Reports in Comp. Chem.*, edited by D. Spellmeyer, volume 2, page 53, Elsevier, Amsterdam, 2006.

[169] J. R. Lacher, E. Emery, E. Bohmfalk, and J. D. Park, J. Phys. Chem. **60**, 492 (1956).

[170] J. R. Lacher, A. Kianpour, F. Oetting, and J. D. Park, Trans. Faraday Soc. **62**, 1500 (1956).

[171] J. R. Lacher, H. B. Gottlieb, and J. D. Park, Trans. Faraday Soc. **58**, 2348 (1962).

[172] J. D. Cox and G. Pilcher, *Thermochemistry of Organic and Organometallic Compounds*, Academic Press, New York, 1970.

[173] Z. B. Alfassi, D. M. Golden, and S. W. Benson, J. Chem. Thermodyn. **5**, 411 (1973).

[174] S. V. Levanova, Y. A. Treger, S. M. Velichko, and A. M. Rozhnov, Russ. J. Phys. Chem. (Engl. Transl.) **50**, 1148 (1976).

[175] E. R. Ritter, J. Chem. Inf. Comp. Sci. **31**, 400 (1991).

[176] L. V. Gurvich, *Elements C, Si, Ge, Sn, Pb, and their compounds*, volume 2 of *Thermodynamic properties of individual substances*, 1991.

[177] J. A. Manion, J. Phys. Chem. Ref. Data **31**, 123 (2002).

[178] C. E. Moore, *NSRDS-NBS-3*, U. S. National Bureau Standards: Gaithersburg, MD, 1970.

[179] C. E. Moore, *NSRDS-NBS-34*, U. S. National Bureau Standards: Gaithersburg, MD, 1970.

[180] A. Karton, B. Ruscic, and J. M. L. Martin, J. Mol. Struc. **811**, 345 (2007).

[181] J. D. Cox, D. D. Wagman, and V. A. Medvedev, *CODATA Key values for Thermodynamics*, Hemisphere Publishing Corp., New York, 1984.

[182] N. S. Shuman, M. A. Ochieng, B. Sztaray, and T. Baer, J Phys Chem A **112**, 5647 (2008).

[183] S. Parthiban and J. M. L. Martin, J. Chem. Phys. **115**, 2051 (2001).

[184] E. J. Prosen, R. Gilmont, and F. Rossini, J. Res. NBS. **34**, 64 (1945).

[185] F. London, J. Phys. Radium **8**, 397 (1937).

[186] H. Hameka, Mol. Phys. **1**, 203 (1958).

[187] H. Hameka, Z. Naturforsch. **A 14**, 599 (1959).

[188] R. Ditchfield, J. Chem. Phys. **56**, 5688 (1972).

[189] R. Ditchfield, Mol.Phys. **27**, 789 (1974).

[190] K. Wolinski, J. F. Hinton, and P. Pulay, J. Am. Chem. Soc. **112**, 8251 (1990).

[191] S. Berger, S. Braun, and H.-O. Kalinowski, *NMR-Spektroskopie der Nichtmetalle, Bd. 4,* ^{19}*F-NMR-Spektroskopie*, Wiley-VCH, 1994.

[192] T. Helgaker, M. Jaszunski, and K. Ruud, Chem. Rev. **99**, 293 (1999).

[193] J. Gauss and J. F. Stanton, Adv. Chem. Phys. **123**, 355 (2002).

[194] U. Fleischer and M. Schindler, Chem. Phys. **120**, 103 (1988).

[195] U. Gross and R. Wolff, J. Fluorine Chem. **94**, 115 (1999).

[196] W. Adcock, D. Lunsmann, J. Peralta, and R. Contreras, Magn. Reson. Chem. **37**, 167 (1999).

[197] K. B. Wiberg and K. W. Zilm, J. Org. Chem. **66**, 2809 (2001).

