Theoretical Studies of Nuclear Magnetic Resonance Chemical Shifts

Dissertation
zur
Erlangung des Doktorgrades (Dr. rer. nat.)
der
Mathematisch-Naturwissenschaftlichen Fakultät
der
Rheinischen Friedrich-Wilhelms-Universität Bonn

von
Julius Benedikt Kleine Büning geb. Stückrath
aus
Moers

Bonn, Januar 2024

Angefertigt mit Genehmigung der Mathematisch-Naturwissenschaftlichen Fakultät der Rheinischen Friedrich-Wilhelms-Universität Bonn

1. Gutachter:	Prof. Dr. Stefan Grimme
2. Gutachter:	Prof. Dr. Thomas Bredow
Tag der Promotion:	19.04.2024
Erscheinungsjahr:	2024

Relativity ideas [...] are therefore of no importance in the consideration of atomic and molecular structure and ordinary chemical reactions [...].[1]

— Paul A. M. Dirac

Publication List

The following publications serve as parts of this thesis.

1. M. Bursch, T. Gasevic, J. B. Stückrath, and S. Grimme. "Comprehensive Benchmark Study on the Calculation of ^{29}Si NMR Chemical Shifts."
Inorg. Chem. **2021**, *60*, 272–285. DOI: 10.1021/acs.inorgchem.0c02907.

2. J. B. Stückrath, T. Gasevic, M. Bursch, and S. Grimme. "Benchmark Study on the Calculation of ^{119}Sn NMR Chemical Shifts."
Inorg. Chem. **2022**, *61*, 3903–3917. DOI: 10.1021/acs.inorgchem.1c03453.

3. J. B. Kleine Büning and S. Grimme. "Computation of CCSD(T)-Quality NMR Chemical Shifts via Δ-Machine Learning from DFT."
J. Chem. Theory Comput. **2023**, *19*, 3601–3615. DOI: 10.1021/acs.jctc.3c00165.

4. J. B. Kleine Büning, S. Grimme, and M. Bursch. "Machine Learning-Based Correction for Spin-Orbit Coupling Effects in NMR Chemical Shift Calculations."
Phys. Chem. Chem. Phys. **2024**, *Advance Article*. DOI: 10.1039/d3cp05556f.

5. D. Kalle, J. Bahr, T. J. Keller, J. B. Kleine Büning, S. Grimme, M. Bursch, S.-S. Jester, and S. Höger. "Modular Bicyclophane-Based Molecular Platforms."
Chem. Eur. J. **2023**, *29*, e202302662. DOI: 10.1002/chem.202302662.

During the time period of this thesis, the following further studies were carried out.

6. I. Georg, M. Bursch, J. B. Stückrath, E. Alig, M. Bolte, H.-W. Lerner, S. Grimme, and M. Wagner. "Building up Strain in One Step: Synthesis of an Edge-Fused Double Silacyclobutene from an Extensively Trichlorosilylated Butadiene Dianion."
Angew. Chem. Int. Ed. **2020**, *59*, 16181–16187. DOI: 10.1002/anie.202006463.

7. Z. Zhang, J. B. Stückrath, S. Grimme, and A. Gansäuer. "Titanocene-Catalyzed [2+2] Cycloaddition of Bisenones and Comparison with Photoredox Catalysis and Established Methods."
Angew. Chem. Int. Ed. **2021**, *60*, 14339–14344. DOI: 10.1002/anie.202102739.

8. T. Gasevic, J. B. Stückrath, S. Grimme, and M. Bursch. "Optimization of the r^{2}SCAN-3c Composite Electronic-Structure Method for Use with Slater-Type Orbital Basis Sets." *J. Phys. Chem. A* **2022**, *126*, 3826–3838. DOI: 10.1021/acs.jpca.2c02951.

9. E. Bartalucci, A. A. Malär, A. Mehnert, J. B. Kleine Büning, L. Günzel, M. Icker, M. Börner, C. Wiebeler, B. H. Meier, S. Grimme, B. Kersting, and T. Wiegand. "Probing a Hydrogen-π Interaction Involving a Trapped Water Molecule in the Solid State." *Angew. Chem. Int. Ed.* **2023**, *62*, e202217725. DOI: 10.1002/anie.202217725.

10. Z. Zhang, D. Slak, T. Krebs, M. Leuschner, N. Schmickler, E. Kuchuk, J. Schmidt, L. I. Domenianni, J. B. Kleine Büning, S. Grimme, P. Vöhringer, and A. Gansäuer. "A Chiral Titanocene Complex as Regiodivergent Photoredox Catalyst: Synthetic Scope and Mechanism of Catalyst Generation." *J. Am. Chem. Soc.* **2023**, *145*, 26667–26677. DOI: 10.1021/jacs.3c08029.

11. T. Gasevic, J. B. Kleine Büning, S. Grimme, and M. Bursch. "Benchmark Study on the Calculation of ^{207}Pb NMR Chemical Shifts." *Inorg. Chem.*, *Manuscript in revision*.

Abstract

Nuclear magnetic resonance (NMR) spectroscopy is among the most important analytical methods for the determination of the chemical structure of matter. Its versatile applicability leads to the necessity of reliable computational techniques for simulating NMR spectra and parameters. Depending on the field of application, these should be general and yield sufficiently accurate results with limited computational demands. Therefore, this thesis contains five articles which deal with the assessment of existing methods for NMR chemical shift calculation, the development of new approaches to achieve a higher accuracy, and the application of these new methods to provide answers to actual open research questions.

The calculation of accurate NMR chemical shifts can be done with density functional theory (DFT) in a relatively efficient way. It is, however, not straightforward to estimate the performance of each density functional approximation (DFA) without rigorous tests, of which there are too few in the literature, especially for nuclei other than ^{1}H and ^{13}C. Therefore, this work focuses on the comprehensive evaluation of DFT methods regarding their suitability for the calculation of ^{29}Si and ^{119}Sn NMR chemical shifts. In the scope of these studies, the SiS146 and the SnS51 benchmark sets are compiled and used for the assessment of various DFAs. Correspondingly, the use of computationally more demanding hybrid functionals does not generally lead to improved results compared to more efficient (meta-)generalized gradient approximation (GGA) DFAs. However, a relativistic treatment (ideally including spin-orbit effects) reveals being indispensable if heavy atoms are close to a light NMR nucleus (^{29}Si) or if the nucleus is heavy itself (^{119}Sn).

Afterward, two machine learning-based correction methods (Δ-ML) are presented for the improved description of ^{1}H and ^{13}C NMR chemical shifts. The NMR prediction performance of DFT can generally be improved with the Δ_{corr}-ML method, which adds a correction term based on highly accurate coupled cluster reference data. Besides that, the Δ_{SO}-ML method is built on reference data obtained from spin-orbit relativistic DFT calculations of organic molecules containing heavy atoms and represents an efficient alternative to relativistic calculations. Both methods are evaluated using the underlying training and test data sets as well as benchmark sets beyond that data and show to consistently outperform the routinely applied linear regression approach. Applying both methods in a multilevel workflow can significantly reduce the deviation to experimental data of ^{13}C NMR chemical shifts. Finally, a joint study with experimentalists on platform molecules for surface decoration is presented, that showcases the usage of the Δ_{corr}-ML method to verify computed ^{1}H NMR chemical shifts, which leads to important new insights into the involved chemical reactions.

Contents

1 Introduction 1

2 Theoretical Background 5
- 2.1 Electronic Structure Theory . . . 5
  - 2.1.1 General Concept . . . 5
  - 2.1.2 Hartree-Fock Theory . . . 7
  - 2.1.3 Electron Correlation Methods . . . 10
  - 2.1.4 Density Functional Theory . . . 12
- 2.2 Further Considerations . . . 13
  - 2.2.1 Solvation and Free Energy . . . 14
  - 2.2.2 Ensemble and Dynamic Effects . . . 15
  - 2.2.3 Relativistic Effects . . . 17
- 2.3 Nuclear Magnetic Resonance . . . 18
  - 2.3.1 Experimental Aspects . . . 18
  - 2.3.2 Computational Aspects . . . 20
- 2.4 Machine Learning . . . 24

3 Comprehensive Benchmark Study on the Calculation of ^{29}Si NMR Chemical Shifts 27

4 Benchmark Study on the Calculation of ^{119}Sn NMR Chemical Shifts 31

5 Computation of CCSD(T)-Quality NMR Chemical Shifts via Δ-Machine Learning from DFT 35

6 Machine Learning-Based Correction for Spin-Orbit Coupling Effects in NMR Chemical Shift Calculations 39

7 Modular Bicyclophane-Based Molecular Platforms 43

8 Conclusion and Outlook 47

A Appendix: Comprehensive Benchmark Study on the Calculation of ^{29}Si NMR Chemical Shifts **51**
A.1 Introduction . . . 53
A.2 Computational Details . . . 54
A.3 Results and Discussion . . . 55
A.3.1 General Considerations . . . 55
A.3.2 Benchmark Set . . . 56
A.3.3 Benchmark of Density Functional Approximations for the Calculation of ^{29}Si NMR Chemical Shifts . . . 57
A.3.4 The Heavy Atom Subset *SiS-H* and Relativistic Effects . . . 62
A.3.5 Structure Dependence of ^{29}Si NMR Calculated Chemical Shifts . . . 64
A.4 Conclusion . . . 66

B Appendix: Benchmark Study on the Calculation of ^{119}Sn NMR Chemical Shifts **71**
B.1 Introduction . . . 73
B.2 Computational Details . . . 74
B.3 Results and Discussion . . . 76
B.3.1 General Considerations . . . 76
B.3.2 Benchmark Set . . . 78
B.3.3 Study of Density Functional Approximations and Relativistic Approaches for the Calculation of ^{119}Sn NMR Chemical Shifts . . . 78
B.3.4 Linear Scaling . . . 83
B.3.5 Conformational Analysis . . . 86
B.3.6 Structure Dependence . . . 88
B.4 Conclusion . . . 89

C Appendix: Computation of CCSD(T)-Quality NMR Chemical Shifts via Δ-Machine Learning from DFT **93**
C.1 Introduction . . . 95
C.2 Methodology . . . 96
C.2.1 ML Data Set . . . 96
C.2.2 Reference Level of Theory . . . 97
C.2.3 Neural Network Architecture and Input Feature Vector . . . 99
C.2.4 Computational Details . . . 101
C.3 Results and Discussion . . . 101
C.3.1 Correction of ^{1}H NMR Chemical Shifts . . . 101
C.3.2 Correction of ^{13}C NMR Chemical Shifts . . . 103
C.3.3 Methodological Influences . . . 105
C.3.4 Performance for External Systems . . . 108
C.3.5 Computational Demands . . . 114
C.4 Conclusion . . . 115

D Appendix: Machine Learning-Based Correction for Spin-Orbit Coupling Effects in NMR Chemical Shift Calculations **117**
D.1 Introduction . . . 119
D.2 Methods . . . 120
D.2.1 Machine Learning Data Set . . . 120
D.2.2 Reference Level of Theory . . . 121
D.2.3 Neural Network Architecture and Input Feature Vector . . . 122
D.2.4 Computational Details . . . 124
D.3 Results . . . 125
D.3.1 Prediction of $\Delta_{SO}\delta$ for ^{13}C NMR . . . 125
D.3.2 Prediction of $\Delta_{SO}\delta$ for ^{1}H NMR . . . 128
D.3.3 Generalizability of the Model . . . 130
D.3.4 Performance for External Test Systems . . . 134
D.4 Conclusion . . . 139

E Appendix: Modular Bicyclophane-Based Molecular Platforms **143**
E.1 Introduction . . . 145
E.2 Results and Discussion . . . 145
E.3 Conclusions . . . 153

Bibliography **155**

List of Figures **197**

List of Tables **199**

Acknowledgements **201**

CHAPTER 1

Introduction

From the very beginning of chemistry, people have been eager to unravel the microscopic structure of matter. Even a century ago, it needed years of research to clearly identify seemingly simple molecules, e.g., in the famous search for the benzene structure, which was first unequivocally conformed in 1929 by X-ray crystallography.[2] Knowledge has been growing rapidly since then but, until today, the elucidation of molecular and solid-state structures remains a non-trivial task. A detailed understanding of the fundamental composition of a material is the key to studying its properties and explore new, similar substances. Among other analytical tools that allow the investigation of the structural characteristics of a compound, nuclear magnetic resonance (NMR) spectroscopy is an especially versatile and powerful one. Being applicable for solids, liquids, gases, and solvated substances, NMR spectroscopy is used throughout a wide range of research areas in chemistry and related fields.[3–7] Thus, it is relevant to various active scientific subjects, such as chemical synthesis development and optimization,[8–10] characterization of new substances,[11–14] understanding of (bio)chemical processes,[15–18] and analytical detection of chemicals in diverse environments.[19–21]

In NMR spectroscopy, a strong external magnetic field is applied, leading to an alignment of the nuclear spins of an analyte's NMR-active atomic nuclei (^{1}H, ^{13}C, ^{15}N, ^{19}F, ^{29}Si, ^{31}P, etc.). Low energy electromagnetic radiation can be used to study the nuclei's response to an energetic excitation, which provides information about their local magnetic environment determined by the surrounding electrons. The recorded spectra can, therefore, be used to generate atom-resolved information about the electronic structure of a substance, making it possible to identify molecular characteristics, such as the presence of aromatic systems or aliphatic chains and the interaction of atoms via covalent bonds or mere spacial proximity. This information is provided by the unique chemical shift δ of a nucleus that determines the spectral position of the corresponding signal. Furthermore, the spin-spin coupling pattern of the signal can provide further insights into structural features within a molecule.

High-performance computing resources are nowadays widely available and usually employed in many chemical studies. With modern quantum mechanical and classical

simulation methods, it is possible to address virtually any physical or chemical process that is relevant to research in a chemical laboratory. For example, very accurate electronic energies can be calculated with coupled cluster (CC) methods for small systems. For larger systems, density functional theory (DFT)[22] can be applied and also allows for subsequent frequency or property calculations, enabling direct comparisons to experimental findings. A detailed introduction to these and other electronic structures methods is provided in Section 2.1. If the investigated systems are large (several hundreds of atoms) or if a large number of calculations is necessary, semiempirical quantum mechanical (SQM) or classical force field (FF) methods, that usually have a lower accuracy, allow a rigorous assessment.

The importance of NMR and the availability of various computational methods leads to the overarching subject of this thesis as logical consequence: The theoretical study of NMR spectroscopy, or, more precisely, of the NMR chemical shift. The connection of a wave function and NMR parameters has been known since the 1950s[23] and a rational analysis of Lodewyk et al. in a work from 2012[24] listed the five main sources of error that should be regarded in an NMR chemical shift calculation. These are electron correlation, conformational flexibility, solvation, heavy atom effects, and rotational-vibrational effects (see Figure 1.1; a more detailed explanation is given in Section 2.3.2).

While the rotational and vibrational contributions are usually expected to cancel due to the relative nature of the chemical shift, the conformational and solvation effects can be taken care of by considering not only the optimal geometry but a full ensemble of structures (discussed further in Sections 2.2.1 and 2.2.2). In this way, a more realistic picture can be created in which differently populated conformations coexist. For instance, the conformer-rotamer ensemble sampling tool (CREST)[25] is capable of using the GFNn-xTB[26,27] ($n = 1, 2$)

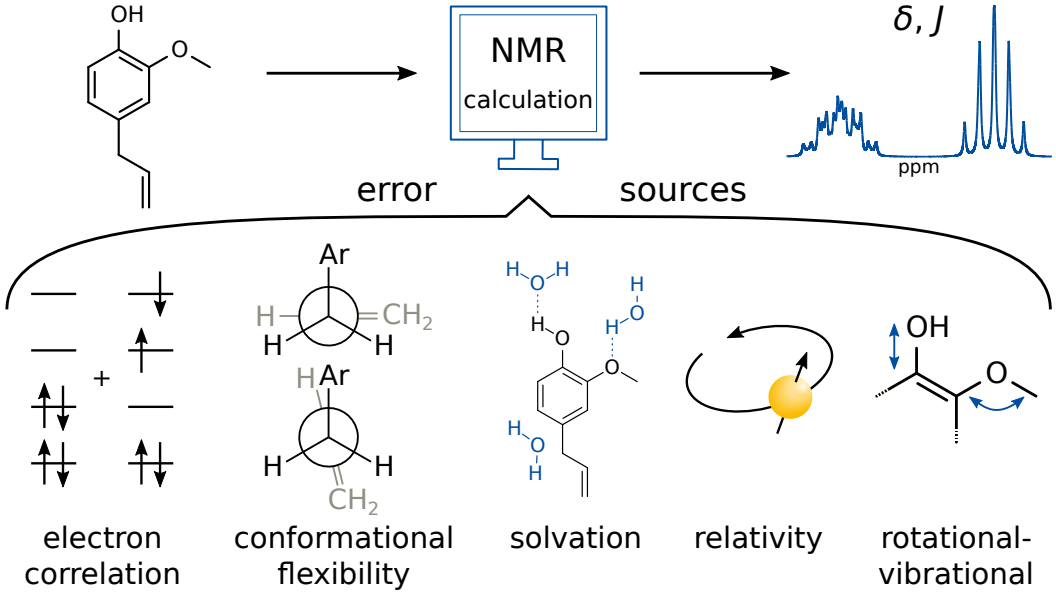


Figure 1.1: Schematic representation of the five most important aspects in NMR calculations.

SQM methods or the GFN-FF[28] force field to generate potentially relevant conformers while applying an implicit solvation model. The resulting ensemble can subsequently be refined using DFT via the command-line energetic sorting tool (CENSO)[29] to yield an ensemble at DFT quality. Thus, it is possible to obtain molecular properties as Boltzmann-weighted average, e.g., NMR spectra[30] or optical rotation.[31]

Regarding electron correlation and the overall usability of an electronic structure method to calculate NMR chemical shifts, DFT has often proven a robust and well-balanced compromise between computational demands and accuracy. For example, studies have been performed on the comparison of different density functional approximations (DFAs) regarding their performance for the calculation of chemical shifts from ^{1}H,[32–34] ^{13}C,[33,35] ^{19}F,[36] ^{31}P,[36–38] or several further main group[39–41] and transition metal[42] nuclei. Analogously, there are investigations for higher levels of theory that precisely compare various aspects of coupled cluster theory and rigorously approach the possible accuracy limit.[43–46]

NMR nuclei other than the most common ones, ^{1}H and ^{13}C, are also of great importance, as they often complement the experimental findings for certain compounds. When it comes to heavy nuclei, Paul Dirac would probably qualify his statement about the importance of relativity for chemistry (see opening quote[1]) today, being aware of the relevance of NMR spectroscopy. Special relativity is indeed necessary to tackle the last of the mentioned error sources in NMR calculations if heavy elements are present.[47,48] Due to the high nuclear charge of heavy atoms, inner-shell electrons reach high velocities that necessitate such a treatment in order to obtain an accurate picture of the electronic structure in the vicinity of the nucleus. It is precisely this region of the electronic structure that contributes most to the final chemical shift value, so a consideration is rational for heavy nucleus NMR, such as for the ^{183}W,[49] ^{195}Pt,[50] ^{199}Hg,[51,52] or ^{207}Pb[49,53] nuclei. But there is even more to it: The relativistic spin-orbit effect of a heavy nucleus on the local magnetic field can propagate through the electron density in chemical bonds and thus affect the chemical shift of other (light) nuclei, which is called heavy atom effect on the light atom (HALA).[48,54] A relativistic treatment can, therefore, be relevant for the calculation of the ^{1}H, ^{13}C and ^{29}Si NMR chemical shifts, too.

In order to shed more light on the computational simulation of NMR chemical shifts in solvated systems using DFT and relativistic approaches, the presented thesis includes two comprehensive studies, focusing on the ^{29}Si (Chapter 3) and the ^{119}Sn nucleus (Chapter 4) in organic compounds. Various density functional approaches will be evaluated and other simulation aspects, such as the influence of the level of theory used for geometry optimization and the effect from conformational averaging, will be discussed. DFT calculations will be augmented with different relativistic approaches to study the relativistic effects in the ^{119}Sn NMR chemical shift and the HALA effect to ^{29}Si in organosilicon compounds.

With the knowledge about the strengths and weaknesses of DFT for calculating NMR chemical shifts, procedures can be developed that systematically aim at eliminating a known deficiency of the original method. The thorough inclusion of electron correlation effects is possible, yet associated with a very high computational effort that limits it to the investigation of small systems. Similarly, the explicit use of approximations to a fully relativistic Hamiltonian (as described in Section 2.2.3) involves additional computational resources as

well and is not as straightforward for a non-expert user as it requires advanced knowledge and – at least currently – particular software packages.

The second aim of this thesis is, therefore, the development of correction methods for some of the aspects mentioned above. For this purpose, the versatile field of machine learning (ML) will serve as toolbox for constructing such techniques. The concept of artificially mimicking biological neurons, which dates back to 1943,[55] has been used for chemical applications since the late 1980s.[56,57] However, this interest has intensely grown in the past decade[58,59] with 25 recorded publications in 2013 and 790 in 2023.[60] Nowadays, there is hardly any area of computational chemistry that has not yet been explored with the use of an ML approach. To name only a few examples, the applications include the general prediction of molecular properties[61] or (conformational) energies,[62,63] accurate structure determination,[64,65] spectra generation,[66,67] the generative construction of new structures,[68,69] and the full prediction of the electronic structure,[70–72] e.g., with neural network potentials.[73,74]

Even when focusing on the computational investigation of NMR spectroscopy, a wide range of ML techniques have been developed based on various flavors. There are methods to directly predict an NMR chemical shift from merely a three-dimensional condensed-phase[75,76] or molecular structure,[77–81] or even to yield the full shielding tensor.[82] Other approaches use machine learning for structural assignment and interpretation of complex (multi-dimensional) spectra.[83–87] However, it is also possible to combine results from quantum chemical calculations with ML, e.g., with the Δ-ML approach, in which a correction (Δ) to a computed baseline value is predicted by the ML model. Thereby, the calculated NMR chemical shifts can be enhanced and corrected with respect to high-level computed[88,89] or experimental[90] reference data.

With the help of a rather simple ML approach described in Section 2.4, two correction methods (Δ-ML) for ^{1}H and ^{13}C NMR chemical shifts based on DFT will be presented in this thesis. The first one (Chapter 5), called Δ_{corr}-ML, corrects a DFT-based result toward coupled cluster accuracy, meaning it predicts the electron correlation contribution to the chemical shift. Then (Chapter 6), a second correction, Δ_{SO}-ML, is introduced to predict the contribution to the ^{1}H and ^{13}C NMR chemical shift originating from the inclusion of spin-orbit relativistic effects. This method proves useful for compounds involving heavy atoms. Finally, in Chapter 7, an application example for the calculation of ^{1}H NMR chemical shifts with DFT will be shown, in which a large chemical shift change during a ring closure reaction of a bicyclophane is investigated. The previously developed Δ_{corr}-ML method is applied to verify the calculated values and helps to identify the crucial computational aspects necessary for an accurate replication of the experimental findings.

In the following, a more detailed introduction to the physical concepts and computational methods will be given (Chapter 2), explaining the fundamentals of electronic structure theory, NMR spectroscopy and machine learning. It is followed by the main part of this thesis, which are the five already mentioned investigations related to the computational simulation of NMR parameters in Chapters 3 to 7 and appendices A to E. Finally, Chapter 8 will provide a summary of the key findings from these studies and place the work into a larger context for future research.

CHAPTER 2

Theoretical Background

This chapter gives an overview of the most important theoretical concepts that are the basis for the work in this thesis. It starts with the most fundamental principles of electronic structure methods in general (Section 2.1) and proceeds with approaches beyond the sole electronic energy (Section 2.2). Afterward, the fundamentals of nuclear magnetic resonance spectroscopy (Section 2.3) and machine learning (Section 2.4) are outlined.

2.1 Electronic Structure Theory

This section provides a brief overview of the concepts of electronic structure and approaches for its calculation. It is mainly based on the explanations given in the textbooks [91] and [92]. If not stated otherwise, all equations in this section are given in atomic units.

2.1.1 General Concept

At a fundamental level, the aim of electronic structure methods is to describe the distribution and movement of electrons in a system, say, a collection of atoms. The electronic structure of molecules can be represented by a wave function Ψ. A given Hamiltonian $\hat{\mathcal{H}}$ can be used to attribute a total energy E to the electronic structure via an eigenvalue problem, which is, for now, the non-relativistic time-independent Schrödinger equation.[93]

$$\hat{\mathcal{H}}\Psi = E\Psi \tag{2.1}$$

The physically exact form of this Hamiltonian consists of five components, which correspond to kinetic ($\hat{T}$) energy contributions of electrons (e) and nuclei (n) and the potential ($\hat{V}$)

energies associated with their pairwise interactions.

$$\hat{\mathcal{H}} = \underbrace{-\frac{1}{2}\sum_{i}^{N_{\mathrm{el}}}\nabla_i^2}_{\hat{\mathcal{T}}_{\mathrm{e}}} \underbrace{-\sum_{A}^{N_{\mathrm{at}}}\frac{\nabla_A^2}{2M_A}}_{\hat{\mathcal{T}}_{\mathrm{n}}} + \underbrace{\sum_{i}^{N_{\mathrm{el}}}\sum_{j>i}^{N_{\mathrm{el}}}\frac{1}{r_{ij}}}_{\hat{\mathcal{V}}_{\mathrm{ee}}} \underbrace{-\sum_{i}^{N_{\mathrm{el}}}\sum_{A}^{N_{\mathrm{at}}}\frac{Z_A}{r_{iA}}}_{\hat{\mathcal{V}}_{\mathrm{en}}} + \underbrace{\sum_{A}^{N_{\mathrm{at}}}\sum_{B>A}^{N_{\mathrm{at}}}\frac{Z_A Z_B}{r_{AB}}}_{\hat{\mathcal{V}}_{\mathrm{nn}}} \tag{2.2}$$

In this notation, N_{el} and N_{at} are the total numbers of electrons and atoms in the system while i, j represent electrons and A, B represent nuclei. M_A and Z_A are the mass and nuclear charge of nucleus A, respectively, and the distance r_{iA} between two objects i and A is given as their Euclidean norm from the position vector $\boldsymbol{r}$.

$$r_{iA} = |\boldsymbol{r}_i - \boldsymbol{r}_A| \tag{2.3}$$

Since even the smallest possible nucleus (the proton) is almost 2000 times heavier than an electron[94] and their dynamics happen on very different time scales, the nuclei "see" a somewhat averaged electron cloud that responds instantly to alternations in the nuclear positions. It is thus, in most situations, reasonable to separate nuclear and electronic movements and assume a fixed scaffold of nuclei, which leads to

$$\hat{\mathcal{T}}_{\mathrm{n}} = 0 \tag{2.4}$$

and

$$\hat{\mathcal{V}}_{\mathrm{nn}} = \text{const.} = E_{\mathrm{nuc}} \tag{2.5}$$

and is known as Born-Oppenheimer approximation.[95] This concept is fundamental and is applied throughout all parts of this thesis. It leaves the electronic Hamiltonian $\hat{\mathcal{H}}_{\mathrm{el}}$, which is object of all electronic structure methods, with three terms.

$$\hat{\mathcal{H}}_{\mathrm{el}} = \hat{\mathcal{T}}_{\mathrm{e}} + \hat{\mathcal{V}}_{\mathrm{ee}} + \hat{\mathcal{V}}_{\mathrm{en}} \tag{2.6}$$

With all the physical relations included in the operator $\hat{\mathcal{H}}$, the other rather complex object in equation (2.1) is the wave function Ψ that contains the actual description of the system of interest. In the ansatz, that is further explained in Section 2.1.2, the wave function comprises the electrons in orbitals ϕ. A spin orbital $\phi_i(\boldsymbol{x})$, regardless if its origin is atomic or molecular, is a four-dimensional one-particle wave function consisting of a spacial orbital $\psi_i(\boldsymbol{r})$ and a spin component $\sigma_i \in \alpha, \beta$.

$$\phi_i(\boldsymbol{x}) = \psi_i(\boldsymbol{r})\sigma_i \tag{2.7}$$

Naturally, atomic orbitals can be described using Slater-type (ST) functions as those are obtained from solving the Schödinger equation for the hydrogen atom.[96] However, it is a common approach to use Gaussian-type (GT) functions instead, as they are computationally more feasible. Exemplary functions are shown below, where $a(\boldsymbol{r})$ are spherical harmonics

determined by the azimuthal quantum number and α is a real scaling factor.

$$\psi^{\mathrm{ST}}(\boldsymbol{r}) = a(\mathbf{r}) \cdot e^{-\alpha|\boldsymbol{r}|} \tag{2.8}$$

$$\psi^{\mathrm{GT}}(\boldsymbol{r}) = a(\mathbf{r}) \cdot e^{-\alpha\boldsymbol{r}^2} \tag{2.9}$$

When Gaussian-type orbitals (GTOs) are used, they are often contracted in order to better mimic the shape of an STO, i.e., the actual contracted orbital is obtained as an arbitrary linear combination of a number of primitive functions.

The atomic orbitals (AOs), now termed χ_i, are used as basic ingredients for an electronic structure calculation. The molecular orbitals (MOs) ψ_i, which are obtained from such a calculation of a molecular system, are linear combinations of all atomic orbitals (LCAO ansatz) with the coefficients $C_{\mu i}$ and N_{AO} being the total number of AOs.

$$\psi_i = \sum_{\mu}^{N_{\mathrm{AO}}} C_{\mu i}\chi_\mu \tag{2.10}$$

During a typical electronic structure calculation, the coefficients $C_{\mu i}$ are optimized such that the resulting wave function Ψ minimizes the energy E in equation (2.1), yielding the final electronic structure described by the final MOs.

2.1.2 Hartree-Fock Theory

One of the most basic electronic structure methods and a starting point for many others is Hartree-Fock (HF) theory. It uses the exact (non-relativistic and time-independent) Hamiltonian from equation (2.2) and computes the energy E variationally. The wave function Ψ is approximated by a Slater determinant Φ^{SD}, including all N spin orbitals ϕ_i occupied by N electrons $\boldsymbol{x}_i$.

$$\Psi^{\mathrm{HF}} \approx \Phi^{\mathrm{SD}} \tag{2.11}$$

$$\Phi^{\mathrm{SD}}(\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_N) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(\boldsymbol{x}_1) & \phi_2(\boldsymbol{x}_1) & \cdots & \phi_N(\boldsymbol{x}_1) \\ \phi_1(\boldsymbol{x}_2) & \phi_2(\boldsymbol{x}_2) & \cdots & \phi_N(\boldsymbol{x}_2) \\ \vdots & \vdots & \ddots & \vdots \\ \phi_1(\boldsymbol{x}_N) & \phi_2(\boldsymbol{x}_N) & \cdots & \phi_N(\boldsymbol{x}_N) \end{vmatrix} \tag{2.12}$$

In that way, the requirement of the wave function to be antisymmetric is fulfilled, which holds for all fermions. Consequently, Ψ will vanish if all quantum numbers of two electrons are identical, which is in accordance with the Pauli exclusion principle. In HF theory, the energy is given as the expectation value of the Hamiltonian.

$$E^{\mathrm{HF}} = \left\langle \Psi^{\mathrm{HF}} \middle| \hat{\mathcal{H}} \middle| \Psi^{\mathrm{HF}} \right\rangle \tag{2.13}$$

Inserting the electronic Hamiltonian $\hat{\mathcal{H}}_{\text{el}}$ from equation (2.6) leads to

$$E^{\text{HF}} = \sum_{i}^{N} \left\langle \phi_i \middle| \hat{h} \middle| \phi_i \right\rangle + \frac{1}{2} \sum_{ij}^{N} \left(\left\langle \phi_i \middle| \hat{J}_j \middle| \phi_i \right\rangle - \left\langle \phi_i \middle| \hat{K}_j \middle| \phi_i \right\rangle \right) + E_{\text{nuc}} \tag{2.14}$$

with the one-electron operator $\hat{h}$ and the two-electron Coulomb and exchange operators $\hat{J}$ and $\hat{K}$, respectively.

$$\hat{h} \left| \phi_i \right\rangle = -\frac{1}{2} \nabla_i^2 - \sum_{A}^{N_{\text{at}}} \frac{Z_A}{r_{iA}} \tag{2.15}$$

$$\hat{J}_j \left| \phi_i \right\rangle = \left(\int \phi_j^* r_{ij}^{-1} \phi_j \, \mathrm{d}\boldsymbol{x}_j \right) \left| \phi_i \right\rangle = \left\langle \phi_j \middle| r_{ij}^{-1} \middle| \phi_j \right\rangle \left| \phi_i \right\rangle \tag{2.16}$$

$$\hat{K}_j \left| \phi_i \right\rangle = \left(\int \phi_j^* r_{ij}^{-1} \phi_i \, \mathrm{d}\boldsymbol{x}_j \right) \left| \phi_j \right\rangle = \left\langle \phi_j \middle| r_{ij}^{-1} \middle| \phi_i \right\rangle \left| \phi_j \right\rangle \tag{2.17}$$

These operators trace back to the ingredients in equation (2.6). $\hat{\mathcal{T}}_{\text{e}}$ and $\hat{\mathcal{V}}_{\text{en}}$ are contained in $\hat{h}$ and $\hat{\mathcal{V}}_{\text{ee}}$ is divided into the classical Coulomb repulsion of two electrons and their exchange interaction, which is a purely quantum mechanical phenomenon and is associated with an energy contribution for the hypothetical exchange of two non-distinguishable particles. Now, all three operators above can be combined in the Fock operator $\hat{f}$.

$$\hat{f} \left| \phi_i \right\rangle = \hat{h} \left| \phi_i \right\rangle + \sum_{j}^{N} \hat{J}_j \left| \phi_i \right\rangle - \hat{K}_j \left| \phi_i \right\rangle \tag{2.18}$$

The eigenvalues of the Fock operator are the orbital energies ε of the corresponding orbital ϕ.

$$\hat{f} \left| \phi_i \right\rangle = \varepsilon_i \left| \phi_i \right\rangle \tag{2.19}$$

All operator expressions shown so far can be reformulated in a matrix form. Furthermore, for closed-shell cases (no unpaired electrons), it can be assumed that all spacial orbitals are doubly occupied and $\alpha = \beta$. Then, only half of the orbitals have to be treated explicitly and half of the exchange terms vanish due to spin orthogonality. This yields the Fock operator as

$$f_{ii} = \varepsilon_i = h_{ii} + \sum_{j}^{N} J_{ij} - K_{ij} = h_{ii} + \sum_{j}^{N/2} 2J_{ij} - K_{ij} \tag{2.20}$$

and the HF energy (2.14) can be expressed by using f_{ii}.

$$E^{\mathrm{HF}} = 2\sum_{i}^{N/2} h_{ii} + \sum_{ij}^{N/2} \Big(2J_{ij} - K_{ij}\Big) + E_{\mathrm{nuc}} \tag{2.21}$$

$$= \sum_{i}^{N/2} (h_{ii} + f_{ii}) + E_{\mathrm{nuc}} \tag{2.22}$$

It is now feasible to insert the basis set expansion (2.10) with N_{AO} atomic orbitals into equation (2.19) and define the overlap matrix $\boldsymbol{S}$, the core-Hamiltonian $\boldsymbol{H}^{\mathrm{core}}$, and the Fock matrix $\boldsymbol{F}$ as $N_{\mathrm{AO}} \times N_{\mathrm{AO}}$ matrices.

$$\hat{f} \sum_{\mu} C_{\mu i} \chi_{\mu} = \varepsilon_i \sum_{\mu} C_{\mu i} \chi_{\mu} \tag{2.23}$$

$$S_{\mu\nu} = \langle \chi_{\mu} | \chi_{\nu} \rangle \tag{2.24}$$

$$H^{\mathrm{core}}_{\mu\nu} = \langle \chi_{\mu} | \hat{h} | \chi_{\nu} \rangle \tag{2.25}$$

$$F_{\mu\nu} = \langle \chi_{\mu} | \hat{f} | \chi_{\nu} \rangle \tag{2.26}$$

These equations lead to an eigenvalue problem that connects the Fock matrix with the orbital coefficients and is known as Roothaan equations (2.27).

$$\boldsymbol{FC} = \boldsymbol{SC\varepsilon} \tag{2.27}$$

Furthermore, the HF energy can be expressed via the definition of a density matrix $P_{\mu\nu}$.

$$E^{\mathrm{HF}} = \frac{1}{2} \sum_{\mu\nu} P_{\nu\mu} \Big(H^{\mathrm{core}}_{\mu\nu} + F_{\mu\nu} \Big) + E_{\mathrm{nuc}} \tag{2.28}$$

$$P_{\mu\nu} = 2\sum_{i}^{N/2} C_{\mu i} C^{*}_{\nu i} \tag{2.29}$$

With this equation at hand, the HF energy can directly be calculated from any linear combination of the atomic orbitals using their coefficients $C_{\mu i}$.

During an actual HF calculation, the solution of the Schrödinger equation (2.1) is approximated by using the variational principle to find the Slater determinant giving the lowest possible energy. This procedure is usually conducted in a self-consistent way, starting with an initial guess for $P_{\mu\nu}$ that corresponds to a starting wave function. The guess can then be used to calculate the Fock matrix via $H^{\mathrm{core}}_{\mu\nu}$ and the two-electron integrals.

$$F_{\mu\nu} = H^{\mathrm{core}}_{\mu\nu} + \sum_{\lambda\sigma} P_{\lambda\sigma} \left(\langle \phi_{\mu}\phi_{\sigma} | r_{12}^{-1} | \phi_{\nu}\phi_{\lambda} \rangle - \frac{1}{2} \langle \phi_{\mu}\phi_{\sigma} | r_{12}^{-1} | \phi_{\lambda}\phi_{\nu} \rangle \right) \tag{2.30}$$

After an intermediate transformation step, the resulting Fock matrix is subsequently diago-

nalized to solve the Roothaan equations (2.27), which yields new orbital coefficients and a new density matrix. This procedure is repeated until a convergence of E and $P_{\mu\nu}$ is reached and the final self-consistent field (SCF) energy solution is obtained. Due to the evaluation of the two-electron integrals, the HF method has a formal scaling of $\mathcal{O}(N^4)$ with the system size N.

2.1.3 Electron Correlation Methods

The major drawback of HF theory is its insufficient treatment of electron correlation effects. In fact, a single-determinant ansatz is not capable of including interactions of two electrons with opposite spin. By definition, the correlation energy E_{corr} of a system is defined as the difference between the energies obtained from a method that considers electron correlation and HF.

$$E_{\text{corr}} = E^{\text{correlated}} - E^{\text{HF}} \tag{2.31}$$

The most straightforward way to overcome this issue is to approximate the wave function with more than one determinant, more precisely, with a linear combination of Slater determinants. This approach is called configuration interaction (CI).

$$|\Psi^{\text{CI}}\rangle = |\Phi_0\rangle + \sum_{ia} c_i^a \, |\Phi_i^a\rangle + \sum_{ijab} c_{ij}^{ab} \, |\Phi_{ij}^{ab}\rangle + \sum_{ijkabc} c_{ijk}^{abc} \, |\Phi_{ijk}^{abc}\rangle + \dots \tag{2.32}$$

In this notation, Φ_0 is a reference wave function – most commonly the HF wave function – and $\Phi_{ij\dots}^{ab\cdots}$ indicates that, with respect to Φ_0, an electron has been moved from the occupied orbital i to the virtual orbital a, a second electron from j to b, and so on. The contributions are weighted by their corresponding coefficients $c_{ij\dots}^{ab\cdots}$. When the series is complete, i.e., when all possible excited determinants are included, the method is called full CI (FCI) and Ψ is an exact solution of equation (2.1). In truncated CI, the series (2.32) is terminated after a desired term, e.g., after Φ_{ij}^{ab}, leaving only singly (S) and doubly (D) excited determinants (CISD). The CI approach has two major drawbacks: Firstly, significant computational resources are demanded as CISD formally scales as $\mathcal{O}(N^6)$ and CISDT as $\mathcal{O}(N^8)$. Secondly, truncated CI methods are not size-consistent, meaning that the energy of a system with two non-interacting particles A and B (because of a sufficiently large distance) does not equal the sum of the separate energies. Size-consistency only holds for FCI.

$$E^{\text{CI}}(A+B) \neq E^{\text{CI}}(A) + E^{\text{CI}}(B) \tag{2.33}$$

Mainly because of this size-inconsistency, the more common choice of a correlated method is coupled cluster (CC) theory. In contrast to CI, the CC wave function is expanded in an

exponential approach via a Taylor series.

$$|\Psi^{\mathrm{CC}}\rangle = e^{\hat{T}} |\Phi_0\rangle \tag{2.34}$$

$$e^{\hat{T}} = 1 + \hat{T} + \frac{1}{2}\hat{T}^2 + \frac{1}{6}\hat{T}^3 + \ldots + \frac{1}{N!}\hat{T}^N \tag{2.35}$$

Furthermore, the cluster operator $\hat{T}$ is the sum of operators of certain degrees of excitation (singly: $\hat{T}_1$, doubly: $\hat{T}_2$, ...).

$$\hat{T} = \hat{T}_1 + \hat{T}_2 + \hat{T}_3 + \ldots + \hat{T}_N \tag{2.36}$$

When the complete $\hat{T}$ is taken into account, the method is identical to FCI, otherwise the cluster operator is truncated after a desired degree of excitation. In contrast to truncated CI, the formulation of equation (2.34) ensures that even for truncated CC, excitations up to order N are implicitly included, making the method size-consistent and thus superior to truncated CI. The scaling behavior remains unchanged (e.g., CCSD scales as $\mathcal{O}(N^6)$). An especially noteworthy coupled cluster method is CCSD(T),[97] which uses a CCSD approach and an additional perturbative inclusion of the triples excitations. It has a formal scaling of $\mathcal{O}(N^7)$ and is still applicable to medium-sized systems with an extraordinary accuracy as long as static correlation is expected to play a minor role. It is, therefore, often referred to as the "gold standard" of quantum chemistry.[98]

Perturbation theory can also directly be applied on top of a non-correlated method, leading to the last group of correlated methods that shall be introduced in this scope: many body perturbation theory (MBPT). It treats the correlation energy as a (small) perturbation $\mathcal{V}$ to the reference Hamiltonian $\hat{\mathcal{H}}_0$, for which a converged result of Ψ_0 is known and expected to be somewhat close to the exact solution.

$$\hat{\mathcal{H}} = \hat{\mathcal{H}}_0 + \mathcal{V} \tag{2.37}$$

In the Møller-Plesset (MP) variant of MBPT, $\hat{\mathcal{H}}_0$ is used from a HF calculation. The total energy is then expanded as a sum of energies from different orders of perturbation. The first energy expression that contains an electron correlation contribution is obtained at the second order (MP2).

$$E^{(0)} + E^{(1)} = E^{\mathrm{HF}} \tag{2.38}$$

$$E^{(0)} + E^{(1)} + E^{(2)} = E^{\mathrm{MP2}}_{\mathrm{tot}} \tag{2.39}$$

The MP2 correlation energy can be calculated non-iteratively by including double excitations.

$$E^{\mathrm{MP2}}_{\mathrm{corr}} = \frac{1}{4} \sum_{ijab} \frac{\left| \langle \phi_i \phi_j | \phi_a \phi_b \rangle - \langle \phi_i \phi_j | \phi_b \phi_a \rangle \right|^2}{\varepsilon_i + \varepsilon_j - \varepsilon_a - \varepsilon_b} \tag{2.40}$$

While a truncation at an arbitrary order of perturbation is possible (MP3, MP4, ...) and all MP methods are size-consistent, the MP2 method is most commonly used, as it has a rather low scaling behavior of $\mathcal{O}(N^5)$. However, due to the starting requirement of HF being an already good approximation to an exact method, MP2 fails if this is not the case, as in homolytic bond dissociations. Nevertheless, it can be valuable if well-behaved systems are considered.

2.1.4 Density Functional Theory

A fundamentally different approach to obtain the electronic structure of a system is density functional theory (DFT). It goes back to the (first) Hohenberg-Kohn theorem[99] that proves that the electronic energy of a system is completely determined by its electron density $\rho(\boldsymbol{r})$. Thus, a unique energy functional $E[\rho]$ exists that only depends on the spacial coordinates of the density and the complexity of the problem is not increased with the system size as it is the case for the use of Slater determinants. However, the functional remains unknown, which goes along with serious drawbacks. Nowadays – and throughout this whole thesis – DFT actually refers to Kohn-Sham (KS-) DFT, an ansatz that uses both a wave function in the form of a Slater determinant and concepts from orbital-free DFT. The electron density can thus be formulated using orbitals ϕ_i with occupation numbers n_i.

$$\rho = \sum_i^{N_\mathrm{el}} n_i \left|\phi_i\right|^2 \tag{2.41}$$

In KS-DFT, the energy functional is divided into different contributions, similar to equation (2.2). The known parts are the kinetic energy T_S (calculated using a Slater determinant) and the Coulomb interaction between electrons and nuclei (E_{ne}) or other electrons (J).

$$T_\mathrm{S} = \sum_i^{N_\mathrm{el}} \left\langle \phi_i \middle| -\frac{1}{2}\nabla^2 \middle| \phi_i \right\rangle \tag{2.42}$$

$$E_{ne}[\rho] = -\sum_A^{N_\mathrm{at}} \int \frac{Z_A \rho(\boldsymbol{r})}{|\boldsymbol{r}_A - \boldsymbol{r}|}\,\mathrm{d}\boldsymbol{r} \tag{2.43}$$

$$J[\rho] = \frac{1}{2}\iint \frac{\rho(\boldsymbol{r})\rho(\boldsymbol{r}')}{|\boldsymbol{r} - \boldsymbol{r}'|}\,\mathrm{d}\boldsymbol{r}\mathrm{d}\boldsymbol{r}' \tag{2.44}$$

Compared to equation (2.6), used in HF, the exchange contribution is missing here. This is due to the fact that it is impossible to reformulate it using $\rho(\boldsymbol{r})$. Therefore, an unknown exchange functional E_X along with an also unknown functional accounting for electron correlation E_C are added, forming the so-called exchange-correlation (XC) functional.

$$E[\rho] = T_\mathrm{S}[\rho] + E_{ne}[\rho] + J[\rho] + E_\mathrm{XC}[\rho] \tag{2.45}$$

$$E_\mathrm{XC}[\rho] = E_\mathrm{X}[\rho] + E_\mathrm{C}[\rho] \tag{2.46}$$

In that way, the two-electron energy contributions from Coulomb and exchange, described by the $\hat{J}$ and $\hat{K}$ operators in equation (2.14) for HF, do not completely cancel if $i = j$, which is unphysical. The resulting deviation is known as self-interaction error (SIE).

The exchange-correlation functional is essentially the ingredient that makes the difference of various DFT methods. These density functional approximations (DFAs) are only to a limited extent systematically improvable, but they can be arranged in different rungs of the "Jacob's ladder" of DFT.[100] Thus, the simplest approximation and the lowest rung is covered by the local density approximation (LDA), which only uses density-dependent terms in E_{XC} and treats electrons as a uniform electron gas. This approach has some application to metal-like systems but a limited relevance for molecules. On the second rung, generalized gradient approximation (GGA) methods are situated. These additionally include terms that depend on the gradient of the density $\nabla\rho$, allowing it to adapt better to the heterogeneous nature of the electronic structure of molecules. Famous examples are PBE[101] or BP86.[102,103] When higher derivatives of ρ or the kinetic energy density are included, the third rung is reached and the methods are called meta-GGA. They can be superior to GGAs and examples are TPSS[104] or r^{2}SCAN.[105]

However, a more significant difference is found on rung four, which includes hybrid DFAs that substitute a certain amount of $E_X[\rho]$ by the exact exchange expression from HF, evaluated with KS orbitals. This approach can help to reduce the SIE and is recommended in cases that are prone to SIE, such as transition states.[22] Popular hybrid functionals are, e.g., B3LYP[106–108] or PBE0.[109] The fifth and last rung, with respect to accuracy, follows the concept of hybrid functionals and also $E_C[\rho]$ is partially substituted by explicitly including electron correlation, usually from MP2. These methods are called double hybrids and the B2PLYP[110] or PWPB95[111] functionals are to be mentioned here. In the context of this thesis, mainly (meta-)GGA and (meta-)hybrid DFAs are considered.

DFT is often a good compromise between accuracy and computational cost. (meta-)GGAs can have a formal scaling of $\mathcal{O}(N^3)$ and hybrid DFAs of $\mathcal{O}(N^4)$, like HF. Still, both types are in general more accurate than HF. Double hybrid functionals lie at the upper end of the accuracy scale but come with increased computational demands with a scaling of $\mathcal{O}(N^5)$.

2.2 Further Considerations

In Section 2.1, it was shown how the total energy of a (molecular) system can be approximated by using the Schrödinger equation (2.1). However, this energy is normally not accessed in an experimental setting as there, processes are usually also determined by temperature and entropic effects. Therefore, this section presents the most important concepts of free energy calculations and nuclear dynamic effects, as well as some approaches for including special relativity in electronic structure calculations. The following is mainly based on reference [92].

2.2.1 Solvation and Free Energy

In chemical laboratories, it is very common to carry out experiments (syntheses, measurements, etc.) in solution. Electronic structure methods, however, do not consider a solvent. In fact, to describe the experimental setting correctly, it is necessary to not only include the molecule of interest in a calculation, but to add at least as many solvent molecules as a saturation of the solvation effects is reached or expected. This approach is called explicit solvation. It is, however, extremely complex and impractical for large systems or for screening calculations of many molecules as an unmanageably large number of conformations of the solvent shell are possible and have to be considered. Nevertheless, it is also possible to include only a small number of solvent molecules explicitly to describe their most important (non-covalent) interactions with the solute in a so-called microsolvation approach.

A much simpler and computationally more accessible way of including solvation effects is an implicit treatment with continuum models. The bulk solvent is, in this case, replaced by an electrostatic potential depending on a dielectric constant ε and contains a cavity in which the solute molecule is placed. The electronic structure of the solute and its charge distribution induces a polarization of the continuum, which in turn influences the molecule's electronic structure. This electronic structure change is associated with a stabilization of the molecule and its solvation (free) energy. Thanks to this general procedure, an implicit solvation model can, in principle, be added easily to most calculations. Some of the most commonly used continuum solvation models – that are also applied in this thesis – are the (conductor-like) polarizable continuum model ((C)PCM),[112,113] the conductor-like screening model (COSMO)[114] with its variant for real solvents (COSMO-RS),[115,116] and the analytical linearized Poisson-Boltzmann (ALPB)[117,118] model.

Besides the surroundings, enthalpic contributions are also missing in an electronic structure calculation compared to a reaction flask. In practice, enthalpies H or, more commonly, free energies G are investigated and are experimentally accessible.

$$H = E + p\Delta V \tag{2.47}$$

$$G = H - T\Delta S \tag{2.48}$$

The other variables are the electronic energy E, the pressure p, the volume V, the temperature T and the entropy S. The microscopic level of a single molecule can be connected to the macroscopic scale via a partition function q, which is the sum over all states s and their energies ε_s (k_B is the Boltzmann constant).

$$q = \sum_s e^{-\varepsilon_s / k_\mathrm{B} T} \tag{2.49}$$

Furthermore, the total energy ε_tot is the sum of all contributions from translational, rotational, vibrational, and electronic degrees of freedom.

$$\varepsilon_\mathrm{tot} = \varepsilon_\mathrm{trans} + \varepsilon_\mathrm{rot} + \varepsilon_\mathrm{vib} + \varepsilon_\mathrm{el} \tag{2.50}$$

The relation applies analogously to H_{tot} and S_{tot}, which can be obtained from a vibrational frequency calculation using any electronic structure method. Thus, the total contribution to the (Gibbs) free energy G_{RRHO} in this rigid-rotor, harmonic oscillator (RRHO) approach[119] is, for a certain temperature, accessible via equation (2.48). In this thesis, a modified approach (mRRHO) is used due to its better robustness.[120]

The solvation process of a molecule, i.e., bringing it from the gas to the solution phase, is also associated with a free energy contribution δG_{solv}, that can be calculated using a solvation model as the difference caused by the use of the model. In total, the free energy obtained from an electronic structure calculation and its subsequent evaluations is the sum of the individual contributions.

$$G = E + G_{\text{RRHO}} + \delta G_{\text{solv}} \tag{2.51}$$

2.2.2 Ensemble and Dynamic Effects

Up to this point, a molecule was always considered a fixed arrangement of atoms, for which one can calculate the total electronic energy as well as free energy contributions. However, for every set of atoms, a total energy can be attributed to every possible structure, forming a $3N_{\text{at}}$-dimensional hypersurface (potential energy surface, PES). Minima on the PES are considered stable structures and first-order saddle points can be associated with transition state structures. These points can be reached with geometry optimizations. At a temperature $T > 0$ K, atoms in molecules can move to some extent and different conformations of the same molecule can be present at the same time. In a computational simulation, this can be described in different ways and the two most important ones for this thesis are a static ensemble approach and a molecular dynamics approach.

Static Conformer Ensemble

A conformer ensemble is a set of different conformations (minima on the PES) of a compound simultaneously present at a certain temperature. Their relative frequencies of occurrence, also called populations P_i, depends on their total free energies G_i relative to the one of the lowest free energy G_{min} and the temperature and can be described by the Boltzmann statistics (R is the universal gas constant). The corresponding Boltzmann weights w_i are obtained relative to all N_P populations.

$$\Delta G_i = G_i - G_{\text{min}} \tag{2.52}$$

$$P_i = e^{-\Delta G_i / RT} \tag{2.53}$$

$$w_i = \frac{P_i}{\sum_j^{N_P} P_j} \tag{2.54}$$

With this approach, most of the properties of a compound, including the total free energy and quantities from derivations of the energy (as presented in Section 2.3), can be Boltzmann

averaged, meaning that the total quantity Q is the weighted sum of the values from all conformers.

$$Q = \sum_{i}^{N_P} w_i Q_i \tag{2.55}$$

In that way, the conformational flexibility of a molecule is replicated by explicit treatment of every important unique structure. For flexible molecules, such as those with long aliphatic alkyl chains, many thermally accessible conformations exist even at room temperature and their consideration might be important to achieve an accurate result. When the molecule is very rigid, only few conformations are populated and it can suffice to explicitly simulate only one representative conformer, such as the one with the lowest free energy.

There are various computational tools that can be used to generate a conformer ensemble from a starting structure. In this thesis, a workflow combining the conformer-rotamer ensemble sampling tool (CREST)[25] and the command-line energetic sorting algorithm (CENSO)[29] for conformer ensembles was used throughout. In the first step (CREST), a metadynamics simulation (see next paragraph), using the GFN-FF[28] force field or the semiempirical quantum mechanical GFN2-xTB[27] method, is performed from a starting structure in order to sample as many minima on the PES as possible. For all candidates, free energies are calculated and the highest-lying conformers are sorted out. Second (CENSO), the ensemble is systematically refined by optimizing the structures at increasingly accurate levels of DFT. Hence, the final ensemble is condensed to the most important conformers.

Molecular Dynamics

A different way of including the diversity of structures given by the PES are molecular dynamics (MD) simulations. In a Born-Oppenheimer MD simulation (the most commonly used type and used herein), the nuclei are considered classical particles and their movements follow Newton's equation of motion. Accordingly, the molecule experiences nuclear displacements to a certain extent associated with a movement on the PES over a given time period (typically some 100 ps or a few ns). The conformational space of a molecule can thus be sampled and the simulation can reach other minima on the PES, while some lie behind energy barriers that are too high to cross. The result of an MD simulation depends on the global settings, which include the temperature. They can be used to replicate a behaviour as close to reality as possible, but it is also possible to artificially alter the PES by using bias potentials, e.g., to overcome high energy barriers, like in metadynamics simulations.[121]

Given the molecular motions are described properly and long enough, MD simulations can also be used to obtain an average of a desired quantity Q that is obtained from an electronic structure calculation. In that case, N_S snapshot structures are taken in equidistant time steps throughout the simulation and the quantity is calculated for each snapshot i.

$$Q = \frac{1}{N_S} \sum_{i}^{N_S} Q_i \tag{2.56}$$

In contrast to the Boltzmann average, this approach represents a time-average of molecular movements and includes non-equilibrium structures. It might, therefore, allow access to other structural features that would not be present in a conformer ensemble but usually requires much greater computational resources as numerous individual calculations are necessary.

2.2.3 Relativistic Effects

In Section 2.1, the whole electronic structure problem was only discussed for the time-independent and non-relativistic Hamiltonian because it is a reasonable approximation in many cases. Paul Dirac himself did not consider special relativity at all important for applications in chemistry (see the opening quote[1]). However, when heavy elements are present, relativistic effects can indeed have a significant influence on the structure and the chemical behaviour of substances. For example, the extraordinarily low melting point of mercury is nowadays explained by relativistic effects[122] and a similar behavior was recently found for gold,[123] for which the unusual color[124] and the high stability of Au(III) compounds[125] are also believed to originate from such effects. To be fair, Dirac was referring to "ordinary chemical reactions", presumably meaning standard organic chemistry with only light elements, and NMR spectroscopy was not known at that time.

In a heavy atom (the term "heavy" strongly depends on the context it is used in, but this work will mostly refer to elements in the fourth period or higher), electrons close to the nucleus, e.g., in the 1*s*-orbital, experience huge Coulomb attraction forces and can reach velocities that are considerably close to the speed of light. This is accompanied by a non-negligible increase of the electron mass. Subsequently, this leads to a contraction of the *s*-orbitals, while *d*- and *f*-orbitals expand and get more diffuse. Finally, this mutation in orbital shapes affects the whole electronic structure of a molecule. Especially for properties that originate from the region close to the atomic nucleus, it can be important to consider a relativistic treatment in the calculation, as will be shown in Section 2.3.2.

In order to incorporate special relativity in an electronic structure calculation, one can consider the time-independent Dirac equation (2.57) as an extension to the Schrödinger equation (2.1).

$$(c\boldsymbol{\alpha} \cdot \boldsymbol{p} + \boldsymbol{\beta} mc^2 + \boldsymbol{V})\Psi = E\Psi \tag{2.57}$$

Here, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ are 4×4 parameter matrices, m is the relativistic electron mass, c is the (constant) speed of light, $c\boldsymbol{\alpha}$ represents the relativistic velocity operator, $\boldsymbol{p} = -i\nabla$ is the momentum operator, and $\boldsymbol{V}$ is an (external) electrostatic potential. Since the Dirac equation is four-dimensional, the wave function has four components as well. Two components each can be combined to yield a *large* (l) and a *small* (s) component.

$$\Psi = \begin{pmatrix} \Psi_\mathrm{s} \\ \Psi_\mathrm{l} \end{pmatrix} \tag{2.58}$$

As the name suggests, the Ψ_s contributes only little to the total wave function. Hence, the

highly computationally demanding four-component problem can be reduced to an effective two-component one by neglecting the *small* component.

Various approaches exist that build on this principle and find an approximation to solving equation (2.57) in order to allow a relativistic treatment in a quantum mechanical calculation. The most common ones are the Douglas-Kroll-Hess (DKH)[126,127] approach and the zeroth-order regular approximation (ZORA)[128,129] to the Dirac equation. The exact transformation of the four-component Dirac equation to two-components (X2C)[130–132] even performs the transformation in an arbitrarily accurate fashion. Furthermore, it is, depending on the used approach, possible to include only specific relativistic contributions. For example, purely scalar-relativistic (SR) effects can be included in a calculation at only few additional computational costs, while including also spin-orbit (SO) effects drastically increases the computational requirements. In most cases in this thesis, the SR- and SO-ZORA methods were applied.

Heavy atoms naturally have a large number of electrons and hence basis functions in a calculation. However, only few of the atomic orbitals – the valence and other high-lying orbitals – contribute significantly to the chemically relevant (bonding) molecular orbitals. Therefore, when the electrons close to the nucleus are not expected to be of great importance, the inner atomic orbitals can be replaced by an effective core potential (ECP) to reduce the computational demands. As these potentials are usually fitted to reference data, a relativistic treatment can be included implicitly. There are ECPs, including an SR or SO treatment, that can straightforwardly be used in a relativistic calculation.

2.3 Nuclear Magnetic Resonance

Nuclear magnetic resonance (NMR) is a nuclear physical phenomenon, which is the underlying effect of a wide range of analytical measurement techniques, such as NMR spectroscopy or magnetic resonance imaging (MRI).[133] NMR spectroscopy can deliver detailed information on the chemical environment of nuclei in molecules or solids and is therefore used on a daily basis in chemical laboratories. The most important physical concepts in an experimental and computational setting are outlined in this section, mainly based on references [134] and [135].

2.3.1 Experimental Aspects

To be NMR-active, an atomic nucleus has to have a non-zero nuclear spin with a quantum number of $I \neq 0$. If such an atom is brought into an external homogeneous magnetic field $\vec{B}_0$, its nuclear spin is aligned parallel to the magnetic field lines. In an NMR experiment, an oscillating magnetic field is then applied perpendicular to the external field (also referred to as radio-frequency pulse). If the pulse frequency matches the resonant frequency of the nucleus (also called Larmor frequency f_L and characteristic for every NMR-active isotope), a deflection of the nuclear spin is caused.

After the pulse, the nucleus begins to relax back to its original alignment and undergoes Larmor precession, determined by its intrinsic gyromagnetic ratio γ. This causes an electromagnetic signal that is received by the detection coil and transformed into a spectrum. The resonant Larmor frequency is proportional to the external magnetic field strength B_0.

$$f_{\mathrm{L}} = \frac{\gamma}{2\pi} \cdot B_0 \tag{2.59}$$

The relaxation of a nucleus is influenced by other magnetic fields, such as those generated by the movement of the surrounding electrons. Hence, the relaxation time varies in different electronic environments of the nucleus, which usually corresponds to an also different chemical environment. In the end, each chemically non-equivalent atom in a molecule corresponds to a signal in the resulting NMR spectrum. Because of the relation in equation (2.59), the overall resolution and the sensitivity of the measurement increase with the magnetic field strength. Therefore, powerful electromagnets made out of a superconducting material cooled with liquid helium are commonly used and reach field strengths of up to $B_0 = 21$ T and more.[134]

In basic NMR experiments, only one desired nucleus at a time is chosen to be investigated. The most common ones are the ^{1}H and ^{13}C nuclei, since they appear in every organic compound. Furthermore, both have a nuclear spin quantum number of $I = 1/2$, which leads to clear signals in the spectrum. The abundance of an isotope in the naturally occurring isotope mixture of an element is an important consideration, too. While the natural abundance of ^{1}H is 99.985%, ^{13}C only occurs with 1.10% and thus has an approximately 100-fold lower sensitivity than ^{1}H NMR.[134] In ^{15}N or ^{17}O NMR spectroscopy with natural abundances of 0.366% and 0.038%, respectively, samples can be enriched with the respective isotope before the measurement. Further examples of commonly investigated NMR nuclei are ^{19}F, ^{29}Si, ^{31}P, ^{35}Cl, ^{119}Sn, and ^{207}Pb. While NMR measurement techniques exist for virtually all states of matter, this thesis only focuses on NMR spectroscopy in solution. Since most organic solvents themselves contain ^{1}H nuclei that produce NMR signals, they are usually enriched with ^{2}H (or D, deuterium) atoms, which are also NMR active but with a different resonant frequency, so they do not appear in the ^{1}H NMR spectrum.

A signal in a typical NMR spectrum conveys two important measures (cf. Figure 2.1). Firstly, the chemical shift δ is related to the position of the signal in the spectrum and given as dimensionless shift to a reference shielding value in ppm units. It is a measure for the degree of shielding a nucleus experiences caused by its local magnetic environment. It is a purely relative quantity that always has to be referenced to some value. For ^{1}H and ^{13}C NMR, the chemical shifts of tetramethylsilane ($Si(CH_3)_4$) are usually used for that purpose and set to $\delta = 0$ ppm. The second measure is the spin-spin coupling constant ^{n}J, that appears for every pair of NMR-active nuclei connected via n bonds. In ^{1}H NMR spectroscopy, especially the $^{3}J(^{1}\mathrm{H},^{1}\mathrm{H})$ couplings determine the shape and splitting of the signals, from which one can gain even more information on the relative structure of the investigated molecule.

There are numerous further NMR experiments and measurement techniques that enable a vast range of possibilities for structure elucidation of molecules, including multi-dimensional

spectra even with different nuclei. However, introducing those would be beyond the scope of this thesis. Some more information on NMR spectroscopy can be found in Chapters 3 to 7 and the respective appendices.

2.3.2 Computational Aspects

As mentioned in Chapter 1, there is great interest in the computational prediction of NMR parameters (shielding and coupling constants) and the simulation of NMR spectra since it can supplement the already great amount of information obtained from an experimental measurement. As worked out by Ramsey between 1950 and 1953,[23] the NMR parameters are obtained as second-order properties from, in principle, any desired electronic structure method. With the three-dimensional external magnetic field vector $\vec{B}_0$ and nuclear magnetic moment $\vec{\mu}_i$ of nucleus i, the elements of the 3×3 shielding tensor σ_{ij} and the spin-spin coupling tensor J_{ij} are obtained as second derivatives of the energy E.

$$\sigma_{ij} = \frac{\partial^2 E}{\partial \mu_i \partial B_{0j}} \tag{2.60}$$

$$J_{ij} = \frac{\partial^2 E}{\partial \mu_i \partial \mu_j} \tag{2.61}$$

Since this thesis mainly focuses on the chemical shielding process, the spin-spin coupling is not considered further. The most important quantity that can be obtained from the shielding tensor σ via equation (2.60) is the isotropic shielding constant σ_{iso}, which contains a paramagnetic and a diamagnetic component.

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \tag{2.62}$$

$$\sigma_{\text{iso}} = \frac{1}{3}(\sigma_{11} + \sigma_{22} + \sigma_{33}) \tag{2.63}$$

In order to be directly comparable to an experimentally obtained value, the calculated shielding constant $\sigma_{A,\text{iso}}$ of a nucleus A requires a reference point. Analogously to the experimental setting, the final chemical shift δ_A is obtained from the isotropic shielding constant relative to that of a reference compound that should be calculated at the same level of theory as the compound of interest.

$$\delta_A = \sigma_{\text{ref,iso}} - \sigma_{A,\text{iso}} \tag{2.64}$$

The spin-spin coupling constants from equation (2.61) and the chemical shifts can finally be used to build a spin Hamiltonian that has to be solved to generate the simulated NMR spectrum. As an example, a simulated ^{1}H NMR spectrum of chloroethane is shown and annotated in Figure 2.1.

The existence of a directed magnetic field is accompanied by a gauge origin problem. Some terms in the energy expression, and consequently the properties obtained via equation (2.60), depend on a somewhat arbitrarily chosen gauge origin. This dependence must vanish, if the molecules are expected to move freely in space or in a solution. Different approaches to cope with this circumstance exist, e.g., the use of gauge-including atomic orbitals (GIAO)[137–139] for the basis set expansion (2.10) – which was used in all calculations in this thesis – or the individual gauge for localized orbitals (IGLO)[140–142] method.

Another artificial effect in an NMR shielding calculation is the absence of an averaging of chemical shifts of chemically equivalent nuclei, which is a dynamic effect. In solution and at a given temperature, the rotational barriers of some bonds in a molecule are small enough to enable a free rotation. Consider an ethane molecule (C_2H_6) as a simple example. A rotation around the C−C bond is usually possible, making all ^{1}H nuclei chemically equivalent, which results in only one signal from all six atoms. However, the calculation yields different values for σ_{iso} depending on the dihedral angle between two hydrogen atoms. Averaging the chemical shifts from snapshots of an MD simulation, as described in Section 2.2.2, would yield the same value for all ^{1}H nuclei. To circumvent this computational effort, chemically equivalent nuclei are identified manually or even automatically[30] and their shift values are averaged before the resulting value or spectrum can be compared to experimental data.

It was already claimed that DFT is often a reliable choice as electronic structure method

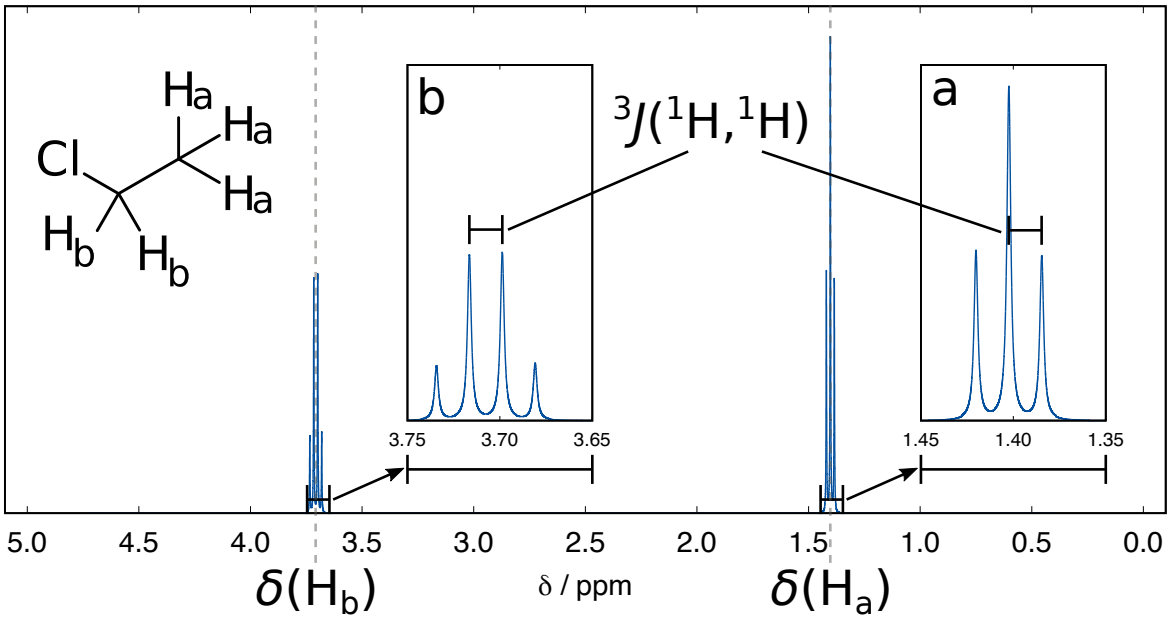


Figure 2.1: Simulated ^{1}H NMR spectrum of chloroethane with NMR parameter calculations at the PBE0/pcSseg-2[109,136] level of theory. The five hydrogen atoms lead to two signals with distinct chemical shifts δ (corresponding to their positions in the spectrum) and coupling patterns caused by the presence of $^3J(^1H,^1H)$ spin-spin couplings.

for the calculation of NMR chemical shifts. Nevertheless, it is important to be aware of the various sources of error that occur in such calculations. Lodewyk et al. declared five of them as the major ones:[24] electron correlation, conformational flexibility, solvent and other intermolecular interactions, heavy atom effects, and rotational-vibrational effects (cf. Figure 1.1). All of them are briefly commented on in the following paragraphs.

Electron Correlation

It is somehow obvious that a better description of electron correlation effects is accompanied by a higher accuracy for calculated properties, not restricted to NMR chemical shifts. DFT intrinsically includes some correlation and is, therefore, usually superior to HF.[33] If the computational resources allow, MP2 or double hybrid DFAs can be used to obtain more accurate NMR chemical shifts.[39] However, for highly accurate shielding constants, coupled cluster may serve as a reference method. It has been shown that CCSD(T) significantly outperforms CCSD, but all methods with even higher excitations usually benefit only slightly.[43]

To account for electron correlation effects without explicitly calculating them, a simple linear regression (LR) technique is commonly applied. When data for a set of shifts is available at a *low* and a *high* level (e.g., two different levels of theory or calculated and experimental data), the parameters α and β can be fitted to function (2.65) and used to scale and improve the data obtained at the *low* level.

$$\delta_{\text{low}} = \alpha \cdot \delta_{\text{high}} + \beta \tag{2.65}$$

$$\delta_{\text{low,scaled}} = \frac{\delta_{\text{low}} - \beta}{\alpha} \tag{2.66}$$

In this thesis, a method based on machine learning (see Section 2.4) was developed to correct ^{1}H and ^{13}C NMR chemical shifts obtained with DFT toward CCSD(T) accuracy. This method is the subject of Chapter 5.

Conformational Flexibility

The influence of considering more than one conformation of a sample molecule does not only apply to the total (free) energy as presented in Section 2.2.2, but also to derived properties. Thus, it can be important to perform NMR chemical shift calculations for all conformers in an ensemble and obtain the Boltzmann-weighted average chemical shift via equation (2.55). This procedure is particularly substantial if the molecule of interest is flexible. The already mentioned workflow using the CREST and CENSO algorithms is a reliable way of generating a conformer ensemble for that purpose. With this workflow, it is possible to automatically simulate an NMR spectrum and minimize the error due to conformational flexibility.[30]

Solvent and other Intermolecular Effects

The electrostatic perturbation of a molecule's electronic structure caused by an implicit solvation model has an indirect effect on the subsequent chemical shielding calculation. For NMR simulations in solution, it is, therefore, advisable to apply a solvation model. The differences in performance of different solvation models is expected to be small. However, a rigorous study of the influence of different solvent models for the prediction of NMR chemical shifts is missing until now.

Applying an explicit solvation model essentially breaks down the solvent influence to intermolecular effects that enhance the description of the real environment in a calculation. The effect of such non-covalent interactions can be significant, especially when π-systems are involved, which might exhibit a ring-current effect that is transfused to another molecule. In a side study to this thesis,[143] a maximum ^{1}H NMR chemical shift change of 1.9 ppm was found for a water molecule approaching a benzene molecule.

Heavy Atom Effects

As mentioned in Section 2.2.3, a relativistic treatment can be necessary in an electronic structure calculation when heavy elements are present. The NMR shielding constant is special in this context, because the local magnetic field experienced by the nucleus is influenced most by the electrons in the low-lying orbitals. These are, in turn, most prone to relativistic effects. Heavy atom effects to NMR parameters can, therefore, be very large. There are two important cases that should be distinguished: The effect on a shielding constant of a heavy nucleus and the effect from a heavy nucleus on the shielding constant of its neighboring nuclei.

When the investigated nucleus itself is heavy, such as in ^{119}Sn or ^{207}Pb NMR spectroscopy, it is intuitive to recognize that a missing relativistic treatment leads to large errors. In Chapter 4, it will be shown that it is essential to include relativistic effects in ^{119}Sn NMR chemical shift calculations in order to get qualitatively correct results. Furthermore, approaches including spin-orbit effects should be generally preferred over simpler scalar-relativistic methods. Since ECPs do not explicitly contain electrons in the core region, the orbitals are of wrong shape, making them unusable for heavy nucleus NMR shift calculations.

If a light nucleus is investigated by NMR spectroscopy but it is connected to a heavy atom, it experiences the heavy atom on the light atom (HALA) effect.[48] It results from the fact that the relativistic effect in the heavy nucleus can propagate via Fermi-contact (FC) interactions, which are magnetic interactions between a nuclear and an electron spin from an electron in an *s*-orbital. It can thus affect the NMR shielding of a light nucleus significantly, when they are connected via electrons in σ-orbitals, usually covalent bonds. However, the effect can also occur through non-covalent bonds[144] and is, in general, especially pronounced for heavy halide substituents due to their strong spin-orbit coupling and large number or electron lone pairs. In Chapter 3, it will be shown that a relativistic treatment is necessary to capture the HALA effect and SO methods outperform SR approaches. Since the light

nucleus is only indirectly affected by the relativistic effects, ECPs can be used.

The subject of Chapter 6 is the development of a second machine learning-based method that predicts the change in ^{1}H and ^{13}C NMR chemical shifts close to heavy atoms caused by including SO relativistic effects.

Rotational-Vibrational Effects

Molecules consistently rotate and vibrate at certain energy levels. Since a molecular vibration essentially means that its geometry is changing over time, the NMR chemical shielding constants of the affected nuclei experience vibrational effects. Although there are approaches that allow a vibrational correction to the shielding constant, it was found that the effect is highly systematic throughout various electronic structure methods, thus it mostly cancels when a relative shift is calculated with equation (2.64).[44] Consequently, vibrational effects are important to consider if absolute shielding constants σ are calculated, but they are negligible for chemical shifts δ.

2.4 Machine Learning

In the past years, machine learning (ML) and artificial intelligence (AI) have entered our everyday life and are nowadays omnipresent. This section shortly presents the fundamentals of the concept of ML, mainly based on reference [145], and puts it in the context of this thesis.

The term "machine learning" describes a principle in data science in which a program is designed such that it *learns* from the data it is provided with and builds an abstract statistical model based on experience. The structure and internal relations of the underlying training data are, therefore, generalized and can be used to make predictions for similar data that is unknown to the model. The diversity of ML techniques is massive and many of them have become part of our daily life (e.g., in recommendation algorithms,[146] picture or voice recognition[147] and generation,[148] and universal chat bots[149]). Machine learning techniques are particularly useful when a problem is too complex to be described by simple mathematical operations or the exact relation between input data and the desired output is simply not known.

Among many others, artificial neural networks (ANNs) can be applied for such problems. Since these have been used in this thesis, all other flavors of ML methods are neglected here. ANNs are inspired by biological neural networks (e.g., the human brain), which consist of numerous neurons that are connected with each other and pass information encoded in electrical pulses. While a single neuron functions as a rather simple object, the entirety of billions of neurons makes up a network that is capable of highly complex (computational) operations. In an ANN, the neurons are also called nodes and are connected to other nodes. Each node represents a mathematical operation that acts on an input value under a certain condition and passes an output value to the next node(s).

Many nodes can be organized together to form a so-called perceptron. In that case, the nodes are arranged in two layers: An input layer, in which every node corresponds to an

input feature of a set of data points, and a layer, in which every node acts on a weighted sum of all inputs from the input layer. For a node i, x_{ij} and w_{ij} are the input values and weights of all input nodes j that are transferred to i. When the layer is fully connected (also called dense), i.e., all nodes are connected to all nodes of the previous layer (here the input), the output h_i of the node is calculated from an activation function ϕ acting on the weighted sum of all inputs with a bias b_i.

$$h_i = \phi\left(\sum_j w_{ij} x_{ij} + b_i\right) \tag{2.67}$$

Following this procedure, more sophisticated ANNs can be constructed. The most common one is the multilayer perceptron (MLP), which consists of an input layer as before, one layer of nodes that form the output and an arbitrary number of hidden layers with nodes that act on their inputs and return their output as given in equation (2.67). If hidden layers are present in an ANN, it is called a deep neural network (DNN). The architecture of an MLP with two hidden layers and a single value output is sketched in Figure 2.2.

The actual *learning* process is the optimization of the weights w_{ij} connecting the neurons given a large amount of training data. This approach requires an initial guess of the starting weights, which are usually chosen randomly. Then, a gradient descent algorithm is applied that is based on backpropagation. That means that each training data point is passed through

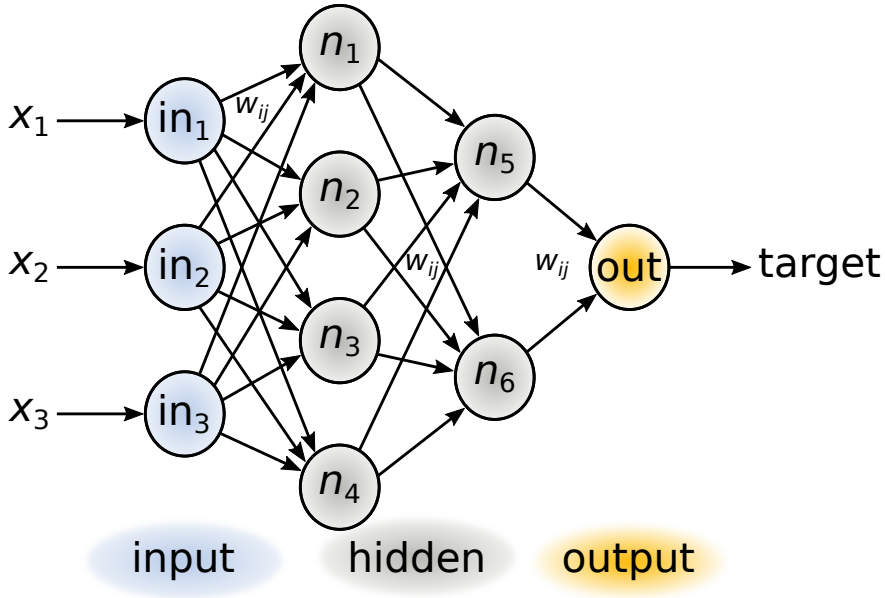


Figure 2.2: Structure of a simple multilayer perceptron. The input layer receives the three-dimensional input feature vector (x_1, x_2, x_3) and passes it to all nodes n_1-n_4 of the first hidden layer, with every arrow indicating the respective weights w_{ij}. Each node applies equation (2.67) to calculate its output that is passed to the second hidden layer consisting of nodes n_5 and n_6. The same is repeated until the output layer is reached (in this case only one node) and the target value(s) are returned.

the model to predict the output value and calculate its error, before it is traced back and the error contributions from each node are used to update the weights accordingly. The overall error is tracked by a loss function (e.g., the mean squared (MSD) or mean absolute deviations (MAD) of the predictions) and the procedure is repeated until a convergence is reached.

MLPs can be used for both classification and regression tasks and can learn to predict several parameters simultaneously. For a classification problem, one node must be present in the output layer for each task and typically uses an activation function that makes a choice based on the finally calculated probability. If an MLP is used for a regression problem – as the ones built in this thesis – the number of nodes in the output layer equals the dimensionality of the regression and the value is usually obtained directly from the node(s) without an additional activation function in order to guarantee the maximum freedom for the result.

There are many technical details that can be varied in an ANN to adapt it optimally to the problem it is used for. The so-called hyperparameters include the number of hidden layers, the number of nodes per hidden layer, the activation function applied in the nodes of the hidden (and output) layers, a dropout rate between the layers, the loss function, and the optimization algorithm. Besides these characteristics of the model itself, its success is also substantially conditional on the training data. A reliable training data set should be uniform and large enough so that the ML model can recognize important connections and should not contain much noise or redundant information.[150] Furthermore, it is important to cross-check the performance of an ML model with a test data set that is unknown to the model during the training process. An excessive focus on the training data can lead to an overfitted model that is not sufficiently general and performs significantly better for the training data than for any other test case.

In this thesis, two regression MLPs were constructed and optimized in a supervised learning approach, meaning that the training data is labeled, so that the target values are known during the training process. The first model, Δ_{corr}-ML, is used to predict the difference (Δ) in a ^{1}H or ^{13}C NMR chemical shift calculated with DFT and a reference value based on CCSD(T), which essentially means to predict the electron correlation contribution to the chemical shift (Chapter 5). Each data point represents one nucleus with features that describe it and its environment and the one-dimensional output is the expected difference $\Delta\delta$. The second model, Δ_{SO}-ML, uses a similar approach, but the data is labeled with chemical shifts from SO relativistic NMR calculations, so it predicts the relativistic contribution to ^{1}H and ^{13}C NMR chemical shifts caused by nearby heavy atoms (Chapter 6).

CHAPTER 3

Comprehensive Benchmark Study on the Calculation of ^{29}Si NMR Chemical Shifts

Markus Bursch,[‡] Thomas Gasevic,[‡] Julius B. Stückrath,[‡] Stefan Grimme[‡]

Received: September 30, 2020
First published: December 15, 2020

Reprinted (adapted) in Appendix A with permission from:
M. Bursch, T. Gasevic, J. B. Stückrath, and S. Grimme, *Inorg. Chem.* **2021**, *60*, 272–285.
DOI: 10.1021/acs.inorgchem.0c02907.

Own manuscript contributions:

- Performing all relativistic calculations (for which ADF was used)
- Evaluation of the results of the relativistic tests
- Writing parts of the manuscript

[‡] Mulliken Center for Theoretical Chemistry, Clausius Institute for Physical and Theoretical Chemistry, University of Bonn, Beringstr. 4, 53115 Bonn, Germany.

Besides ^{1}H and ^{13}C NMR spectroscopy, the ^{29}Si nucleus is a powerful probe widely used for (organo-)silicon chemistry as its NMR chemical shift exhibits a strong dependence on the chemical environment. There is, thus, a great research interest in the computational simulation of ^{29}Si NMR parameters for prediction and verification purposes with the goal of structure elucidation.[151,152] As mentioned earlier, DFT is often used for NMR shift calculations as a well-balanced compromise between the factors of accuracy and computational demands. Therefore, a comprehensive benchmark set for the evaluation of different computational methods regarding their ability to predict ^{29}Si NMR parameters was compiled in this study and named SiS146. It consists of 100 Si-containing structures with sizes between 5 and 234 atoms, including 146 experimental ^{29}Si NMR chemical shifts from solution measurements (+828 to –400 ppm, see Figure A.1) with a wide range of bonding motifs and chemical surroundings of the ^{29}Si nucleus. It is, hence, more universal than preceding studies that mostly focused on a special group of organosilicon compounds.[153,154] Furthermore, the set was divided into the "light" *SiS-L* subset that only contains elements with $Z \leq 18$ and the "heavy" *SiS-H* subset featuring heavy elements to study the HALA effect on the ^{29}Si nucleus as light atom.[48] The structures of all compounds in the benchmark set were generated via a workflow using GFN2-xTB and the final PBEh-3c[155] level of theory with CPCM for implicitly treating the solvent. Subsequently, NMR shielding calculations were performed using 22 different density functionals and HF in combination with the def2-TZVP[156] or pcSseg-2[136] basis sets, while relativistic effects were treated via ECPs or the SR- or SO-ZORA approaches.

In general, the influence of the type of basis set (def2-TZVP *vs.* pcSseg-2) was found to be of minor importance, although the pcSseg-2 basis set yielded slightly lower mean absolute deviations than def2-TZVP. However, for clarity, only the results using def2-TZVP are discussed in the following. When the different functionals are evaluated for the *SiS-L* subset (Figure A.2), a somewhat "reversed Jacob's ladder" behaviour was observed, meaning that the (meta-)GGA functionals outperformed the tested hybrid functionals in general (class mean absolute deviations of $\mathrm{MAD_{GGA}} = 10.9$ ppm and $\mathrm{MAD_{hybrid}} = 16.7$ ppm). The best performing DFAs of both types are B97-D3[157] (MAD = 7.2 ppm) and TPSSh[158,159] (MAD = 10.3 ppm). As a common and simple empirical correction, linear regression with the new data can be used to improve the results. When the chemical shifts are scaled according to the regression function, MADs can be reduced by 8% (0.8 ppm, KT2[160]) to 53% (15.4 ppm, M06-2X[161]). This technique showed to be particularly beneficial for methods with an initially poor performance. Apart from the density functional, the basis set convergence behavior was investigated for different sizes of the def2-XVP, pcSseg-X, pc-X,[162] and cc-pVXZ[163,164] families (see Figure A.4). In all cases, a clear convergence was observed with an optimal cost/accuracy ratio for triple-ζ basis sets. Thus, the def2-TZVP and pcSseg-2 basis sets are the preferred choice as they yielded accurate and robust results with a minimum risk of SCF convergence issues.

For the *SiS-H* subset, the suitability of different relativistic DFT approaches was assessed for ^{29}Si NMR chemical shifts in the presence of heavy atoms. Throughout the whole range of different DFAs, a negligible change in the results was observed for a scalar-relativistic

treatment (Figure A.6a). In contrast, including spin-orbit effects significantly reduces the MAD by up to 36.0 ppm (B3LYP) with respect to the ECP-based def2-TZVP results. The SO contribution to the ^{29}Si NMR chemical shift reaches extreme values when halogen atoms are bound to the Si atom. In the case of SiI_4, with an experimental value of –346.6 ppm, only the SO-ZORA approach yields the qualitatively correct result (–343.1 ppm with PBE0/TZP[109,165]) while all other approaches lead to deviations above 400 ppm. Finally, the influence of geometry optimization on the ^{29}Si NMR chemical shift was tested. Five representative structures from each subset were reoptimized at different (GGA and hybrid) DFT levels as well as at the GFN2-xTB and GFN-FF levels. While the influence of the DFA is negligible, structure optimization with the semiempirical and force field methods yielded clearly increased deviations. However, it was shown that this effect can partially be compensated when a linear regression correction is used.

It can be concluded that DFT is a powerful tool for the calculation of ^{29}Si NMR chemical shifts and the B97-D3, TPSSh, and PBE0 functionals can be generally recommended for that purpose. A linear regression correction, e.g., from the data of the SiS146 set, can improve the results of functionals with poor performance or when low-cost geometry optimizations were performed. If heavy elements are in the vicinity of the ^{29}Si nucleus, a (spin-orbit) relativistic treatment is inevitable and should be applied in any case.

CHAPTER 4

Benchmark Study on the Calculation of ^{119}Sn NMR Chemical Shifts

Julius B. Stückrath,[‡] Thomas Gasevic,[‡] Markus Bursch,[‡] Stefan Grimme[‡]

Received: November 4, 2021
First published: February 18, 2022

Reprinted (adapted) in Appendix B with permission from:
J. B. Stückrath, T. Gasevic, M. Bursch, and S. Grimme, *Inorg. Chem.* **2022**, *61*, 3903–3917.
DOI: 10.1021/acs.inorgchem.1c03453.

Own manuscript contributions:

- Preparation of the calculations
- Reviewing all structures and data of the benchmark set
- Evaluation of the results
- Writing the manuscript

[‡] Mulliken Center for Theoretical Chemistry, Clausius Institute for Physical and Theoretical Chemistry, University of Bonn, Beringstr. 4, 53115 Bonn, Germany.

Since NMR spectroscopy of certain heteronuclei can be crucial for the characterization of chemical compounds of the respective nucleus, the focus in Chapter 3 was on ^{29}Si NMR chemical shifts. However, the same holds for the versatile field of (organo)tin chemistry in which substances are routinely examined by ^{119}Sn NMR spectroscopy.[166–168] There is, thus, a natural need of reliable computational methods for the prediction of ^{119}Sn NMR parameters,[169–171] especially because of the large chemical shift range and time-consuming measurements for heavy nuclei. Hence, various DFT approaches were assessed in a comprehensive benchmark set called SnS51. It is made up from 50 tin-containing compounds with 4 to 209 atoms and 51 experimentally determined ^{119}Sn NMR chemical shifts obtained from various studies in solution and covering the whole typical range from +2448 to –2204 ppm (see Figure B.2). In contrast to the organosilicon compounds in Chapter 3, Sn−Sn bonds are rare and tin is more often involved in coordinative bonds. The compounds were chosen such that a broad variety of bonding patterns and coordination numbers is represented (Figure B.3). In a multilevel workflow, conformer ensembles were generated with CREST for every compound and optimized with CENSO at the final r^{2}SCAN-3c[172] level of theory. In total, 219 unique conformers were obtained and the chemical shifts were calculated as Boltzmann-weighted averages. Besides the assessment of different (meta-)GGA and hybrid density functional approximations, a focus is placed on the relativistic treatment in the calculations, which is expected to be essential for the heavy ^{119}Sn nucleus.[48,53,169,173]

15 different DFAs were compared for the calculation of ^{119}Sn NMR chemical shifts in conjunction with the scalar-relativistic SR-X2C approach and the x2c-TZVPall-s[174,175] basis set. The "reversed Jacob's ladder" behavior observed for the ^{29}Si NMR case mentioned before was not found for ^{119}Sn NMR. Instead, all DFAs performed similarly and, on average, consistently underestimate the experimental value. The MADs lie between 130.2 (M06[161]) and 219.5 ppm (M06-L[176]) and insignificantly better results were obtained with the hybrid functionals (MAD_{GGA} = 184.1 ppm, MAD_{hybrid} = 158.4 ppm). As long as no ECPs are used (they are unsuitable for heavy nucleus NMR shieldings due to an inherently wrong physical description), the relativistic Hamiltonian has only a limited effect on the result when a scalar approach is chosen. Averaged over all functionals, the mean absolute deviation for X2C ($MAD_{SR\text{-}X2C}$ = 172.1 ppm) is slightly lower than that for ZORA ($MAD_{SR\text{-}ZORA}$ = 187.0 ppm). More importantly, the deviations decrease significantly when the spin-orbit variant of ZORA is used with the TZP basis set ($MAD_{SO\text{-}ZORA}$ = 101.2 ppm). This step also drastically reduces the errors from two compounds with multiple Sn−I bonds (**45** and **48**) that are outliers when evaluated for the SR methods, showing that the pronounced heavy atom neighbor effect from heavy halogens observed for ^{29}Si NMR also appears for ^{119}Sn. The overall best performance is achieved with SO-ZORA/PBE0 with an MAD of 87.6 ppm. On top of that, the linear regression correction technique used before proved useful for ^{119}Sn NMR as well. When a linear function is fitted on the data of the SnS51 set, MADs were reduced by up to 51% (from 102.2 to 49.6 ppm for SO-ZORA/B3LYP).

In addition to the investigations of the property prediction, the influence from molecular structures was investigated. First, the evaluations for the SO-ZORA/revPBE[177] level were replicated using only the conformer lowest in free energy rather than the ensemble average

(see Figure B.7). In the vast majority of systems, the results were not affected considerably by neglecting all other conformers. The exception with the largest ensemble influence of 43.0 ppm originates from two competing coordinated bonding motifs of 1,1-dithiolate ligands that lead to differences in the ^{119}Sn NMR chemical shift of around 300 ppm. Finally, different levels of theory were studied for geometry optimizations and – in line with the findings for ^{29}Si NMR – no significant changes were observed for the use of different DFAs, neither for any relativistic treatment (Figure B.9). When the efficient GFN2-xTB and GFN-FF methods are used, larger deviations are observed. However, these errors can again partially be compensated by applying a linear regression correction.

In conclusion, it can be stated that the application of DFT for the calculation of ^{119}Sn NMR chemical shifts can lead to accurate results and the linear regression parameters from the SnS51 set can improve the results further. The revPBE and PBE0 functionals perform well and it is strongly recommended to use a relativistic treatment including spin-orbit coupling, otherwise severe errors are probable. If the computational resources are sparse, low-cost methods can be applied for geometry optimization and influences from conformational flexibility are usually expected to be small.

CHAPTER 5

Computation of CCSD(T)-Quality NMR Chemical Shifts via Δ-Machine Learning from DFT

Julius B. Kleine Büning,[‡] Stefan Grimme[‡]

Received: February 9, 2023
First published: June 1, 2023

Reprinted (adapted) in Appendix C with permission from:
J. B. Kleine Büning and S. Grimme, *J. Chem. Theory Comput.* **2023**, *19*, 3601–3615.
DOI: 10.1021/acs.jctc.3c00165.

Own manuscript contributions:

- Development of the machine learning model and data set
- Performing all calculations
- Evaluation of the results
- Writing the manuscript

[‡] Mulliken Center for Theoretical Chemistry, Clausius Institute for Physical and Theoretical Chemistry, University of Bonn, Beringstr. 4, 53115 Bonn, Germany.

The importance of NMR spectroscopy for chemistry on a daily basis has been outlined in Chapter 1. It was further stated that DFT is often used as compromise between accuracy and computational cost for the calculation of NMR parameters. However, the typical errors caused by an insufficient treatment of electron correlation, which can be expected from DFT-based NMR calculations, often impair the agreement of calculated and experimental results.[33,41,178] It is therefore desirable to approach the accuracy of coupled cluster calculations, ideally with minimal additional computational demands. For this purpose, a correction method based on machine learning called Δ_{corr}-ML (Δ-ML in Appendix C) was developed to tackle one of the five main error sources in NMR parameter computation (see Section 2.3.2), namely the electron correlation. The empirical model aims at predicting the difference $\Delta_{corr}\delta$ between a ^{1}H or ^{13}C NMR chemical shift from a DFT calculation to a *high-level* reference value obtained from a CCSD(T)/pcSseg-2[136,179] calculation with a basis set extrapolation scheme. It is built upon a data set containing 1000 structures (optimized at the r^{2}SCAN-3c level of theory and geometrically distorted) with 7090 ^{1}H and 4320 ^{13}C chemical shifts. These structures include small, mostly organic, molecules with sizes up to 22 atoms and hydrogen and carbon atoms in a huge variety of bonding motifs. Heteroatoms are included up to chlorine. The underlying artificial neural network is a multilayer perceptron for which the input feature vector is obtained from a *low-level* DFT calculation, making it a Δ-ML approach. It includes descriptors from three different categories: *geometric* (obtained solely from the three-dimensional molecular structure), *electronic* (properties from the converged DFT density matrix), and *magnetic* (properties from the converged DFT shielding tensor). While many other ML approaches use molecular data,[180–182] this MLP is based on atom-wise input features.