[198] S. Cai, Z. Chen, and H. Wan, J. Phys. Chem. A **106**, 1060 (2002).

[199] W. Adcock, J. E. Peralta, and R. H. Contreras, Magn. Reson. Chem. **41**, 503 (2003).

[200] P.-O. Åstrand and K. Ruud, Phys. Chem. Chem. Phys. **5**, 5015 (2003).

[201] H. Fukaya and T. Ono, J. Comput. Chem. **25**, 15 (2004).

[202] X. Zhao, J. Devries, R. Mcdonald, and B. Sykes, J. Magn. Res. **187**, 88 (2007).

[203] D. Sundholm, J. Gauss, and A. Schäfer, J. Chem. Phys. **105**, 11051 (1996).

[204] C. J. Jameson, A. K. Jameson, and P. M. Burrell, J. Chem. Phys. **73**, 6013 (1980).

[205] C. J. Jameson, A. K. Jameson, and J. Honarbakhsh, J. Chem. Phys. **81**, 5266 (1984).

[206] J. Gauss, Chem. Phys. Letters **191**, 614 (1992).

[207] J. Gauss and J. F. Stanton, J. Chem. Phys. **102**, 251 (1995).

[208] J. Gauss and J. F. Stanton, J. Chem. Phys. **104**, 2574 (1996).

[209] J. Gauss, J. Chem. Phys. **116**, 4773 (2002).

[210] J. P. Perdew, Phys. Rev. B **33**, 8822 (1986).

[211] A. D. Becke, J. Chem. Phys. **98**, 5648 (1993).

[212] R. Ahlrichs, M. Bär, M. Häser, H. Horn, and C. Kölmel, Chem. Phys. Letters **162**, 165 (1989), TURBOMOLE V5-7-1, Current version, see http://www.turbomole.com.

[213] Y. Morino and E. Hirota, B. Chem. Soc. Jpn. **31**, 423 (1958).

[214] J. F. Stanton, C. L. Lopreore, and J. Gauss, J. Chem. Phys. **108**, 7190 (1998).

[215] J. F. Stanton and J. Gauss, Int. Rev. Phys. Chem. **19**, 61 (2000).

[216] T. Helgaker, J. Gauss, P. Jørgensen, and J. Olsen, J. Chem. Phys. **106**, 6430 (1997).

[217] A. Halkier, P. Jørgensen, J. Gauss, and T. Helgaker, Chem. Phys. Letters **274**, 235 (1997).

[218] C. J. Jameson, Chem. Rev. **91**, 1375 (1991).

[219] D. Sundholm and J. Gauss, Mol. Phys. **92**, 1007 (1997).

[220] M. Bühl, *Encyclopedia of Computational Chemistry*, page 1835, Wiley, Chichester, 1998.

[221] E. Dumont and P. Chaquin, Chem. Phys. Lett. **435**, 354 (2007).

[222] J. Cheng, W. Zhu, Y. Tang, Y. Xu, Z. Li, K. Chen, and H. Jiang, Chem. Phys. Lett. **422**, 455 (2006).

[223] T. Heine, C. Corminboeuf, G. Grossmann, and U. Haeberlen, Angew. Chem. Int. Ed. **45**, 7292 (2006).

[224] A. K. Jameson and C. J. Jameson, Chem. Phys. Lett. **134**, 461 (1987).

[225] P. Buzek, P. v. R. Schleyer, and S. Sieber, Chemie in unserer Zeit **26**, 116 (1992).

[226] M. Bühl, J. Gauss, M. Hofmann, and P. v. R. Schleyer, J. Am. Chem. Soc. **115**, 12385 (1993).

[227] P. v. R. Schleyer, J. Gauss, M. Bühl, R. Greatrex, and M. Fox, J. Chem. Soc., Chem. Commun. , 1766 (1993).

[228] S. Sieber, P. v. R. Schleyer, and J. Gauss, J. Am. Chem. Soc. **115**, 6987 (1993).