The performance of the Δ_{corr}-ML model was first evaluated on the test data set (12.5% of the data) for the correction of ^{1}H NMR chemical shifts using PBE0/pcSseg-2 as the *low-level* method. In this test, a reduction of the MAD between DFT and CCSD(T) of 81% (from 0.130 to 0.025 ppm) was observed, while also the error spread and standard deviation were reduced significantly (see Figure C.2). A potential overfitting effect was precluded by evaluating the method for its own training set as well, which yielded only a slightly better MAD reduction of 84%. Furthermore, the performance of the Δ_{corr}-ML method was compared to a simple linear regression approach that is commonly used and has proved useful in Chapters 3 and 4. In all tests, linear regression was consistently outperformed by Δ_{corr}-ML. In similar evaluations of the ^{13}C data, an even better performance was found with a reduction of the PBE0/pcSseg-2 MAD by 92% (from 5.39 to 0.42 ppm) and a drastically narrowed error spread (Figure C.3). This generally better performance for ^{13}C compared to ^{1}H is a result from the fact that a hydrogen nucleus is less shielded due to its smaller electron cloud, which makes it more prone to its local magnetic environment. Most computational approaches have to deal with this obstacle, not excluding the presented method.

The Δ_{corr}-ML model was constructed such that virtually any computational approach that provides the necessary chemical shifts and descriptors can serve as *low-level* method – not necessarily DFT – and different variants were investigated (see Figure C.4). Most importantly, this revealed a consistent performance throughout all tested *low-level* methods.

A large basis set (basically converged at triple-ζ quality) can be beneficial as it enhances the quality of the *electronic* and *magnetic* input features. However, the choice of the density functional is of overall much greater importance. Despite partially large differences for the uncorrected results of different DFAs, a similar corrected MAD was found for all methods except M06, which seems to yield too unsystematic errors for ^{13}C NMR chemical shifts. Finally, the Δ_{corr}-ML method applied to PBE0/pcSseg-2 was evaluated on truly external test cases from three benchmark studies. Throughout all tests, the general performance was retained and significantly reduced deviations were found for both ^{1}H and ^{13}C NMR chemical shifts with respect to highly accurate computed reference data (CCSD(T) with a slightly different basis set than used for training).[33,41] Not surprisingly, it was shown that a somewhat reduced accuracy has to be expected when the investigated molecules differ considerably from those in the training set. In the last benchmark study,[34] the Δ_{corr}-ML approach was directly evaluated against experimental data and a clear error reduction was achieved, although only one (electron correlation) out of the previously mentioned five main error sources was addressed.

To conclude, the presented Δ_{corr}-ML method provides robust and reliable corrections for ^{1}H and ^{13}C NMR chemical shifts toward a CCSD(T) quality that can be used to improve simulated NMR spectra at negligible computational costs. It represents a promising contribution to efficient NMR prediction workflows.

CHAPTER 6

Machine Learning-Based Correction for Spin-Orbit Coupling Effects in NMR Chemical Shift Calculations

Julius B. Kleine Büning,[‡] Stefan Grimme,[‡] Markus Bursch[§]

Received: November 15, 2023
First published: January 9, 2024

Reproduced with permission from the PCCP Owner Societies from:
J. B. Kleine Büning, S. Grimme, and M. Bursch, *Phys. Chem. Chem. Phys.* **2024**, *Advance Article*.
DOI: 10.1039/d3cp05556f.

Own manuscript contributions:

- Development of the machine learning model and data set
- Preparation of the 17HAC benchmark set
- Performing all calculations
- Evaluation of the results
- Writing the manuscript

[‡] Mulliken Center for Theoretical Chemistry, Clausius Institute for Physical and Theoretical Chemistry, University of Bonn, Beringstr. 4, 53115 Bonn, Germany.
[§] Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr, Germany.

In Chapters 3 and 4, the focus of the investigation of DFT methods for NMR chemical shift calculation was placed on heavy elements (tin and other heavy elements in organosilicon compounds). The interest in (organic) compounds featuring some heavy atoms is huge, ranging from the use in catalysts[183–185] and electronic devices[186–189] to biological importance due to toxicity[190] or other important functions.[191–196] For interpreting the NMR measurement results as well as for computationally simulating spectra, heavy atoms and their direct vicinity require special considerations. For instance, ^{1}H and ^{13}C NMR chemical shifts can take on extreme values that originate from relativistic spin-orbit (SO) effects transmitted from the heavy atom to a light atom: the HALA effect.[48] In fact, such heavy atom effects represent one of the five main error sources for NMR shift calculation presented in Section 2.3.2. The relativistic contributions to the NMR chemical shift are not included in "standard" (non-relativistic) DFT and computationally demanding relativistic approaches have to be applied, such as ZORA or X2C, as used in the earlier chapters. Alternatively, this work follows the ideas of the Δ_{corr}-ML[197] method (Chapter 5), presenting a novel machine learning-based correction method for the efficient prediction of spin-orbit effects to NMR chemical shifts, called Δ_{SO}-ML.

As its predecessor, the Δ_{SO}-ML method uses an MLP with two hidden layers and the hyperparameters were adjusted. For the data set, small to medium-sized molecules with organic backbones and at least one out of 17 different heavy atoms ($Z \geq 17$ from groups 12 to 17) were used to create 6388 structures with a total number of 38740 ^{13}C and 64436 ^{1}H NMR chemical shifts from nuclei in different distances to a heavy atom (Figure D.3). For each shift, the reference SO contribution $\Delta_{\text{SO}}\delta$ was calculated at the ZORA-PBE0/TZ2P[109,165] level of theory as difference between an SO and a scalar-relativistic (SR) calculation. Most input features used in the Δ_{corr}-ML method were also used in the new approach and the *geometric* and *electronic* descriptors were extended with information about heavy atoms. When the Δ_{SO}-ML method is evaluated on the test data set (12.5% of the data), a significant and consistent decrease of the deviations between the baseline DFT results and the reference values is observed. For the ^{13}C nucleus, the MAD and RMSD are reduced by 85% and 87%, the mean signed deviation almost vanishes, and the error spread is narrowed drastically (see Figure D.4). Like for the Δ_{corr}-ML method, a very similar yet slightly less pronounced behavior was found for the ^{1}H case with decreases of 68% and 71% of the test set MAD and RMSD, respectively. Since a linear regression model is too simple to capture the SO contribution to NMR in a comparably general way as the Δ_{SO}-ML approach, the latter is generally superior to such a regression for both nuclei.

Interestingly, the generalizability of the Δ_{SO}-ML method was found to hold for the use of different *low-level* DFT methods and even relativistic approaches (see Figure D.6). Due to a rather strong correlation between descriptors from different levels of theory, the performance for ^{13}C NMR does not change significantly when the *low-level* method is changed from SR-PBE0/TZ2P to SR-PBE/TZ2P or SR-r^{2}SCAN0/TZ2P.[198] Even when using an effective core potential for including relativistic effects rather than an explicit ZORA approach, the average deviations increase only slightly. The behavior is similar for ^{1}H with generally larger deviations, but in all tested cases, retraining the model with data from the chosen

low-level method recovers the original performance. To investigate the training data bias of the new approach, it was evaluated with the use of benchmark sets outside the training data, including the SnS51[199] set (Chapter 4) and a new study on ^{207}Pb NMR chemical shifts,[200] for which a comparable error reduction was observed for the near ^{13}C and ^{1}H nuclei.

Finally, a combination of Δ_{SO}-ML and Δ_{corr}-ML was tested on a newly compiled benchmark set called 17HAC, containing experimental ^{13}C NMR chemical shifts from molecules with all 17 included heavy elements. Each correction reduces the errors significantly, but only the application of both leads to the remarkably low MAD of 5.7 ppm, corresponding to a 50% lowering of the overall errors. Furthermore, a bismabenzene compound[201] was showcased as example for an exceptionally large SO contribution rather far away from the heavy atom (three bonds). Also here, the qualitative trends are correctly predicted by the Δ-ML approaches. Due to the still limited amount of training data, the computational demands required for the training and prediction process are negligible compared to the necessary DFT calculations. In summary, the Δ_{SO}-ML method represents an efficient approximation to relativistic NMR shielding calculations and even with a limited accuracy, it can serve as a powerful tool for screening applications or in composite method schemes for the calculation of ^{1}H and ^{13}C NMR chemical shifts.

CHAPTER 7

Modular Bicyclophane-Based Molecular Platforms

Daniel Kalle,[‡] Joshua Bahr,[‡] Tristan Johannes Keller,[‡] Julius B. Kleine Büning,[§] Stefan Grimme,[§] Markus Bursch,[¶] Stefan-Sven Jester,[‡] and Sigurd Höger[‡]

Received: August 15, 2023
First published: September 18, 2023

Reprinted (adapted) in Appendix E with permission from:
D. Kalle, J. Bahr, T. J. Keller, J. B. Kleine Büning, S. Grimme, M. Bursch, S.-S. Jester, and S. Höger, *Chem. Eur. J.* **2023**, *29*, e202302662. DOI: 10.1002/chem.202302662.

Own manuscript contributions:

- Performing all NMR shielding calculations
- All applications of the Δ_{corr}-ML method
- Computational modeling of the adsorbed structures
- Writing parts of the theory section of the manuscript

[‡] Kekulé Institute of Organic Chemistry and Biochemistry, University of Bonn, Gerhard-Domagk-Str. 1, 53121 Bonn, Germany.
[§] Mulliken Center for Theoretical Chemistry, Clausius Institute for Physical and Theoretical Chemistry, University of Bonn, Beringstr. 4, 53115 Bonn, Germany.
[¶] Max-Planck-Institut für Kohlenforschung, Kaiser-Wilhelm-Platz 1, 45470 Mülheim an der Ruhr, Germany.

The design and exploration of new nanomaterials is a vast and active field of research with numerous potential applications especially in materials science.[202–206] One important area is the modification of two-dimensional materials with specific molecular functionalities using the so-called platform approach.[202] It allows the precise and ordered decoration of a surface, e.g., of a metal or graphite, with a desired property provided by certain molecular moieties at a specific distance to the surface. In the case of graphite, a pillar carrying the functionality can be connected perpendicularly onto a platform that will then adsorb on the surface. The characteristics of the material can be optimized by variations in the length of the pillar or the size of the platform.[207] In addition, long alkyl chains can be attached to the platform to optimize its solubility and adsorption properties.[208,209] The rigidity of the pillar is, on the other hand, responsible for holding the functionality in place.[210] In the experimental part of this work (performed by the collaborators), different variants of platforms made from phenylene-based bicyclophanes were investigated.[211,212] The compounds are adsorbed on highly oriented pyrolytic graphite (HOPG) and examined using scanning tunneling microscopy (STM) to systematically study the influence of the molecule design on the adsorption.

After the final step of the synthesis of the bicyclophanes (see Figures E.1 and E.2), an essential difference between those without a pillar (compounds **1a**, **2a**, **3a**) and those with a pillar (compounds **1b**, **2b-d**, **3b-d**) was observed. More precisely, a large shift in the ^{1}H NMR signals from the central phenylene unit appeared for the reaction of **6** to **1b** but not for **4** to **1a**. To study this phenomenon further, model compounds **7** (open form) and **8** (closed, final form) with truncated side chains were prepared and subsequently investigated via ^{1}H NMR spectroscopy, showing a very similar shift with a maximum value of +1.59 ppm in the t position (Figure E.3). These model systems were then studied computationally by generating conformer ensembles with CREST and refining them with CENSO at the final r^{2}SCAN-3c level of theory. For the open form (**7**), the majority of conformers exhibited a structure in which the central phenylene unit is stacked between two phenyl rings from the side chains of the molecule and the ^{1}H nuclei are, thus, expected to experience large influences from external ring current effects. This presumption was confirmed by nucleus independent chemical shielding (NICS) calculations at the r^{2}SCAN0/def2-TZVPP level from a cutout structure that revealed a strong shielding of –3.1 ppm at the t position. In the closed form (**8**), the structure does not allow a spacial proximity of the central phenylene unit to other benzyl rings anymore, thus, the shielding effect vanishes. While this structural change explains qualitatively the large shift of the ^{1}H nucleus t, the calculated values are shifted to much too negative values. To exclude post-DFT correlation effects as reason for this failing, the Δ_{corr}-ML[197] method, introduced in Chapter 5, was applied and predicted insignificantly small corrections to the calculated ^{1}H NMR chemical shifts. Therefore, an MD simulation was performed and randomly taken snapshots were used for NMR shift averaging rather than the Boltzmann average of the conformers. In this way, a much better agreement with the experimental data was obtained, with a shift of +1.68 ppm (DFT + Δ_{corr}-ML) for the t position.

The investigation of the three different platform designs (**1**, side chains only in “west/east”

directions, **2**, side chains only in "north/south" directions, and **3**, side chains in all directions) by STM showed their consequences for organization on the HOPG surface (Figure E.5). In the case of **1**, a two-dimensional pattern stable under STM conditions was only observed without a pillar (for **1a**, but not for **1b**). In contrast, for platform design **2**, the compound carrying a pillar functionalized with a triisopropylsilyl (TIPS) unit (**2b**) exhibited an adsorption pattern very similar to its unfunctionalized counterpart (**2a**) with short pillar–pillar distances. To investigate a potential bending of the pillar toward the surface, a slightly modified structure of **2b** on a graphene sheet was modeled at the GFN2-xTB level of theory. It was found that the pillar points upright into the solution phase without bending. However, in the experiment, replacing the TIPS functionality by a significantly larger moiety (**2d**) inhibits the formation of a two-dimensional arrangement. Only with ten side chains per molecule in every direction (**3**) similar patterns were observed for all pillar modifications **3a-d**.

All in all, this study of different platform and pillar modifications showed that the design of the bicyclophane compounds determines their adsorption patterns and that a high number of alkyl chains per molecule allows the tolerance for large functionalizations. The experimental findings were underpinned by computational studies of the molecular structures, including ^{1}H NMR shielding calculations using the new Δ_{corr}-ML method and a structural investigation with GFN2-xTB.

CHAPTER 8

Conclusion and Outlook

It is barely possible to overrate the importance of nuclear magnetic resonance spectroscopy to research in chemistry and beyond. The large amount of structural information that can be obtained from just one measurement makes it very powerful and versatile and renders it indispensable to the everyday life of many researchers, both in academia and industry. With modern computational chemistry methods, it is nowadays also possible to perform NMR calculations on a daily basis to support experimental findings and expand the knowledge gained from these. As a matter of course, it is desirable to always apply the methods that are best suited for the problem of interest. However, for a new substance, it is often not easy to estimate the variance of results obtained, e.g., from the use of different density functionals or relativistic approximations.

For that reason, one aim of this thesis was to extend to the general understanding of the DFT computation of ^{29}Si and ^{119}Sn NMR chemical shifts and their influencing factors, as was announced in Chapter 1. Significant progress has been made regarding the use of DFT and relativistic approximations in the studies presented in Chapters 3 and 4. For example, it is not necessarily advantageous to use a higher rung functional for calculating ^{29}Si or ^{119}Sn NMR chemical shifts. It is rather important to choose an appropriate DFA, such as the B97-D3 GGA or the TPSSh meta-hybrid functional with MADs for ^{29}Si NMR in the *SiS-L* subset of the SiS146 set of 6.9 and 9.0 ppm, respectively (applying the pcSseg-2 basis set).

When heavy elements (from the fourth and higher periods of the periodic table) are involved, it is generally important to include some explicit treatment of relativistic effects for computing NMR parameters and all-electron calculations should be preferred over ECPs. Furthermore, it is usually not sufficient to only account for scalar-relativistic effects, but including spin-orbit effects is often required to achieve a reasonable agreement with experimental data. An adequate approach, such as SO-ZORA, is therefore suggested, especially if the degree of heavy atom and HALA effects cannot be estimated beforehand. For ^{29}Si NMR chemical shifts including heavy atoms in the vicinity to the ^{29}Si nucleus of interest, good results are expected for revPBE and PBE0 with MADs on the *SiS-H* subset of 11.6 and 13.6 ppm, respectively (SO-ZORA, TZP basis set). For the ^{119}Sn nucleus, which is a

heavy element itself, SO relativistic effects are essential and again, the SO-ZORA method with revPBE or PBE0 in a TZP basis can be recommended with MADs as low as 96.8 and 87.6 ppm, respectively (note the much larger absolute scale for ^{119}Sn than for ^{29}Si, which cover about 5000 ppm and 1300 ppm, respectively; cf. Figures B.2 and A.1). However, most other tested DFAs exhibited only slightly worse performances and can be applied safely as well.

In general, the predictive power of NMR chemical shift calculations can be improved with an empirical linear scaling approach. It was shown that this technique also functions well for ^{29}Si and ^{119}Sn NMR chemical shift calculations and is particularly beneficial if the underlying method has a poor performance. Furthermore, efficient semiempirical quantum mechanical methods can be used for geometry optimization when a linear scaling factor, fitted on higher level data, is applied to remedy structural deficiencies. This approach is a useful way to reduce the computational effort in multilevel workflows. For organotin compounds (Chapter 4), it was shown that the effect from conformational averaging might be significant, but those systems are few. However, the significance of this influence is not always obvious, so an at least crude investigation of the conformational space is recommended.

It is good scientific practice to use newly gained knowledge to build new concepts and methods that refine existing models and allow better predictions. The understanding of the weaknesses of DFT for calculating NMR chemical shifts and the importance of relativistic effects in calculations involving heavy atoms lead to the development of two correction methods presented in Chapters 5 and 6. With an ever more growing importance of machine learning techniques for (computational) chemistry and, thus, the simple accessibility of development frameworks, such as TensorFlow, ML is an appropriate tool for the desired correction methods.

The constructed Δ_{corr}-ML method proved to reduce the discrepancy between ^{1}H and ^{13}C NMR chemical shifts calculated with DFT and with CCSD(T) significantly: The MADs evaluated on the test data set using PBE0/pcSseg-2 – which is one of the best performing functionals in the aforementioned studies – as the baseline DFT method were reduced by 81% (to 0.025 ppm) for ^{1}H and 92% (to 0.42 ppm) for ^{13}C NMR chemical shifts. The already mentioned well-established linear scaling approach is consistently outperformed. Furthermore, it is generally applicable to any desired density functional by solely calculating the respective chemical shifts in the data set (which contains only small molecules) once and retraining the model. The performance of the approach is only loosely dependent on the functional or basis set used as baseline method. The use of PBE0/pcSseg-2 was evaluated for three different test sets of experimentally accessible molecules and, despite including systems that differ significantly from those in the training data set, the performance of the Δ_{corr}-ML method remained reasonable and robust as no outliers occurred.

Using a similar procedure, the Δ_{SO}-ML method was introduced to predict relativistic spin-orbit contributions to the ^{1}H and ^{13}C NMR chemical shift caused by near heavy atoms. For that purpose, a data set was constructed including (metal) organic compounds featuring 17 different heavy (main group) elements and the corresponding relativistic contributions as difference $\Delta_{\text{SO}}\delta$ between an SO and an SR calculation. On the test data set, about 85% of this

quantity was recovered by the method for ^{13}C and 70% for ^{1}H NMR. Just like its predecessor, it was designed to be generally applicable and can be based on any desired reasonable DFA. The dependence of $\Delta_{SO}\delta$ on the method it was calculated with was found to be rather small, allowing other relativistic approaches, including the use of ECPs, to serve as *low-level* methods, too. Moreover, a pretrained model can be applied to other baseline methods even without retraining and still yield reasonable results. The suitability of the Δ_{SO}-ML approach was verified by testing it for systems outside its training and test data sets, including the SnS51[199] set and a new collection of organolead compounds.[200] In combination with Δ_{corr}-ML, the deviation to experimental ^{13}C NMR shift data in the new 17HAC set was reduced by about 50%. Both correction methods come at virtually no additional computational costs.

Developing new theoretical methods is only meaningful if they find applications to research questions from other (experimental) fields. In a study on platform molecules with significance for materials science, NMR spectroscopy was used to study the reaction process and investigate the products. An unusually large change in the ^{1}H NMR chemical shift was observed and associated with a substantial structure change that was only incorrectly reproduced by DFT calculations on a conformer ensemble. Here, an only minor correction predicted by the Δ_{corr}-ML method could exclude the intrinsic DFT error as main reason for the deviation, which was finally eliminated with an MD approach.

In conclusion, the scientific advances that have been made in this thesis include new insights in the computational simulation of NMR chemical shifts with different quantum chemical methods (for ^{29}Si and ^{119}Sn), the development of new approaches for the correction of these methods (for ^{1}H and ^{13}C), and the application of the new approaches to help answering open questions in other (experimental) research areas.

In future studies, more electronic structure methods should be investigated for their ability to calculate NMR chemical shifts of other relevant nuclei. Even though many aspects are similar for different NMR nuclei, exceptions are not necessarily obvious, such as the slightly better performance of GGAs than hybrid functionals for ^{29}Si, which was opposed by the findings for ^{119}Sn NMR chemical shift calculations. Furthermore, other electronic structure methods, such as double hybrid functionals or MP2, have not or not sufficiently been investigated on the SiS146[213] and SnS51[199] sets. A comprehensive evaluation of these methods would enable the exploration of the limits of accuracy that are possible for medium-sized systems. Regarding this limit, efficient approximations to MP2 and coupled cluster are of high interest, such as the domain-based local pair natural orbital (DLPNO)[214] technique, which drastically reduces the computational demands of correlated wave function-based methods[215–217] and can also be applied to double hybrid density functionals.[218] These methods have a great potential to yield highly accurate yet affordable NMR parameters and the approach is already available for DLPNO-MP2.[219] Still, a comprehensive assessment and an extension to DLPNO-CC is yet to come.

Besides trying to tackle the electron correlation error in NMR simulation by using more accurate electronic structure methods, the Δ_{corr}-ML method demonstrated an empirical correction approach to that goal. However, both the Δ_{corr}-ML and the Δ_{SO}-ML approach are far from being in their final stage and rather represent a starting point for further

development. From a technical point of view, the underlying ML infrastructure can be optimized using the recent (or future) developments and features in the field in order to achieve the most efficient and robust performance. In addition, the data sets for both methods are designed such that they are straightforwardly expandable by "just" performing the reference CCSD(T) or SO-ZORA calculations for new molecules. It will be interesting to both aim for better generality by including more diverse systems and to specialize the model by including many systems of a particular class of compounds. The Δ_{corr}-ML method can further be extended to other NMR nuclei, such as ^{19}F or ^{29}Si, which, besides the required data, necessitates a reevaluation of the input features and potentially a search for new ones. For ^{1}H and ^{13}C NMR as well, it might be advantageous to evaluate further atomic and molecular representations, both from the three-dimensional structure and from the *low-level* calculation. In the case of the Δ_{SO}-ML approach, the future focus should not only lie on the expansion to other light atoms, but also on including more heavy atoms in the training data, especially transition metals. Finally, both Δ-ML methods can be used to build a new low-cost composite method for efficient NMR shift prediction, such as the one described in reference [220]. Such a method could even be based on a semiempirical method as baseline. Very efficient methods with a possibly poor initial performance will benefit most from the corrections and represent a promising concept for data screening or the calculation of large systems.

Since this thesis focuses on the computation of NMR chemical shifts, not much attention was drawn to spin-spin coupling constants. Nevertheless, they are the key for the simulation of an NMR spectrum. In principle, all studies in this thesis can be extended to coupling constants and there already are respective studies[169,221,222] and ML-based prediction methods.[223] Adapting the Δ_{corr}-ML and Δ_{SO}-ML methods for spin-spin coupling constants would enrich this research area further.

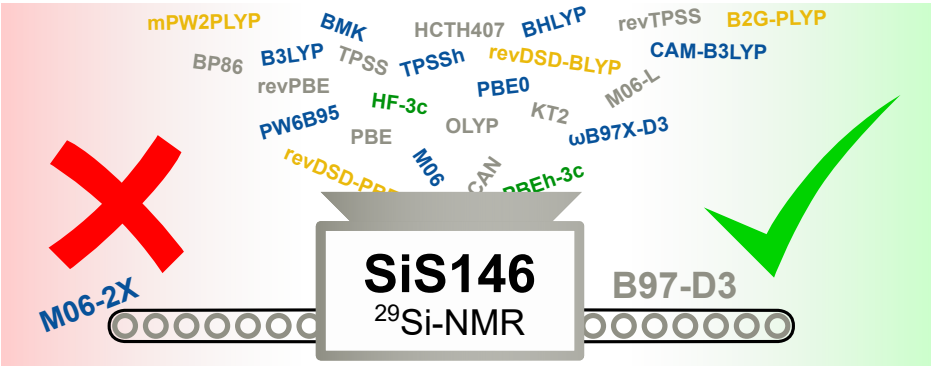
In the bigger picture with regard to the simulation of NMR spectra, more accurate solvation approaches should be studied, especially with focus on the influence of solvent-solute interactions. Explicit treatment of solvent molecules for NMR calculations requires rigorous solvent cluster generation and multilevel schemes but can result in valuable insights.[224,225]

In closing, it can be stated that various contributions to the understanding of NMR chemical shift computation have been achieved in this thesis, including assessment, development, and application of established and new methods. Thus, the acquired knowledge can be of importance for future theoretical and experimental studies. It is further desirable to continue research in the field of NMR parameter prediction and, in particular, for the presented methods in order to drive forward the scientific exploration of this highly relevant analytic technique.

APPENDIX A

Appendix: Comprehensive Benchmark Study on the Calculation of ^{29}Si NMR Chemical Shifts

M. Bursch, T. Gasevic, J. B. Stückrath, and S. Grimme, *Inorg. Chem.* **2021**, *60*, 272–285. DOI: 10.1021/acs.inorgchem.0c02907.



A comprehensive benchmark set termed SiS146 consisting of 146 experimentally determined ^{29}Si NMR chemical shifts is compiled. Diverse structural motifs including hypo-, hypercoordinate and transition metal bound silicon atoms are represented. 26 common density functional approximations are tested for the performance on the calculation of ^{29}Si NMR chemical shifts. Further, the influence of relativistic approaches, choice of AO basis sets, and structure dependence are investigated.

Abstract

A comprehensive and diverse benchmark set for the calculation of ^{29}Si NMR chemical shifts is presented. The SiS146 set includes 100 silicon containing compounds with 146 experimentally determined reference ^{29}Si NMR chemical shifts measured in nine different solvents in a range from –400 to +828 ppm. Silicon atoms bound to main group elements as well as transition metals with coordination numbers of 2-6 in various bonding patterns including multiple bonds and coordinative and aromatic bonding are represented. The performance of various common and specialized density functional approximations including (meta-)GGA, hybrid, and double-hybrid functionals in combination with different AO basis sets and for differently optimized geometries is evaluated. The role of scalar-relativistic effects is further investigated by inclusion of the zeroth order regular approximation (ZORA) method into the calculations. GGA density functional approximations (DFA) are found to outperform hybrid DFAs with B97-D3 performing best with an MAD of 7.2 ppm for the subset including only light atoms ($Z < 18$), while TPSSh is the best tested hybrid functional with an MAD of 10.3 ppm. For ^{29}Si cores in the vicinity of heavier atoms, the application of ZORA proved indispensable. Inclusion of spin-orbit effects into the ^{29}Si NMR chemical shift calculation decreases the mean absolute deviations by up to 74% compared to calculations applying effective core potentials.

A.1 Introduction

Nuclear magnetic resonance (NMR) spectroscopy is one of the most powerful analytical methods used on a daily basis throughout various chemical research fields providing much information on the geometrical and electronic structure of molecules and solids. Besides the most common ^{1}H NMR spectroscopy, other atomic nuclei are of great interest for special research questions, one of which is the ^{29}Si nucleus. The vast applicability of NMR spectroscopy has evoked a general interest in the theoretical calculation of chemical shifts.[30,135,151,152,226,227] As for many computational applications, density functional theory (DFT) methods represent an efficient compromise between accuracy and computational demand also for NMR calculations using the gauge-including atomic orbital (GIAO) method.[137,138,228,229] Still, it is not trivial to find the most efficient choice out of the large amount of existing density functionals, especially for nontheoretical chemists. Benchmark sets can help to gain an overview over the performance of the most common computational methods.

Common organosilicon compounds such as alkyl silanes, silanols, and silyl ethers have already been studied with respect to the calculation of ^{29}Si NMR chemical shifts.[230–235] Nevertheless, more complex molecules containing, e.g., multiple bonded, transition metal bonded, and low-coordinate silicon are mainly studied in an experimental context or in very specialized studies[153,154,236,237] and lack a thorough comprehensive assessment. Compounds like triply bonded silicon compounds have been studied regarding the prediction of ^{29}Si NMR shifts. In 2005[238] and 2011,[239] three different functionals were tested for tetra-, penta-, and hexacoordinate organosilicon compounds highlighting the good performance of the HCTH407[240] density functional approximation (DFA). Related studies by Zhang et al.[241] and Fedorov et al.[242] included relativistic and solvent effects in the computational treatment of tetracoordinate organosilicon compounds.

Regarding the high demand for ^{29}Si NMR chemical shift calculations, a comprehensive benchmark study, covering the chemical versatility of silicon compounds, is yet missing. Therefore, in this work a large benchmark set comprised of 146 experimentally determined ^{29}Si NMR chemical shifts is presented including aromatic, hypervalent, multiple bonded, and tetrahedral coordinated silicon compounds as well as cyclosilanes, silylenes, and organosilicon transition metal complexes. Various common DFAs and basis sets are tested for the calculation of ^{29}Si NMR chemical shifts, and the basis set size and structure dependence of the ^{29}Si NMR chemical shift calculations are investigated. The role of relativistic effects is further evaluated for a subset of silicon compounds with heavy elements in the vicinity of the silicon atom. An empirical linear scaling approach to improve the calculated chemical shifts is tested. These investigations may enable computational and experimental chemists to make a profound DFA choice for a reliable calculation of this property.

A.2 Computational Details

All quantum chemical calculations were conducted applying the ORCA 4.2.1,[243,244] ADF2019.303,[245,246] and xTB 6.3.3[247] program packages. For preoptimization of the structures the GFN2-xTB[27] extended tight-binding method was applied with the GBSA implicit solvation model.[248] Geometries were further optimized at the PBEh-3c[155] composite level (*GRID4*) in combination with the CPCM[112] implicit solvation model for the respective solvents (for details see the SI). For all other noncomposite DFT geometry optimizations, the D4 London dispersion correction scheme[249–251] was applied. Structures were verified as minima on the potential energy surface by the absence of imaginary frequencies below $i\omega = -30$ cm^{-1} in the numerical harmonic frequencies calculation. NMR chemical shielding tensors were calculated with ORCA applying the gauge-including atomic orbital (GIAO) method,[137] CPCM implicit solvation, and either Ahlrichs' def2-TZVP[156] or Jensen's pcSseg-2[136] Gaussian-type orbital (GTO) basis set. Further basis sets used for the basis set convergence study are MINIX,[252,253] def2-mXVP (X = S, TZ), def2-XVP(P) (X = S, TZ, QZ), pcSseg-n (n = 0-3), pc-n (n = 0-3),[162] and cc-pVXZ(-PP)[163,164,254] (X = D, T, Q). The RIJCOSX[255] scheme with *GRID5* and *GRIDX6* options was applied for the chemical shielding tensor calculations, and RIJ was used to accelerate geometry optimizations. Both were conducted with matching def2/J and def2-TZVP/C[256] auxiliary basis sets as implemented in ORCA. For double-hybrid calculations, the grid of the response operator on the right-hand side of the Z-vector equations was set to 2, and no frozen core approximation was applied. Stuttgart-Dresden def2 effective core potentials (ECPs)[254,257] were used if not stated else. For cc-pVXZ(-PP) basis set calculations, SK-MCDHF-RSC[164] ECPs were used. The scalar-relativistic (SR) version of the zeroth order regular approximation (ZORA)[128,258] to account for relativistic effects was used in combination with matching recontracted ZORA/SARC-def2-TZVP basis sets.[259,260] For all SR-ZORA calculations conducted with ORCA, tightened radial grid options were applied (*Grid7*, *IntAcc 10*). For chemical shift calculations with ADF, the Slater-type orbital (STO) ZORA/TZP[165] basis set and the COSMO implicit solvation model[261] were applied. In ADF, both SR- and spin-orbit (SO-)ZORA[129] relativistic approximations were used. All ^{29}Si NMR chemical shifts were calculated with reference to tetramethylsilane (TMS). The reference shielding tensors were recalculated for the respective solvents used for the experimental measurements. A list of all tested method combinations is shown in Table A.1.

Chemical shifts were calculated according to eq (A.1) with $\sigma(^{29}\mathrm{Si}_{\mathrm{TMS}})$ being the ^{29}Si NMR shielding constant for TMS calculated on the structure reoptimized in the respective solvent.

$$\delta(^{29}\mathrm{Si}) = \sigma(^{29}\mathrm{Si}_{\mathrm{TMS}}) - \sigma(^{29}\mathrm{Si}) \tag{A.1}$$

Table A.1: Tested methods, basis sets and relativistic approximations.

	method	basis sets	relativistic treatment
composite	HF	def2-TZVP, pcSseg-2	ECP, SR-ZORA
	HF-3c[253]	MINIX	ECP
	PBEh-3c[155]	def2-mSVP	ECP
(meta-)GGA	PBE[101]	def2-TZVP, pcSseg-2, TZP	ECP, SR/SO-ZORA
	revPBE[177]	def2-TZVP, pcSseg-2, TZP	ECP, SR/SO-ZORA
	OLYP[262]	def2-TZVP, pcSseg-2, TZP	ECP, SR/SO-ZORA
	BP86[102,103]	def2-TZVP, pcSseg-2, TZP	ECP, SR/SO-ZORA
	B97-D3[157]	def2-TZVP, pcSseg-2	ECP, SR-ZORA
	TPSS[104]	def2-TZVP, pcSseg-2	ECP, SR-ZORA
	revTPSS[263,264]	def2-TZVP, pcSseg-2	ECP, SR-ZORA
	M06-L[176]	def2-TZVP, pcSseg-2	ECP, SR-ZORA
	KT2[160]	def2-TZVP, pcSseg-2, TZP	ECP, SR/SO-ZORA
	SCAN[265]	def2-TZVP, pcSseg-2	ECP, SR-ZORA
	HCTH407[240]	def2-TZVP, pcSseg-2	ECP, SR-ZORA
(meta-)hybrid	PBE0[109]	def2-TZVP, pcSseg-2, TZP	ECP, SR/SO-ZORA
	B3LYP[106–108]	def2-TZVP, pcSseg-2, TZP	ECP, SR/SO-ZORA
	CAM-B3LYP[266]	def2-TZVP, pcSseg-2	ECP, SR-ZORA
	BHLYP[267]	def2-TZVP, pcSseg-2	ECP, SR-ZORA
	BMK[268]	def2-TZVP, pcSseg-2	ECP, SR-ZORA
	ωB97X-D3[269–271]	def2-TZVP, pcSseg-2	ECP, SR-ZORA
	PW6B95[272]	def2-TZVP, pcSseg-2	ECP, SR-ZORA
	TPSSh[158,159]	def2-TZVP, pcSseg-2	ECP, SR-ZORA
	M06[161]	def2-TZVP, pcSseg-2	ECP, SR-ZORA
	M06-2X[161]	def2-TZVP, pcSseg-2	ECP, SR-ZORA
double-hybrid	revDSD-PBEP86[273–275]	def2-TZVP, pcSseg-2	ECP
	revDSD-BLYP[275,276]	def2-TZVP, pcSseg-2	ECP
	B2GP-PLYP[277]	def2-TZVP, pcSseg-2	ECP
	mPW2PLYP[278]	def2-TZVP, pcSseg-2	ECP

A.3 Results and Discussion

A.3.1 General Considerations

The calculation of chemical shifts involves some difficulties, and a few general computational aspects have to be considered. These include the role of the geometry optimization method (equilibrium structure), the application of implicit or explicit solvation models, the influence of molecular conformations, and finally the choice of the best density functional

approximation (DFA). The quality of geometries obtained with a certain DFA must not necessarily coincide with the quality of properties calculated with the same. Further, geometry optimizations are computationally demanding, and thus more sophisticated DFAs are often not applicable. Multilevel approaches are a widely accepted compromise. Efficient DFT based composite methods like PBEh-3c[155] or B97-3c[279] yield excellent molecular geometries at low to moderate cost and are generally recommended.

Solvation effects should be considered when highly polar, charged, or structurally affected molecules are investigated. Specifically, ^{1}H NMR is prone to the account of solvation. Nevertheless, electrostatic effects are often sufficiently described by implicit solvation models like COSMO[114] or PCM.[280] Explicit solvation should be considered, if strongly bound solute-solvent complexes are to be expected. Otherwise, explicit solvation may limit the applicability of some methods due to the increased computational demand, in particular, regarding sampling. Furthermore, the role of molecular conformations can become important if structural motifs, that have large a impact on the chemical environment of the core under investigation, are thermodynamically accessible under the given reaction conditions. Nevertheless, in common, saturated compounds, this influence is small and far less pronounced compared to more exposed cores like ^{1}H. Fully automated conformer and rotamer sampling tools like the recently developed CREST[25] program can be applied to find such conformations, and thus critical cases could be investigated individually. Finally, these points highly depend on the system under investigation and have to be considered on a case by case basis. Nevertheless, the routine consideration of molecular conformations and application of implicit solvation models is recommended.

Finally, the most crucial point that has to be assessed is the choice of the DFA for the computation of chemical shielding tensors (chemical shifts). At this point, for general use, not only accuracy and numerical precision but also the robustness and applicability of the method should be considered. To do so, a sufficiently large and representative benchmark set with a consistent level of high quality reference data has to be compiled to test a wide range of commonly applied DFAs.

A.3.2 Benchmark Set

The benchmark set consists of 146 experimentally determined ^{29}Si NMR chemical shifts of 100 diverse Si-containing compounds with chemical shifts in a range from –400 to +828 ppm (Figure A.1) and molecule sizes between 5 and 234 atoms. Solvents include chloroform, dichloromethane, benzene, chlorobenzene, bromobenzene, toluene, acetonitrile, cyclohexane, and THF. Coordination and bonding patterns of the Si-atom range from coordination numbers of 2 to 6 including single, double, triple, coordinative, and aromatic bonding patterns. Si bonds to main group elements as well as transition metals are represented. Selected examples are depicted in Figure A.1, and a detailed overview of all compounds can be found in the Supporting Information. Molecular geometries are optimized consistently at the PBEh-3c(CPCM) level of theory applying the CPCM implicit solvation model for the respective solvent of the experimental ^{29}Si NMR chemical shift measurement. The benchmark set is

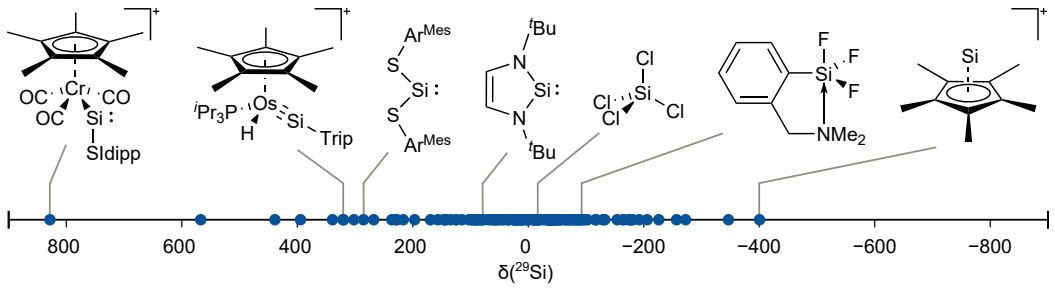


Figure A.1: Scale of investigated ^{29}Si NMR chemical shifts in ppm.

divided into two categories: (i) a set containing only elements up to argon ($Z = 18$), further denoted as "light" subset *SiS-L* (compounds **1-55**), and (ii) a set of structures containing elements with $Z > 18$ adjacent to the Si atom, further denoted as "heavy" subset *SiS-H* (compounds **56-100**, $SiCl_4$ is also included as it is known to be prone to spin-orbit coupling effects).

A.3.3 Benchmark of Density Functional Approximations for the Calculation of ^{29}Si NMR Chemical Shifts

Overall, the performance of 25 conventional DFAs, 11 (meta-)GGA, 10 (meta-)hybrid, and four double-hybrid functionals was assessed. Further, the two composite methods HF-3c and PBEh-3c as well as the Hartree-Fock theory were tested. All DFAs were applied in conjunction with either the def2-TZVP or pcSseg-2 basis sets with ECPs, and explicit relativistic ZORA treatment was evaluated.

The Light Atom Subset *SiS-L*

As for def2-TZVP and pcSseg-2 with ECPs, very similar trends and qualitative results are observed, and only the def2-TZVP values will be discussed in the following. For both basis sets, no typical "Jacob's ladder"[100,281] behavior (better performance with increased theoretical level) is observed. In fact, the lower level (meta-)GGA class functionals yield better results (class mean absolute deviation, $MAD_{GGA} = 10.9$ ppm) than the tested (meta-)hybrid ($MAD_{hybrid} = 16.7$ ppm) functionals (Figure A.2, Table A.2). Specifically, the GGA type B97-D3 (MAD = 7.2 ppm, $R^2 = 0.993$) and HCTH407 (MAD = 8.1 ppm, $R^2 = 0.991$) functionals show good results for the *SiS-L* subset. The KT2 functional that was specifically designed for NMR chemical shielding calculations also performs well with an MAD of 8.8 ppm and $R^2 = 0.989$. KT2 already yielded good results for ^{29}Si NMR chemical shifts in previous studies for a smaller set of molecules.[282] For B97-D3/def2-TZVP, bad results are obtained for $[(Idipp)(Me)Si{=}Si(Idipp)]^{+}$[283] (**45**, Idipp = $C[N(C_6H_3\text{-}2,6\text{-}i\text{Pr}_2)CH]_2$) with a deviation of 28.9 ppm for the silicon atom (shift 89; exp. shift = 115.2 ppm) in the formal

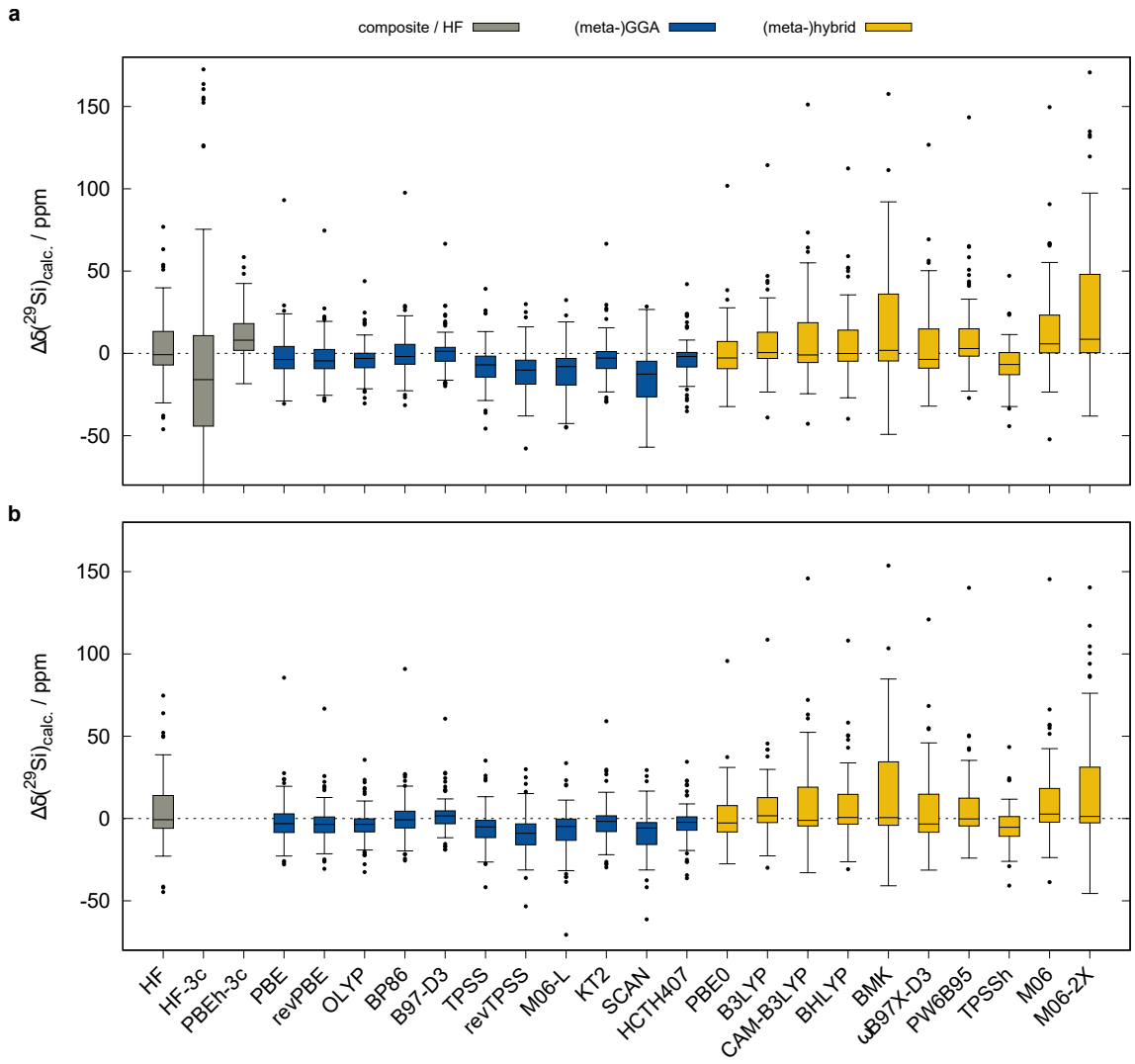


Figure A.2: Box plots for all assessed composite/HF, (meta-GGA), and hybrid DFAs applying the def2-TZVP (**a**) or pcSseg-2 (**b**) basis set (except for composite methods that use specialized basis sets) for the *SiS-L* subset. The central black line represents the median value, and the whiskers represent the most distant points within 1.5 times the interquartile range. Black dots represent outliers outside this threshold.

oxidation state of II. In this case the ^{29}Si NMR chemical shift may be quite sensitive to deviations in the double bond length from the effective one in solution. Further, partly dynamic hyperconjugative effects involving the methyl ligand may influence the chemical shift of the respective silicon atom. The largest deviation is observed for the silylene silicon atom in Kira's cyclosilylene 2,2,5,5-tetrakis(trimethylsilyl)silacyclopentane-1,1-diyl[284] (**49**) with a deviation of 66.6 ppm (shift 94; exp. shift = 567.4 ppm) which may also correspond to deviations of the silylene bond angle compared to an effective angle in the experiment. In

Table A.2: Mean deviation (MD), Mean absolute deviation (MAD), linear scaled MAD (MAD_{scaled}) in ppm, and determination coefficient R^2 for all tested methods with the def2-TZVP and pcSseg-2 basis sets on the *SiS-L* subset.

SiS-L	def2-TZVP				pcSseg-2			
	MD	MAD	R^2	MAD_{scaled}[a]	MD	MAD	R^2	MAD_{scaled}
HF	4.46	14.34	0.986	9.62	4.68	14.17	0.985	9.69
HF-3c[b]	−8.39	45.07	0.863	34.56	–	–	–	–
PBEh-3c[c]	10.49	12.95	0.986	10.25	–	–	–	–
PBE	−1.77	10.33	0.991	7.13	−1.54	9.26	0.991	7.06
revPBE	−2.86	9.70	0.991	7.06	−2.74	8.82	0.991	7.02
OLYP	−3.59	8.36	0.992	6.94	−3.68	7.98	0.992	7.01
BP86	0.25	9.49	0.992	6.57	0.66	8.45	0.992	6.60
B97-D3	1.31	7.19	0.993	6.24	1.88	6.89	0.993	6.43
TPSS	−7.17	10.77	0.989	8.34	−5.41	9.23	0.991	7.49
revTPSS	−10.71	13.78	0.986	9.86	−9.16	12.27	0.989	8.75
M06-L	−11.86	14.58	0.980	12.26	−7.18	10.91	0.986	9.90
KT2	−2.68	8.80	0.989	8.05	−1.92	8.14	0.989	7.96
SCAN	−16.01	18.81	0.975	14.07	−8.48	11.51	0.987	9.54
HCTH407	−2.94	8.09	0.991	7.32	−2.94	7.80	0.991	7.41
PBE0	0.35	12.05	0.992	7.21	0.61	11.14	0.992	7.14
B3LYP	4.94	11.33	0.993	6.41	5.17	10.84	0.993	6.51
CAM-B3LYP	7.93	16.49	0.990	8.62	8.18	15.63	0.990	8.48
BHLYP	6.24	13.34	0.992	7.65	6.52	12.78	0.992	7.59
BMK	13.78	22.45	0.982	11.58	12.77	21.35	0.984	10.96
ωB97X-D3	4.72	16.70	0.988	9.32	5.00	15.86	0.988	9.24
PW6B95	9.20	13.61	0.990	8.15	4.25	11.87	0.991	6.92
TPSSh	−5.90	10.30	0.990	7.98	−4.38	8.99	0.992	7.30
M06	13.43	18.80	0.989	8.88	9.07	15.43	0.992	7.51
M06-2X	26.25	32.03	0.966	16.65	17.26	25.67	0.979	12.51

[a] The linear scaling approach principally yields MD = 0 and does not influence the R^2 values.
[b] MINIX basis.
[c] def2-mSVP basis.

both cases, the discussed ^{29}Si NMR chemical shifts seem to be problematic for a majority of the tested methods. The generally worst performing GGA is SCAN with an MAD of 18.8 ppm and $R^2 = 0.975$. From the (meta-)hybrid class of functionals, TPSSh performs best (MAD = 10.3 ppm, $R^2 = 0.990$) but still significantly worse than B97-D3. M06-2X yields by far the worst results in this class with a very poor MAD of 32.0 ppm ($R^2 = 0.966$). The popular B3LYP and PBE0 functionals yield reasonable MADs of 11.3 ($R^2 = 0.993$) and 12.1 ppm ($R^2 = 0.992$), respectively. For the *SiS-L* subset, the pcSseg-2 basis set generally yields slightly improved results compared to def2-TZVP. The most remarkable improvement is

obtained for SCAN, decreasing the MAD from 18.8 to 11.5 ppm and increasing R^2 to 0.987.

Double-hybrid density functional approximations (DH DFAs) that were shown to perform well for the calculation of ^{1}H, ^{13}C, and ^{31}P NMR properties were also tested. Unfortunately, the drastically increased memory demand of the calculations proved problematic, specifically for the larger structures included in the SiS146. Nevertheless, they only yield no to small improvement over the well performing (meta-)GGA DFAs, even though a direct comparison should be made with care considering the smaller number of data points. For the *SiS-L* subset the best DH DFA (DSD-BLYP) yields an MAD of 8.3 ppm ($R^2 = 0.992$), and mPW2PLYP as the worst yields an MAD of 11.5 ppm ($R^2 = 0.961$). All data for the DH DFAs can be found in the Supporting information and is not further discussed.

Linear Scaling

Based on a sufficiently large data set, an empirical, linear scaling approach can be applied to derive method dependent scaling parameters α and β that represent the slope and the intercept of the linear best fit with respect to the experimental data. As shown in previous works,[32,37,285–287] α and β can be applied according to eq (A.2) to correct for systematic errors.

$$\delta_{calc,scaled} = \frac{(\delta_{calc} - \beta)}{\alpha} \tag{A.2}$$

We applied this approach to all tested functionals on the *SiS-L* subset. Overall, the MAD can be reduced by 0.8 ppm (8%, KT2) to 15.4 ppm (53%, M06-2X) with 31% improvement on average for the def2-TZVP basis set (Table A.2). The effect of linear scaling for B97-D3 and M06-2X is depicted in Figure A.3.

Basis Set Dependence

As the choice of the applied basis set always inherits the question of efficiency and applicability, we investigated the basis set dependence of the chemical shifts on the *SiS-L* subset with the Ahlrichs def2-X (X = SVP, mSVP, TZVP, mTZVP, TZVPP, QZVP, QZVPP), Jensen's polarization-consistent pcSseg-*n* and pc-*n* (n = 0-3), and Dunning's correlation-consistent cc-pVXZ (X = D, T, Q) basis set families for B97-D3. For all tested families, a rapid convergence is observed regarding the mean absolute deviations and determination coefficients R^2 on the *SiS-L* subset (Figure A.4). Increasing the basis set beyond triple-ζ quality does not improve the results significantly and leads to drastically increased computation times. Thus, the def2-TZVP and pcSseg-2 basis sets can be recommended as both offer a good compromise between speed and performance. Nevertheless, for pc-*n* and pcSseg-*n* basis sets beyond $n = 2$, increasing SCF convergence problems were observed. A better performance of Dunning's correlation consistent cc-pVXZ(-PP) basis sets compared to def2-X and pcSseg-*n* basis sets, as reported for ^{13}C NMR and ^{1}H NMR,[35] was not observed. Further, it should be noted that specifically the pc-*n* and pcSseg-*n* basis sets are only available up to Kr ($Z = 36$),

and hence for the heavier elements, the corresponding def2-X basis set was applied. The small MINIX basis set produces large errors and a low determination coefficient of 0.865.

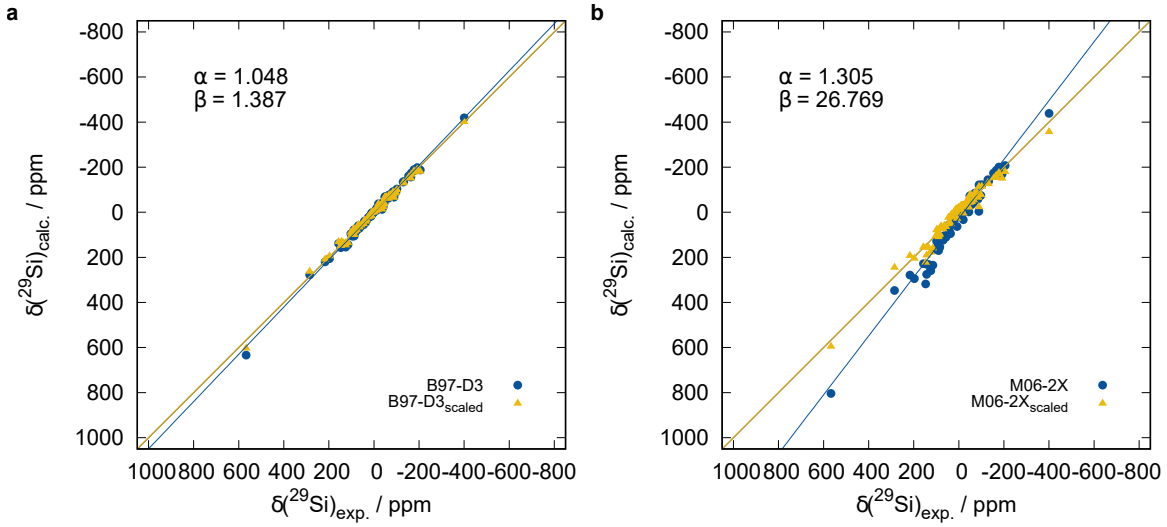


Figure A.3: Correlation plots with respect to the experimental chemical shifts for B97-D3 (**a**) and M06-2X (**b**) before and after linear scaling correction applying the def2-TZVP basis set on the *SiS-L* subset.

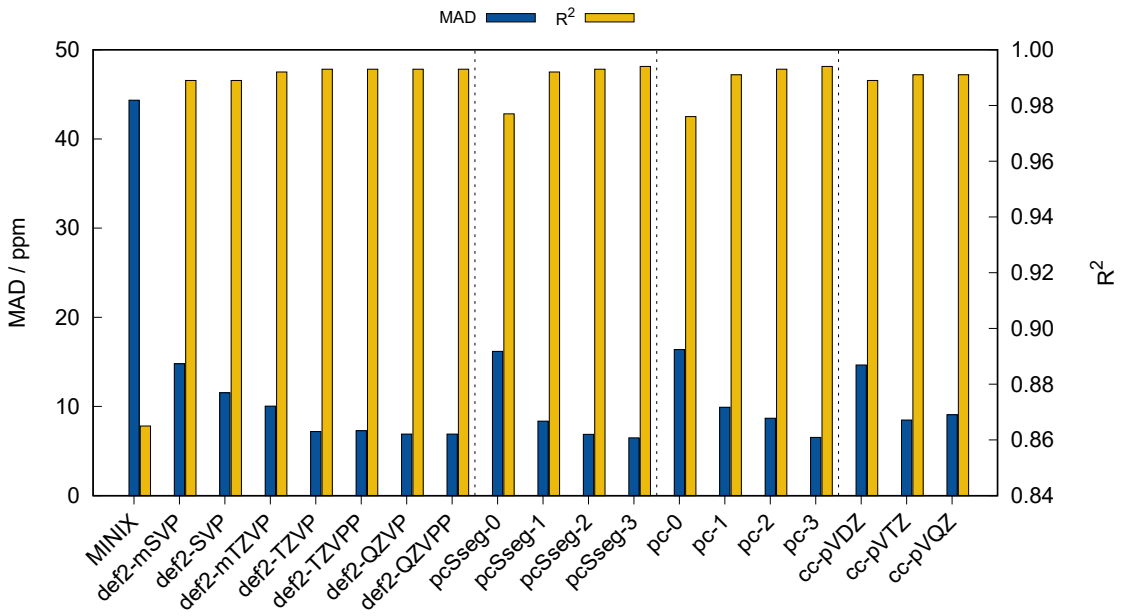


Figure A.4: MAD in ppm and determination coefficient R^2 for the *SiS-L* subset calculated with B97-D3 and various AO basis sets.

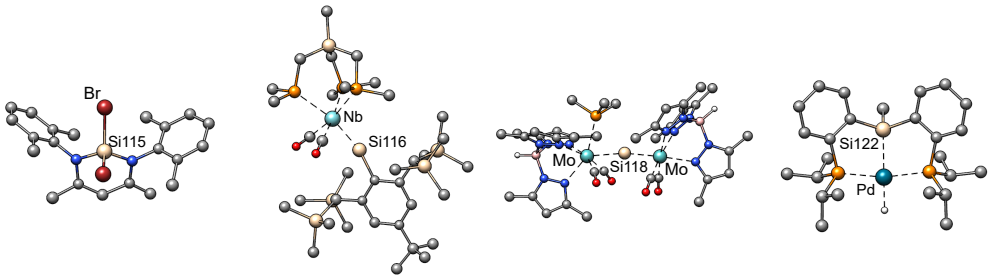
A.3.4 The Heavy Atom Subset *SiS-H* and Relativistic Effects

The vicinity of the relatively light silicon atom to a heavier atom requires consideration of relativistic effects on its chemical shift. Generally, such so-called heavy atom effects on the light atom (HALA) include scalar or spin-free relativistic (SFR) and the often dominant spin-orbit (SO) HALA effects[48,54,288–290] that can have a huge impact on the quality of the chemical shift calculation.[291] Popular treatments to account for HALA effects are (SO-)ECPs or the widely used (SO-)ZORA[129,292] approximation. As ECPs and the scalar-relativistic (SR) version of ZORA are the most common and easily accessible methods for the calculation of NMR shielding constants, we herein focus on to what extent scalar-relativistic treatments can improve the results for ^{29}Si NMR chemical shift calculations. Nevertheless, it must be stressed that particularly pronounced SO-HALA effects require more sophisticated theory, and such cases should be assessed individually.

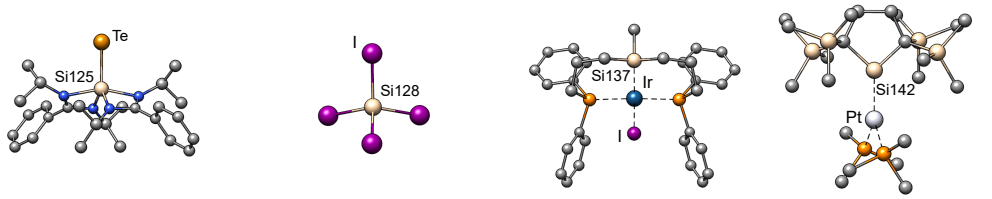
Besides the ECP and SR-ZORA calculations for all tested DFAss we applied the two-component SO-ZORA for seven GGA and hybrid functionals for the *SiS-H* subset consisting of 44 molecules with heavy atoms ($Z > 18$) in direct neighborhood to the silicon atom as well as for $SiCl_4$ (Table A.3). Selected structures and calculated ^{29}Si NMR chemical shifts are depicted in Figure A.5.

For the *SiS-H* subset, HALA effects are expected to play an important role for the ^{29}Si NMR chemical shifts. To investigate the HALA effects in more detail, the commonly used SR-ZORA and SO-ZORA approaches were further evaluated in conjunction with the Slater-type orbital TZP basis set as implemented in the ADF2019.303 program package[245,246] (Table A.3).

For the *SiS-H* subset, no improvement upon application of scalar relativistic ZORA in conjunction with the def2-TZVP basis set is observed (Figure A.6a,b). Instead mostly a slightly increased MAD is observed except for the hybrid DFAs M06, PW6B95, ωB97X-D3, BMK, and CAM-B3LYP (Table A.3). SR-ZORA with the TZP basis set also yields no clear trend of improvement, increasing the MAD for KT2 by 5.1 ppm which may be attributed to disturbed error compensation due to the construction of the KT2 functional being explicitly fitted to NMR shielding constants.[160] For B3LYP on the other hand, an MAD improvement by 3.1 ppm is observed. While the effect of both SR-ZORA approaches is generally small, the introduction of SO-ZORA drastically decreases the MADs by 30.1 ppm (71%, KT2) to 36.0 ppm (68%, B3LYP) with respect to the ECP/def2-TZVP results. Thus, the SO-ZORA treatment is strongly recommended to obtain reasonable ^{29}Si NMR chemical shifts for compounds with expected HALA effects. An example with tremendous spin-orbit relativistic effects is tetraiodosilane (**83**, SiI_4) with an experimental shift of –346.6 ppm. Here ECP and SR-ZORA based PBE0 calculations yield drastically wrong chemical shifts of 91.9 to 105.0 ppm, and only the SO-ZORA approach yields a good result of –343.1 ppm. Overall, robust GGA DFAs like BP86, revPBE or OLYP can be recommended for the use with SO-ZORA and a TZP basis set for the calculation of ^{29}Si NMR chemical shifts of HALA affected silicon atoms (*SiS-H* $MAD_{SO\text{-}ZORA\text{-}BP86/TZP} = 12.3$ ppm, $R^2 = 0.993$). The PBE0 hybrid functional may represent a valuable alternative if explicit Fock-exchange is needed (*SiS-H* $MAD_{SO\text{-}ZORA\text{-}PBE0/TZP} = 13.6$ ppm, $R^2 = 0.995$). For SO-ZORA-BP86/TZP, the largest remaining deviation on the *SiS-H* subset is observed for



	70	**71**	**73**	**77**
$\delta_{exp.}$	−55.1 ppm	267.8 ppm	438.9 ppm	60.8 ppm
$\delta_{ECP/def2\text{-}TZVP}$	−16.2 ppm	293.4 ppm	523.5 ppm	68.4 ppm
$\delta_{SR\text{-}ZORA/def2\text{-}TZVP}$	−13.6 ppm	304.6 ppm	536.3 ppm	48.0 ppm
$\delta_{SR\text{-}ZORA/TZP}$	−9.4 ppm	271.3 ppm	514.0 ppm	67.0 ppm
$\delta_{SO\text{-}ZORA/TZP}$	−48.7 ppm	278.2 ppm	501.6 ppm	64.5 ppm



	80	**83**	**92**	**97**
$\delta_{exp.}$	−116.5 ppm	−346.6 ppm	25.4 ppm	394.6 ppm
$\delta_{ECP/def2\text{-}TZVP}$	−52.8 ppm	91.9 ppm	105.9 ppm	382.9 ppm
$\delta_{SR\text{-}ZORA/def2\text{-}TZVP}$	−50.7 ppm	89.6 ppm	113.1 ppm	371.7 ppm
$\delta_{SR\text{-}ZORA/TZP}$	−46.2 ppm	105.0 ppm	102.9 ppm	364.3 ppm
$\delta_{SO\text{-}ZORA/TZP}$	−111.0 ppm	−343.1 ppm	43.5 ppm	400.9 ppm

Figure A.5: Selected molecules with heavy atoms in the vicinity of the Si atom. Chemical shifts were calculated with the PBE0 functional. The shift that deviates least from the experimental reference is marked in blue. Hydrogen atoms attached to carbon are omitted for clarity.

Filippou's cationic chromio silylene complex $[(\eta^5\text{-}C_5Me_5)(CO)_3Cr{-}Si(SIdipp)]^+$ (**58**, SIdipp = 1,3-bis(2,6-diisopropylphenyl)imidazolidin-2-ylidene)[293] with a calculated shift of 775.9 ppm and a deviation of 52.7 ppm. Further, a relatively large deviation is observed for the iridium complex [IrIH(biPSi)] (**92**, biPSi = κ-*P,P,Si*−Si(Me){$(CH_2)_3PPh_2$}$_2$)[294] with a calculated chemical shift of 70.3 ppm and a deviation of 44.9 ppm. For the *SiS-L* subset, the effect of SO-ZORA is much smaller with 1% (KT2) to 25% (BP86) improvement by SO-ZORA compared to the ECP results (Figure A.6c,d). Thus, nonrelativistic chemical shift calculations on non-HALA affected silicon atoms should be sufficiently accurate for most purposes. For the *SiS-L* subset, a slight deterioration of the results was observed applying the SR-ZORA/def2-TZVP approach. Even though an empirical linear scaling correction cannot be applied reasonably to the ECP data due to strong scattering of the data points and very low determination coefficients (cf. Figure A.6b,d), it can further improve the results of the ZORA relativistic calculations.

Table A.3: Mean absolute deviation (MAD) in ppm and determination coefficient (R^2) for several relativistic approaches for the *SiS-H* subset with expected strong HALA effects.

SiS-H	ORCA				ADF			
	ECP/def2-TZVP		SR-ZORA/def2-TZVP[a]		SR-ZORA/TZP[b]		SO-ZORA/TZP[b]	
	MAD / ppm	R^2	MAD / ppm	R^2	MAD / ppm	R^2	MAD / ppm	R^2
HF	63.01	0.794	70.75	0.778	–	–	–	–
HF-3c[c]	80.17	0.827	–	–	–	–	–	–
PBEh-3c[d]	52.72	0.840	–	–	–	–	–	–
PBE	44.32	0.859	46.94	0.858	44.13	0.852	12.10	0.993
revPBE	43.86	0.860	46.80	0.858	44.40	0.851	11.58	0.994
OLYP	43.15	0.859	46.43	0.857	46.09	0.847	12.44	0.994
BP86	45.00	0.859	45.56	0.858	44.45	0.852	12.25	0.993
B97-D3	44.80	0.858	48.63	0.856	–	–	–	–
TPSS	43.87	0.856	48.86	0.863	–	–	–	–
revTPSS	45.87	0.853	47.43	0.862	–	–	–	–
M06-L	50.68	0.844	51.38	0.864	–	–	–	–
KT2	42.39	0.860	47.52	0.859	47.49	0.843	12.30	0.994
SCAN	49.83	0.847	58.13	0.872	–	–	–	–
HCTH407	44.03	0.856	47.68	0.852	–	–	–	–
PBE0	49.25	0.863	51.44	0.860	47.50	0.853	13.60	0.995
B3LYP	52.63	0.861	53.80	0.859	49.56	0.853	16.66	0.997
CAM-B3LYP	63.14	0.862	63.00	0.862	–	–	–	–
BHLYP	61.48	0.861	63.84	0.857	–	–	–	–
BMK	69.85	0.863	67.16	0.861	–	–	–	–
ωB97X-D3	58.73	0.863	57.69	0.863	–	–	–	–
PW6B95	57.80	0.866	56.78	0.838	–	–	–	–
TPSSh	43.97	0.858	48.62	0.864	–	–	–	–
M06	63.93	0.861	58.03	0.839	–	–	–	–
M06-2X	83.79	0.856	84.56	0.820	–	–	–	–

[a] ZORA-def2-TZVP and/or SARC-def2-TZVP recontracted basis sets within ORCA.
[b] ZORA/TZP basis set within ADF.
[c] MINIX basis.
[d] def2-mSVP basis.

A.3.5 Structure Dependence of ^{29}Si NMR Calculated Chemical Shifts

To estimate the influence of the geometry optimization level on the calculated ^{29}Si NMR chemical shift, ^{29}Si NMR chemical shifts are calculated for five structures from each, the *SiS-L* and *SiS-H* subsets, that were reoptimized on different theoretical levels (Figure A.7). These are, B97-3c from the 3c-composite method family, the GFN2-xTB extended tight-binding method, the GFN-FF[28] generic force field, TPSS-D4/def2-TZVP, and B3LYP-D4/def2-SVP.

Overall, the MAD varies only slightly depending on the geometry optimization method when changing the DFA. The largest deviation in the DFA class is observed on B3LYP-D4/def2-SVP structures with an MAD change of 1.9 ppm for PBE0 (2.1 ppm for PBE) compared to the results at the reference PBEh-3c structures. Application of the semiempirical

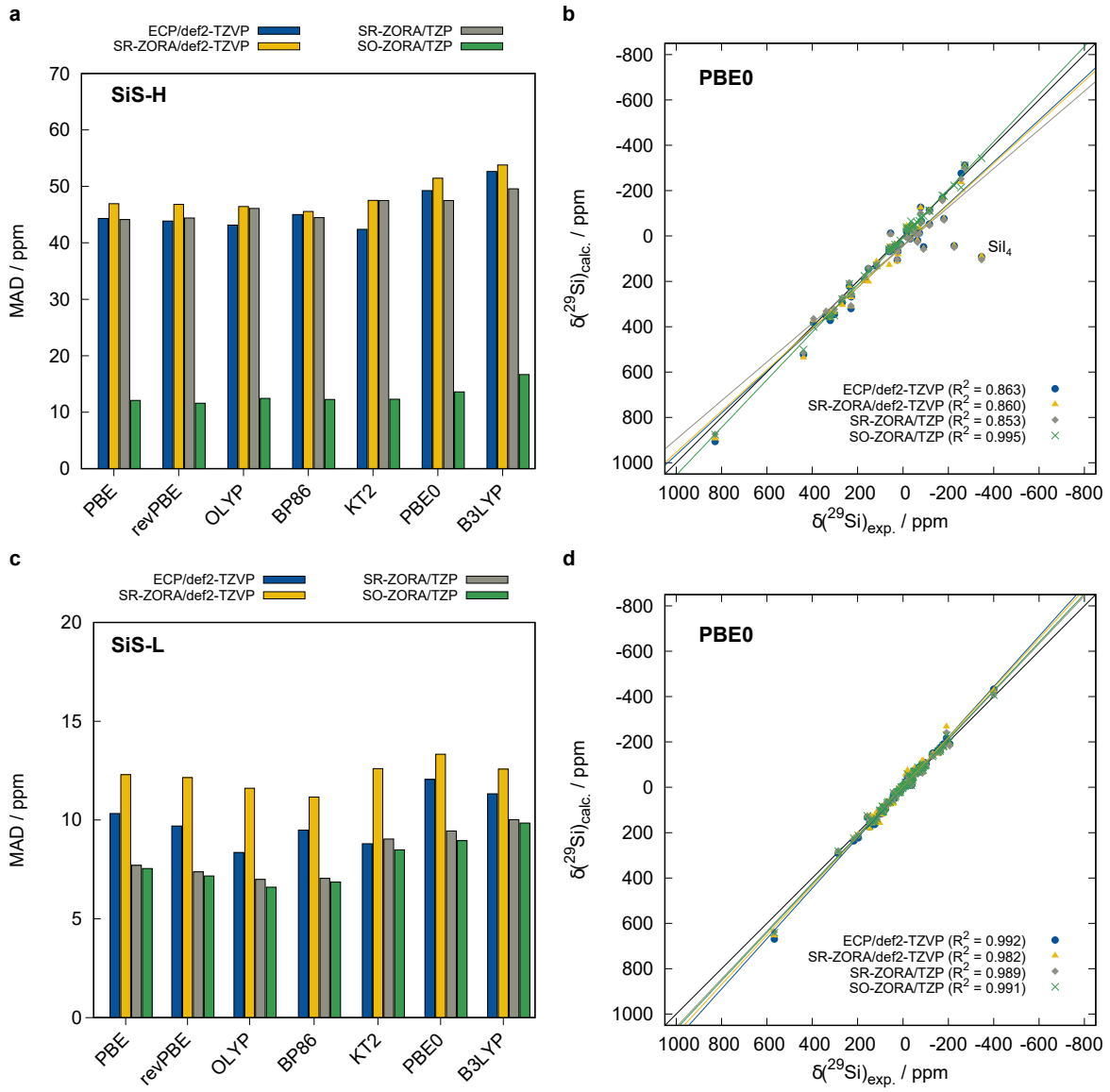


Figure A.6: Mean absolute deviations for various relativistic approaches for the **a**) *SiS-H* and the **c**) *SiS-L* subsets. Correlation plots of experimentally obtained ^{29}Si NMR chemical shifts and those calculated with the PBE0 functional combined with different relativistic approaches **b**) *SiS-H* and **d**) *SiS-L*.

quantum mechanical (SQM) extended tight-binding method GFN2-xTB for geometry optimization results in significantly increased MADs, i.e., by 15.8 ppm for PBE0 and 11.9 ppm for PBE, respectively, more than doubling the MAD for the evaluated subset. Even worse results are obtained for geometries optimized with the GFN-FF generic force field (FF) which mainly results from deviations in bonding angles of low coordinated silicon atoms. Nevertheless,

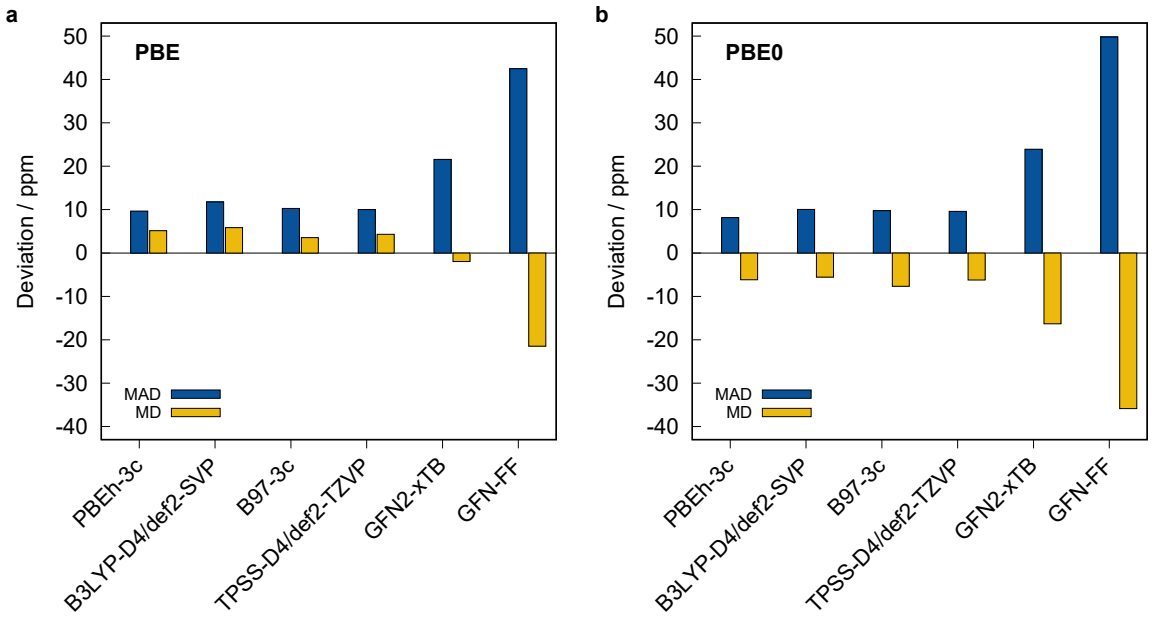


Figure A.7: Deviations for **a**) PBE and **b**) PBE0 ^{29}Si NMR chemical shifts calculated on differently optimized geometries for a selection of ten representative structures applying the SO-ZORA treatment and the TZP basis set. All deviations are in ppm.

the errors introduced by geometry optimization with a low-cost SQM or FF method can be partly compensated by linear scaling (Figure A.8) which is strongly recommended in this case. For the SO-ZORA/BP86 DFA applied on GFN-FF structures, the *SiS-H* subset MAD can be reduced by 19.8 ppm (47%) resulting in an MAD_{scaled} of 22.0 ppm (Figure A.8d). Thus, even though an optimization at the low-cost DFT level is recommended to obtain the best results, substitution of the computationally demanding geometry optimization by SQM or FF methods may be applied in combination with a linear scaling correction to obtain reasonable results. This approach may be applied to efficiently calculate ^{29}Si NMR chemical shifts of large structures. Respective scaling parameters for selected, well performing methods based on the SiS146 set for use on SQM or FF method optimized geometries are provided in Table A.4. Structures that are chemically destroyed upon optimization with the respective SQM/FF method were excluded from the evaluation.

A.4 Conclusion

The ^{29}Si NMR spectroscopy is a highly valuable analytical tool for academic and industrial researchers due to the well-defined ^{29}Si core with a spin of $1/2$. It is sensitive to details of the chemical surroundings, and the high range of observed chemical shifts is suited as a reliable indicator for chemical coordination and bonding patterns. Thus, an efficient computational prediction of these chemical shifts is highly desirable. Nevertheless, several previous studies

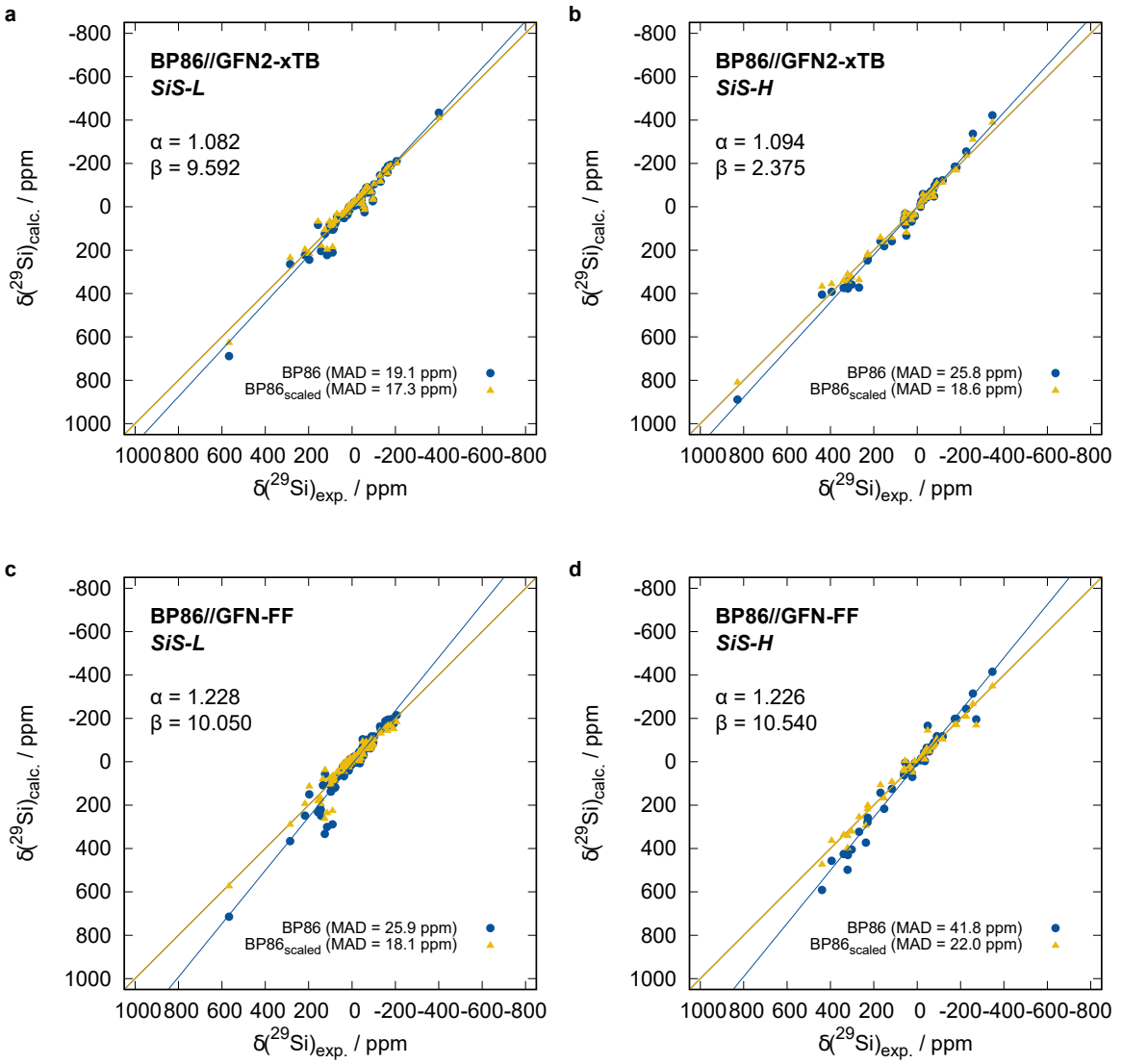


Figure A.8: Uncorrected and linear scaling corrected ^{29}Si NMR chemical shifts (SO-ZORA/BP86/TZP) at GFN2-xTB and GFN-FF structures for the *SiS-L* and *SiS-H* subsets. Respective MADs are given in parentheses.

only covered a narrow range of already well studied structural motifs or only a small number of methods regarding the calculation of ^{29}Si NMR chemical shifts.

Therefore, a comprehensive benchmark set termed SiS146 containing 146 experimentally determined ^{29}Si NMR chemical shifts for 100 small to large sized compounds with diverse bonding situations at the silicon center was composed. For this benchmark set, the performance of 26 common DFT based methods with several AO basis sets was tested for the

Table A.4: Scaling parameters derived from the SiS146 set for ^{29}Si NMR chemical shift calculations on GFN2-xTB and GFN-FF geometries.

	GFN2-xTB				GFN-FF			
SiS-L	MAD	MAD_{scaled}	α	β	MAD	MAD_{scaled}	α	β
BP86/def2-TZVP	20.75	17.17	1.130	11.189	29.33	18.86	1.282	11.249
B97-D3/def2-TZVP	21.00	17.21	1.094	11.754	25.99	17.99	1.234	11.765
PBE0/def2-TZVP	22.18	17.49	1.159	13.667	32.42	20.41	1.323	12.034
B3LYP/def2-TZVP	26.34	18.71	1.183	17.130	32.82	19.95	1.342	16.740
BP86/pcSseg-2	20.57	17.23	1.116	11.244	28.27	18.78	1.267	11.491
B97-D3/pcSseg-2	20.88	17.53	1.080	11.958	25.19	17.90	1.220	12.207
PBE0/pcSseg-2	21.61	17.32	1.145	13.623	31.18	20.39	1.308	12.261
B3LYP/pcSseg-2	26.07	18.85	1.170	17.034	31.86	19.92	1.328	16.950
SO-ZORA-BP86/TZP	19.05	17.25	1.082	9.592	25.92	18.10	1.228	10.050
SO-ZORA-PBE0/TZP	20.33	18.01	1.108	13.261	31.67	22.11	1.227	9.037
	GFN2-xTB				GFN-FF			
SiS-H	MAD	MAD_{scaled}	α	β	MAD	MAD_{scaled}	α	β
SO-ZORA-BP86/TZP	25.79	18.59	1.094	2.375	41.78	21.96	1.226	10.540
SO-ZORA-PBE0/TZP	26.50	16.53	1.142	9.087	43.14	26.76	1.250	12.297

calculation of ^{29}Si NMR chemical shifts. It is sectioned into two subsets: the *SiS-L* containing only compounds composed of light elements ($Z < 18$), and the *SiS-H* subset including compounds with heavier elements. For the *SiS-H* subset, the role of explicit treatment of relativistic effects (SR/SO-ZORA) is evaluated. Further, the basis set convergence and geometry optimization method dependence are investigated exemplarily.

Overall, the B97-D3 GGA functional performs best (with the def2-TZVP basis set) with an MAD of 7.2 ppm and a high determination coefficient of 0.993 for the *SiS-L* subset. Generally, the tested (meta-)GGA DFAs outperform hybrid functionals with TPSSh producing the smallest MAD of 10.3 ppm in the hybrid class. These results do not follow a trend of improved results for higher rungs of Jacob's ladder that is known for, i.e., thermochemical calculations but is in agreement with prior studies on other elements.[39] The worst results are obtained with the M06-2X DFA, yielding a high MAD of 32.0 ppm. Application of Jensen's pcSseg-2 basis set generally yields slightly smaller (mostly about 1 ppm) MADs compared to the def2-TZVP basis set but is not generally available for heavier elements. For the worse methods, a linear scaling approach may be applied to significantly improve the results. For both the def2-X and the pcSseg-*n* basis sets, the results proved to be sufficiently converged with respect to the basis set size at the triple-ζ level (def2-TZVP, pcSseg-2), representing a good compromise between computational cost and accuracy.

For the *SiS-H* subset, the application of ECPs proved insufficient due to in part pronounced HALA effects. While scalar-relativistic ZORA approaches do not yield significant improvements, spin-orbit ZORA improves the calculated ^{29}Si NMR chemical shifts drastically. The

SO-ZORA treatment reduces the MADs by up to 74%, which corresponds to an MAD decrease of, e.g., 32.3 ppm for revPBE compared to the ECP result. Thus, an explicit spin-orbit relativistic treatment for HALA prone ^{29}Si cores is highly recommended.

The DFT geometry optimization level has only minor influence on the quality of the results as long as no serious structural issues are observed. For a representative selection of structures from the SiS146 set, the variations in the MADs compared to those calculated on PBEh-3c geometries only amount up to about 2 ppm. Geometry optimization by semiempirical or force field methods may be applied in combination with a linear scaling correction of the shifts to obtain reasonable MADs in combination with a well performing ^{29}Si NMR chemical shift calculation DFA.

In conclusion, robust GGA methods like B97-D3/def2-TZVP can be used reliably for the calculation of ^{29}Si NMR chemical shifts based on PBEh-3c geometries. The triple-ζ basis sets def2-TZVP and pcSseg-2 provide a good cost to accuracy ratio for efficient chemical shift calculations. For ^{29}Si cores in the vicinity of heavier atoms ($Z > 18$), explicit spin-orbit relativistic treatment, e.g., by the SO-ZORA approach, is indispensable to obtain reliable results. In this context, i.e., the SO-ZORA-BP86/TZP method proved to be a reliable choice with a relatively small MAD of 12.3 ppm and a high correlation coefficient of 0.993 for the *SiS-H* subset. Other GGAs like revPBE perform comparably well. If a hybrid DFA is needed, SO-ZORA-PBE0/TZP can be used alternatively with an MAD of 13.6 ppm and $R^2 = 0.995$. The presented SiS146 benchmark set provides a comprehensive, diverse, and robust test set for further method development and enables computational chemists to make a profound method choice for calculating ^{29}Si NMR chemical shifts.

Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.inorgchem.0c02907.

- Detailed information and results for all tested computational methods (PDF)
- Optimized geometries and input data for the full benchmark set (ZIP)

Acknowledgement

The authors thank Fabian Bohle and Karola Schmitz for fruitful discussions, all experimental groups for synthesizing manifold silicon containing molecules accompanied by high quality ^{29}Si NMR measurements, and Prof. Yitzhak Apeloig for bringing the topic of this study to our attention. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) in the framework of the "Gottfried Wilhelm Leibniz Prize" awarded to S.G. Financial support by the "Fonds der Chemischen Industrie (FCI)" to J.B.S. is gratefully acknowledged.

Conflict of Interests

The authors declare no competing financial interest.

APPENDIX B

Appendix: Benchmark Study on the Calculation of ^{119}Sn NMR Chemical Shifts

J. B. Stückrath, T. Gasevic, M. Bursch, and S. Grimme, *Inorg. Chem.* **2022**, *61*, 3903–3917.
DOI: 10.1021/acs.inorgchem.1c03453.

A comprehensive benchmark set termed SnS51 consisting of 51 experimentally determined ^{119}Sn NMR chemical shifts is compiled. 15 commonly used density functional approximations and three relativistic Hamiltonians are assessed for chemical shift computations. Further, the roles of considering Boltzmann weighted conformer-ensembles and the use of relativistic Hamiltonians for geometry optimization are demonstrated. Empirical linear scaling is shown to further improve computed ^{119}Sn NMR chemical shifts.

Abstract

A new benchmark set termed SnS51 for assessing quantum chemical methods for the computation of ^{119}Sn NMR chemical shifts is presented. It covers 51 unique ^{119}Sn NMR chemical shifts for a selection of 50 tin compounds with diverse bonding motifs and ligands. The experimental reference data are in the spectral range of $\pm$2500 ppm measured in seven different solvents. 15 common density functional approximations, two scalar- and one spin-orbit relativistic approach are assessed based on conformer-ensembles generated using the CREST/CENSO scheme and state-of-the-art semiempirical (GFN2-xTB), force field (GFN-FF), and composite DFT methods (r^{2}SCAN-3c). Based on the results of this study, the spin-orbit relativistic method combinations of SO-ZORA with PBE0 or revPBE functionals are generally recommended. Both yield mean absolute deviations from experimental data below 100 ppm and excellent linear regression determination coefficients of $\leq$ 0.99. If spin-orbit calculations are not affordable, the use of SR-ZORA with B3LYP or X2C with ωB97X or M06 may be considered to obtain qualitative predictions if no severe spin-orbit effects, for example, due to heavy nuclei containing ligands, are expected. An empirical linear scaling correction is demonstrated to be applicable for further improvement, and respective empirical parameters are given. Conformational effects on chemical shifts are studied in detail but are mostly found to be small. However, in specific cases when the ligand sphere differs substantially between conformers, chemical shifts can change by up to several hundred ppm.

B.1 Introduction

Nuclear magnetic resonance (NMR) spectroscopy is of great importance to a large number of experimental chemists, especially for characterizing newly synthesized compounds. The NMR chemical shift is very sensitive to the chemical and structural environment of the respective nucleus which enables the reliable identification of specific structural motifs. Although the most commonly examined NMR active nuclei are ^{1}H and ^{13}C, other heteronuclei can be analyzed to yield indications for complex bonding situations in new compounds.[168,295,296] In general, inorganic elements form a plethora of diverse and exotic chemical structures, involving various coordination numbers, bonding motifs, and ligands. Therefore, NMR spectroscopy of inorganic elements, such as the textsuperscript119Sn nucleus[166,167,297], is of high interest in various fields of structural chemistry. Additionally, for heavy, NMR less sensitive nuclei, experiments are often challenging and very time-consuming. Further, the range of possible chemical shifts is large, and thus, computer-aided methods for predicting chemical shifts of such nuclei reliably are highly desirable.

Kohn-Sham density functional theory (DFT) is a versatile tool for the efficient calculation of various molecular properties.[298] It has also proven as a robust choice for the calculation of nuclear magnetic shielding tensors[30,33,39,41,135,226,299,300] if gauge-including atomic orbitals (GIAOs)[137–139,301] are used. For heavy nuclei, such as tin, electrons close to the nucleus are susceptible to scalar- and spin-orbit relativistic (SO) effects and respective corrections become indispensable to correctly describe NMR properties. Some frequently used approaches to include relativistic effects in DFT are the second-order Douglas-Kroll-Hess method (DKH2),[126,127] the zeroth-order regular approximation (ZORA),[128,258] and the exact transformation of the four-component Dirac equation to two-components (X2C).[130–132,302,303] In the context of heavy nuclei, and specifically ^{119}Sn NMR, it has been shown that the inclusion of a SO treatment can improve the agreement of calculated and experimental shifts,[173] especially when other heavy atoms are in the vicinity of the Sn nucleus.[169,304] The performance of some density functional approximations (DFAs) for the calculation of ^{119}Sn NMR chemical shifts was assessed previously for simple tetracoordinated tin compounds without relativistic treatment,[305,306] alkynyltin compounds,[53] tinhalides,[169] aryltin chlorides and hydrides[170] as well as metallostannylenes.[171] However, each of these studies is restricted to either a rather specialized field of tin chemistry or rather small compounds. Further, none of these studies represents a comprehensive assessment of various quantum chemical methods and relativistic treatments.

In line with our previous study on ^{29}Si NMR chemical shift calculations,[213] we herein present a newly compiled comprehensive benchmark set featuring 51 ^{119}Sn NMR chemical shifts in various chemical environments of molecular compounds in solution. A wide range of diverse experimentally determined tin-containing compounds is covered in order to enable a general evaluation of different DFAs and relativistic Hamiltonians. Large organometallic tin complexes are also represented for which the ligands can affect the structural environment of the Sn nucleus directly or play a major role in the overall stability. Their description is important for the chemical shift prediction, therefore reinforcing the

relevance of explicitly including such cases in this benchmark study. Overall, this work aims to provide guidance for the choice of a reliable method combination of DFA and relativistic treatment for the efficient calculation of ^{119}Sn NMR chemical shifts.

B.2 Computational Details

Quantum chemical and force field (FF) calculations were performed with the xTB 6.3.3,[247,248] TURBOMOLE 7.5.1[307–309] and AMS2020.102[245,246] (ADF module) program packages. The initial structures were obtained from experimentally determined X-ray crystal structures when available and were preoptimized with the generic force field GFN-FF[28] applying the ALPB[118] implicit solvation model for the respective solvent from the NMR measurement. For a detailed list of references and solvents see the next sections and the Supporting Information (SI), Table S5. Compounds **5**, **7**, **18**, **45** and **46** were preoptimized with the semiempirical quantum mechanical (SQM) extended tight-binding method GFN2-xTB[27]. For each compound, a conformer-rotamer ensemble (CRE) was generated with the CRE sampling tool CREST 2.11[25,310] applying GFN-FF or GFN2-xTB if the initial structure is not correctly described by GFN-FF. The CREs were subsequently refined with the command line energetic sorting algorithm CENSO 1.0.0[29,311] and afterward checked for identical conformers to yield final conformer ensembles, their free energies, and the respective Boltzmann weights. Within this scheme, geometry optimizations were performed using the efficient r^{2}SCAN-3c[172] composite DFT method and the COSMO[114] implicit solvation model. Enhanced numerical integration grid settings (radsize 10) were used. Thermostatistical free energy contributions were obtained applying the modified rigid-rotor harmonic-oscillator (mRRHO)[119] based single-point Hessian[120] approach using GFN2-xTB. Solvation free energies were calculated with COSMO-RS[115,116] using the COSMOtherm 19.0 program package[312] (G_{solv} option, BP_TZVP_C30_1601 parametrization, $T = 298.15$ K, $p = 1$ atm). All structures were verified as minima on the potential energy surface by the absence of imaginary vibrational frequencies below $i\omega = -30$ cm^{-1} in the harmonic frequency calculation. All geometry optimizations, single-point energy, and harmonic frequency calculations in TURBOMOLE were accelerated by the resolution of the identity (RI) approximation for Coulomb integrals[255,313,314] with matching auxiliary basis sets.

All ^{119}Sn NMR shielding constants were calculated using GIAOs and the COSMO solvation model. In TURBOMOLE, all NMR shielding calculations[315,316] were performed with the scalar-relativistic (SR) exact two-component X2C method[317] applying the Gaussian-type orbital (GTO) recontracted all-electron basis set x2c-TZVPall-s[174,175] and the 5a integration grid. For the NMR shielding calculations in AMS, the Slater-type orbital (STO) all-electron triple-ζ ZORA/TZP[165] basis set and the COSMO[261] solvation model were applied. In AMS, relativistic effects were included by the scalar- and spin-orbit relativistic versions of ZORA.[129,318,319] Overall, a total number of 15 different DFAs and their combination with three relativistic approaches as listed in Table B.1 was assessed if available.

NMR chemical shifts δ were calculated from nuclear magnetic shielding constants σ accord-

Table B.1: Assessed density functional approximations (DFAs) and scalar- (SR) and spin-orbit (SO) relativistic approaches for ^{119}Sn NMR shielding calculations in the SnS51 benchmark set (TM = TURBOMOLE, AMS = ADF module in AMS2020).

		relativistic treatment		
		X2C (TM)	ZORA (AMS)	
class	DFA	SR	SR	SO
(meta-)GGA	PBE[101]	✓	✓	✓
	revPBE[177]	✓	✓	✓
	OLYP[262]	✓	✓	✓
	BP86[102,103]	✓	✓	✓
	B97-D[157]	✓		
	TPSS[104]	✓		
	M06-L[176]	✓		
	KT2[160]	✓	✓	✓
	r^{2}SCAN[105,320]	✓		
(meta-)hybrid	PBE0[109]	✓	✓	✓
	B3LYP[106–108]	✓	✓	✓
	ωB97X[269]	✓		
	TPSSh[158,159]	✓		
	M06[161]	✓		
	M06-2X[161]	✓		

ing to eq (B.1) with $\sigma(^{119}\mathrm{Sn}_{\mathrm{SnMe_4}})$ being the shielding constant of the reference compound tetramethyltin ($\mathrm{SnMe_4}$).

$$\delta(^{119}\mathrm{Sn}) = \sigma(^{119}\mathrm{Sn}_{\mathrm{SnMe_4}}) - \sigma(^{119}\mathrm{Sn}) \tag{B.1}$$

Thermally averaged chemical shifts from a conformer ensemble were obtained by a Boltzmann-weighted average of each conformer's chemical shift according to eq (B.2) with Boltzmann weights w_i of conformer i given by

$$w_i = \frac{\exp\left(-\Delta G_i / RT\right)}{\sum_j^{\mathrm{CRE}} \exp\left(-\Delta G_j / RT\right)} \tag{B.2}$$

with the free energy difference of conformer i to the lowest conformer $\Delta G_i = G_i - G_{\mathrm{lowest}}$, molar gas constant R, and absolute temperature T. If not stated otherwise, the temperature was set to 298.15 K. For the conformational analysis of compound **46**, a temperature range between 173.15 and 298.15 K was considered.