[229] P. Buzek, P. v. R. Schleyer, H. Vancik, Z. Mihalic, and J. Gauss, Angew. Chem. **104**, 470 (1994).

[230] C. Ochsenfeld, S. P. Brown, I. Schnell, J. Gauss, and H. W. Spiess, J. Am. Chem. Soc. **123**, 2597 (2001).

[231] S. P. Brown, T. Schaller, U. P. Seelbach, F. Koziol, C. Ochsenfeld, F.-G. Klärner, and H. W. Spiess, Angew. Chem., Int. Ed. Engl. **40**, 717 (2001).

[232] D. R. Price and J. F. Stanton, Org. Lett. **4**, 2809 (2002).

[233] E. Prochnow, A. A. Auer, and K. Banert, J. Phys. Chem. A. **111**, 9945 (2007).

[234] G. A. Olah, *Cage Hydrcarbons*, Wiley, New York, 1990.

[235] G. A. Olah, G. K. S. Prakash, J. G. Shih, V. V. Krishnamurthy, G. D. Mateescu, G. Liang, G. Sipos, V. Buss, T. M. Gund, and P. v. R. Schleyer, J. Am. Chem. Soc. **107**, 2764 (1985).

[236] J. Gauss, unpublished results .

[237] G. Rasul, G. A. Olah, and G. Prakash, Proc. Natl. Acad. Sci. U.S.A. **101**, 10868 (2004).

[238] A. A. Auer and J. Gauss, Phys. Chem. Chem. Phys. **3**, 3001 (2001).

[239] J. Kraitchman, Am. J. Phys. **21**, 17 (1953).

[240] C. C. Constain, J. Chem. Phys. **29**, 864 (1958).

[241] P. Pyykkö, Mol. Phys. **99**, 1617 (2001).

[242] W. H. Flygare, Chem. Rev. **74**, 653 (1974).

[243] J. Gauss, K. Ruud, and T. Helgaker, J. Chem. Phys. **105**, 2804 (1996).

[244] K. Mathisen and P. Siegbahn, Chem. Phys. **90**, 225 (1984).

[245] T. Jungkamp and J. Seinfeld, Chem. Phys. Lett. **257**, 15 (1996).

[246] J. Orlando, G. Tyndall, and T. Wallington, Chem. Rev. **103**, 4657 (2003).

[247] C. Murray, E. Derro, T. Sechler, and M. Lester, J. Phys. Chem. A **111**, 4727 (2007).

[248] O. Setokuchi, M. Sato, and S. Matuzawa, J. Phys. Chem. A **104**, 3204 (2000).

[249] H. Yu and A. Varandas, Chem. Phys. Lett. **334**, 173 (2001).

[250] P. Denis, M. Kieninger, O. Ventura, R. Cachau, and G. Diercksen, Chem. Phys. Lett. **365**, 440 (2002).

[251] W. Fabian, J. Kalcher, and R. Janoschek, Theor. Chem. Acc. **114**, 182 (2005).

[252] A. Mansergas, J. Anglada, S. Olivella, M. Ruiz-López, and M. Martins-Costa, Phys. Chem. Chem. Phys. **9**, 5865 (2007).

[253] K. Suma, Y. Sumiyoshi, and Y. Endo, Science **308**, 1885 (2005).

[254] M. Heckert, M. Kállay, and J. Gauss, Mol. Phys. **103**, 2109 (2005).

[255] J. F. Stanton, M. E. Varner, M. E. Harding, and J. Gauss, Chem. Phys. **346**, 53 (2008).

[256] J. Noga and R. J. Bartlett, J. Chem. Phys. **86**, 7041 (1987).

[257] E. Hirota, Bull. Chem. Soc. Jpn. **31**, 130 (1958).

[258] I. A. Maretina and B. A. Trofimov, Russ. Chem. Rev. **69**, 591 (2000).