For the structure dependence study (see Section B.3.6), the energetically lowest-lying conformer of each CRE was reoptimized on the TPSS-D4(COSMO)/def2-TZVP[104,156,249–251] level with the respective Stuttgart-Dresden def2 effective core potentials (ECPs)[254,257] and the `m5` numerical integration grid. Further reoptimizations were performed on the TPSS-D4(COSMO)/x2c-TZVPall level with SR-X2C in TURBOMOLE[321] as well as the TPSS-D4(COSMO)/TZP level with SR- and SO-ZORA in AMS.[322] Furthermore, the structures were optimized on the semiempirical GFN2-xTB(ALPB) level and with the GFN-FF(ALPB) force field in xTB. Subsequent NMR shielding calculations were performed with AMS on the revPBE(COSMO)/TZP and PBE0(COSMO)/TZP levels of theory with SO-ZORA.

B.3 Results and Discussion

B.3.1 General Considerations

The quantum chemical computation of NMR chemical shifts is not trivial. Besides the choice of a reliable method combination for the geometry optimization, the correct description of the conformational space for flexible molecules is challenging, yet important. In general, differently populated conformations can have a crucial impact on the overall NMR chemical shifts observed in the experiment.[25,30] Further, the choice of a reliable level of theory (DFA, basis set, Hamiltonian, solvation model) for the NMR property calculation is the key for high accuracy. These methods may also differ from those chosen for geometry optimization in commonly applied multilevel approaches. Another important aspect is the role of solvation effects on structures and NMR chemical shifts. Even though implicit solvation models such as COSMO or PCM[112,113] are typically sufficient to cover electrostatic influences caused by the solvent, in some cases, explicit solvent molecules can influence the structure of the target compound. As such cases can not be treated in a black-box type approach, in the following, only cases for which implicit models can be considered sufficient will be discussed.

Usually, several conformers of a molecule in a temperature-dependent equilibrium exist, each of which individually contributes to the observable. Therefore, a detailed analysis of energetically favorable conformers is necessary in order to simulate an observable satisfactorily. As a sophisticated conformational sampling is usually prohibitively costly with DFT methods, recent SQM and FF methods such as GFN2-xTB and GFN-FF are used within the CREST approach. The initially generated CRE at a lower theory level should be further refined by calculation of conformational free energies and geometries using efficient composite DFT methods (e.g., PBEh-3c,[155] B97-3c[279], r^{2}SCAN-3c[172]) within the CENSO approach. The recently developed r^{2}SCAN-3c method has proven to yield reasonable structures and conformational energies at a comparably low computational effort. Based on these refined conformer ensembles, more advanced, yet more computationally demanding method combinations can be employed. Triple-ζ quality basis sets have proven to be sufficiently converged for heteronuclei NMR shifts and are therefore used exclusively in this study.[213] Nevertheless, for heavy nuclei the careful choice of the relativistic treatment regarding the Hamiltonian (non-/scalar-/spin-orbit relativistic) is important and can have a great impact on the results.

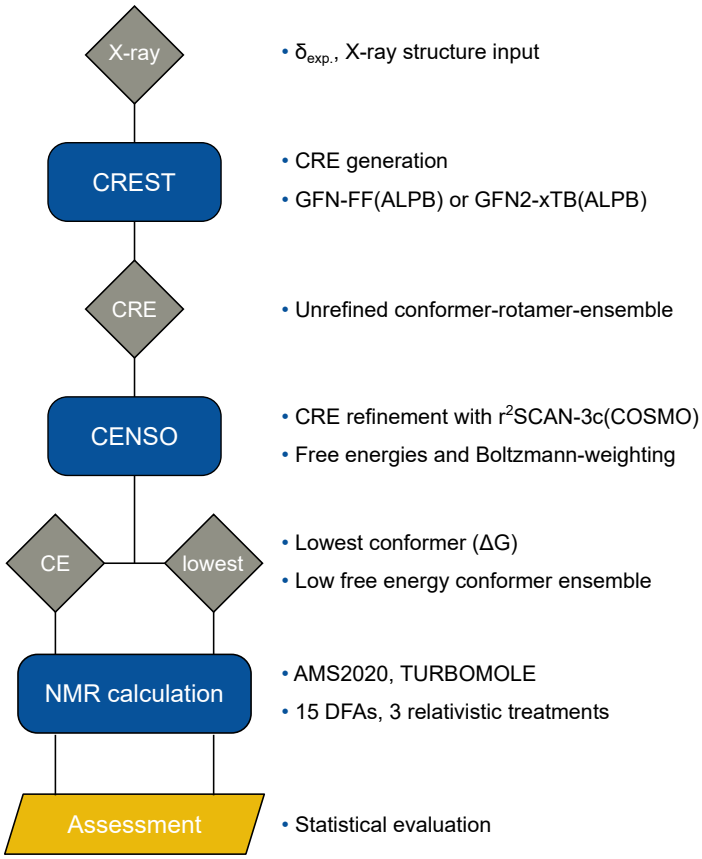


Figure B.1: Schematic protocol used for the SnS51 benchmark study.

Initial tests on the application of ECPs at the NMR active nuclei yielded qualitatively wrong results, which is in line with the expectation due to the removal of all nodes in the core region and the erroneous shape of the resulting pseudo-orbitals. Therefore, no further results regarding ECPs at the ^{119}Sn nucleus are discussed in this study. All the mentioned aspects make a well-founded selection of a suitable method for the prediction of ^{119}Sn NMR chemical shifts difficult. To provide guidance for a reliable choice, we assess a selection of different combinations of prominent DFAs and relativistic approaches on a newly compiled comprehensive benchmark set termed SnS51. Furthermore, the role of relativistic treatments in the geometry optimization as well as a possible application of SQM- or FF-based geometries are evaluated. The general influence of considering conformer ensembles compared to a single structure approach based on the lowest identified conformer is also investigated. The schematic protocol applied to generate the molecular structures used in our benchmark study is depicted in Figure B.1.

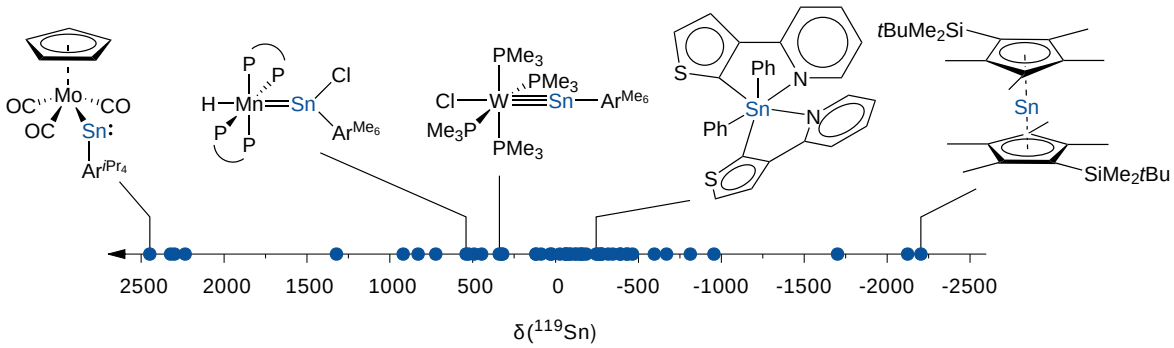


Figure B.2: Overview of the investigated compounds and their ^{119}Sn NMR chemical shifts in ppm (referenced to $SnMe_4$).

B.3.2 Benchmark Set

The new SnS51 benchmark set consists of 50 tin-containing compounds (42 neutral, 4 positively, 4 negatively charged; ordered according to their descending experimental ^{119}Sn NMR chemical shifts). For these compounds, overall 219 unique conformers with various tin bonding motives including single, double and triple as well as coordinate and aromatic bonds are evaluated. Tin coordination numbers range from 2 to 7 and molecule sizes between 4 and 209 atoms. The investigated ^{119}Sn nuclei are bound to light and heavy main group elements as well as to transition metals. The total number of 51 experimental ^{119}Sn NMR shifts ranges from 2448 to –2204 ppm that were measured in seven different solvents: benzene, chloroform, tetrahydrofuran, dichloromethane, acetonitrile, carbon disulfide, or cyclohexane. An exemplary overview of included compounds is depicted in Figure B.2. All further compounds are shown in Figure B.3, and their reference data are listed in Table B.2. More detailed information can be found in Table S5 in the Supporting Information.

The performance of 15 different DFAs is assessed. These include nine (meta-)generalized gradient approximation (GGA) and six (meta-)hybrid DFAs. All functionals are applied alongside three different relativistic Hamiltonians if available in the respective quantum chemistry packages. The assessed relativistic treatments are SR-X2C and SR-ZORA as well as SO-ZORA (see Table B.1 for a list of all method combinations).

B.3.3 Study of Density Functional Approximations and Relativistic Approaches for the Calculation of ^{119}Sn NMR Chemical Shifts

Unless otherwise stated, all statistical evaluations are based on NMR chemical shifts calculated as the Boltzmann weighted average of all conformers populated at room temperature. The overall performance of the different methods is depicted in Figure B.4, and all statistical measures are listed in Table B.3.

All DFAs yield systematically too negative ^{119}Sn NMR chemical shifts in combination with both tested scalar-relativistic Hamiltonians SR-X2C (Figure B.4a) and SR-ZORA (Figure B.4b)

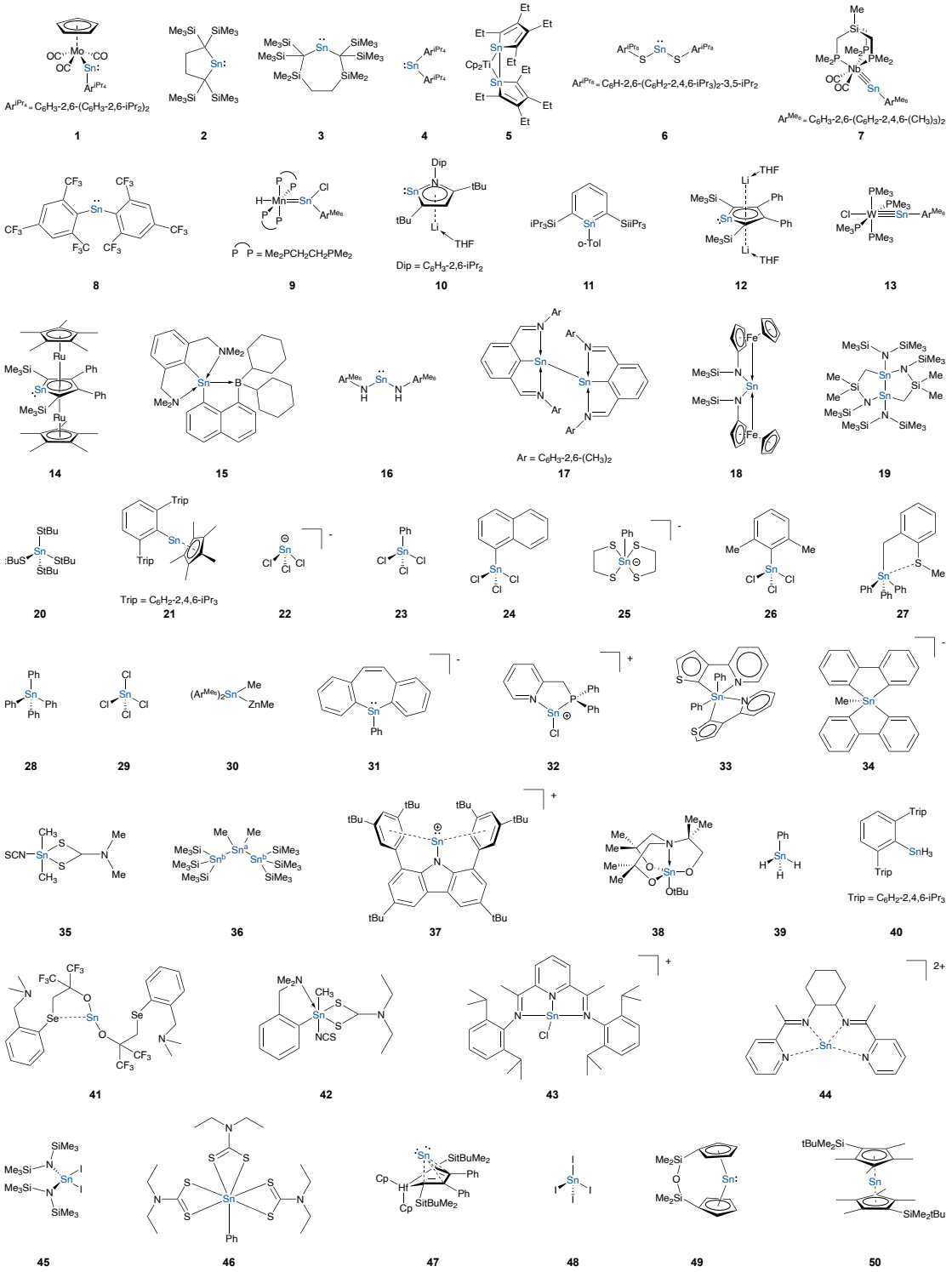


Figure B.3: Compounds included in the SnS51 benchmark study. Tin atoms are highlighted in blue.

Table B.2: All compounds in the SnS51 set with their identification number (the corresponding structures can be found in Figure B.3), the solvent used in the experimental measurement, the experimental ^{119}Sn NMR chemical shift δ_{exp}, and the reference.

No.	Solvent	δ_{exp} (ppm)	Ref.	No.		Solvent	δ_{exp} (ppm)	Ref.
1	benzene	2 448	[323]	**27**		chloroform	−115.5	[324]
2	benzene	2 323	[325]	**28**		benzene	−127	[170]
3	benzene	2 299	[326]	**29**		benzene	−148.6	[170]
4	benzene	2 235	[327]	**30**		benzene	−155	[328]
5	benzene	1 322	[329]	**31**		benzene	−164	[330]
6	benzene	919.4	[331]	**32**		THF	−184	[332]
7	THF	829.7	[333]	**33**		chloroform	−245.5	[334]
8	cyclohexane	723	[335]	**34**		benzene	−253	[336]
9	THF	540	[337]	**35**		chloroform	−270.1	[338]
10	benzene	524.2	[339]	**36**	Sn^a	THF	−277	[340]
11	benzene	490.7	[341]		Sn^b		−464	
12	benzene	446.4	[342]	**37**		dichloromethane	−278.2	[343]
13	THF	340.1	[344]	**38**		benzene	−317	[345]
14	benzene	324.5	[346]	**39**		benzene	−345	[170]
15	benzene	319	[347]	**40**		benzene	−388.3	[348]
16	benzene	118.9	[349]	**41**		benzene	−392	[350]
17	benzene	118	[351]	**42**		chloroform	−427.3	[338]
18	benzene	88	[352]	**43**		THF	−435.1	[353]
19	benzene	32.2	[354]	**44**		benzene	−595.8	[355]
20	chloroform	24.3	[356]	**45**		benzene	−669.6	[357]
21	benzene	−26	[348]	**46**		dichloromethane	−813	[358]
22	THF	−60.3	[353]	**47**		benzene	−955.6	[359]
23	benzene	−61	[170]	**48**		CS_2	−1 701	[360]
24	benzene	−62.3	[170]	**49**		benzene	−2 124	[361]
25	acetonitrile	−82.4	[362]	**50**		benzene	−2 204	[363]
26	benzene	−90.5	[170]					

(76% of all calculated chemical shifts included in Figure B.4 and Table B.3 are below the experimental value). Within the group of scalar-relativistic treatments, SR-X2C and SR-ZORA yield average MADs of $MAD_{SR\text{-}X2C} = 172.1$ ppm and $MAD_{SR\text{-}ZORA} = 187.0$ ppm, respectively. For SR-X2C, the M06 hybrid DFA yields the smallest MAD of 130.2 ppm ($R^2 = 0.975$). The best-performing GGA using SR-X2C is BP86 with an MAD of 158.2 ppm ($R^2 = 0.972$) closely followed by B97-D with an MAD of 160.4 ppm ($R^2 = 0.972$). For the DFAs tested in combination with SR-ZORA, the smallest MAD is obtained with the hybrid B3LYP yielding an MAD of 163.6 ppm ($R^2 = 0.968$). Here, the best-performing GGA is again BP86 with an MAD of 174.8 ppm ($R^2 = 0.970$). It is to note that within the applied computational settings and basis sets, the SR-X2C-based method combinations generally yield better results compared to those calculated with SR-ZORA. Furthermore, the very different performance of the Minnesota-type DFAs is remarkable. While M06-L and M06-2X

(54% HF exchange) are the worst-performing GGA and hybrid DFAs applying SR-X2C, respectively, M06 (27% HF exchange) yields good results with the lowest standard deviation (SD) of 144.3 ppm. Overall, no general benefit of using hybrid DFT compared to the much less computationally demanding (meta-)GGAs is observed when using SR-X2C even though the average class MAD is reduced from $MAD_{SR\text{-}X2C/GGA} = 184.1$ ppm to $MAD_{SR\text{-}X2C/hybrid} = 158.4$ ppm. A typical "Jacob's ladder" behavior[100,281] is therefore not observed for the calculation of ^{119}Sn NMR chemical shifts, which is in line with our findings for ^{29}Si NMR chemical shifts[213] and is also in agreement with previous studies on other nuclei.[34,39,41] Nevertheless, all scalar-relativistic approaches yield severe deviations for the iodine bound tin nuclei in compounds **45** ($SnI_2(N(SiMe_3)_2)_2$)[357] and **48** (SnI_4, cf. Figure B.5).[360]

The electrons in the heavy iodine atom experience severe relativistic effects, and furthermore, halogens that are close to the nucleus of interest are known to increase the *s*-character of bonding orbitals of the respective nucleus thus evoking strong spin-orbit coupling effects.[144,169,296] This influence is generally known as heavy neighbor atom (HNA) effect.[47,48,54] To allow for an unbiased statistical evaluation, these cases (outliers) are excluded from the statistics for SR approaches. The importance of explicit inclusion of spin-orbit effects for these compounds is demonstrated upon application of the SO-ZORA Hamiltonian (Figure B.4c). Both former outliers are described correctly using any DFA with SO-ZORA treatment. Furthermore, the average method MAD is greatly reduced by 41% and 46% to $MAD_{SO\text{-}ZORA} = 101.2$ ppm. Moreover, for both assessed hybrid DFAs PBE0 and B3LYP, the systematic error approaches zero with MDs of –10.4 and –0.6 ppm. The lowest MAD in the SO-ZORA group is obtained with PBE0 yielding an MAD of 87.6 ppm ($R^2 = 0.994$) and an excellent determination coefficient of 0.994. In the GGA class, revPBE represents a good alternative with a low MAD of 96.8 ppm ($R^2 = 0.990$) and a similar SD of 115.1 ppm. For the SO-ZORA methods, the iodine-containing compounds **45** and **48** still yield deviations above the average of the tested compounds, but are within a much more narrow range (cf. Figure B.4, lower outliers) and thus do not severely bias the statistics. For instance, excluding the two data points also from the SO-ZORA evaluation would yield MADs of 89.6 and 85.1 ppm for revPBE and PBE0, respectively. For lighter nuclei such as ^{29}Si, we found that the explicit inclusion of SO effects is only necessary when other heavy nuclei are close to the NMR active one.[213] On the contrary, computed ^{119}Sn NMR chemical shifts profit from application of the SO-ZORA treatment even in the absence of heavy neighbor atoms.

Overall, based on these observations, some recommendations can be given based on these observations. If available, a SO relativistic approach should be used in general for the reliable calculation of ^{119}Sn NMR chemical shifts. The SO-ZORA/PBE0 combination yields good results over the large ^{119}Sn NMR shift scale of almost ±2500 ppm. If the application of a hybrid DFA is not affordable, the SO-ZORA/revPBE approach represents an efficient alternative. Even though some SR-X2C-based methods also yield comparably small MADs, such as M06 or ωB97X, their general use may not be recommended if any indication for a pronounced role of SO effects is given. This is specifically the case when HNA effects due to the vicinity of other heavy nuclei or in general an increased *s*-character

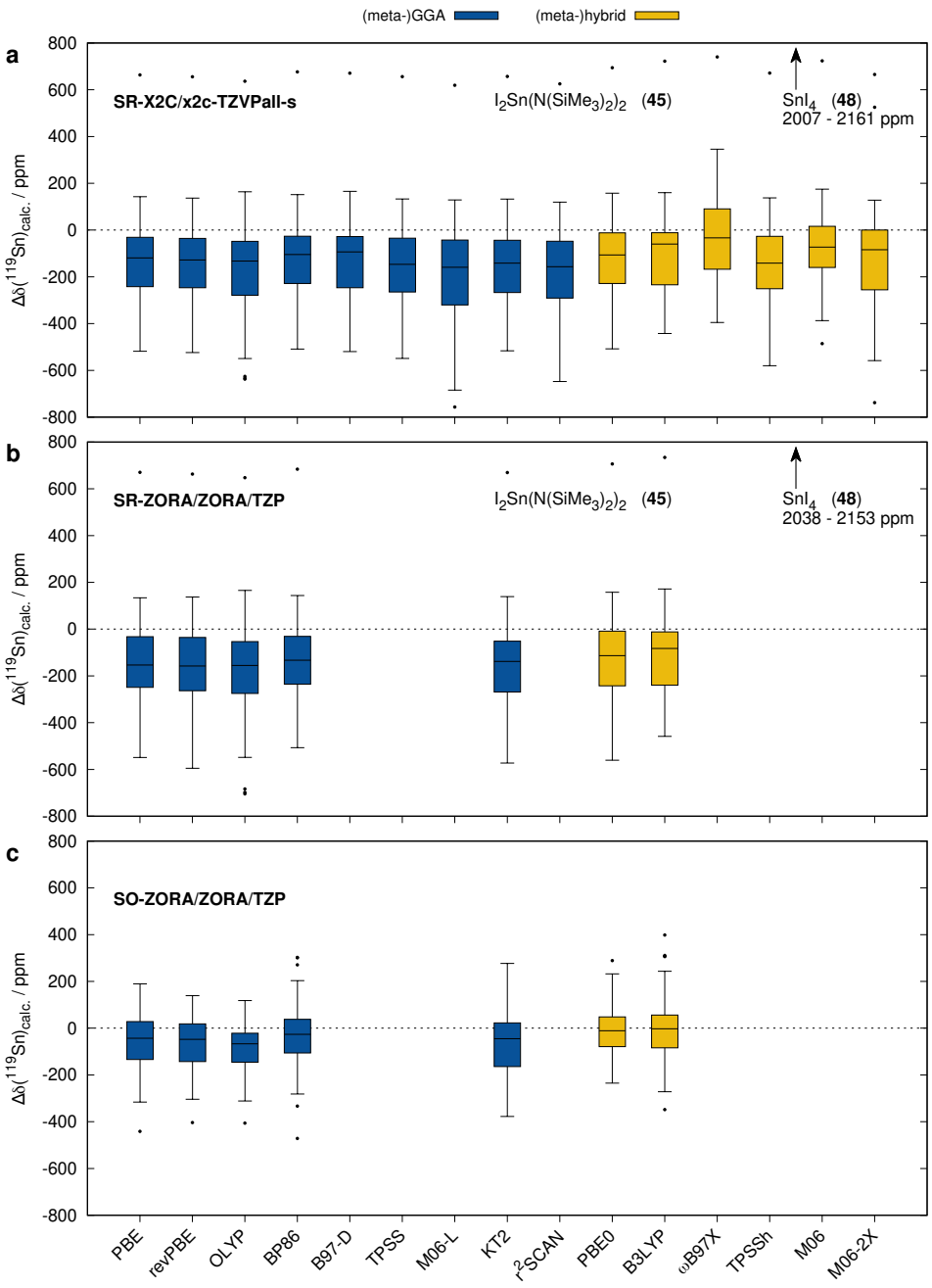
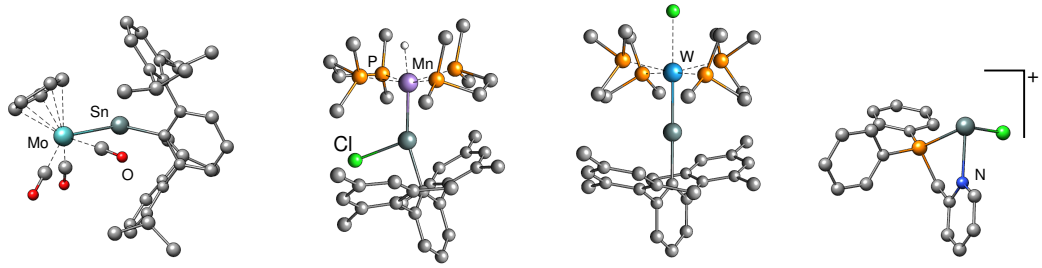
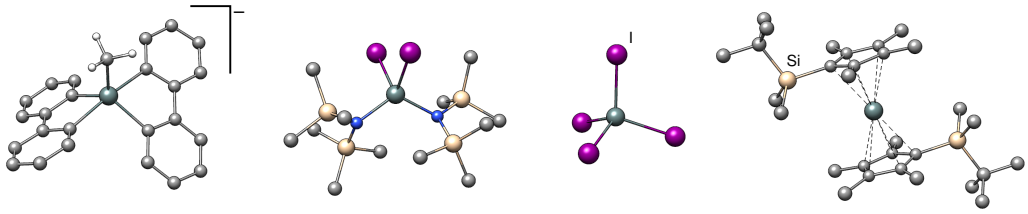


Figure B.4: Box plots showing the ^{119}Sn NMR shift deviation between calculation and experiment ($\Delta\delta = \delta_{\text{calc}} - \delta_{\text{exp}}$) for all tested density functional approximations in combination with the three different relativistic approaches SR-X2C (**a**), SR-ZORA (**b**), and SO-ZORA (**c**). The central lines represent the median values, the boxes the range of 25% to 75% of the data, and the whiskers all points within 1.5 times the interquartile range. Outliers are depicted by black dots, especially conspicuous are the compounds **45** and **48** for the scalar-relativistic methods.



	1	**9**	**13**	**32**
$\delta_{exp.}$	2448 ppm	540 ppm	340 ppm	–184 ppm
$\delta_{SR\text{-}X2C}$	1929 ppm	228 ppm	36 ppm	–329 ppm
$\delta_{SR\text{-}ZORA}$	1853 ppm	185 ppm	30 ppm	–421 ppm
$\delta_{SO\text{-}ZORA}$	2267 ppm	511 ppm	412 ppm	–196 ppm



	34	**45**	**48**	**50**
$\delta_{exp.}$	–253 ppm	–670 ppm	–1701 ppm	–2204 ppm
$\delta_{SR\text{-}X2C}$	–323 ppm	–14 ppm	362 ppm	–2464 ppm
$\delta_{SR\text{-}ZORA}$	–315 ppm	–7 ppm	348 ppm	–2407 ppm
$\delta_{SO\text{-}ZORA}$	–324 ppm	–813 ppm	–2105 ppm	–2468 ppm

Figure B.5: Selected molecules from the SnS51 set and ensemble-based ^{119}Sn NMR chemical shifts calculated applying the revPBE DFA. The best matching value is highlighted in blue. Due to the large deviation of the SR-X2C and SR-ZORA values for compounds **45** and **48**, these data points have been excluded from all statistics of scalar-relativistic approaches.

near the nucleus of interest are expected.[296] The following general ranking regarding the best-performing method combinations from each DFA class by decreasing MAD can be given: SR-ZORA/BP86 (174.8 ppm) > SR-ZORA/B3LYP (163.6 ppm) > SR-X2C/BP86 (158.2 ppm) > SR-X2C/M06 (130.2 ppm) ≫ SO-ZORA/revPBE (96.8 ppm) > SO-ZORA/PBE0 (87.6 ppm).

B.3.4 Linear Scaling

Empirical linear scaling was previously used successfully to correct errors of NMR chemical shifts[32,37,285–287] also for ^{29}Si NMR.[213] Subsequently, it is applied to derive method-dependent scaling parameters for the correction of systematic errors in the calculation of ^{119}Sn NMR chemical shifts. The parameters α and β represent the slope and the intercept of the linear best fit with respect to experimental reference data and are applied according to eq (B.3).

Table B.3: Ensemble-based mean deviation (MD), mean absolute deviation (MAD), standard deviation (SD) of the MD and root mean square deviation (RMSD) from experimental values in ppm as well as determination coefficient R^2 for all DFAs in combination with the three relativistic approaches. Note that due to the nature of the linear scaling approach, the scaled MD is always zero. The scaled MAD (MAD_{scaled} in ppm) and MAD of only the lowest conformer ($MAD_{lowconf}$ in ppm) are discussed in the following sections.

	SR-X2C/x2c-TZVPall-s‡						
DFA	MD	MAD	SD	RMSD	R^2	MAD_{scaled}	$MAD_{lowconf}$
PBE	−142.6	170.3	156.7	210.7	0.970	127.1	172.8
revPBE	−153.9	178.4	163.3	223.1	0.968	131.7	181.0
OLYP	−182.4	199.4	186.8	259.7	0.961	148.2	202.1
BP86	−126.6	158.2	150.8	195.7	0.972	121.2	160.2
B97-D	−135.1	160.4	151.8	202.0	0.972	123.0	162.8
TPSS	−162.5	186.6	166.7	231.6	0.968	133.8	189.1
M06-L	−212.9	219.5	198.6	289.8	0.965	141.2	221.7
KT2	−156.4	181.3	160.7	223.1	0.969	129.0	183.5
r^2SCAN	−188.0	202.8	180.0	259.0	0.965	141.5	205.3
PBE0	−136.1	166.4	166.4	213.7	0.967	140.5	169.0
B3LYP	−113.7	150.4	155.3	191.2	0.971	128.8	153.5
ωB97X	−60.4	130.9	151.2	161.4	0.975	120.5	132.4
TPSSh	−158.8	183.9	169.9	231.3	0.966	138.2	186.4
M06	−91.6	130.2	144.3	169.6	0.975	113.5	132.6
M06-2X	−142.8	188.7	218.8	259.4	0.943	171.7	192.2
	SR-ZORA/ZORA/TZP‡						
DFA	MD	MAD	SD	RMSD	R^2	MAD_{scaled}	$MAD_{lowconf}$
PBE	−160.7	187.4	165.2	229.3	0.968	132.4	189.4
revPBE	−172.7	196.0	174.0	243.9	0.966	137.2	198.2
OLYP	−200.3	216.7	199.9	281.5	0.957	155.0	219.0
BP86	−144.4	174.8	156.3	211.6	0.970	126.5	176.7
KT2	−164.4	192.4	169.5	234.9	0.967	134.4	194.4
PBE0	−147.3	178.2	175.5	227.8	0.964	148.2	178.3
B3LYP	−127.3	163.6	162.9	205.4	0.968	138.3	165.7
	SO-ZORA/ZORA/TZP						
DFA	MD	MAD	SD	RMSD	R^2	MAD_{scaled}	$MAD_{lowconf}$
PBE	−51.0	102.5	129.3	137.8	0.990	70.7	103.4
revPBE	−63.1	96.8	115.1	130.3	0.990	70.2	97.9
OLYP	−89.4	104.2	106.9	138.6	0.989	79.3	106.5
BP86	−33.9	109.1	147.6	150.0	0.991	68.6	109.9
KT2	−59.5	106.1	127.6	139.7	0.988	79.6	107.0
PBE0	−10.4	87.6	115.3	114.6	0.994	55.7	90.2
B3LYP	−0.6	102.2	145.6	144.2	0.995	49.6	104.1

‡For all SR methods, compounds **45** and **48** have been excluded from the statistics.

$$\delta_{\text{calc,scaled}} = \frac{\delta_{\text{calc}} - \beta}{\alpha} \tag{B.3}$$

With this equation, a scaled ^{119}Sn NMR chemical shift is obtained for every compound, and the new MAD ($\text{MAD}_{\text{scaled}}$) for every computational method is given in Table B.3. It is observed that the linear scaling approach decreases the mean absolute deviation for every tested method combination by at least 8% (130.9 to 120.5 ppm with SR-X2C/ωB97X) up to 51% (102.2 to 49.6 ppm with SO-ZORA/B3LYP). In general, this approach can be applied to improve the results obtained from any method combinations, also including less accurate methods such as SO-ZORA/BP86 which is the worst-performing SO-ZORA method (37% improvement, MAD reduction from 109.1 to 68.6 ppm) and SR-X2C/M06-L as the worst-performing SR-X2C method (36% improvement, MAD reduction from 219.5 to 141.2 ppm). The linear scaling procedure for these examples is depicted in Figure B.6. Provided that the underlying data is extensive enough, this approach has a high probability of correcting a specific calculated chemical shift in the right direction. At this point, it is specifically important to remove serious outliers such as compounds **45** and **48** for the SR methods to avoid a strong bias of the scaling parameters. Such severe errors due to the missing physical description of important effects may not be corrected in a simple empirical linear scaling approach.

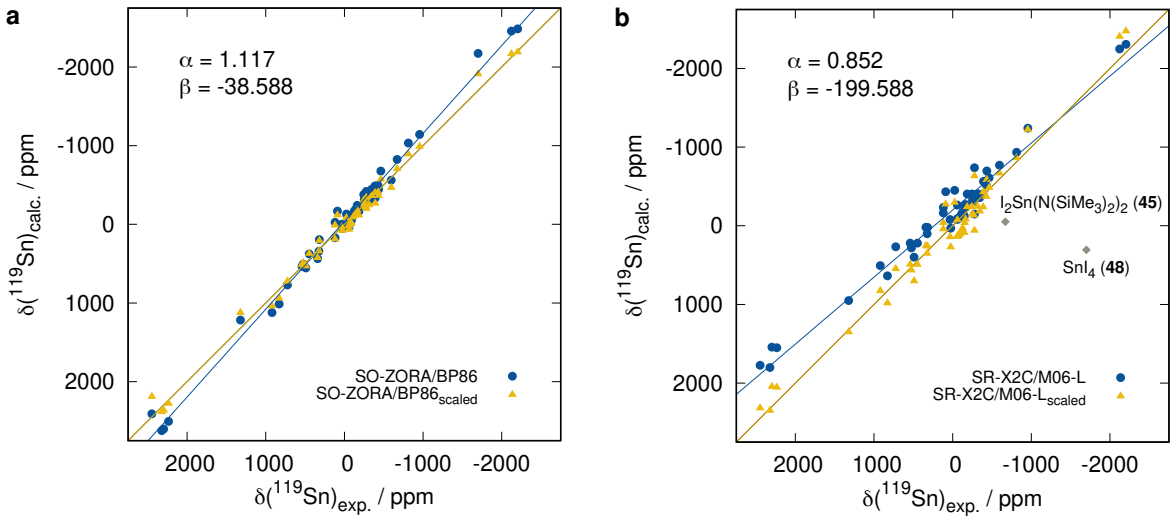


Figure B.6: Linear scaling applied to ensemble-based (**a**) SO-ZORA/BP86 and (**b**) SR-X2C/M06-L calculated ^{119}Sn NMR chemical shifts. For SR-X2C/M06-L, both iodine-containing compounds **45** and **48** were excluded from the scaling to avoid a strong bias due to the very high deviations.

B.3.5 Conformational Analysis

Up to this point, for all compounds an ensemble of distinctly Boltzmann populated conformers was considered. Accordingly, the number of computations and thus the overall computational demand depends on the number of considered conformers. Therefore, for the sake of efficiency, it is still questionable whether this approach is necessary and how well the systems are described by only taking the lowest free energy conformer into account. If only this conformer is considered in the evaluation, the overall MADs of the assessed methods increase only slightly (minimum 0.2 ppm for SR-ZORA/PBE0, maximum 3.5 ppm for SR-X2C/M06-2X), regardless of the applied density functional or relativistic approach (see Table B.3, last column). As a representative example, Figure B.7 shows the ^{119}Sn NMR chemical shift deviations with respect to experimental data obtained with SO-ZORA/revPBE from the ensemble-averaged and lowest-conformer approaches. The majority of the investigated compounds do not show a significant ensemble dependence, and in a few cases, the description deteriorates by considering only the lowest conformer. This is due to the usually structurally rigid bonding region around the tin nucleus that is typically buried inside the first ligand sphere. Therefore, conformational effects may be generally less pronounced compared to more exposed nuclei such as ^{1}H.

The largest influence of the ensemble on the ^{119}Sn NMR chemical shift (+43.0 ppm) is observed for compound **46**[358] featuring a tin center surrounded by one phenyl and three 1,1-dithiolate ligands. The denticity of the latter ones can change, and an equilibrium

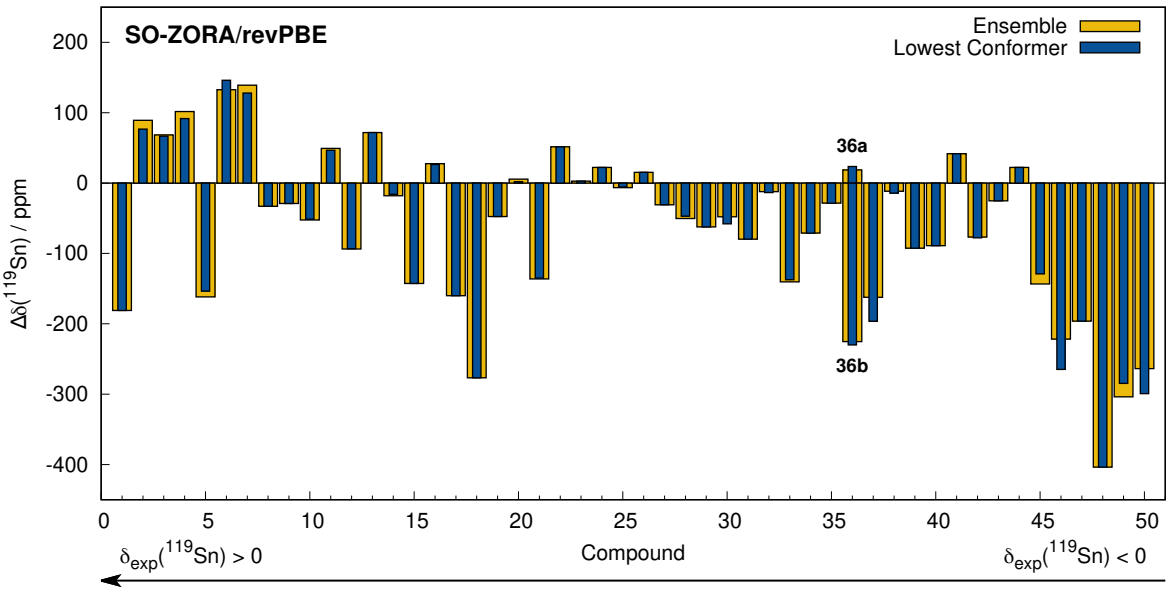


Figure B.7: Deviations from experiment of ^{119}Sn NMR chemical shifts calculated at the SO-ZORA/revPBE level to experiment for the energetically lowest conformer (blue) and the Boltzmann weighted ensemble (yellow). The experimentally obtained shift decreases from compound **1** to **50** (except for **36b**).

between the monodentate and bidentate bonding motif has been reported. The refined conformer ensemble of compound **46** consists of seven unique conformers which can be grouped according to bi- (tin coordination number of 7) and monodentate bonding of the 1,1-dithiolate ligand (tin coordination number of 6). Two examples for these bonding patterns are shown in Figure B.8a. When all three 1,1-dithiolate ligands act as bidentates (motif **A**), a pseudopentagonal bipyramidal coordination of the tin atom is obtained. Decoordination of one equatorial 1,1-dithiolate ligand results in the octahedrally coordinated bonding motif **B**. Due to the drastic change in the direct coordination sphere of the tin nucleus, the respective ^{119}Sn NMR chemical shift differs by approximately 300 ppm between the conformers (see Table S17 in the Supporting Information). Furthermore, the experimental study of compound **46** includes two NMR measurements yielding –813 ppm at 25 °C and –888 ppm at –100 °C.[358] In order to reproduce this trend, temperature-dependent Boltzmann populations of the conformers were calculated, and the resulting ^{119}Sn NMR chemical shift on the SO-ZORA/PBE0 level is shown in Figure B.8b. While at –100 °C the ensemble is clearly dominated by motif **A**, its Boltzmann weight decreases with temperature which leads to an increased (less negative) chemical shift. This is qualitatively in accordance with the experiment, although we note that the computed temperature effect is much smaller than the experimentally observed. Nevertheless, it is clear that conformational effects on ^{119}Sn NMR chemical shifts can be significant for special cases and that our approach enables such temperature dependent calculations.

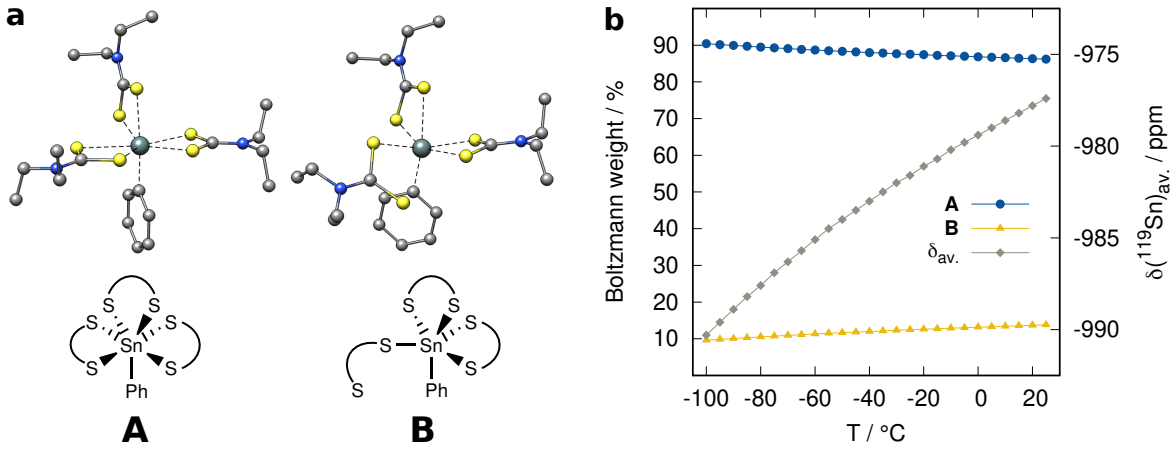


Figure B.8: (**a**) Molecular and Lewis structures of conformer 1 and 2 of compound **46** representing the two bonding motifs **A** and **B** (hydrogen atoms have been omitted). (**b**) Dependence of the total Boltzmann weights of the two bonding motifs on the temperature as well as the full ensemble-based ^{119}Sn NMR chemical shift calculated with SO-ZORA/PBE0. The experimental values are $\delta_{\text{-100°C}}$ = –888 ppm and $\delta_{\text{25°C}}$ = –813 ppm.

B.3.6 Structure Dependence

The accuracy of a calculated NMR chemical shift is generally affected by the molecular geometry and in particular sensitive to covalent bond lengths.[364,365] Therefore, we studied the influence of different geometry optimization levels for the energetically lowest conformer of each compound. The focus lies on the one hand on the inclusion of relativistic effects and on the other hand on the evaluation of the efficient SQM and FF methods GFN2-xTB and GFN-FF. These methods are specifically valuable when a DFT-based geometry optimization is not affordable anymore due to the size of the investigated molecule or when molecular dynamics treatments are required.

Figure B.9 depicts the mean (absolute) deviations of the ^{119}Sn NMR chemical shifts on the SO-ZORA/revPBE and SO-ZORA/PBE0 levels after reoptimization with different methods. The TPSS-D4/def2-TZVP geometry optimization level with standard ECPs (MAD of 107.4/86.9 ppm for revPBE/PBE0) shows a very similar performance as r^{2}SCAN-3c (MAD of 97.9/90.2 ppm), which was used by default in the presented study. Comparing all different variants of including relativistic effects in the TPSS-D4 calculation (SR-X2C as well as SR- and SO-ZORA), no systematic improvement is observed (MADs of 110.2-113.2/84.5-90.6 ppm for revPBE/PBE0). Accordingly, applying the computationally more demanding SO-ZORA for geometry optimization does not yield a systematic improvement of the finally computed

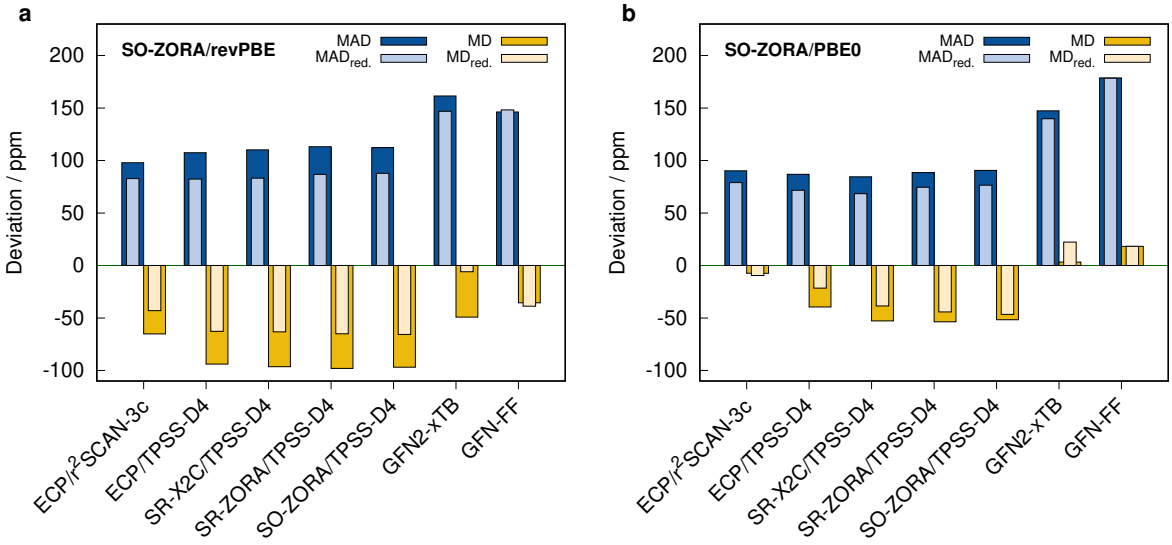


Figure B.9: MD and MAD for SO-ZORA in combination with (**a**) revPBE and (**b**) PBE0 for the lowest-lying conformer optimized as before (ECP/r^{2}SCAN-3c) and with different relativistic approaches in combination with TPSS-D4 as well as with the GFN2-xTB semiempirical and the GFN-FF force field approach. Due to failure in geometry optimization and strong outliers, only 48 instead of 51 ^{119}Sn NMR chemical shifts have been considered for GFN2-xTB, 36/40 (revPBE/PBE0) for GFN-FF, and 50 for ECP/TPSS-D4 with PBE0. The MD$_{red.}$ and MAD$_{red.}$ quantities only include compounds with reliable results for all methods (35/38 data points for revPBE/PBE0).

^{119}Sn NMR chemical shifts compared to geometry optimization with a fast ECP treatment. This observation underlines the excellent performance of r^{2}SCAN-3c for geometry optimization and in general the robustness of ECP-based geometry optimizations. In line with our previous study on ^{29}Si NMR chemical shifts, the choice of the method for the NMR chemical shift calculation has a much larger impact on the result than the geometry optimization method as long as a reasonably robust DFT approach, such as r^{2}SCAN-3c or TPSS-D4, is used.

Larger deviations are observed for the use of geometries optimized with GFN2-xTB (MAD of 161.5/147.4 ppm for revPBE/PBE0) and GFN-FF (MAD of 146.2/178.7 ppm). It is important to note that not all geometry optimizations using SQM or FF methods were successful. In some cases, especially the GFN-FF potential energy surface differs severely from the DFT-based one, resulting in qualitatively wrong structures or large deviations. Those cases have been excluded from the statistics to avoid an artificial bias. In order to still provide a fair method comparison, a second evaluation only considers those compounds that yielded reliable chemical shifts for all methods. In this case, the $MAD_{red.}$ is 146.9/139.8 ppm for GFN2-xTB and 148.2/178.3 ppm for GFN-FF.

The ^{119}Sn NMR chemical shifts calculated with SO-ZORA/PBE0 indicate a similar behavior of the deviations as for ^{29}Si NMR chemical shifts, which is a decreasing MAD in the order GFN-FF > GFN2-xTB > DFT.[213] As in the preceding study, a compensating linear scaling approach can also be applied here. Figure B.10 depicts the respective correlation plots with scaling parameters α and β for SO-ZORA/PBE0 and geometries on the GFN2-xTB and GFN-FF levels (the revPBE analogous figure can be found in Figure S1 in the Supporting Information). Upon application of linear scaling, the MAD decreases from 147.4 to 102.8 ppm for GFN2-xTB and from 178.7 to 144.0 ppm for GFN-FF. Consequently, the low-cost GFN methods in combination with a simple linear scaling on the SnS51 set represent a promising approach for the combination of SQM geometries and DFT chemical shift calculations for large systems. Such combinations have the potential of reaching similar accuracies as some of the tested fully DFT-based approaches but at a reduced computational cost.

B.4 Conclusion

The reliable calculation of ^{119}Sn NMR chemical shifts is of great importance to researchers working with new organotin compounds in order to facilitate the experimental measurement, signal assignment, and structure elucidation. The presented extensive benchmark study includes 50 representative tin-containing compounds (51 chemical shifts) with various bonding situations and provides an overview of the performance of commonly applied density functional approximations and relativistic approaches with respect to the calculation of ^{119}Sn NMR chemical shifts. Furthermore, structural influences on this property were studied via the assessment of conformer ensembles as well as different levels of theory for the geometry optimizations.

For the computation of ^{119}Sn NMR chemical shifts, no typical performance ordering

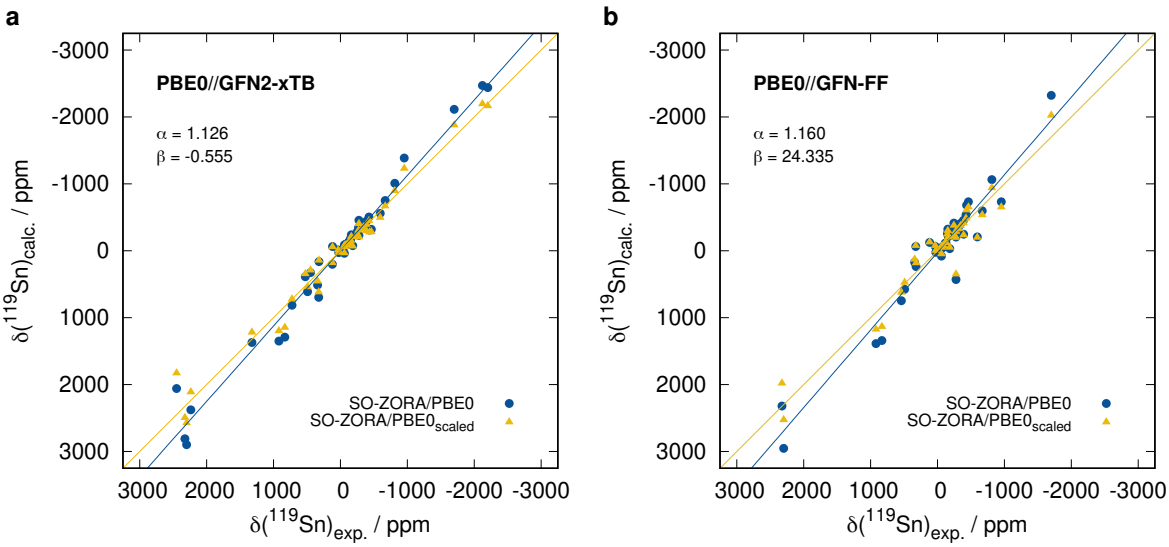


Figure B.10: Linear scaling correction applied to SO-ZORA/PBE0 ^{119}Sn NMR chemical shifts on the lowest-lying conformers optimized with (**a**) GFN2-xTB and (**b**) GFN-FF. The uncorrected/corrected MAD values are 147.4/102.8 ppm for GFN2-xTB and 178.7/144.0 ppm for GFN-FF.

behavior according to "Jacob's ladder" is observed, meaning that higher-rung hybrid DFAs do not generally outperform (meta-)GGA type functionals. More importantly, the explicit inclusion of relativistic effects applying a spin-orbit relativistic Hamiltonian is strongly recommended, as it leads to a significant improvement over the scalar-relativistic variants of ZORA and X2C. The application of ECPs yields qualitatively wrong results and should be avoided in this context. In general, covering spin-orbit effects by, for example, SO-ZORA is indispensable when heavy atoms such as iodine are bound to the respective tin nucleus. Especially SO-ZORA/revPBE and SO-ZORA/PBE0 are recommended with small total MADs of 96.8 and 87.6 ppm, respectively. The robust performance of these approaches is in line with our previous study on the calculation of ^{29}Si NMR chemical shifts. If the SO-ZORA approach can not be applied (e.g., because it is computationally unaffordable for a certain application or it is simply not available in the applied quantum chemistry code), we recommend to at least use a scalar-relativistic approach like SR-X2C or SR-ZORA with the respective optimized basis sets. The PBE family of functionals (PBE, revPBE, PBE0) performs well in all three relativistic categories, whereas further reliable choices are BP86 and B3LYP for SR-ZORA and BP86, ωB97X, and M06 for SR-X2C. Further, all assessed methods can be corrected empirically by applying a simple linear scaling approach which is especially worthwhile for the ones with a low performance. Respective scaling parameters are given in this study. The findings of the presented work should serve as a guidance for choosing a quantum chemical treatment that can be applied reliably to diverse tin compounds that go far beyond the examples tested here (cf. Supporting Information, Section 4.1).

The role of conformer ensembles for the ^{119}Sn NMR chemical shift calculation was investigated in detail. It was found that the effect of including conformers is usually small as most

structural divergence is found for the outer ligand periphery and not at the bonding site of the tin atom. Nevertheless, specific systems were found in which the effect of different bonding motifs near the Sn nucleus significantly influences the results, for example, with changes of a few hundred ppm for different conformers. This is generally expected for situations with loosely bound ligands allowing a change in the coordination sphere. In any case, the applied CREST/CENSO approach generally allows for a reliable sampling of the conformational space also for inorganic main group elements, and hence, conformer ensembles in chemical shift calculations can be routinely investigated.

For the geometry optimization, the inclusion of explicit relativistic Hamiltonians did not lead to a significant improvement in the subsequent chemical shift calculations. The efficient r^2SCAN-3c composite DFT scheme using default ECPs within its modified mTZVPP basis set is therefore recommended. The application of efficient semiempirical or force field methods such as GFN2-xTB and GFN-FF causes larger deviations but can still be of great use for more extended systems when applying a linear scaling of the ^{119}Sn NMR chemical shifts based on the new SnS51 data set.

In conclusion, the presented comprehensive benchmark study covers a large diversity of tin compounds and many commonly used computational methods, which makes it a firm basis for new perceptions about ^{119}Sn NMR chemical shift calculations. The benefit of SO approaches such as SO-ZORA with efficient density functionals such as PBE0 or revPBE is demonstrated. With typical errors below 100 ppm, we provide researchers a well-founded method recommendation and a broad data set for testing future developments.

Supporting Information

The Supporting Information is available free of charge at
https://pubs.acs.org/doi/10.1021/acs.inorgchem.1c03453.

- Further computational details, and tabulated data (PDF)
- Cartesian coordinates in XYZ format (ZIP)

Acknowledgements

The authors thank Prof. Dr. Yitzhak Apeloig and Prof. Dr. Philip P. Power for bringing the topic of this study to our attention. The authors further thank Fabian Bohle for fruitful discussions. J.B.S. is most grateful to the "Fonds der Chemischen Industrie (FCI)" for financial support. This work was supported by the Deutsche Forschungsgemeinschaft (DFG) in the framework of the "Gottfried Wilhelm Leibniz Prize" awarded to S.G.

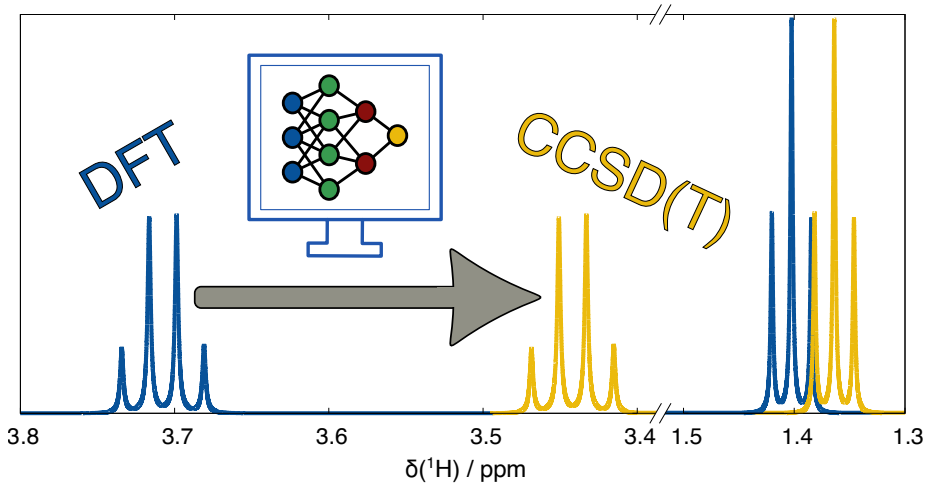
Conflict of Interests

The authors declare no competing financial interest.

APPENDIX C

Appendix: Computation of CCSD(T)-Quality NMR Chemical Shifts via Δ-Machine Learning from DFT

J. B. Kleine Büning and S. Grimme, *J. Chem. Theory Comput.* **2023**, *19*, 3601–3615.
DOI: 10.1021/acs.jctc.3c00165.



Abstract

NMR spectroscopy undoubtedly plays a central role in determining molecular structures across different chemical disciplines, and the accurate computational prediction of NMR parameters is highly desirable. In this work, a new Δ-machine learning approach is presented to correct DFT-computed NMR chemical shifts using input features from the calculation and in addition highly accurate reference data at the CCSD(T)/pcSseg-2 level of theory with a basis set extrapolation scheme. The model is trained on a data set containing 1000 optimized and geometrically distorted structures of small organic molecules comprising most elements of the first three periods and containing data for 7090 ^{1}H and 4230 ^{13}C NMR chemical shifts. Applied to the PBE0/pcSseg-2 method, the mean absolute deviation (MAD) on the internal NMR shift test set is reduced by 81% for ^{1}H and 92% for ^{13}C at virtually no additional computational cost. For twelve different DFT functional and basis set combinations, the MAD of the ML-corrected NMR shifts ranges from 0.021 to 0.039 ppm (^{1}H) and from 0.38 to 1.07 ppm (^{13}C). Importantly, the new method consistently outperforms the simple and widely used linear regression correction technique. This behavior is reproduced on three different external benchmark sets, confirming the generality and robustness of the correction scheme which can easily be applied in DFT-based spectral simulations.

C.1 Introduction

The use of nuclear magnetic resonance (NMR) spectroscopy for structure elucidation is essential for the everyday life of chemists. The method is as important for the routine study of small and simple analytes as it is for large, challenging structures in complex environments. Because of the immense importance of the information obtained from NMR spectroscopy, the computation of chemical shielding tensors via quantum mechanics (QM)-based methods is part of the standard repertoire of quantum chemistry codes.

One of the most accurate and still feasible theories for calculating chemical shielding constants of fixed, gas-phase molecular structures is the coupled cluster approach, which should include at least singles, doubles, and perturbative triples contributions (CCSD(T)).[43] Since the computational cost of such methods is too high for most chemically interesting systems, density functional theory (DFT) has emerged as an efficient compromise[24,220] that has the potential to achieve reasonable accuracy for ^{1}H and ^{13}C NMR shielding constants,[33,41,178] especially when double hybrid density functionals are applied.[39] However, the accuracy of DFT is often too low for reliable spectra predictions.

The field of machine learning (ML) within computational chemistry has grown very rapidly in the past decade.[59] Application examples are found in various areas, such as molecular property prediction,[61] crystal structure determination via NMR chemical shifts,[366] prediction of vibrational spectra,[367] and a complete general density functional approximation (DM21)[71]. Since quantum chemistry-based methods for predicting NMR parameters can have serious drawbacks, such as insufficient accuracy or high computational cost (depending on the systems of interest), the use of artificial neural networks (ANN) and machine learning for this purpose has become increasingly popular in recent years. NMR chemical shift computations purely via ML approaches have been performed for condensed-phase[75,76] and molecular systems for which graph neural networks[77,78] or kernel ridge regression, as implemented in the IMPRESSION[79,80] software, can be a suitable choice.

However, quantum chemistry and ML need not be antagonists but can work together in the concept of Δ-machine learning, where a QM-calculated value, e.g., from DFT, is corrected via an ANN. This is particularly advantageous when the targeted deviation is small and has the benefit of QM-based information being transferred to the ANN. In the context of NMR, the ANN can be trained to provide the expected deviation (Δ) of the chemical shift from the DFT approach to the reference data, which can be obtained experimentally[90] or by computations at a higher level of theory.[88] As far as the application for NMR chemical shifts is concerned, the quality of a (Δ)-ML model depends quite significantly on the algorithm chosen.[180] Therefore, various types of machine learning approaches have been investigated, including graph convolutional networks,[181] transfer learning,[368] and message passing.[182] The use of machine learning for NMR parameters is generally not limited to shielding constants or chemical shifts but can also help in NMR-based structural assignment[84,86] or analysis of 2D experiments.[87]

In this work, we present a general Δ-ML model capable of predicting the deviation of ^{1}H and ^{13}C NMR chemical shift values of an arbitrary low-level (DFT) method to the highly

accurate CCSD(T) method with a complete basis set extrapolation. In contrast to the use of experimentally obtained reference values, this allows the straightforward generation of large amounts of reference data and the disentanglement of electron correlation from other sources of NMR shift deviations. The workflow is evaluated for a range of DFT levels but is not limited to those shown in this work. We provide a ready-to-use prelearned model for each of the DFT levels shown and the infrastructure for rapidly building new model on other levels of theory.

C.2 Methodology

C.2.1 ML Data Set

To obtain a well-controllable correction model based on a large and reliable database, we decided to use computed high-level reference data rather than experimentally determined NMR chemical shifts as the prediction target (for more details, see Section C.2.2). The systems included in the data set were chosen to be small enough to be calculated on the high-accuracy CCSD(T) level of theory using a triple-ζ all-electron basis set. Still, they were chosen to be relatively large to ensure the best possible transferability to larger systems. Since NMR parameters represent an electronically rather localized property, a system size of up to 22 atoms (propane dimer, system **72**) or 66 electrons (e.g., thymine, system **50**) is considered sufficiently large for this purpose. It can be expected that long-range effects appearing in practical applications for large systems are covered by the underlying DFT treatment.

In ML applications in chemistry, the limited amount of training data is often a major obstacle to a successful model. Since the reference data of the presented model are taken from calculations rather than experiments, the molecular structure does not necessarily have to be reasonable, and a nearly infinite space of possible structural candidates is available. We selected 100 small organic molecular systems that were structurally optimized at the r^{2}SCAN-3c[172] composite DFT level of theory. The systems were selected primarily to assemble a chemically very diverse data set, containing nearly all elements in the first three rows of the periodic table. Subsequently, the structure of each molecule was randomly geometrically distorted several times to varying degrees to produce nine different non-equilibrium structures for each parent molecule (see Figure C.1 for an example; for more information on the distortion technique, see the Supporting Information, Section 1.1). The relative energies of these distorted structures were chosen in a range between 1 and a maximum of 50 kcal·mol^{-1} above the optimized structure. Suspected chemically equivalent nuclei were not averaged because they may differ significantly in the distorted structures. With this approach, the compiled training data set contains a total of 1000 structures (based on 100 different molecules) including 7090 data points for ^{1}H (only hydrogen atoms bound to carbon atoms are considered) and 4230 for ^{13}C NMR chemical shifts. The relevant data discussed in Section refsubsec:ann were obtained directly from the Cartesian structure files and the DFT calculation output. A detailed list of the systems included can be found in the

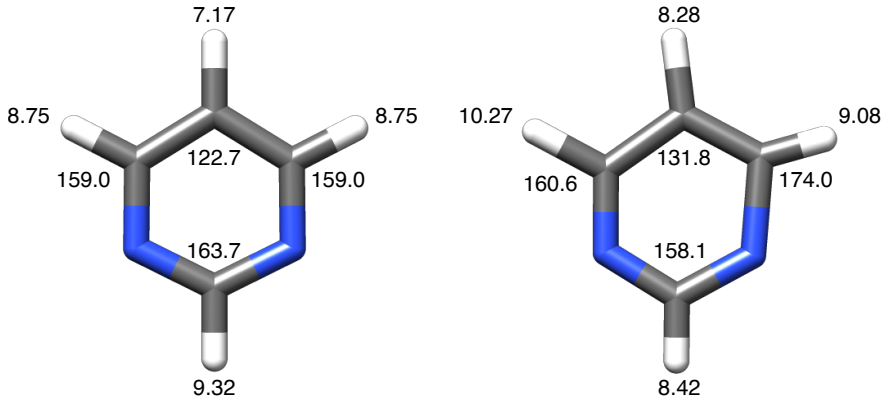


Figure C.1: Example compound (pyrimidine, **014**) showing the optimized (left) and a distorted (+41.7 kcal·mol^{-1}) structure (right) and the respective ^{1}H and ^{13}C NMR chemical shifts in ppm at the reference CCSD(T)/TZ+ level of theory (see below).

Supporting Information, Table S2.

C.2.2 Reference Level of Theory

Since the actual NMR experiment takes place in a rather complex environment, mainly involving a solvent and a full ensemble of possible conformations of the analyte molecule at a finite temperature, the full computational simulation is challenging.

Within the five main sources of error for NMR chemical shift predictions formulated by Lodewyk et al.[24], the electron correlation contribution, captured, e.g., by CCSD(T), is only one, albeit important, aspect. Further error sources are rotational-vibrational effects (which are expected to play a minor role when dealing with relative chemical shifts), heavy-atom effects, and solvation effects. The latter two are not considered in this study (due to the lack of heavy atoms in the data set and the lack of a solvation model for the CCSD(T) shielding calculations) but can in principle be added via standard solvation and relativistic models. The last major error source is conformational flexibility, which is considered, e.g., by the CREST[25] (conformer-rotamer ensemble sampling tool) approach in combination with CENSO[29] (command-line energetic sorting of conformer-rotamer ensembles) that can be used to calculate ensemble-based NMR spectra.[30] In addition, it has been suggested that other dynamic effects are of great importance especially for solid state NMR shift predictions,[369] where their contribution to the final value can be larger than the contribution of the electron correlation.[370] In these cases, (path integral) molecular dynamics simulations can be used to treat the dynamic effects.

However, the presented method does not target dynamic effects or the outcome of the whole CREST/CENSO workflow, which is the result of many individual calculations for geometry optimizations, free energy contributions, chemical shielding calculations, etc. Rather, it aims to improve the quality of the calculated NMR chemical shift for a given molecular structure to improve the signal position in the resulting spectra, thereby reducing

the complexity of the problem to a mere chemical shift correction using high-level computed reference data.

One of the most accurate methods for calculating NMR chemical shielding tensors of the systems mentioned above is the coupled cluster approach with at least singles and doubles contributions and perturbative triples correction (CCSD(T)).[43,44,46,179] While the improvement over CCSD from taking into account the triples contributions perturbatively in CCSD(T) is still significant, the computationally much more demanding step to full triples or the CCSDT-n methods does not pay off in terms of accuracy and efficiency.[43] Therefore, CCSD(T) is used as an accurate and reliable reference for the presented ML approach. Unfortunately, its chemical shielding tensor cannot be considered converged to the complete basis set (CBS) limit when only triple-ζ (TZ) basis sets are used.[44,46,179] Therefore, we use a basis set extrapolation scheme based on the protocol of Kupka et al.[371] using extrapolated DFT values. According to their results, the error in the chemical shielding constant caused by the basis set incompleteness error (BSIE) is assumed to be the same for each of the 47 methods tested, including many DFAs and CCSD(T). Thus, the chemical shielding constant at the CCSD(T) level extrapolated to the CBS $\sigma_{\mathrm{CC}}^{\mathrm{CBS}}$ results from the CCSD(T)/TZ value $\sigma_{\mathrm{CC}}^{\mathrm{TZ}}$ and the DFT-based extrapolated CBS correction $\Delta\sigma_{\mathrm{DFT}}^{\mathrm{CBS}}$ as

$$\sigma_{\mathrm{CC}}^{\mathrm{CBS}} = \sigma_{\mathrm{CC}}^{\mathrm{TZ}} + \Delta\sigma_{\mathrm{DFT}}^{\mathrm{CBS}} \tag{C.1}$$

with

$$\Delta\sigma_{\mathrm{DFT}}^{\mathrm{CBS}} = \sigma_{\mathrm{DFT}}^{\mathrm{CBS}} - \sigma_{\mathrm{DFT}}^{\mathrm{TZ}} \quad . \tag{C.2}$$

The CBS correction is obtained from DFT chemical shielding calculations using the same TZ basis set $\sigma_{\mathrm{DFT}}^{\mathrm{TZ}}$ and the CBS-extrapolated value $\sigma_{\mathrm{DFT}}^{\mathrm{CBS}}$. The total NMR shielding constant calculated at a DFT level can be extrapolated to the complete basis set limit for both correlation-consistent and polarization-consistent basis sets.[372–374] In the approach mentioned above,[371] the method uses the average of the values of the two functionals B3LYP and BHandH (both TZ and CBS extrapolated values) to obtain $\Delta\sigma_{\mathrm{DFT}}^{\mathrm{CBS}}$. However, the use of these explicit functionals is stated to be somehow arbitrary, and in principle, the use of one single functional can be considered sufficient and saves computational effort. The CBS extrapolation at the DFT level in this work was performed via a two-parameter fit to the inverse cubic function[375]

$$\sigma_{\mathrm{DFT}}^{x} = \frac{A}{x^3} + B \quad , \tag{C.3}$$

where $\sigma_{\mathrm{DFT}}^{x}$ is the shielding constant calculated at basis set level x with parameters A and $B = \sigma_{\mathrm{DFT}}^{\mathrm{CBS}}$. For the DFT/CBS approach of the chosen reference method, (only) the BHLYP/pcSseg-n[106,107,136] density functional approximation was chosen to stay close to the original method.[371] The shielding constants were extrapolated via the pcSseg-n ($n = 2, 3, 4$) basis sets with triple-ζ, quadruple-ζ (QZ), and pentuple-ζ (5Z) quality and – to be consistent with the literature for polarization-consistent basis sets[371] – the corresponding basis set levels were chosen to be $x = 4, 5, 6$, respectively.[376] Considering the high amount of Hartree-Fock

exchange in BHLYP (50%) and the use of the pcSseg-*n* basis sets, the presented approximation to CCSD(T)/CBS is similar to a composite scheme proposed in 2015 by Reid et al.[377]

Moreover, the presented extrapolation scheme proved to be robust in an internal test with canonical CCSD(T)/CBS(pcSseg-3/4) reference values calculated for a small test set based on the study of Teale et al.[178] In the following, the reference method will be abbreviated as CCSD(T)/TZ+ to indicate that, especially for the ^{1}H NMR shielding, it should be considered a CCSD(T)/TZ approach with an additional CBS correction rather than a true CCSD(T) method in the complete basis set limit.

C.2.3 Neural Network Architecture and Input Feature Vector

The presented artificial neural network was built as a multilayer perceptron within the framework of TensorFlow 2.7.0 and designed to be easily fine-tunable concerning the key hyperparameters. However, the model was chosen to consist of two hidden layers, and a grid search was carried out to determine the most efficient settings. Thus, the default model has 120 or 80 nodes (for ^{1}H and ^{13}C, respectively) in the first and eight in the second layer, a dropout rate of 0.15 (first layer), and uses the `adam` optimizer with the mean absolute deviation (MAD) as loss function and the softmax activation function in both layers. These settings were used throughout this work unless stated otherwise. The implementation in a Python framework and the data set are available at `https://github.com/grimme-lab/ml4nmr`.

The goal of the ML model is to learn and predict the deviation $\Delta\delta$ of an NMR chemical shift δ_{low} computed at a *low* computational level of theory (i.e., DFT with a triple-ζ basis set) and the *high-level* reference δ_{high} computed with the CCSD(T)/TZ+ approach presented above. For the use as target, the calculated NMR shielding constant σ is converted into a chemical shift δ using a reference compound calculated at the same level of theory (which is tetramethylsilane (TMS), if not stated otherwise). In the context of the model, each atom represents an individual data point that is affected by its environment but does not explicitly depend on the molecule to which it belongs (no explicit molecule-specific information is included). All available information is collected in the input feature vector, whose dimensionality depends on the nature of the nucleus of interest. Besides the reference value, each data point contains 24 variables in the case of ^{1}H and 31 variables for ^{13}C NMR shifts.

In general, input variables can be divided into three categories: *geometric*, *electronic*, and *magnetic* properties (a complete list of all included information can be found in the Supporting Information, Table S1). The information grouped under *geometric* properties contains the coordination number $\mathrm{CN}(A)$ of each atom as introduced in the D3 dispersion correction model[378] which is defined as

$$\mathrm{CN}(A) = \sum_{B \neq A}^{N_{\text{at}}} \frac{1}{1 + \exp\left(-k_1\left(\frac{k_2 R_{AB}^{\text{cov}}}{R_{AB}} - 1\right)\right)} \tag{C.4}$$

with the number of atoms in the molecule N_{at}, the distance R_{AB} between atoms A and B,

the sum of the covalent radii $R_{AB}^{\mathrm{cov}} = R_A^{\mathrm{cov}} + R_B^{\mathrm{cov}}$, and the parameters $k_1 = 16$ and $k_2 = \frac{4}{3}$. The covalent radii were taken from Ref. [378] and slightly modified by –0.15 Å for Li and –0.05 Å for C, respectively. Also included are the bond length of a given hydrogen atom to its neighboring carbon atom and the number of hydrogen, carbon, nitrogen, or oxygen atoms bonded directly to the atom in question or to all neighbors of the atom in question. Further, the use of more sophisticated geometry-based descriptors from atom-centered symmetry functions (ACSF)[379] and the smooth overlap of atomic positions (SOAP)[380] has been considered. Since these input features did not lead to a generally improved performance of the Δ-ML model, they are not included here. Detailed information and the tests can be found in the Supporting Information, Section 4.1. It should be noted that geometric information is also indirectly included by some of the *electronic* and *magnetic* descriptors that strongly depend on the three-dimensional structure of a compound.

The idea of including *electronic* properties is to implicitly consider the converged density matrix from the preceding DFT calculation. This is achieved by using descriptors that depend on the density matrix, such as atomic charges. Besides (Mulliken and Löwdin) charges, atom-resolved shell populations are also considered, in particular for the *s*-, *p*-, and *d*-orbitals as well as the standard deviation of the three components of the *p*-orbital populations. In addition, the bond orders (Löwdin and Mayer models) are considered as the sum and/or average of all neighboring atoms as well as Mayer's total valence quantity.

The *magnetic* properties mainly contain information about the chemical shielding tensor, which is calculated as a second-order property subsequent to the *low-level* energy calculation. Since the quantities derived from the shielding tensor are not uniformly defined in the literature,[381,382] the quantities used in the present work are given in Table S1 with the total NMR shielding tensor $\boldsymbol{\sigma}$ defined as

$$\boldsymbol{\sigma} = \begin{pmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{pmatrix} \tag{C.5}$$

with

$$\sigma_{11} \leq \sigma_{22} \leq \sigma_{33} \quad . \tag{C.6}$$

The total isotropic shielding constant σ_{iso} is $1/3$ the trace of the shielding tensor:

$$\sigma_{\mathrm{iso}} = \frac{\sigma_{11} + \sigma_{22} + \sigma_{33}}{3} \quad . \tag{C.7}$$

The span Ω, skew κ, asymmetry η, and anisotropy Δ result from $\boldsymbol{\sigma}$ and σ_{iso} as defined in Table S1. Incidentally, the total isotropic shielding tensor results from the combination of its dia- and paramagnetic parts $\sigma_{\mathrm{iso,dia}}$ and $\sigma_{\mathrm{iso,para}}$, which are also included in the input vector.

C.2.4 Computational Details

Most of the compounds in the data set were created manually, then a conformer ensemble was generated with the conformer-rotamer ensemble sampling tool (CREST),[25,310] version 2.11.2, using the GFN-FF[28] force field and the GFN2-xTB[27] semiempirical tight-binding method. The ensembles were refined using the command-line energetic sorting of conformer-rotamer ensembles (CENSO)[29,311] algorithm, version 1.1.2. Some of the structures were taken from external sources;[383,384] in these cases, the ensemble sampling step was skipped. Further details are provided in the Supporting Information, Table S2. The conformer with the lowest free energy of the ensemble or a special other selected structure was then reoptimized with the TURBOMOLE 7.5.1[307–309] program package using the r^{2}SCAN-3c[172] density functional composite method that is based on the r^{2}SCAN[105] meta-GGA functional and uses the modified mTZVPP basis set. The `m4` grid and radial grid size 10 and the resolution of the identity scheme for Coulomb integrals (RIJ)[255] were used throughout these geometry optimizations without making use of a solvation model.

All NMR shielding constant calculations have been performed using the gauge-including atomic orbital (GIAO)[137,138] method that has been adapted for both DFT[139] and coupled cluster theory.[179,385,386] The DFT shielding calculations have been performed with the ORCA 5.0.3[243,244,387] program package applying the PBE,[101] KT2,[160] revTPSS,[104,263] PBE0,[109] BHLYP,[106,107] M06,[161] r^{2}SCAN0,[105,198] LC-BLYP,[388] and ωB97X-V[389] density functional approximations in different combinations with the Karlsruhe-type def2-TZVP,[156] the correlation-consistent cc-pVTZ,[163,390] and Jensen's polarization-consistent segmented contracted pcSseg-n ($n = 1, 2, 3, 4$) basis sets.[136] Different resolution of the identity (RI) acceleration approaches were applied, namely RIJ[255] (Coulomb integrals, applied for PBE, KT2, revTPSS), RIJK[228,391] (Coulomb and exchange integrals, applied for PBE0, BHLYP, M06, r^{2}SCAN0), and RIJCOSX[228,392] (chain-of-spheres algorithm for exchange, applied for LC-BLYP), no RI approximation was applied for ωB97X-V. The RIJK approximation was in general preferred over RIJCOSX because it was found to be more robust and efficient for very small systems[393] such as those included in the data set. Auxiliary basis sets were generated automatically by the `AutoAux` algorithm.[394] NMR chemical shielding calculations on the CCSD(T)[97,179,395,396] level were performed with the CFOUR[397,398] program package version 2.1 applying the triple-ζ pcSseg-2 basis set[136] (the basis set files were downloaded from the basis set exchange library[399]).

C.3 Results and Discussion

C.3.1 Correction of ^{1}H NMR Chemical Shifts

As a first step, the performance of the Δ-ML model is investigated for the ^{1}H nucleus with the *low-level* method being PBE0/pcSseg-2. The structures contained in the data set described in Section C.2.1 are divided into a training set (87.5% of the data points) and a test set (12.5% of the data points), which is unknown to the neural network and used to

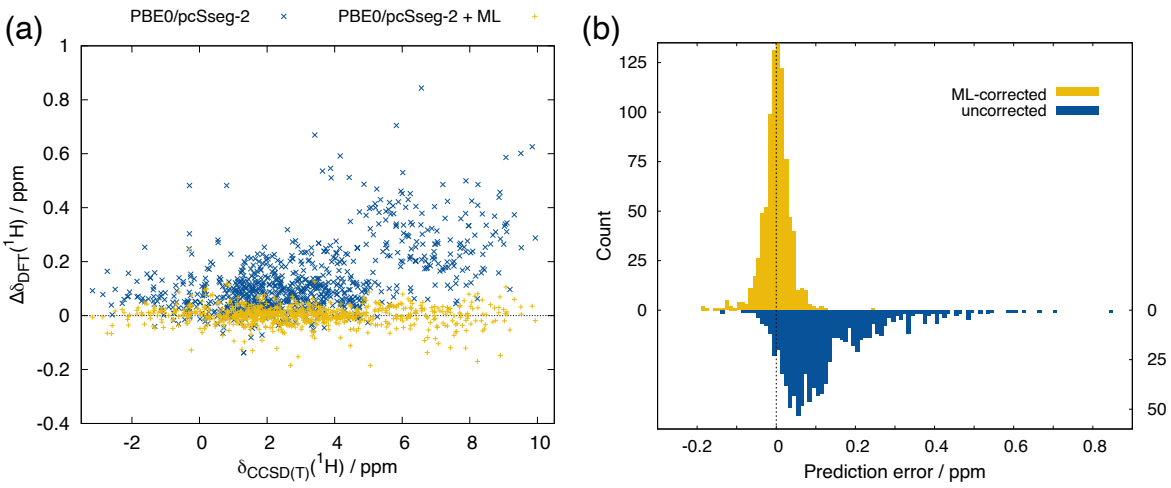


Figure C.2: Comparison of ^{1}H NMR chemical shifts calculated at the PBE0/pcSseg-2 level to the CCSD(T)/TZ+ reference: (a) Deviations $\Delta\delta$ of the uncorrected and corrected method, (b) distribution of errors (test set, 12.5% of the data points, $\Delta\delta_{DFT} = \delta_{DFT} - \delta_{CCSD(T)}$, bar width: 0.01 ppm).

evaluate its performance. The deviations in chemical shift from the original and corrected calculations are plotted against the reference values in Figure C.2(a). A clear reduction of the respective deviation is observed throughout the entire range of typical ^{1}H NMR chemical shifts (≈ –2 to +10 ppm). The mean absolute deviation (MAD) of the uncorrected chemical shifts of the test set calculated with PBE0/pcSseg-2 compared to CCSD(T)/TZ+ is 0.130 ppm and is reduced by 81% to 0.025 ppm by applying the Δ-ML correction. Furthermore, as can be seen in Figure C.2(b), the error spread of the predicted NMR chemical shifts is much more narrow and Gaussian-shaped with the Δ-ML correction (standard deviation (SD): 0.034 ppm). Without the correction, the *low-level* DFT method has a larger error range while it strongly overestimates the chemical shift (SD: 0.121 ppm).

However, it must be stated that the Δ-ML model is not purely deterministic but contains some randomness in building up the model. In addition, the shuffling of the data set before the division is done randomly and will have an impact on the final result. These technical influences are discussed in the following.

Technical Influences Lead to Fluctuations

Randomness included in setting the initial weights on the nodes while building the ML model introduces fluctuations to its performance. To account for this, the results (e.g., the MAD of the test data set) shown further on (and the ones shown above) represent the statistical mean of ten prediction runs with different random seeds in the model building procedure. For the examples in Figures C.2 and C.3, the random seed was fixed to the value of 0. Another bias in the ML model arises from the arrangement of the data between the training and test data sets, which is discussed in detail in the Supporting Information, Section 3.3. It is guaranteed that each chemical shift in a structure is only assigned to the

training *or* the test data set.

So far, the performance of the Δ-ML model has only been evaluated on the test data set (12.5% of the total data set), which is unknown to the model during the training process. This needs to be compared to the performance on the training data set to estimate the bias of the trained model. An overfitted model performs much better on the training than on the test set, indicating that it is not sufficiently generalized. In the example given above, the MAD of the ^{1}H chemical shifts in the training data set calculated with PBE0/pcSseg-2 is 0.145 ppm and is reduced by 84% to 0.023 ppm. This represents a slight but generally acceptable deterioration of the model on the test data set (MAD reduction of 81%).

Comparison to Linear Regression

To set the performance of the developed Δ-ML approach in comparison to another common method for correcting errors in chemical shielding calculations, a simple scaling approach was applied using the same training and test set. In this approach, the training data is fitted to a linear function with parameters α and β

$$\delta_{\text{low}} = \alpha \cdot \delta_{\text{high}} + \beta \quad . \tag{C.8}$$

The approach has been used in past studies[24,32,285,287,400,401] for the correction and improvement of ^{1}H and ^{13}C NMR chemical shifts. Furthermore, the procedure seems to be general as it was also used in a study on ^{31}P NMR shifts[37] as well as in our studies of the ^{29}Si[213] and ^{119}Sn[199] nuclei. In the example of PBE0/pcSseg-2 as the *low-level* method shown above, the MAD evaluated on the test set was reduced from 0.130 to 0.025 ppm (–81%) by applying the Δ-ML correction. Performing a linear regression (LR) on the training set ($\alpha = 1.0307, \beta = 0.0496$ ppm, $R^2 = 0.9981$) and applying it to shift the calculated results of the test set leads to an MAD of 0.077 ppm (–41%). For the error spread, the trend is even more pronounced, as the SD with the LR-correction is 0.100 ppm (no correction: 0.121 ppm, Δ-ML: 0.034 ppm). Thus, the simple regression model proves to be helpful also on the presented data set for ^{1}H NMR shifts but is clearly outperformed by the more sophisticated Δ-ML model.

C.3.2 Correction of ^{13}C NMR Chemical Shifts

The focus of this study lies on the two most commonly measured nuclei, which are ^{1}H and ^{13}C. All evaluations that have been shown for hydrogen have also been performed for carbon to compare the performance of the presented Δ-ML approach for the two different experiments. The deviations of all data points in the test data set for the original DFT and the corrected chemical shifts are plotted against the *high-level* reference values, which lie in the typical ^{13}C NMR chemical shift range of ≈ –50 to +250 ppm and are presented in Figure C.3(a). Compared to ^{1}H, the clear overestimation of the chemical shift by DFT is again prominent but much more pronounced and in general, the deviations $\Delta\delta_{\text{DFT}}$ grow

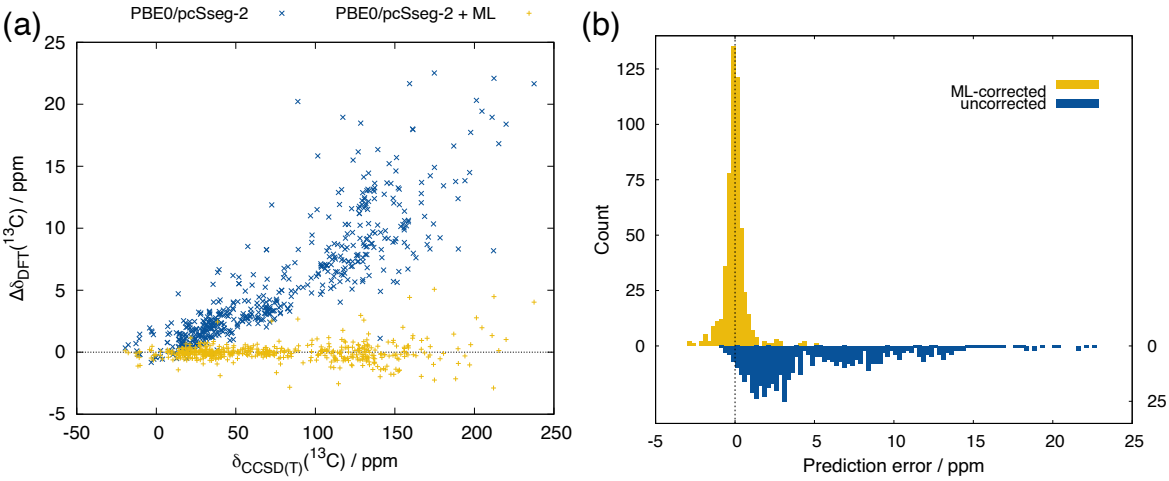


Figure C.3: Comparison of ^{13}C NMR chemical shifts calculated at the PBE0/pcSseg-2 level to the CCSD(T)/TZ+ reference: (a) Deviations $\Delta\delta$ of the uncorrected and corrected method, (b) distribution of errors (test set, 12.5% of the data points, $\Delta\delta_{DFT} = \delta_{DFT} - \delta_{CCSD(T)}$, bar width: 0.25 ppm).

systematically with the chemical shift. This is corrected by the Δ-ML model, reducing the original MAD of 5.39 ppm obtained by evaluating PBE0/pcSseg-2 on the test data set to 0.42 ppm (–92%, mean of ten runs, SD: reduced from 4.69 to 0.66 ppm). A linear regression was also performed for the ^{13}C data (α = 1.0768, β = –0.3088 ppm, R^2 = 0.9983) and achieves a reduction of the test set MAD to 1.48 ppm (–73%, SD: 2.11 ppm). Although a near-linear dependence of the error of the chemical shift on its reference value is apparent (cf. Figure C.3(a)), the neural network is able to adapt more efficiently to the systematic dependencies in the underlying data.

Furthermore, the error spread reduced by the Δ-ML model is depicted in Figure C.3(b) and underpins the quality of the correction approach. The standard deviation of errors in the test set is 4.69 ppm for the uncorrected DFT results and is reduced to 0.66 ppm for the Δ-ML corrected values.

Analogously to hydrogen, the performance of the Δ-ML model on the actual training data set is investigated. For this, the MAD obtained by the uncorrected PBE0/pcSseg-2 method is 5.84 ppm, and the correction gives a reduction to 0.42 ppm (–93%). Since this is almost the same relative error reduction as for the test set, the model is well-generalized and does not seem to suffer from overfitting.

The overall improvement of the calculated ^{13}C NMR chemical shift by the Δ-ML method is larger than it is for ^{1}H NMR (this will also become apparent in Section C.3.3). These findings are a result of the nuclear magnetic shielding phenomenon of the ^{1}H nucleus being more complex than for ^{13}C, which is more shielded from its environment and thus less prone to small deviations in, for instance, the surrounding electronic structure. The better performance of the Δ-ML model for ^{13}C than for ^{1}H is thus a marker for the overall physical complexity that makes it hard to address all relevant influences and their interconnections,

"even" for a machine learning approach.

C.3.3 Methodological Influences

Evaluation of Different *Low-Level* DFT Methods

The technical procedure of the presented Δ-ML correction approach was chosen such that it is in principle applicable to any *low-level* DFT method. For technical reasons, the desired *low-level* functional currently has to be available in the ORCA 5.0.3 program package. For any particular combination of density functional approximation (DFA) and basis set, the whole set of structures has to be calculated once to build up the data set that is needed to train the Δ-ML model. A small number of different DFAs and basis sets was exemplarily evaluated for this purpose. The chosen DFAs include the PBE and KT2 generalized gradient approximations (GGA), the revTPSS meta-GGA (performed well on the NMRH148 benchmark[34]), and the LC-BLYP long-range corrected GGA (particularly good performance on a thorough ^{13}C NMR study[35]) as well as the PBE0, M06, r^{2}SCAN0, and ωB97X-V (range-separated) hybrid functionals in combination with the pcSseg-2 basis sets. In addition to pcSseg-2, the triple-ζ basis sets def2-TZVP and cc-pVTZ as well as the double-ζ pcSseg-1 and quadruple-ζ pcSseg-3 basis sets were tested for PBE0. Figure C.4 shows the performance of the presented Δ-ML approach based on the different method combinations chosen as *low-level* methods compared to the CCSD(T)/TZ+ reference values for both ^{1}H and ^{13}C NMR chemical shifts.

Throughout all investigated classes of DFAs, the Δ-ML correction reduces the mean absolute deviation on the test set drastically without any significant impairment for large initial errors. Thus, after applying the correction, all tested methods are roughly on the same performance level. The different applied triple-ζ basis sets do not show significant differences in the overall behavior for the ^{1}H nucleus. However, while the MAD on the test data set of the PBE0 functional decreases slightly by changing from pcSseg-2 to pcSseg-1 (0.130 to 0.106 ppm), the accuracy gain by the correction decreases both for the linear regression as well as for the Δ-ML method (–41% to –10% for LR and –81% to –63% for Δ-ML). This indicates that applying the smaller pcSseg-1 basis benefits from error compensation, which is not systematic and thus more difficult to correct for, independent from the kind of regression model. Interestingly, for the small organic molecules in this study, the basis set completeness of the pcSseg-*n* series seems to be converged at the triple-ζ level, as no significant difference is observed for pcSseg-3. In the case of the ^{13}C nucleus, the initial errors of using the conventional def2-TZVP and cc-pVTZ basis sets are smaller than with pcSseg-2, but both correction schemes achieve roughly the same final level of deviations for all basis sets. Thus, the relative error reduction is smaller, but the final result is somehow independent of the starting *low-level* basis set.

Compared to the different tested basis sets, the error range caused by the use of different density functionals (with the pcSseg-2 basis set) is larger with a much broader distribution of uncorrected MADs for ^{13}C (from 2.94 ppm (KT2) to 10.08 ppm (M06)) than for ^{1}H (from 0.110 ppm (ωB97X-V) to 0.195 ppm (PBE)). For both nuclei, the KT2 and revTPSS functionals

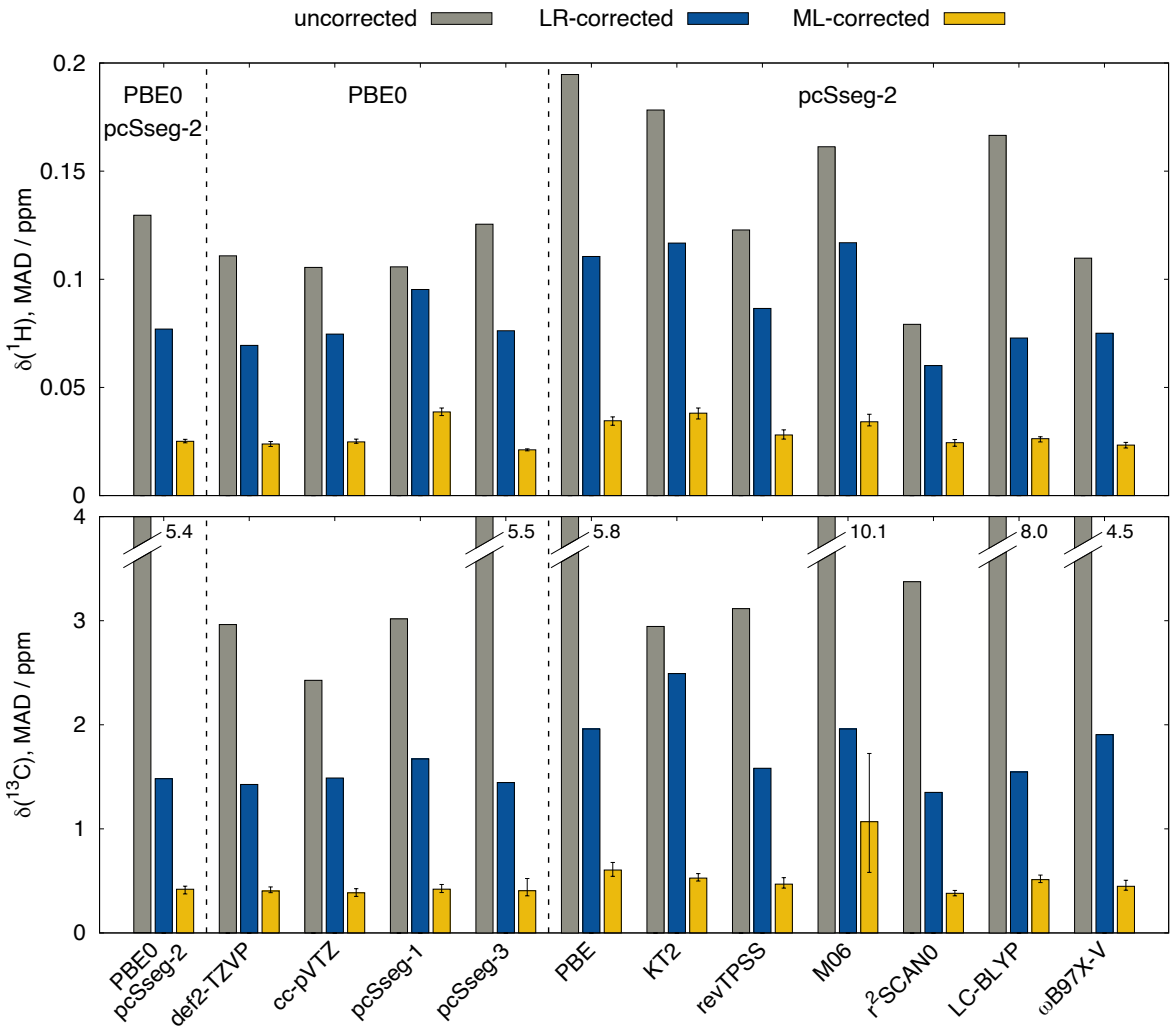


Figure C.4: Mean absolute deviations (MADs) of ^{1}H and ^{13}C NMR chemical shifts with different *low-level* method combinations compared to PBE0/pcSseg-2 (left). For each method, the plain DFT results (gray) are compared to applying a simple linear regression correction (LR, blue) and the presented Δ-ML approach (yellow, mean MAD of 10 runs, the error bars reflect the minimum and maximum values).

exhibit a similar yet less pronounced behavior as the pcSseg-1 basis set mentioned before, again indicating significant amounts of error compensation from which these DFAs in the pcSseg-2 basis set achieve generally high accuracy. After correction with the Δ-ML method, all ^{1}H NMR MADs are in a similar range from 0.023 ppm (ωB97X-V) to 0.038 ppm (KT2). For ^{13}C, the M06 functional seems to be an outlier with an MAD of 1.07 ppm, while the other DFAs are in the range of 0.38 ppm (r^{2}SCAN0) to 0.60 ppm (PBE).