[259] V. G. Kunde, A. C. Aikin, R. A. Hanel, D. E. Jennings, W. C. Maguire, and R. E. Samuelson, Nature **292**, 686 (1981).

[260] J. Cernicharo, A. M. Heras, A. G. G. M. Tielens, J. R. Pardo, F. Herpin, M. Guélin, and L. B. F. M. Waters, Astrophys. J. **546**, L123 (2001).

[261] R. Tay, G. F. Metha, F. Shanks, and D. McNaughton, Struct. Chem. **6**, 47 (1995).

[262] K. Matsumura, R. D. Suenram, F. J. Lovas, and T. Tanaka, J. Mol. Spectrosc. **240**, 120 (2006).

[263] R. J. Berry and M. D. Harmony, Struct. Chem. **1**, 49 (1990).

[264] P. Botschwina and C. Puzzarini, J. Mol. Spectrosc. **208**, 292 (2001).

[265] R. J. McMahon, R. J. Halter, R. J. Fimmen, R. J. Wilson, S. A. Peebles, R. L. Kuczkowski, and J. F. Stanton, J. Am. Chem. Soc. **122**, 939 (2000).

[266] E. Arié and J. W. C. Johns, J. Mol. Spectrosc. **155**, 195 (1992).

[267] S. Thorwirth, M. E. Harding, D. Muders, and J. Gauss, J. Mol. Spectr. **251**, 220 (2008).

[268] C. Puzzarini, M. Heckert, and J. Gauss, J. Chem. Phys. **128**, 194108 (2008).

[269] G. Cazzoli, C. Puzzarini, L. Fusina, and F. Tamassia, J. Mol. Spectrosc. **247**, 115 (2008).

[270] P. Botschwina, Mol. Phys. **103**, 1441 (2005).

[271] G. Cazzoli, L. Cludi, M. Contento, and C. Puzzarini, J. Mol. Spectrosc. **251**, 229 (2008).

[272] M. Le Guennec, J. Demaison, G. Wlodarczak, and C. J. Marsden, J. Mol. Spectrosc. **160**, 471 (1993).

[273] R. J. Halter, R. L. Fimmen, R. J. McMahon, S. A. Peebles, R. L. Kuczkowksi, and J. F. Stanton, J. Am. Chem. Soc. **123**, 12353 (2001).

[274] N. C. Craig, P. Groner, and D. C. McKean, J. Phys. Chem. A **110**, 7461 (2006).

[275] D. McNaughton and D. N. Bruget, J. Mol. Struct. **273**, 11 (1992).

[276] D. A. Matthews, J. Vázquez, and J. F. Stanton, Mol. Phys. **105**, 2659 (2007).

[277] X. Zhang, M. R. Nimlos, G. B. Ellison, M. E. Varner, and J. F. Stanton, J. Chem. Phys. **126**, 174308 (2007).

[278] I. M. Konen, E. X. J. Li, M. I. Lester, J. Vázquez, and J. F. Stanton, J. Chem. Phys. **125**, 074310 (2006).

[279] J. M. L. Martin, P. R. Taylor, and T. J. Lee, Chem. Phys. Let. **275**, 414 (1997).

[280] J. M. L. Martin, T. J. Lee, P. R. Taylor, and J.-P. Francois, J. Chem. Phys. **103**, 2589 (1995).

[281] B. Turner, Astrophys. J. **163**, L35 (1971).

[282] F. Wyrowski, P. Schilke, and C. Walmsley, Astron. Astrophys. **341**, 882 (1999).

[283] F. Wyrowski, P. Schilke, S. Thorwirth, K. Menten, and G. Winnewisser, Astrophys. J. **586**, 344 (2003).

[284] W. Langer, F. Schloerb, R. Snell, and J. Young, Astrophys. J. **239**, L125 (1980).

[285] S. Thorwirth, H. S. P. Müller, and G. Winnewisser, J. Mol. Spectrosc. **204** (2000).