Although the error range of the different functionals is small after applying the Δ-ML correction, the functional with the best final result for ^{1}H NMR is ωB97X-V and r^{2}SCAN0 for

^{13}C NMR. The overall best results for ^{1}H NMR are achieved with PBE0/pcSseg-3 serving as the *low-level* method (MAD of 0.021 ppm), while for ^{13}C, no other tested combination outperforms r^{2}SCAN0/pcSseg-2 (MAD of 0.38 ppm; 0.41 ppm for PBE0/pcSseg-3). At least as far as the basis set completeness is concerned, this indicates that the presented Δ-ML correction improves with improved input features, bringing the initial chemical shift already closer to the CCSD(T)/TZ+ reference. The dependence on the quality of the initial density functional is less clear. However, when the relative improvement by the Δ-ML method is addressed, the correction is especially valuable for *low-level* methods with large deviations such as LC-BLYP or M06. However, the latter is somehow special in the case of ^{13}C NMR. It is a highly parametrized DFA that was optimized for organometallic (thermo)chemistry[161] but not for the calculation of NMR chemical shielding constants. Although the Δ-ML correction does work here, it produces highly fluctuating MADs (ranging from 0.58 to 1.72 ppm within ten runs) caused by variations in the generated Δ-ML models. This behavior shows that the ^{13}C input features generated with M06/pcSseg-2 are not optimal and raises awareness for the problem of non-determinism in ML approaches. It is suggested to always check the reliability of a trained model before using it in relevant applications.

Role of the Input Features

The importance of the input features on the performance of the Δ-ML model for both ^{1}H and ^{13}C NMR is investigated by systematically omitting the groups of features that were introduced in Section C.2.3 in the training and prediction process. The results of this investigation are summarized in Table C.1.

The best results are obtained when all available descriptors are used in the framework of the Δ-ML model, which means that there is little redundant information fed into the model that could cause problems related to overfitting. For the ^{1}H nucleus, there is virtually no preference for the descriptor categories if only one of them is omitted. The MAD of the corrected values of the test set increases slightly from 0.025 to 0.032/0.033 ppm. This indicates that large influences might come from few descriptors, as omitting a whole category only causes a small loss of accuracy. Still, the effect is significant so that all descriptors contribute to the overall performance of the model. The same conclusions can be drawn for the ^{13}C chemical shift model, where the decrease in accuracy is also slight and in the same region for each descriptor category (MAD on the test set rises from 0.42 to 0.50-0.53 ppm).

If, by implication, only one group of input features is used for the model, the response of the model for the two investigated nuclei is different. Both nuclei now show opposing trends regarding the importance of the descriptor categories. For the ^{1}H nucleus, the *electronic* information seems to be of lower importance compared to the other categories, as the model shows the lowest accuracy when only these input features are used. The accuracy is higher when only *geometric* or *magnetic* features describe the model. In contrast, the model used for correcting ^{13}C NMR shifts reaches the best performance if only the *electronic* descriptors are included and the largest decrease in accuracy occurs, when only *geometric* descriptors are used. The *electronic* descriptor category contains atomic charges, shell populations, and

Table C.1: MAD (test data set) of the NMR chemical shifts obtained from PBE0/pcSseg-2 calculations and subsequent correction with the Δ-ML model. Every line only contains a certain subset of the descriptor categories mentioned above (linear regression-corrected and uncorrected values for comparison). The ML-based values are obtained as the statistical mean of ten runs with different random seeds.

Descriptor category included?			MAD [ppm]	
Geometric	Electronic	Magnetic	^{1}H	^{13}C
yes	yes	yes	0.025	0.42
no	yes	yes	0.032	0.50
yes	no	yes	0.033	0.52
yes	yes	no	0.033	0.53
yes	no	no	0.051	1.21
no	yes	no	0.057	0.76
no	no	yes	0.049	1.08
no	no	no	0.086	3.64
linear regression correction			0.077	1.48
uncorrected			0.130	5.39

bond orders. The amount of possible populations and bond orders of a carbon atom is much more diverse than for hydrogen. This and the higher number of descriptors included for ^{13}C results in an overall higher importance of the *electronic* input features for ^{13}C than for ^{1}H NMR.

Using no descriptors at all still shows a significant improvement compared to the uncorrected method, which can be interpreted as a much simpler fit comparable to the mentioned linear regression approach, yet being less accurate. The overall trend shows improved performance for every additional class of descriptors. This emphasizes the importance of the initial DFT calculation and the information it provides about the electronic structure and the magnetic shielding tensor. The presented Δ-ML model benefits strongly from this information rather than merely using input data that is derived from the three-dimensional structure of the molecule of interest.

C.3.4 Performance for External Systems

A fair evaluation of the presented NMR correction method on unknown, large, and realistic molecular systems is not straightforward. Since the model is trained to predict calculated CCSD(T) values, a comparison to experimental data is prone to all sources of error mentioned earlier apart from electron correlation. To ensure an appropriate assessment of the Δ-ML approach, it should thus be compared to values obtained with the CCSD(T)/TZ+ method or

to other high-level computed chemical shifts rather than experimental values. Unfortunately, this limits the investigated systems to sizes not larger than the ones included in the data set presented above. In the following, the performance of the new method applied to the PBE0/pcSseg-2 DFT level of theory is evaluated on three different benchmark sets for ^{1}H and ^{13}C NMR chemical shifts that contain molecules that mostly differ from those included in the training data set (some duplications are coincidental). As a reference, the presented CCSD(T)/TZ+ method as well as other CCSD(T)/quadruple-ζ levels and experimental values are used. For more information on the benchmark studies, see the Supporting Information, Section 4.2.

Benchmark Set of Flaig et al.

A benchmark study for ^{1}H and ^{13}C NMR chemical shifts was published in 2014 by Flaig et al.[33] and provides data calculated at the CCSD(T)/cc-pVQZ level of theory. 39 ^{1}H and 39 ^{13}C NMR chemical shifts are evaluated here, originating from molecules that are mostly smaller in size than the ones in the training set, making the benchmark study still suitable for comparison.

The performance of the Δ-ML method and the linear regression approach (Figure C.5, first row) is comparable to the performance on the data set presented above for both ^{1}H and ^{13}C NMR. The MAD is reduced by 77/87% for ^{1}H/^{13}C resulting in the final values of 0.046/1.00 ppm. This slightly larger deviation compared to the test data set is expected and exhibits the small bias of the trained model toward its training data. To exclude another bias toward the reference method that the model is trained to predict, the chemical shifts are furthermore evaluated against the calculated reference data given in the benchmark study (CCSD(T)/cc-pVQZ). In doing so, a very similar trend is observed with an MAD reduction of 79/86% to 0.041/1.10 ppm by applying the Δ-ML correction, showing that the accuracy gain does work generally also toward other computed high-level reference data. In every test case, the correction by the linear regression (LR) approach is significantly outperformed by Δ-ML.

It is interesting to note that the reference method CCSD(T)/TZ+ still differs slightly (0.023/0.51 ppm) from the reference given in the benchmark study, and it remains uncertain which method is closer to the true basis set limit. This means that any hypothetical perfect correction model trained on the same data would not be able to achieve better accuracy. In the above example, the Δ-ML model is therefore 0.018 ppm above the best possible MAD for ^{1}H and 0.59 ppm for ^{13}C.

NS372 Benchmark Set

The NS372 benchmark set compiled by Schattenberg et al.[41] contains 117 molecules that are mostly smaller than those in the training and test data set. A total number of 70 ^{1}H and 93 ^{13}C NMR chemical shifts will be evaluated below.

Figure C.6 again shows the performance of the Δ-ML and the linear regression approaches

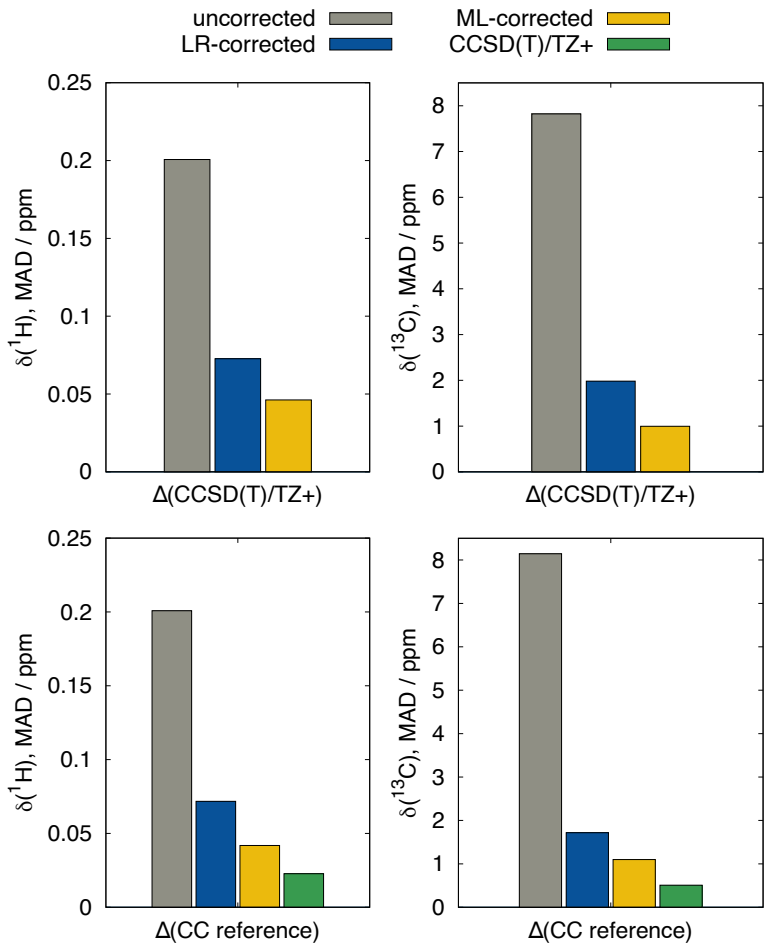


Figure C.5: Performance of the Δ-ML correction model on the benchmark of Flaig et al. with PBE0/pcSseg-2 as the *low-level* method. Comparison against the ^{1}H and ^{13}C reference values calculated with the CCSD(T)/TZ+ approach and with the level of theory used in the original literature,[33] which is CCSD(T)/cc-pVQZ.

with respect to the two reference methods. Although Δ-ML still reduces the mean absolute deviation significantly, this improvement is weaker compared to the benchmark study of Flaig et al. presented in Figure C.5 (MAD reduction to 0.075/2.37 ppm (–67/–76%) for ^{1}H/^{13}C). Furthermore, the LR correction is less clearly outperformed by the Δ-ML approach. This indicates that both models are less suitable to describe the chemical structures in the NS372 benchmark set because the large majority of these compounds are considerably smaller in size than those in the training data set. The NS372 set further contains structural characteristics that are completely absent in the training data such as charged molecules and compounds without C–H bonds. When those systems are excluded from the benchmark study, the performance of both LR and Δ-ML correction improves slightly yet systematically (see Supporting Information, Tables S9 and S10). Conversely, this means, as was to be

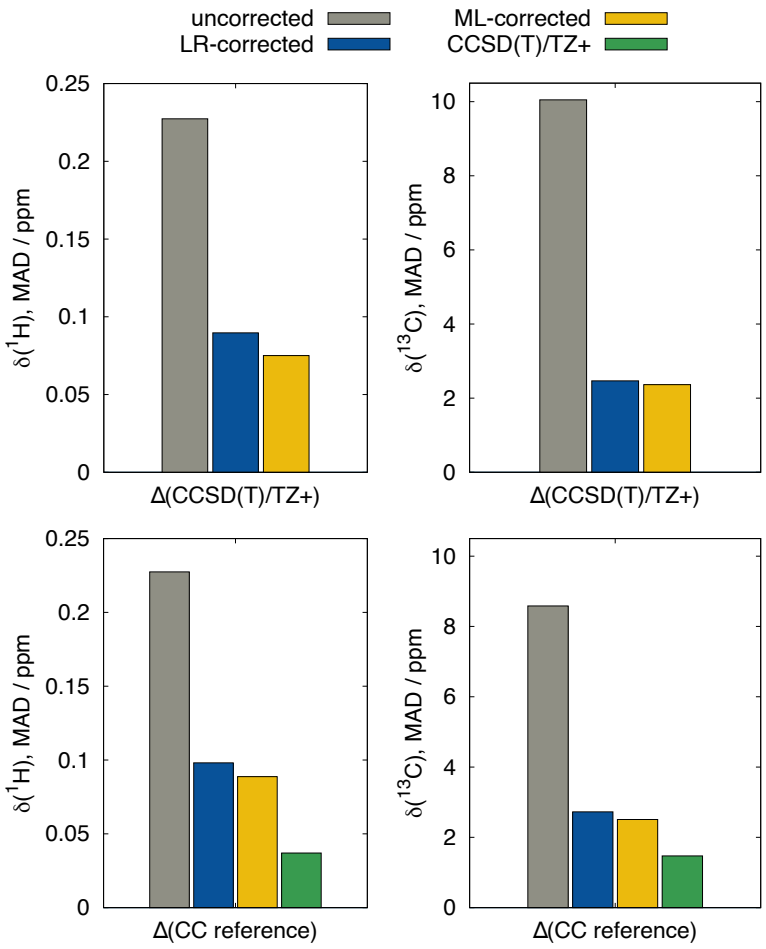


Figure C.6: Performance of the Δ-ML correction model on the NS372 benchmark with PBE0/pcSseg-2 as the *low-level* method. Comparison against the ^{1}H and ^{13}C reference values calculated with the CCSD(T)/TZ+ approach and with the level of theory used in the original literature,[41] which is CCSD(T)/pcSseg-3.

expected, that the model will improve if the training data is extended to better represent the test molecules. The presented workflow uses a sufficiently general pretrained Δ-ML model and still provides the flexibility to adjust it to specialized problems by optimizing the training data set, which "only" requires the calculation of chemical shielding constants on the CCSD(T)/TZ+ reference level of theory for new compounds.

Note that the deviation of chemical shifts calculated at the CCSD(T)/TZ+ level and the literature's reference method CCSD(T)/pcSseg-3 is larger (0.037/1.47 ppm) than in the study of Flaig et al., showing that the NS372 set contains some cases that are challenging for the basis set extrapolation scheme. As a matter of course, any empirical correction model can only work if the reference values are reliable.

NMRH148 Benchmark Set

The NMRH148 data set from de Oliveira et al.[34] consists of 72 small organic compounds, for which experimental ^{1}H NMR reference data is available. A total number of 143 ^{1}H and 179 ^{13}C NMR chemical shifts are evaluated hereinafter.

The correction of both the LR and the Δ-ML approaches (Figure C.7, left and middle) function as expected and seen before with a performance similar to the one obtained in the study of the benchmark set of Flaig et al. (MAD reduction to 0.037/0.66 ppm (–78/–89%) for ^{1}H/^{13}C). The better outcome compared to the evaluation of the NS372 set can be explained by the better representation of the compounds of the NMRH148 set due to larger structural similarities with the training data. Omitting all chemical shifts of the 18 compounds that overlap with the training data set only leads to a slight increase in the deviation of the corrected chemical shifts (final MAD of 0.046/0.83 ppm, reduction by 71/86%, see Supporting Information, Table S11 for details). The Δ-ML model is therefore adequately generalized to successfully predict NMR chemical shift predictions of similar compounds.

As mentioned before, a direct comparison of the presented method with experimentally obtained NMR chemical shifts is difficult due to the missing consideration of solvation effects and conformational flexibility. However, comparing the different ^{1}H NMR chemical shift prediction approaches in the gas phase to the experiment (Figure C.7, right) reveals a systematic error reduction toward a theoretical minimum obtained by exact NMR shielding calculations. This behavior is promising and allows for error decomposition into a part related to the electronic structure method and other parts beyond the static, isolated molecule

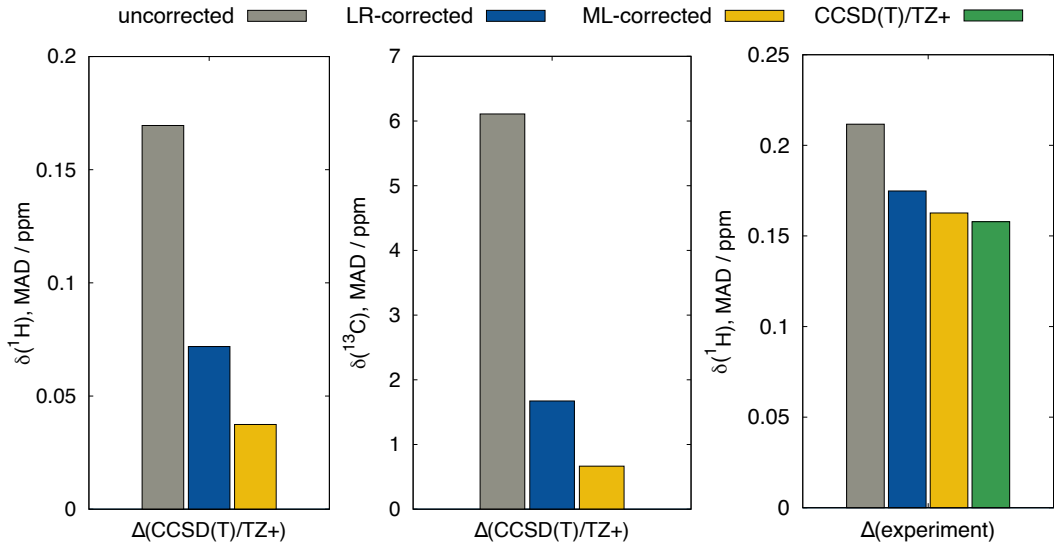


Figure C.7: Performance of the Δ-ML correction model on the NMRH148 benchmark with PBE0/pcSseg-2 as the *low-level* method. Comparison against the ^{1}H and ^{13}C reference values calculated with the CCSD(T)/TZ+ approach and against the experimental ^{1}H NMR chemical shifts.

approximation. Since the compounds in the NMRH148 set are expected to be mostly rather rigid, the solvation treatment might be a major missing part in this example.

NMR Spectrum of Butadiene

The presented Δ-ML approach can be used to correct fully quantum chemically calculated NMR spectra as well. In fact, the original motivation for this work was to develop a method that improves NMR spectra that are generated with the CREST and CENSO workflows.[29] A conformer ensemble generated with this tool for which NMR chemical shifts and spin-spin coupling constants have been calculated can be used to automatically construct a full NMR spectrum.[30] There are a few literature examples of gas-phase NMR spectra,[402] and the butadiene molecule was chosen as a showcase. The compound is similar to those in the training data set but is not a part of it. As the molecule is relatively rigid, there is only one relevant conformer obtained from a CENSO run with the final PBE0/pcSseg-2//r^{2}SCAN-3c level of theory. Its corresponding NMR spectrum is shown in Figure C.8 and compared to the experimental spectrum published by Zuschneid et al.[403]

Since butadiene only exhibits three distinct signals in the ^{1}H NMR spectrum, it is rather clear. The calculated coupling pattern matches the experimental one very well and is in this case not influenced by the correction of the chemical shift. The Δ-ML correction shifts all

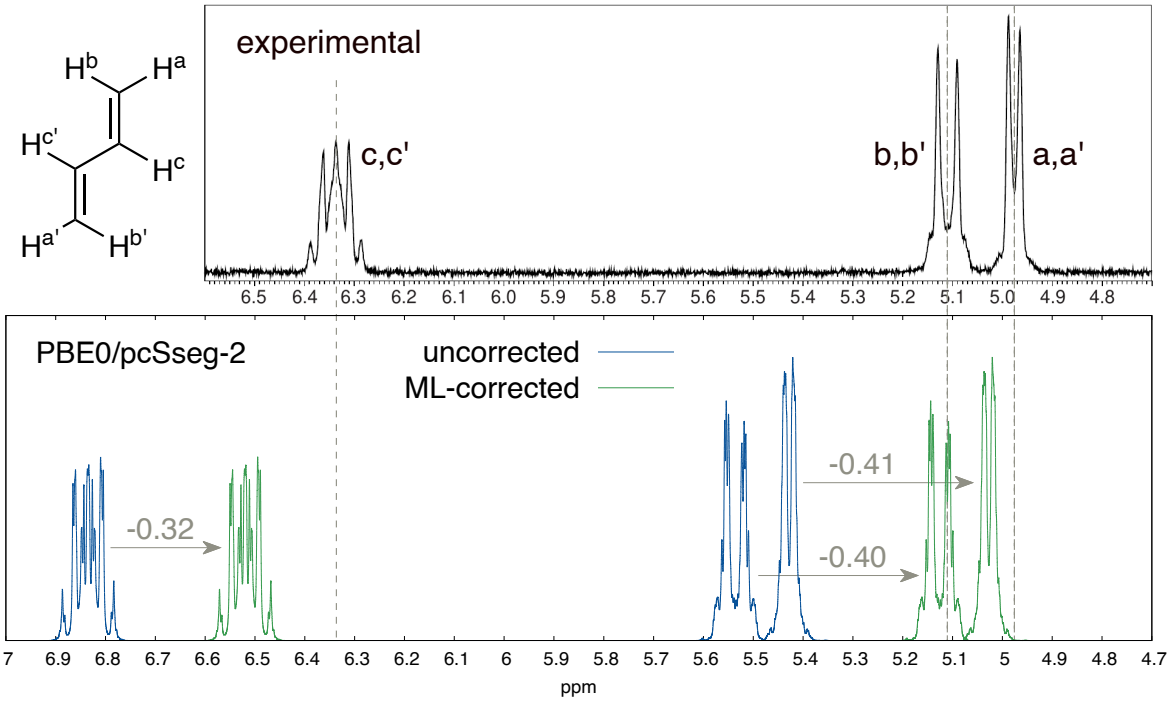


Figure C.8: Experimental gas-phase spectrum of butadiene (top, taken from Zuschneid et al.[403]) and a fully coupled ^{1}H NMR spectrum resulting from chemical shielding and spin-spin coupling calculations (bottom).

signals in the right direction such that the signals of nuclei **a** and **b** are very well reproduced and the signal **c** experiences a significant error reduction. In conclusion, this showcases the use of the presented Δ-ML correction for application in NMR spectra simulation, which will be even more useful with an additional solvent correction.

C.3.5 Computational Demands

Considering the computational resources that are consumed by a CCSD(T)/pcSseg-2 nuclear shielding calculation, the presented Δ-ML correction comes at negligible extra computational cost. A comparison of the computation times of every step necessary for this study is presented in Figure C.9.

The CCSD(T)/pcSseg-2 calculation is by far the most demanding in the entire procedure and therefore the limiting factor for the considerable system size. The BHLYP/pcSseg-n ($n = 2, 3, 4$) calculations needed for the basis set extrapolation come at considerably high costs for DFT methods, but these and the CCSD(T) calculations have to be performed only once per structure independent from the *low-level* method and are given for 1000 structures in the presented data set. Given that the Δ-ML model is used to correct DFT-based chemical shifts that can be obtained in minutes, the correction step of a few seconds on a usual laptop is negligible. If a new model is to be trained using another *low-level* method, the whole data set has to be calculated using that desired method. Subsequent training of the model will take a few minutes on a usual laptop. The disadvantage of the relatively small data

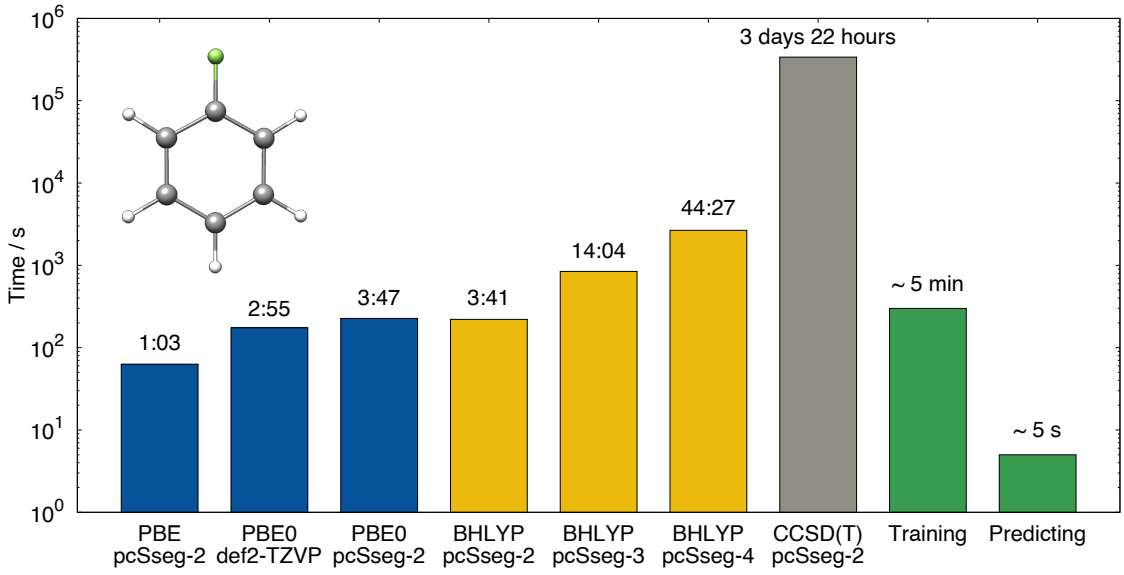


Figure C.9: Timings for serial NMR shielding calculations of fluorobenzene (structure **060_00**) on an Intel® Xeon® CPU E3-1270 v5 @ 3.60 GHz (values over the bars in minutes; the ML procedures (green) were done on a usual laptop PC). DFT calculations have been performed with ORCA 5.0.3 and the CCSD(T) calculation with CFOUR 2.1.

set thus becomes an advantage, because it makes the method practically free of additional computational costs.

C.4 Conclusion

The efficient and accurate calculation of NMR parameters is a great challenge in the field of computational chemistry. The presented new method aims at combining the efficiency of DFT with the accuracy of coupled cluster theory by predicting their respective deviations using a machine learning approach. The newly compiled data set contains 1000 optimized and distorted structures of a variety of small molecules built from carbon, hydrogen, and most elements of the first three periods resulting in a total number of 7090 ^{1}H and 4230 ^{13}C NMR nuclei. For all data points, the target chemical shift value has been calculated with the CCSD(T)/pcSseg-2 level of theory with a complete basis set extrapolation correction, termed CCSD(T)/TZ+. The Δ-ML model has been constructed such that it provides the maximum generality and flexibility, with the possibility to be optimized toward special applications through extensions of the data set. It is applicable to any desired density functional approximation and basis set that are able to provide NMR chemical shielding constants and the necessary *electronic* and *magnetic* descriptors.

When using PBE0/pcSseg-2 as the *low-level* DFA to be corrected, the Δ-ML model achieves a reduction of the mean absolute deviation evaluated on the test data set of 81% for ^{1}H and 92% for ^{13}C NMR chemical shifts accompanied by a drastically narrower error spread. The computational demands are not noticeably increased in this process. Throughout all tested DFA and basis set combinations, final MADs lie in a range of 0.021-0.039 ppm (^{1}H) and 0.38-1.07 ppm (^{13}C). The much less sophisticated correction by linear regression is significantly outperformed in all cases.

The evaluation of the new correction method for realistic systems and experimental data is not straightforward, since only high-level calculated data can serve as a consistent reference. Still, the chemical shift correction performance for PBE0/pcSseg-2 was tested in three benchmark studies with different individual challenges. The Δ-ML correction shows a consistent and substantial improvement in the quality of the predicted chemical shifts in all tests and surpasses the linear regression correction as well. The representation of the tested set of compounds by the molecules in the training data set affects the prediction performance of the method and limits its applicability. However, no outliers are produced by the proposed ML model indicating sufficient robustness in practical applications (cf. Figures C.2 and C.3). Nevertheless, the prediction accuracy can be improved by extending the data set with respective structures and it is possible to estimate the quality of the model by analyzing the conformity of the data sets used for training and testing.

With the new approach, fully simulated NMR spectra can be corrected in a physically reasonable way by reducing the error caused by missing electron correlation. We aim at incorporating the correction into the CENSO[29] program, where the calculated chemical shifts of every conformer in a given ensemble are corrected to obtain the corrected Boltzmann-

averaged NMR spectrum. However, the method can also be used in combination with other techniques that include dynamic effects such as molecular dynamics simulations. Contributions to the shielding constant due to solvation or relativistic effects can be added to the ML-corrected value regardless of the correction procedure.

It is important to note that this work presents a general workflow rather than a completed method. In subsequent work, the applicability can and will be enhanced by extending the training data set, fine-tuning the ML architecture, allowing other software and non-DFT methods (e.g., HF, MP2) for feature generation, and extending the scope to other NMR-relevant nuclei. Furthermore, the search for suitable input features is not finished especially as they may not be the same for all nuclei. The ^{13}C NMR chemical shift showed some correlation with information derived from the DFT density matrix, which might hold true for other heavy nuclei.

In conclusion, the presented Δ-ML correction for DFT-based NMR chemical shifts toward coupled cluster quality has the potential to significantly improve simulated NMR data obtained from standard DFT calculations. Although being an empirical ML-based correction employing a rather small number of training data points, it proved to be general and roust also on external data indicating that the method will improve as the underlying data and knowledge basis grows.

Supporting Information

The Supporting Information is available free of charge at
https://pubs.acs.org/doi/10.1021/acs.jctc.3c00165.

- Additional information and analyses regarding some special aspects that were discussed here (PDF)
- Raw data of all evaluations presented here and in the Supporting Information (XLSX)
- Archive of all relevant chemical structures in Cartesian coordinate (`.xyz`) format (ZIP)

Acknowledgements

The authors thank Fabian Bohle for important discussions and bringing the topic of an ML correction for NMR to our attention. J.B.K.B. is grateful for financial support from the "Fonds der Chemischen Industrie (FCI)". S.G. acknowledges funding by the German Research Foundation (DFG) through the SPP 2363: "Utilization and Development of Machine Learning for Molecular Applications – Molecular Machine Learning".

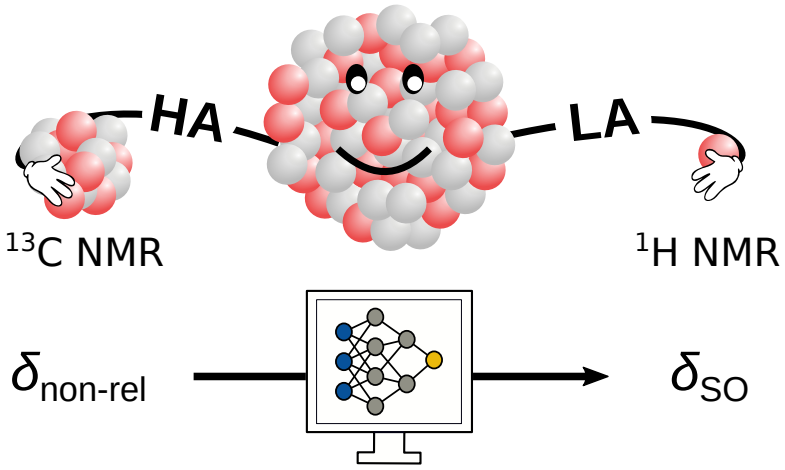
Conflict of Interests

The authors declare no competing financial interest.

APPENDIX D

Appendix: Machine Learning-Based Correction for Spin-Orbit Coupling Effects in NMR Chemical Shift Calculations

J. B. Kleine Büning, S. Grimme, and M. Bursch, *Phys. Chem. Chem. Phys.* **2024**, *Advance Article*. DOI: 10.1039/d3cp05556f.



The relativistic spin-orbit contributions to ^{13}C and ^{1}H NMR chemical shifts in the vicinity of heavy atoms are computed using a novel Δ-machine learning approach at virtually no extra computational cost.

Abstract

As one of the most powerful analytical methods for molecular and solid-state structure elucidation, NMR spectroscopy is an integral part of chemical laboratories associated with a great research interest in its computational simulation. Particularly when heavy atoms are present, a relativistic treatment is essential in the calculations as these influence also the nearby light atoms. In this work, we present a Δ-machine learning method that approximates the contribution to ^{13}C and ^{1}H NMR chemical shifts that stems from spin-orbit (SO) coupling effects. It is built on computed reference data at the spin-orbit zeroth-order regular approximation (ZORA) DFT level for a set of 6388 structures with 38740 ^{13}C and 64436 ^{1}H NMR chemical shifts. The scope of the methods covers the 17 most important heavy *p*-block elements that exhibit heavy atom on the light atom (HALA) effects to covalently bound carbon or hydrogen atoms. Evaluated on the test data set, the approach is able to recover roughly 85% of the SO contribution for ^{13}C and 70% for ^{1}H from a scalar-relativistic PBE0/ZORA-def2-TZVP calculation at virtually no extra computational costs. Moreover, the method is transferable to other baseline DFT methods even without retraining the model and performs well for realistic organotin and -lead compounds. Finally, we show that using a combination of the new approach with our previous Δ-ML method for correlation contributions to NMR chemical shifts, the mean absolute NMR shift deviations from non-relativistic DFT calculations to experimental values can be halved.

D.1 Introduction

Nuclear magnetic resonance (NMR) spectroscopy is a highly valuable analytic tool for structure elucidation and has become a standard method that is used on a daily basis throughout various chemical disciplines.[7,404,405] Besides experimental analysis, the computation of NMR parameters can further yield detailed insight into chemical phenomena and complex bonding situations. In particular, density functional theory (DFT) has proven to be a reliable and efficient choice for the calculation of NMR parameters.[33,39,41,42,178,199,213,227,298] Nevertheless, the complex physical relationship between these parameters, the electronic structure, and the chemical environment of the investigated compound remains challenging for quantum chemical methods.[24,406,407]

There are five main sources of error in quantum chemical NMR prediction as claimed by Lodewyk et al.,[24] which are electron correlation, solvation effects, conformational flexibility, rotational-vibrational, and relativistic effects. The latter become specifically relevant when NMR properties of heavy elements (which we refer to as having an atomic number larger than 18 and including Cl) are computed or if such elements are present in close vicinity to the nucleus under consideration. Heavy elements can play major roles in various application areas, including catalysis,[183–185] batteries,[186,187] and optoelectronics.[188,189] Furthermore, there is a great research interest in biology and biochemistry due to the toxicity of some heavy elements (e.g., Ni, Cd, Hg, Pb)[190] or their essential role in biochemical processes, as for Zn[191–193] and Se.[194–196]

The most crucial relativistic effects originate from spin-orbit (SO) coupling and can be essential even for qualitative modeling of the heavy atom (HA) itself,[53,199,408] but also for the adjacent lighter atoms by the heavy atom on the light atom (HALA) effect.[48,54,213] There are several physics-based methods to incorporate relativistic effects into quantum chemical calculations such as the Douglas-Kroll-Hess (DKH)[126,127] method, the exact transformation of the four-component Dirac equation to two components (X2C),[130–132] and the zeroth-order regular approximation (ZORA).[128,129] However, the spin-orbit variants of these methods in combination with NMR shielding tensor calculations typically become computationally unfeasible for larger compounds due to their high computational demand. Further, such methods are not available in many chemical software packages. Accordingly, more efficient and easily accessible methods to include SO-relativistic effects in the calculation of NMR parameters are highly desirable. Such effects are typically neglected in most low-cost approaches, as done in a recently published correction scheme for efficient NMR chemical shielding prediction[409] and a study on halogenated natural products.[410]

One approach to solve this issue and fill this methodological gap can be an empirical model based on machine learning (ML). The field of ML in chemistry has evolved rapidly in the past decade and besides approaches that tackle the complete electronic structure of a quantum chemical system,[72,411,412] several techniques for the calculation of NMR chemical shifts have been developed.[81,413,414] Especially for NMR-aided structure assignment, the popular DP4 method[415] was improved with an ML approach called DP4-AI[85] and an ML-based technique for structure assignment from two-dimensional NMR spectra has been

proposed.[87] ML approaches can exploit their full potential for highly accurate predictions if they are combined with DFT and use features from a converged electronic structure as input (Δ-ML). This has been shown to yield highly accurate electronic energies[416] and NMR chemical shifts[88–90,197] at costs not significantly higher than for the underlying *low-level* method.

There is evidence that in order to achieve a good prediction quality for ^{13}C NMR chemical shifts of carbon atoms attached to heavy atoms, it is important to account for both correlation and heavy atom effects.[24,417,418] In a test on the *o*-bromochlorobenzene molecule,[419] the pragmatic combination of a non-relativistic second-order Møller-Plesset perturbation theory (MP2) calculation and a SO contribution calculated with DFT yielded the best results compared to experimental data and was the only tested method to achieve a qualitatively correct chemical shift ordering of the six ^{13}C nuclei. We are therefore confident that a combination of separate correlation and spin-orbit corrections will be beneficial for the efficient computation of reasonably accurate NMR chemical shifts. We previously proposed an ML-based correction method that obtains the (beyond DFT) correlation contribution to NMR chemical shifts based on coupled cluster (CCSD(T)) reference data,[197] which we now call Δ_{corr}-ML. In this work, we present an efficient and highly transferable approach called Δ_{SO}-ML to compute the spin-orbit relativistic contribution to ^{13}C and ^{1}H NMR chemical shifts. This new approach is validated for a large number of unique chemical shifts computed at the SO-ZORA-DFT level and is exemplarily applied to ^{13}C NMR chemical shifts in experimentally accessible organotin and -lead compounds and in a set of heavy metal-organic compounds with experimental reference data.

D.2 Methods

D.2.1 Machine Learning Data Set

As in our previous work on the correction of ^{1}H and ^{13}C NMR chemical shifts using machine learning,[197] quantum chemical ab initio data serve as target for the model presented herein. The use of experimental data would increase the overall complexity thus making the data set less suitable for applying an ML correction procedure. Furthermore, the target was chosen such that it keeps an unadulterated focus on one specific component of the chemical shielding constant calculation, spin-orbit coupling. A clear distinction of SO effects from various other error sources in experimental data would be impossible and such an approach would prevent a targeted elimination of the SO error.

The data set is one of the central parts in every ML model and is often particularly challenging to compile for ML applications in chemistry when reference data is sparse. Since it is largely responsible for the performance of the model, we focus on the most common bonding situations in classical (metal-)organic compounds. Further, a focus is set on heavy non-radioactive elements of groups 12 to 17. This includes most *p*-block elements except for noble gasses and group 12 transition metals as their chemistry is comparably dominated by *p*-orbitals.[48] As NMR parameters tend to be spatially local, the reference molecules can be

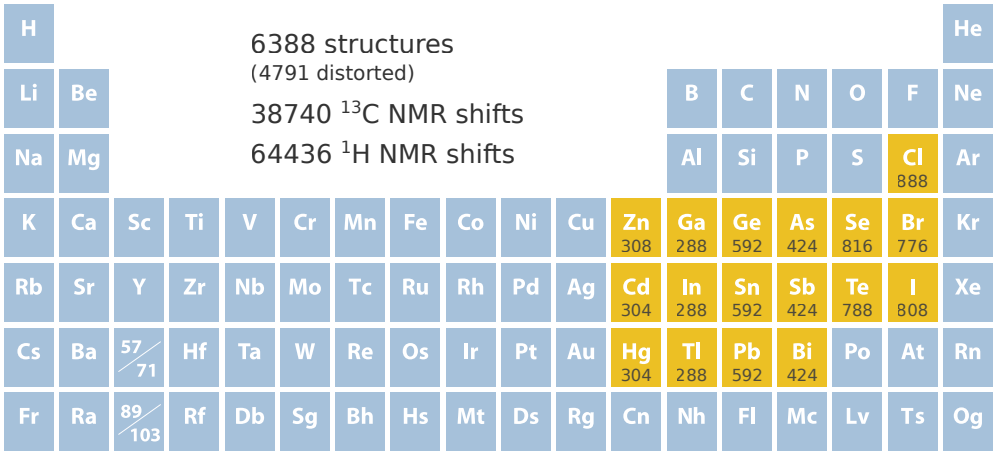


Figure D.1: Key specifications of the data set. Included heavy elements are marked in yellow followed by the number of structures hat contain them.

chosen rather small (3-46 atoms). This allows for the inclusion of many different bonding motifs, covering a wide chemical space with a sufficiently large amount of samples. The data set consists of 1597 unique molecules, in which at least one heavy atom ($Z \geq 17$) is covalently bound to a carbon atom. These molecules were created manually, starting with the methyl compounds mentioned in Section D.3.3, and subsequent substitution of the ligands with larger aliphatic, aromatic, and functional residues that are typically found in compounds of the respective heavy element. Analogous structures are included for all elements within the same group and some more complex compounds were added. The structures were selected such that they are chemically reasonable, yet they do not have to be accessible in an experimental setting. To enrich the data, three geometrically distorted structures, one out of each energy window of 2.5-5.0 kcal·mol^{-1}, 10.0-15.0 kcal·mol^{-1}, and 30.0-40.0 kcal·mol^{-1} above the optimized structures (at the r^{2}SCAN-3c[172] level) were added (for more information on the distortion procedure, see ref. [197] and the Supporting Information). The overall 6388 structures include data points for 38740 ^{13}C and 64436 ^{1}H NMR chemical shifts, which is illustrated in Figure D.1. The set includes 2264 structures containing Cl, Br, or I; 1440 structures containing Se or Te; 1260 structures containing As, Sb, or Bi; 1680 structures containing Ge, Sn, or Pb; 804 structures containing Ga, In, or Tl; and 868 structures containing Zn, Cd, or Hg.

D.2.2 Reference Level of Theory

The target of the Δ_{SO}-ML approach presented here is the contribution to the chemical shielding constant that originates from the inclusion of spin-orbit relativistic effects with a suitable computational method. Scalar-relativistic approaches that either employ effective core potentials (ECPs) or explicitly use a scalar-relativistic (SR) Hamiltonian are available in many quantum chemical program packages such as ORCA. As all-electron approaches with

SR-Hamiltonian can be regarded the more general and reliable choice for the calculation of properties with relevant nucleus effects, we chose an efficient all-electron SR-DFT calculation as origin of the ML input features to predict the additive spin-orbit (SO) correction with our Δ_{SO}-ML model. This contribution $\Delta_{SO}\delta$ is calculated from chemical shifts δ obtained by including the different levels of relativity:

$$\Delta_{SO}\delta = \delta_{SO} - \delta_{SR} \quad . \tag{D.1}$$

In this work, we determine the target $\Delta_{SO}\delta$ from two-component SR/SO-ZORA (zeroth-order regular approximation) calculations at the PBE0[109,420] hybrid DFT level of theory with the Slater-type triple-ζ TZ2P[165] basis set. Note that the performance of an ML approach is directly influenced and limited by the quality – especially the noisiness – of the reference data.[150] PBE0 is a generally robust functional that usually yields good NMR properties and especially the SO-relativistic variant has proven reliable performance in our previous studies on ^{29}Si[213] and ^{119}Sn[199] NMR chemical shifts. Furthermore, in contrast to full four-component relativistic methods, SO-ZORA-PBE0 is still feasible for the medium-sized (>40 atoms) molecules included in the data set. The transferability of our approach based on the PBE0/TZ2P data to other density functionals and basis sets is further discussed in Section D.3.3.

To make it easily accessible, the presented method is built onto a scalar-relativistic baseline level of theory calculated using Gaussian-type orbitals with the ORCA program package (although it is in principle not limited to it). The SR-PBE0/ZORA-def2-TZVP level of theory serves as *low-level* method for most evaluations shown below.

D.2.3 Neural Network Architecture and Input Feature Vector

With the data set and the target values $\Delta_{SO}\delta$ at hand, an ML model can be constructed. For this purpose, the data set is randomly divided into a training set to build the model and a test set that serves as basis for all evaluations made in Section D.3. The data is processed in an atom-wise fashion but it is ensured that all atoms from an individual structure are attributed to the same data set (shuffling mode `structures`, see Supporting Information, Section 2.2 for details). Data from a sample molecule and its *low-level* NMR shielding calculation (currently only possible via the ORCA 5 program package) can finally be used to predict $\Delta_{SO}\delta$. The complete workflow is illustrated in Figure D.2.

The regression artificial neural network used herein is similar to the one used for our previous Δ_{corr}-ML model for correlation contributions of the chemical shift.[197] The same multilayer perceptron architecture with two hidden layers was used in TensorFlow 2.12 and the input feature vector was modified to adapt it to the SO contribution problem. After initial testing, the hyperparameters were set to 300/12 nodes for the first/second hidden layer for ^{13}C (384/80 for ^{1}H) with a dropout rate of 0.1 for the first layer for ^{13}C (0.15 on first and 0.1 on second layer for ^{1}H) and the adam optimizer. The activation function on all layers was set to GELU (Gaussian Error Linear Unit) for ^{13}C and the sigmoid function

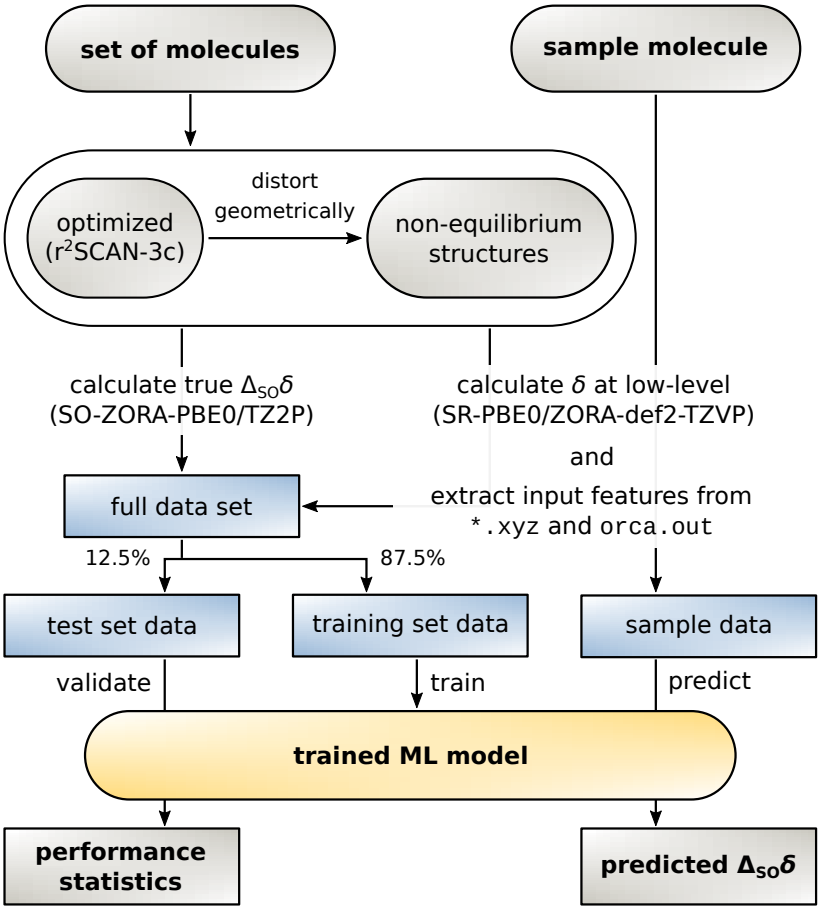


Figure D.2: Workflow of the Δ_{SO}-ML method used in this work. A set of organic molecules with heavy elements is structurally optimized and geometrically distorted structures are created. For these, reference and *low-level* data are calculated, from which input features are extracted that make up the data set. $^{7}/_{8}$ of the data set are used to train the ML model, $^{1}/_{8}$ for validation. For a sample molecule, $\Delta_{SO}\delta$ can then be predicted by the model from the *low-level* data.

was used for ^{1}H. The distribution of the SO contribution values to the chemical shift in the data set is very heterogeneous. Most of the atoms are not in direct vicinity to a heavy atom, so $\Delta_{SO}\delta$ is small but few atoms exhibit a very large value. To make the model focus on the important large values while not placing too much weight on unaffected atoms, the root mean squared deviation (RMSD) was chosen as loss function and showed to be superior to the mean absolute deviation (MAD) and the mean squared deviation (MSD). For the same reason, the RMSD is suited better than the MAD for the evaluations below.