[286] M. E. Sanz, M. McCarthy, and P. Thaddeus, J. Chem. Phys. **122**, 194319 (2005).

[287] S. Thorwirth, H. S. P. Müller, and G. Winnewisser, Phys. Chem. Chem. Phys. **3**, 1236 (2001).

[288] A. Fayt, C. Vigouroux, F. Willaert, L. Margulès, L. Constantin, J. Demaison, G. Pawelke, E. B. Mkadmi, and H. Bürger, J. Mol. Struct. **695**, 295 (2004).

[289] R. DeZafra and P. Mallinson, Mol. Phys. **36**, 827 (1978).

[290] E. Fliege, H. Dreizler, and B. Kleibömer, J. Mol. Struct. **97**, 225 (1983).

[291] L. Tack and S. Kukolich, J. Chem. Phys. **78**, 6512 (1983).

[292] P. Botschwina, Mol. Phys. **103**, 1441 (2005).

[293] H. Spahn, H. S. P. Müller, T. F. Giesen, J.-U. Grabow, M. E. Harding, J. Gauss, and S. Schlemmer, Chem. Phys. **346**, 132 (2008).

[294] L. Bizzochi, C. Degli Esposti, and P. Botschwina, J. Mol. Spectrosc. **225**, 145 (2004).

[295] M. C. McCarthy, E. S. Levine, A. J. Apponi, and P. Thaddeus, J. Mol. Spectrosc. **203**, 75 (2000).

[296] G. Winnewisser, R. A. Creswell, and M. Winnewisser, Z. Naturforsch. A: Phys. Sci. **33**, 1169 (1978).

[297] H. P. Lüthi, J. H. Ammenter, J. Almlöf, and K. Fægri, J. Chem. Phys. **77**, 2002 (1982).

[298] W. Klopper and H. P. Lüthi, Chem. Phys. Lett. **262**, 546 (1996).

[299] H. Koch, P. Jørgensen, and T. Helgaker, J. Chem. Phys. **104**, 9528 (1996).

[300] H. P. Lüthi, J. Mol. Struct.: THEOCHEM **388**, 299 (1996).

[301] Z.-F. Xu, Y. Xie, W.-L. Feng, and H. F. Schaefer III, J. Phys. Chem. A **107**, 2716 (2003).

[302] S. Coriani, A. Haaland, T. Helgaker, and P. Jørgensen, Chem. Phys. Chem. **7**, 245 (2006).

[303] A. Haaland and J. Nilssen, Acta Chem. Scand. **22**, 2653 (1968).

[304] A. Haaland, J. Lusztyk, D. P. Novak, J. Brunvoll, and K. B. Starowieysky, J. Chem. Soc., Chem. Commun. , 54 (1974).

[305] P. Seiler and J. D. Dunitz, Acta Ccrystallogr. **38**, 1741 (1982).

[306] F. Takusagawa and T. M. Koetzle, Acta Ccrystallogr. **35**, 1074 (1979).

[307] C. P. Broc and Y. Fu, Acta Ccrystallogr. B **53**, 928 (1997).

[308] G. Burns, R. Daoud, and J. Vaigl, LAM: An Open Cluster Environment for MPI, in *Proceedings of Supercomputing Symposium*, pages 379–386, 1994.

[309] J. M. Squyres and A. Lumsdaine, A Component Architecture for LAM/MPI, in *Proceedings, 10th European PVM/MPI Users' Group Meeting*, number 2840 in Lecture Notes in Computer Science, pages 379–387, Venice, Italy, 2003, Springer-Verlag.

[310] M. E. Harding, T. Metzroth, J. Gauss, and A. A. Auer, J. Chem. Theor. Comp. **4**, 64 (2008).

[311] M. Häser, R. Ahlrichs, H. P. Baron, P. Weis, and H. Horn, Theor. Chim. Acta **83**, 455 (1993).