The information included as input is of central importance for the quality and performance of the ML model.[150] In the case of Δ_{SO}-ML, the input feature vector is constructed such that it contains information about the *geometric* (solely from the three-dimensional structure), *electronic* (from the converged density matrix of the DFT single-point calculation), and

magnetic (from the DFT NMR shielding constant calculation) surrounding of each atom of interest. The majority of the descriptors of these categories was taken from the Δ_{corr}-ML model and some were omitted. A set of atom-centered symmetry functions[379] (ACSF) was further added.

Furthermore, a new range of descriptors was added that contains *geometric* and *electronic* information about heavy atoms in the vicinity of the atom of interest. These include:

- the total number of heavy atoms bound to the nucleus of interest via one to five covalent bonds,
- the average atomic mass of all atoms within one to five covalent bonds,
- the atomic number of the heavy atom(s),
- the coordination number (from the D3 model) of the heavy atom(s), and
- the atomic, and *s*-, *p*-, and *d*-orbital Mulliken populations of the heavy atom.

The latter ones are arranged in sets of five descriptors per heavy atom in the first covalent bond shell. The inclusion of the atomic charge and the *s*- and *p*-orbital populations was motivated by the findings of Vícha et al. on the spin-orbit heavy-atom (HA) effect on the light atom (LA).[48] These suggest that for most heavy elements the chemical shift contribution originating from spin-orbit coupling has a fixed sign. Accordingly, the HA effect on the LA is always shielding or deshielding (this information is covered by the atomic number of the HA). The cause of this trend lies in the electronic configuration, as formally empty valences shells of the HA (e.g., p^0, d^0) typically lead to a deshielding mechanism whereas partially filled subshells (e.g., p^2, p^4) result in a shielding effect. In some cases, however, different contributions occur simultaneously with comparable magnitudes so that the sign of $\Delta_{SO}\delta$ may vary, e.g., depending on the oxidation state of the HA. Therefore, the atomic charge and orbital populations of the HA are included as descriptors. A detailed list of the complete input feature vector is provided in the Supporting Information, Table S2.

D.2.4 Computational Details

The compounds in the data set were chosen and created manually and a selected structure was pre-optimized at the semiempirical tight-binding GFN2-xTB[27] level using the xTB 6.6.0[247] program package. Subsequent geometry optimizations where performed with the TURBOMOLE 7.7.1[307–309] program package using the r^{2}SCAN-3c[105,172,249] composite DFT method. Throughout the geometry optimizations, the resolution of the identity approach for Coulomb integrals (RIJ)[255] was applied and the `m4` grid and a radial grid size of 10 were used. For the tests on the experimentally accessible structures in Section D.3.4, conformer ensembles were generated using the conformer-rotamer ensemble sampling tool (CREST),[25,310] version 2.12, using the GFN-FF[28] force field and GFN2-xTB with the ALPB[118] solvation model (solvent as in the experimental measurement). The ensembles were further refined

with the command-line energetic sorting (CENSO)[29,311] algorithm, version 1.2.0, at the final level r^{2}SCAN-3c+G_{solv}(COSMO-RS)[115,116,312]+G_{mRRHO}(GFN2-xTB)[119,120,421] //r^{2}SCAN-3c(DCOSMO-RS).[422]

All NMR shielding constant calculations in this study were performed via the gauge-including atomic orbital (GIAO)[137–139] approach using the ORCA 5.0.4[243,244,387] program package for calculations with Gaussian-type orbital (GTO) basis sets and the ADF module of the AMS 2022.103[246] program package for Slater-type orbital (STO) basis sets. For the *low-level* shielding calculations (ORCA), the Hamiltonian of the scalar-relativistic (SR) zeroth-order regular approximation (ZORA)[128,258] was used in combination with the PBE[101] general gradient approximation (GGA) and the PBE0[109] and r^{2}SCAN0[198] hybrid density functionals, together with the GTO triple-ζ ZORA-def2-TZVP[156,423] basis set for all atoms with $Z \leq 36$ and the SARC-ZORA-TZVP[259,260,424] basis set for all atoms with $Z > 36$. For PBE0, the calculations were also done without ZORA applying the def2-TZVP[156] basis set with the def2 effective core potentials (ECP).[254,257] For the NMR shielding calculations in Section D.3.4, the CPCM[112] implicit solvation model was used. All calculations applied the RI scheme with the chain-of-spheres approximation for the exchange (RIJCOSX)[228,392] in combination with the auxiliary SARC/J basis set. The `defgrid3` grid and the `tightscf` convergence settings were used throughout. For the *high-level* reference values (ADF), the NMR shielding calculations were performed each with the scalar (SR) and spin-orbit (SO) variant of ZORA[129,318,319] and the PBE0 functional using the STO polarized triple-ζ ZORA/TZ2P[165] basis set (the "ZORA/" prefix for the basis set is from now assumed for all calculations with ADF and omitted for clarity). The numerical grid quality was set to `verygood`. For the compounds of the methyl subset, the same calculation settings were applied, but with using the DZ, DZP, TZP, and QZ4P basis sets[165] and the PBE, BLYP,[102,106] mPW,[425] B3LYP,[106,108] and mPW1PW[425] functionals.

In the ML training, prediction, and evaluation procedures mentioned in the following, statistical fluctuations are to be expected that originate from the randomized weight initialization when the model is build. All statistics presented for the performance of the Δ_{SO}-ML model are therefore obtained as the mean value of ten training runs if not stated otherwise.

D.3 Results

D.3.1 Prediction of $\Delta_{SO}\delta$ for ^{13}C NMR

Before focusing on the new Δ_{SO}-ML correction, it is worth investigating the data set itself with the computed *low-level* SR NMR shifts and the reference $\Delta_{SO}\delta$ values. As a rather short-range effect, the SO contribution is expected to be small for many C/H atoms, but can be extreme in direct vicinity of a heavy atom. The complete data for the ^{13}C NMR shifts is depicted in Figure D.3(b) including information about the chemical distance (= number of covalent bonds) of each nucleus to the next heavy atom. Thus, ^{13}C nuclei that are directly bound to a HA experience by far the largest spin-orbit coupling effects and these values are most scattered over a wide range of chemical shifts. Still, the SO effect of a HA can

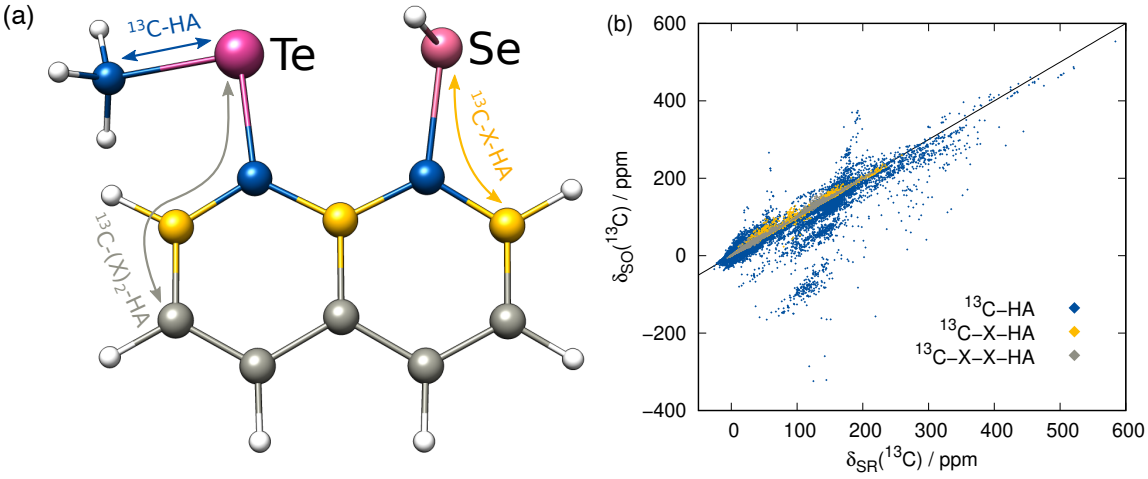


Figure D.3: (a) Example molecule (part of the data set) showing the division of the ^{13}C data set into three distance categories with one (blue), two (yellow), and three or more (gray) bonds between the HA and the ^{13}C nucleus (H atoms are white). (b) Complete data set with 38740 ^{13}C NMR chemical shifts showing the relation between the purely scalar-relativistic NMR shift δ_{SR} calculated with the *low-level* PBE0/ZORA-def2-TZVP method and the relativistic reference $\delta_{SO} = \delta_{SR} + \Delta_{SO}\delta$ with the SO contribution calculated at the PBE0/TZ2P level.

propagate which can have significant consequences for the ^{13}C nuclei even three covalent bonds away. Although the nuclei closest to the HA have the potential to lead to the largest errors in a calculation, the proposed ML model should not only be capable of reproducing the rough magnitude of the effect, but should consequently predict a small $\Delta_{SO}\delta$ value for the weakly affected nuclei.

As mentioned earlier, the PBE0/ZORA-def2-TZVP method applying scalar-relativistic SR-ZORA is chosen as *low-level* method for the following investigations. Several metrics demonstrating the performance of the Δ_{SO}-ML model evaluated for the test data set (12.5% of the data points) for ^{13}C and ^{1}H (discussed in the next section) nuclei are listed in Table D.1. The data for one of the runs is exemplarily shown in Figure D.4 in more detail.

The Δ_{SO}-ML correction clearly succeeds to predict the SO contribution to the ^{13}C NMR chemical shifts with good accuracy. The mean absolute deviation (MAD) from 4843 chemical shifts in the test data set (unknown to the ML model while training) is reduced by 85% from 7.26 to 1.07 ppm and the mean (signed) deviation (MD), which is slightly positive when only SR is applied, essentially reaches zero. More importantly, the root mean square deviation (RMSD), which emphasizes large deviations, is equally reduced (87%, from 21.88 to 2.76 ppm). Thus, it can be concluded that roughly 85% of the SO contribution is recovered by the ML model. Analysis of one of the training runs in Figure D.4(a) shows that accuracy is maintained even for the extreme cases of spin-orbit coupling effects on ^{13}C nuclei directly bound to HAs. The large negative values of $\Delta_{SO}\delta$ occur especially when several heavy halogen atoms are present, such as in a CI_3 moiety.

Furthermore, extremely large errors can be avoided with the Δ_{SO}-ML correction as the

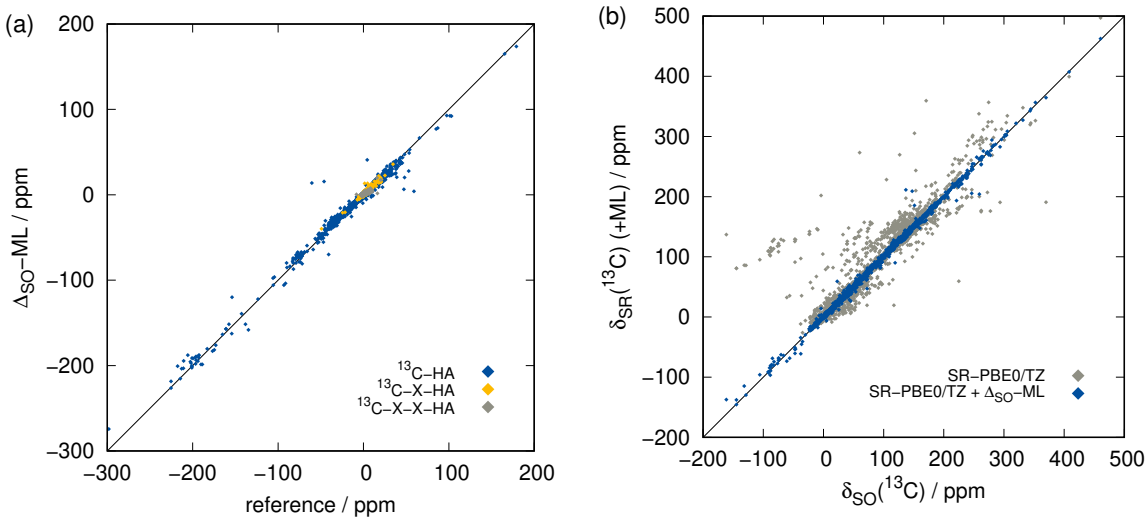


Figure D.4: Comparison of the ML-predicted SO contributions to the reference (SR/SO)-ZORA-PBE0/TZ2P ones for the ^{13}C NMR test set. (a) Values of $\Delta_{SO}\delta$, color-coded according to their distance to the next heavy atom (see Figure D.3(a)). (b) Total chemical shift δ neglecting SO coupling (gray) and adding the ML-predicted $\Delta_{SO}\delta$ (blue).

maximum negative and positive errors are reduced drastically from –179.07/+298.32 to –56.34/+87.55 ppm. This is promising because in this way the ML correction reduces the probability of a complete qualitative failure of a chemical shift prediction. The overall chance for outliers is therefore reduced significantly as underpinned by Figure D.4(b).

It is important to note that an empirical prediction method for spin-orbit effects to NMR chemical shifts is not a straightforward task. It is indeed a rather systematic phenomenon as

Table D.1: Statistics of the test data set before and after applying the Δ_{SO}-ML correction to the scalar-relativistic (SR) baseline SR-PBE0/def2-TZVP values in ppm (MSD in ppm^2), reference: SO-PBE0/TZ2P. Mean over ten training runs, more details on the metrics are given in the Supporting Information.

	^{13}C		^{1}H	
Error metric	only SR	SR+ML	only SR	SR+ML
MAX (<0)	−179.07	−56.34	−19.569	−4.772
MAX (>0)	298.32	87.55	6.042	10.825
MD	2.84	−0.05	−0.185	−0.005
MAD	7.26	1.07	0.281	0.090
MSD	478.82	7.60	0.645	0.056
RMSD	21.88	2.76	0.803	0.236
SD	21.67	2.75	0.782	0.236

it strongly depends on the type and the atomic number of the HA as well as the periodic table main group it belongs to. So, when the underlying data only features a certain chemical subgroup of molecules, a simple linear regression approach, which is often used to correct calculated NMR chemical shifts and which we used to compare the performance of the Δ_{corr}-ML model to, can be used to approximate the SO contribution to ^{13}C NMR chemical shifts.[426,427] However, this is only possible if all molecules in the data set contain the same number and type of heavy toms.[24] Conversely, in the data set presented here, many different HAs and amounts of HAs are present and therefore, there is not even a slight linear connection between the computed scalar-relativistic chemical shift value and the missing SO contribution (see the Supporting Information, Figures S3 to S5). Hence, a linear regression correction approach fails for the SO contribution, as it predicts $\Delta_{SO}\delta \approx 0$ in most cases. However, the additional computational costs of a prediction of $\Delta_{SO}\delta$ by the a pre-trained ML model is – as for the linear regression approach – negligible (few seconds). This consolidates the significance of the Δ_{SO}-ML model as a general low-cost method to predict the SO contribution of NMR chemical shifts.

D.3.2 Prediction of $\Delta_{SO}\delta$ for ^{1}H NMR

The focus of this work lies on the prediction of SO contributions to ^{13}C NMR chemical shifts since in organic compounds, usually the carbon atoms are connected to heavy heteroatoms. However, the ^{1}H nucleus can experience SO contributions from even further HAs as it is more prone to environmental changes and thus can be affected by propagation of the SO contribution via more than three covalent bonds. Compared to heavier elements, the hydrogen atom only comprises a very thin electron shell and the ^{1}H nucleus is thus less shielded from electronic and magnetic fields in its surroundings. This complicates especially the theoretical description of core properties such as NMR parameters. Achieving the same accuracy for ^{1}H NMR as for ^{13}C NMR is therefore difficult, which we already observed for the Δ_{corr}-ML correction and there is no reason to assume that this behavior is different in the case of Δ_{SO}-ML. This is even exacerbated by the fact that the hydrogen atoms are usually further away from the HA than carbon atoms meaning that the $\Delta_{SO}\delta$ values are smaller and can be less systematic. The data set for ^{1}H NMR SO contributions shows a significant number of large values not only for very close ^{1}H nuclei, but also for those bound to a HA via three covalent bonds (see Supporting Information, Figure S2). Hence, the distance criterion is weaker for ^{1}H than it is for ^{13}C, representing a bigger challenge for the Δ_{SO}-ML model.

The greater complexity compared to ^{13}C NMR is confirmed by the somewhat weaker performance of the Δ_{SO}-ML approach applied to the *low-level* method SR-PBE0/ZORA-def2-TZVP for ^{1}H NMR indicated by the metrics in Table D.1 and the detailed analysis of a training run in Figure D.5. Compared to the ^{13}C data, the performance of the ML approach for ^{1}H NMR is indeed worse, but the functionality is still retained. The Δ_{SO}-ML method predicts qualitatively correct values within the whole data range even in the extreme regions (Figure D.5(a)) and scattering of the data including the ML-predicted SO contributions is

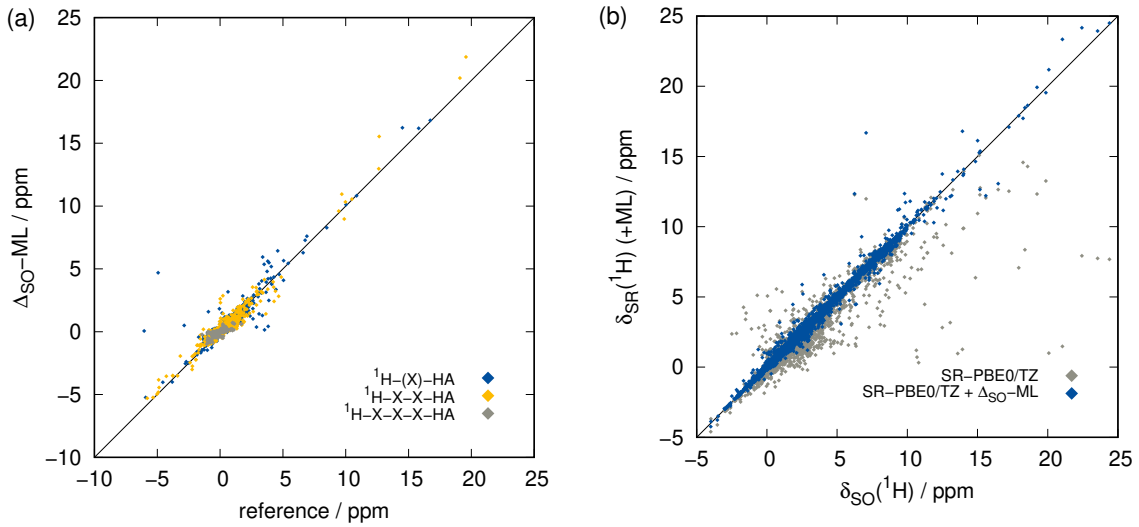


Figure D.5: Comparison of the ML-predicted SO contributions to the reference (SR/SO)-ZORA-PBE0/TZ2P ones for the ^{1}H NMR test set. (a) Values of $\Delta_{SO}\delta$, color-coding: ^{1}H bound to a HA directly or via two (blue), three (yellow), or four or more (gray) covalent bonds. (b) Total chemical shift δ neglecting SO coupling (gray) and adding the ML-predicted $\Delta_{SO}\delta$ (blue).

reduced significantly compared to the purely SR values (Figure D.5(b)). It is noticeable that the data of nuclei in close and medium vicinity to a HA is spread over the whole data range, only the ^{1}H nuclei at least four bonds away from the HA are loosely restricted to a region of small $\Delta_{SO}\delta$ values and small prediction errors. The overall improvement achieved by the correction is also substantiated by the metrics in Table D.1. That is, the MAD resulting from the 8055 data points in the test set is reduced by 68% from 0.281 to 0.090 ppm and the overall chance for large errors is reduced, too, as indicated by the 71% decrease of the RMSD from 0.803 to 0.236 ppm. In contrast to the ^{13}C NMR case, the MD for the purely SR ^{1}H NMR chemical shifts is slightly negative, but it is nevertheless basically eliminated. Unfortunately, there is at least one large outlier in the predicted $\Delta_{SO}\delta$ values leading to an increased positive maximum error of 10.825 ppm. Taking into account the overall reduced error spread and range (reduced from 26.611 to 15.597 ppm) as well as the small RMSD, this can be considered an artifact that occurs only very rarely.

The fundamental non-linear correlation between the scalar-relativistic NMR chemical shift and its missing spin-orbit coupling contribution seems to be general and can at least be transferred form ^{13}C to ^{1}H NMR (see Supporting Information, Figures S6 to S8). The linear regression technique is therefore unusable also for the ^{1}H NMR case. To conclude, despite the lower performance of the Δ_{SO}-ML approach for ^{1}H compared to ^{13}C NMR, it still accomplishes a decent improvement, especially if the negligible extra computational expenses and efforts are considered.

D.3.3 Generalizability of the Model

Method Dependence of $\Delta_{SO}\delta$

Since the presented prediction method is supposed to be generally valid, it would be beneficial if the spin-orbit contribution to the chemical shift $\Delta_{SO}\delta$ calculated with DFT does not depend strongly on the functional and basis set that are used for its computation. This can be expected due to the "doubly relative" nature of $\Delta_{SO}\delta$ (δ: difference between two shielding constants, Δ: difference between SR and SO). To test this dependence, a small test set called methyl subset, which is also included in the ML data set, has been investigated with different DFT levels of theory. It contains compounds of all heavy atoms saturated with methyl groups (except for the halogens), more precisely, the molecules CH_3A^{I}, CHA^{I}_3, $(CH_3)_2A^{II}$, $(CH_3)_3A^{III}$, and $(CH_3)_4A^{IV}$ (with A^{I} = Cl, Br, I; A^{II} = Zn, Cd, Hg, Se, Te; A^{III} = Ga, In, Tl, As, Sb, Bi; A^{IV} = Ge, Sn, Pb). As mentioned earlier, the PBE0/TZ2P level of theory was chosen for the calculation of the reference $\Delta_{SO}\delta$ values. The average deviations of those from other functionals and basis sets are summarized in Table D.2. In all test molecules, all carbon nuclei are in direct vicinity to the HA, and all hydrogen nuclei are connected to the HA via two covalent bonds.

Variation of the basis set size is found to have an almost negligible effect on $\Delta_{SO}\delta$ when still a triple- or quadruple-ζ size is sustained with maximum mean absolute deviations of 0.40 ppm for ^{13}C (TZP) and 0.013 ppm for ^{1}H (QZ4P). These values lie below the typically expected errors for density functional approximations (DFA) of roughly 3-8 ppm for ^{13}C NMR and 0.1-0.3 ppm for ^{1}H NMR.[33] When the basis set size is reduced to double-ζ (DZ, DZP), significantly larger deviations are observed. Therefore, TZ2P is considered a well-

Table D.2: Deviation in ppm of $\Delta_{SO}\delta$ calculated with various functional/basis set combinations evaluated to the reference level of theory PBE0/TZ2P on the methyl subset of compounds. M(A)D = Mean (absolute) deviation.

		^{13}C		^{1}H	
Functional	Basis set	MD	MAD	MD	MAD
PBE0	DZ	−1.28	1.28	0.022	0.045
PBE0	DZP	−0.60	0.64	0.016	0.025
PBE0	TZP	−0.28	0.40	0.006	0.012
PBE0	QZ4P	0.11	0.30	0.003	0.013
PBE	TZ2P	−0.82	0.84	0.004	0.053
BLYP	TZ2P	−1.74	1.98	0.012	0.065
mPW	TZ2P	−0.94	0.94	0.000	0.063
B3LYP	TZ2P	−1.05	1.51	0.010	0.019
mPW1PW	TZ2P	−0.12	0.26	−0.003	0.006
PBE0	TZ2P	0.00	0.00	0.000	0.000

balanced and reasonably large basis set for the purpose of this work. The use of different DFAs exhibits a more pronounced effect on the value of $\Delta_{SO}\delta$. While for ^{13}C, no dependence on the functional class (GGA/hybrid) is observed, in the ^{1}H case, the deviation from the hybrid PBE0 to the GGAs is larger than for the hybrids DFAs. Nevertheless, even when BLYP is used, which has the largest MAD compared to PBE0 (1.98 ppm for ^{13}C, 0.65 ppm for ^{1}H), the functional differences are still very low. According to these data, the SO prediction method is expected to be generalizable to methods other than PBE0/TZ2P with a small residual method inconsistency error.

Transferability of the Δ_{SO}-ML Method

The claim made earlier, that the reference $\Delta_{SO}\delta$ contribution can basically be predicted via any *low-level* DFT method, has yet to be proven. Therefore, three other example DFT levels of theory were investigated regarding their use as baseline methods for the Δ_{SO}-ML approach. These include SR-ZORA in conjunction with PBE as the GGA variant in the same functional family as PBE0 and r^{2}SCAN0 as a different hybrid DFA that performed well for both ^{1}H and ^{13}C NMR chemical shifts with respect to canonical CCSD(T) evaluated on the data set of the Δ_{corr}-ML model.[197] Furthermore, an approach without the explicit treatment of scalar-relativistic effects is investigated, namely PBE0/def2-TZVP, which uses the def2 effective core potentials (ECP) for elements with $Z > 36$ and thus implicitly comprises some amount of relativity. Furthermore, the question arises whether it is necessary to recompute all chemical shifts of the data set and retrain the Δ_{SO}-ML model if methods other than SR-PBE0 are used.

Answers to these questions can be obtained from analyzing the performance differences of the mentioned methods for the test data set depicted in Figure D.6. First, the previously examined performance using the SR-PBE0 method is clearly visible when compared to the *uncorrected* data (without any SO contribution) and the prediction qualities for all other tested methods are of equal dimension. By taking a closer look, it is surprising to find that in the case of ^{13}C NMR, the performance of the Δ_{SO}-ML method is not reduced noticeably when SR-PBE or SR-r^{2}SCAN0 are used as *low-level* methods. This means that the *electronic* and *magnetic* input features obtained from PBE and r^{2}SCAN0 chemical shielding calculations closely resemble those from a PBE0 calculation (all *geometric* descriptors are identical). A computationally more affordable level of theory such as PBE can therefore easily be used to reconstruct the SO contribution to the ^{13}C NMR chemical shift at the PBE0 level without changing the ML model. Subsequently, it is not surprising that the Δ_{SO}-ML approaches for PBE and r^{2}SCAN0 retain their performance when the model is trained on data obtained from these respective DFAs. The situation changes slightly, when not the functional, but the relativistic approximation is changed. The use of the simpler ECP variant of PBE0/def2-TZVP and the standard instead of the ZORA Hamiltonian results in a slightly increased RMSD of 3.67 ppm (compared to 2.76 ppm for SR-PBE0/ZORA-def2-TZVP) and a larger error spread. This indicates that the input features differ more severely between the ECP and SR-ZORA approaches than between the different functionals. For some further insights,

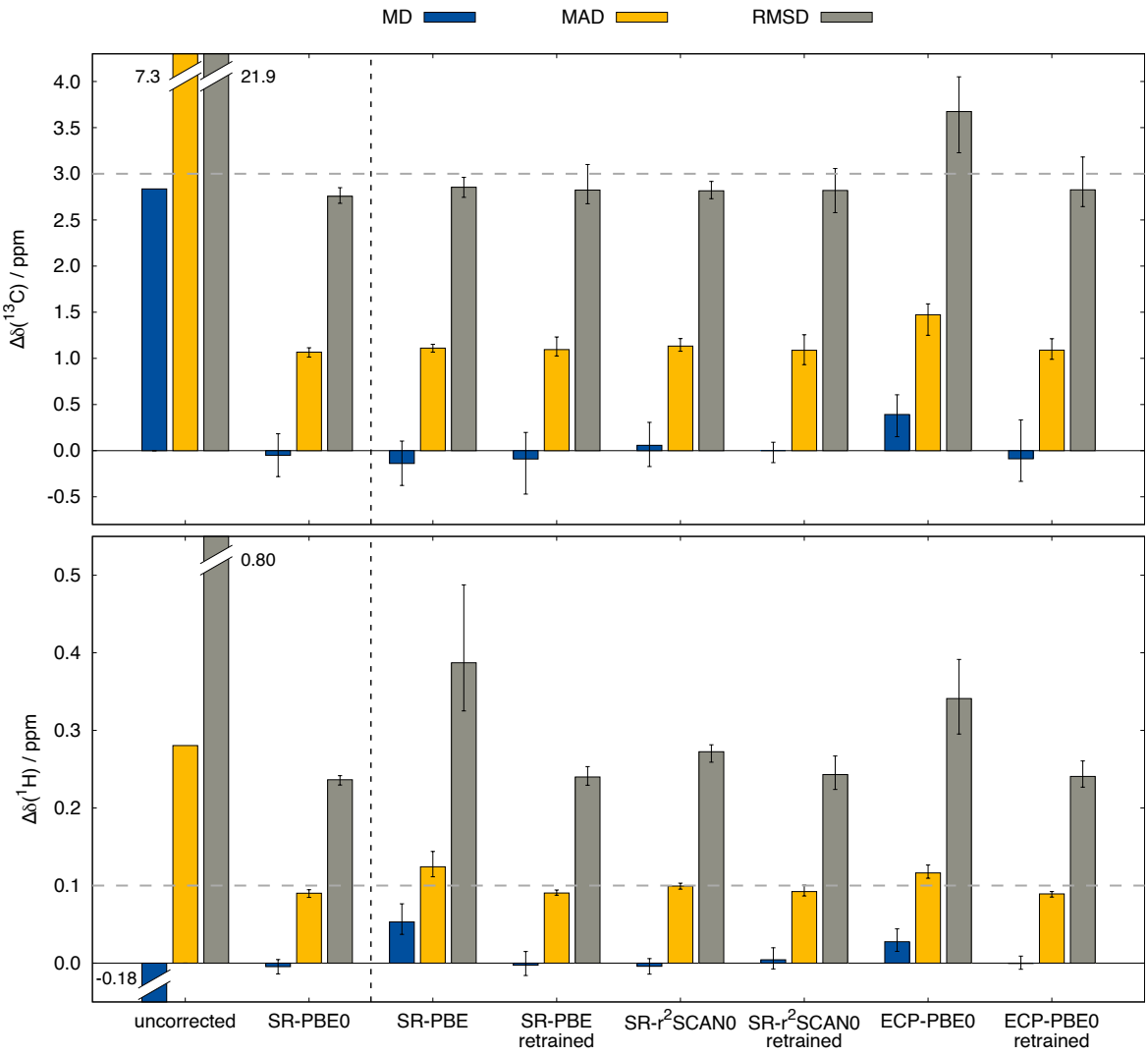


Figure D.6: Comparison of different *low-level* methods as baseline of the Δ_{SO}-ML approach and their metrics evaluated for the ^{13}C and ^{1}H test data sets. In all cases, the def2-TZVP basis set in the scalar-relativistic ZORA (SR) framework or with the def2-ECPs (ECP) has been used with the respective functional, *uncorrected* refers to $\Delta_{SO}\delta = 0$. Each functional's data results from using the ML model trained on SR-PBE0 data, while *retrained* indicates that the ML model has been recreated using the training data calculated at the respective level of theory. The horizontal dashed line indicates the minimum expectable DFT method error.

a selection of six *electronic* and *magnetic* ^{13}C descriptors is depicted in Figure D.7 showing their correlation when obtained from the different calculations via the ECP and the SR-ZORA approach. In all cases, a more or less strongly pronounced correlation is found which explains the ability of the Δ_{SO}-ML method to predict reasonable results even without retraining of the model. Since the correlation has an approximately linear character, the ML model is capable of adapting to the different data when it is retrained and thus recovers its

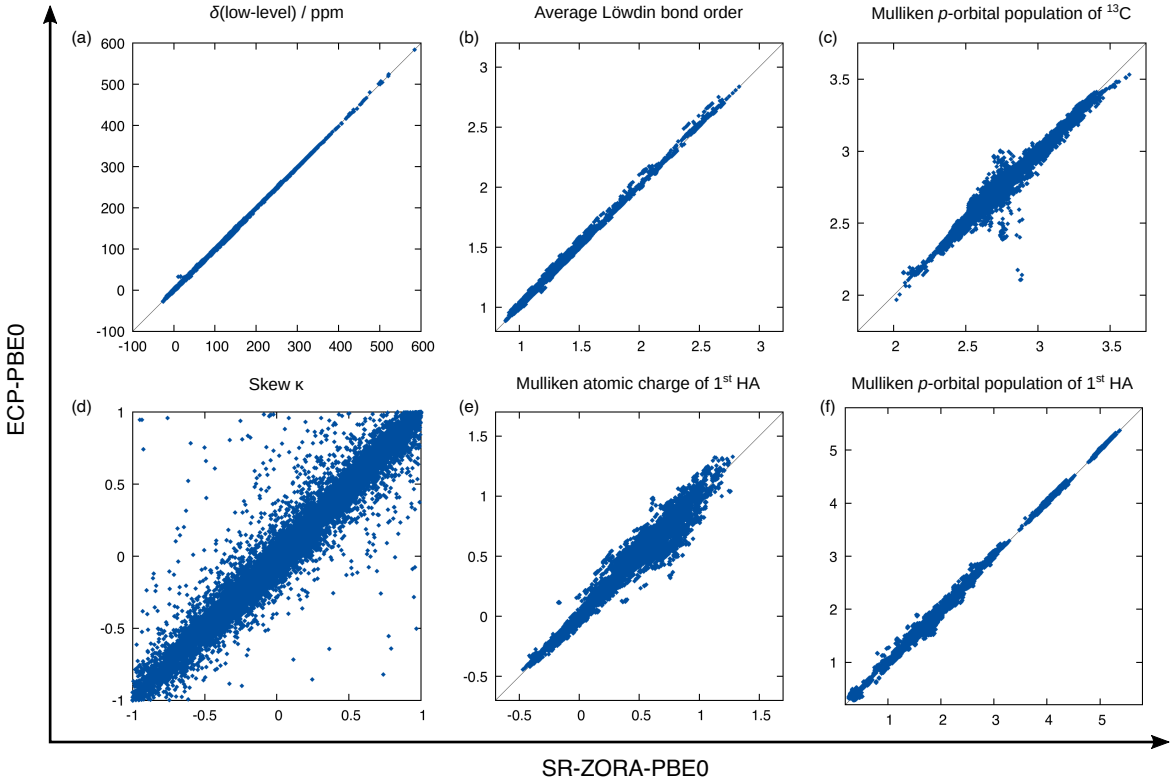


Figure D.7: Correlation plots of a selection of descriptors in the ^{13}C input feature vector as calculated at the PBE0 level with the ECP variant (def2-TZVP) against the SR-ZORA variant (ZORA-def2-TZVP). (a) δ calculated at *low-level*, (b) average bond order of all bonds of the respective ^{13}C, (c) *p*-orbital population at the ^{13}C, (d) skew $\kappa = \frac{3(\sigma_{iso}-\sigma_{22})}{\sigma_{33}-\sigma_{11}}$ (from shielding tensor σ), (e) atomic charge of the neighboring HA, (f) *p*-orbital (valence) population of the neighboring HA.

initial performance (RMSD of 2.83 ppm). The features from the ECP-PBE0 calculation are therefore not necessarily less suited for use in the ML model.

Analysis of the results for ^{1}H NMR reveals the expected behavior for the more complex circumstances mentioned above. In contrast to ^{13}C, there are significant performance losses if other functionals are used to generate the ML input features. While the initial RMSD of 0.236 ppm for SR-PBE0/ZORA-def2-TZVP is only slightly increased to 0.272 ppm for SR-r^{2}SCAN0, the loss is more drastic for SR-PBE with an RMSD of 0.387 ppm, which is even higher than the RMSD of ECP-PBE0 (0.341 ppm). Apparently, the ^{1}H nucleus is more prone to differences in the input features and the variations between data from different functionals is larger than for ^{13}C NMR. This might, to some extent, originate from the much smaller typical chemical shift range for ^{1}H NMR (about 0-12 ppm) for which deviations of a few tenths of ppm are already substantial. Nonetheless, also for ^{1}H NMR the performance of the original SR-PBE0 method can be recovered when the model is trained on the corresponding data. Thus, RMSDs of 0.240, 0.243, and 0.241 ppm can be achieved for SR-PBE, SR-r^{2}SCAN0,

and ECP-PBE0, respectively. Despite the limited number of investigated DFT levels of theory, we feel confident that the presented behaviour of the Δ_{SO}-ML method is of general nature, making it a powerful tool for the low-cost assessment of SO contributions to computed NMR chemical shifts. Even when the base method SR-PBE0 cannot be applied, the $\Delta_{SO}\delta$ values predicted from lower-level DFT methods are reliable enough for a rough estimation and can serve as diagnostic tool to detect possible severe SO contributions and avoid large computational errors.

D.3.4 Performance for External Test Systems

To evaluate the Δ_{SO}-ML method in real-world applications, it was tested on three different sets that are independent from the data set used for training and testing described herein. These are the SnS51[199] set containing various organotin compounds and its successor for organolead chemistry[200]. In addition, a new set has been compiled for HALA effects on ^{13}C nuclei comprising experimental ^{13}C NMR chemical shifts from structures with all 17 heavy elements included in the training data of the Δ_{SO}-ML method, which we call 17HAC. Finally, the particularly difficult example of a bismabenzene compound is showcased in more detail.

Performance for Organotin and Organolead Compounds

In comprehensive benchmark studies, we investigated various DFT methods regarding their ability to predict ^{119}Sn and ^{207}Pb NMR NMR chemical shifts. Most of the compounds from these studies were now used to investigate the HALA effect caused by the presence of Sn/Pb atoms and the predictive power of the Δ_{SO}-ML approach on this quantity. For this purpose, the conformers lowest in Gibbs free energy from both benchmark sets were recalculated at the reference level of theory used herein (SO-ZORA-PBE0/TZ2P) for a purely computational evaluation (for more technical details, see the Supporting Information, Section 3.2.1). Thus, a data set containing 817 ^{13}C NMR chemical shifts in Sn-containing compounds and 1415 in Pb-containing compounds (for ^{1}H NMR: 1170 and 2059, respectively) was analyzed.

The results for ^{13}C NMR are shown in Figure D.8 and segmented into subgroups of atoms with different distances between the heavy and the light atom. It is obvious that in both cases (Sn and Pb), the (virtually costless) Δ_{SO}-ML prediction helps to reduce the relativity-related errors drastically. It stands out that this is especially significant for ^{13}C nuclei directly bound to the HA, which is the clear strength of the method. In the case of organolead compounds, this category undergoes the by far most pronounced HALA effects (RMSD of 29.55 ppm if no SO contribution is included). For ^{13}C nuclei connected to Pb via more than one bond, the initial errors are much smaller due to a notably smaller SO contribution making it more difficult to cover the effect by the correction method. For a distance of three or more bonds, Δ_{SO}-ML does not yield a useful prediction anymore, which, in practice, would not stand out as the average SO contibution of this category (MAD of 2.21 ppm) is below the typical error of DFT in general. The analysis of the organotin compounds suggests a very similar behavior with the main difference that the SO contribution form Sn as HA is generally smaller. Still,

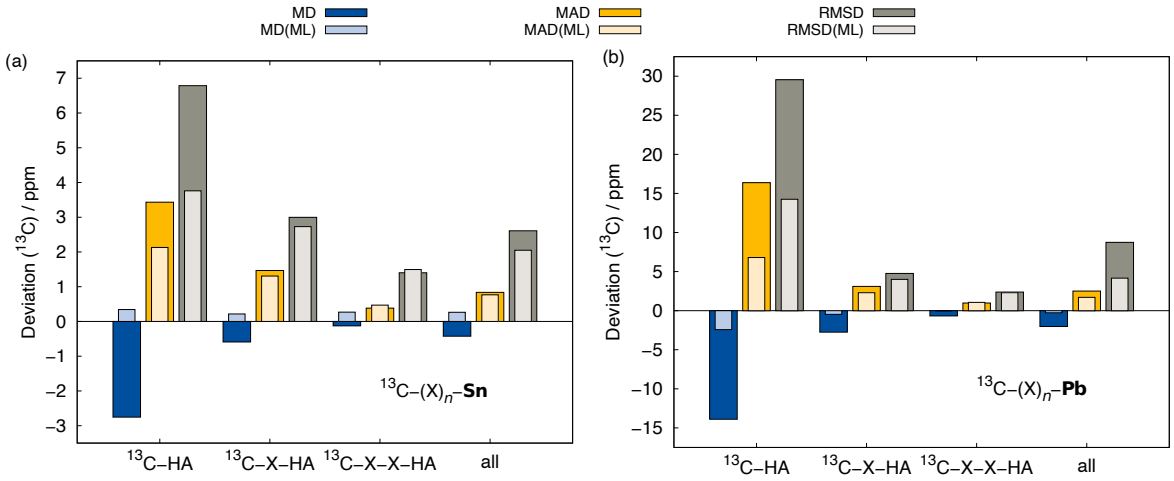


Figure D.8: Comparison of ^{13}C NMR metrics for the (a) C−Sn and (b) C−Pb test structures without any SO contribution (PBE0/ZORA-def2-TZVP, full colors) and with the ML-predicted $\Delta_{SO}\delta$ values (brighter colors). The data is averaged over ^{13}C nuclei bound to a HA via one (C−HA), two (C−X−HA), and three or more bonds (C−X−X−HA) and over the full data (all).

the Δ_{SO}-ML method predicts large SO contributions reasonably well and also improved the overall statistics, but the generally smaller $\Delta_{SO}\delta$ values are lost in the DFT-related noise sooner.

Similar trends are observed for the ^{1}H NMR data, whereas a more pronounced HALA effect on nuclei connected to Sn/Pb via two and three bonds was noticed (see Figure S9 in the Supporting Information). However, the overall SO contributions are again significantly smaller so that the Δ_{SO}-ML correction will be less important than the choice of an appropriate density functional in these cases. Nevertheless, it succeeds in predicting large SO contributions in ^{1}H nuclei close to Pb atoms.

Performance on the 17HAC Test Set

When heavy elements are involved, it is obvious that any inclusion of SO contributions to chemical shifts should reduce the deviation to experimental data. However, we have so far only tested the Δ_{SO}-ML approach with respect to theoretical reference data. As it is important to validate the method against experimental data, too, a new benchmark set was constructed that consists of 63 mostly organic molecules featuring all 17 HAs (at least three molecules per heavy element) included in the training data set. In total, 236 experimental ^{13}C NMR shifts were collected from nuclei in different distances form the HA and with different degrees of SO effects to the ^{13}C nucleus (more details are given in the Supporting Information, Section 3.2.2). To systematically address all typical sources of error mentioned in the beginning, the following workflow was applied for all compounds. First, a conformer search was performed as described in Section D.2.4 to integrate the conformational flexibility of the systems. The plain DFT results were then obtained as the Boltzmann-average of

the ^{13}C NMR chemical shifts calculated with PBE0/ZORA-def2-TZVP with the implicit CPCM solvent model to incorporate solvent effects. Subsequently, in order to tackle the electron correlation and relativistic effects to the NMR shifts, the Δ_{SO}-ML and Δ_{corr}-ML (retrained on *low-level* PBE0/ZORA-def2-TZVP data) corrections were applied to calculations without a solvent model and the contributions were added. The last main error source – the rotational-vibrational effects – are expected to mostly cancel when chemical shifts are calculated relative to a reference compound.[44] Some exemplary results from this workflow are shown in Figure D.9 and the final statistics are given in Table D.3.

The analysis in Table D.3 shows that the two Δ-ML corrections tackle different quantities. Upon including the SO effects via Δ_{SO}-ML, the RMSD is reduced drastically, because the focus of the correction lies in detecting large SO-HALA effects which leads to a clear decrease

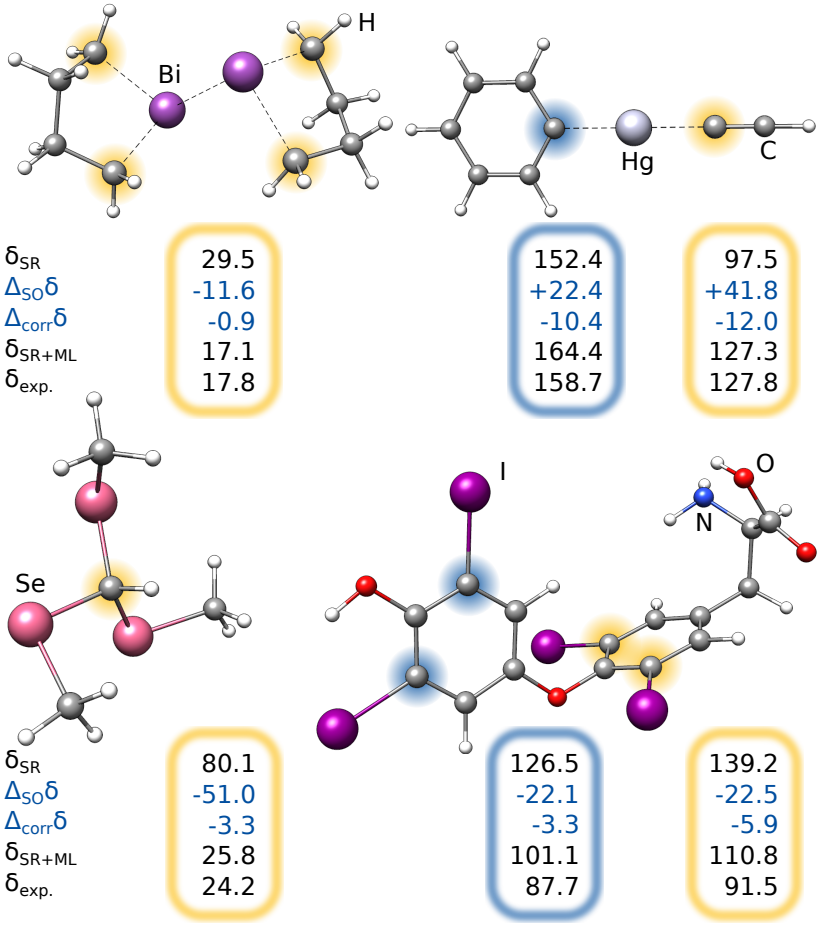


Figure D.9: Four example molecules from the 17HAC test set (showing the lowest-energy conformer) and selected ^{13}C NMR chemical shifts calculated with the *low-level* DFT method PBE0/ZORA-def2-TZVP before (SR) and after addition of the Δ-ML contributions (SR+ML) and compared to experimental values.

Table D.3: Statistics for the 17HAC benchmark set evaluated with the baseline PBE0/ZORA-def2-TZVP(CPCM) level (SR) and with systematic addition of the ML-predicted spin-orbit and/or correlation contributions.

Error metric	SR	$+\Delta_{SO}$-ML	$+\Delta_{corr}$-ML	+both
MAX (<0)	−37.32	−15.75	−50.38	−18.47
MAX (>0)	112.62	73.11	104.58	61.54
MD	9.99	8.59	6.33	4.94
MAD	11.41	8.93	8.81	5.73
RMSD	20.81	12.48	18.80	9.37

of the large errors in these cases. On the other hand, the Δ_{corr}-ML corrects for a rather systematic correlation-related error which is usually not as large for single cases, but smaller, yet significant, for the majority of the ^{13}C nuclei. Therefore, the RMSD is only slightly reduced, but the MAD is smaller than when only Δ_{SO}-ML is used. Nevertheless, the best results are achieved when both corrections are applied, yielding a roughly halved value for all statistical quantities (MAD reduced by 50%, RMSD by 55%). Hence, a systematic treatment of the typical error sources in the computation of NMR chemical shifts does lead to a systematic decrease of the deviation to experimental data.

Showcase: Bismabenzene

For the majority of the test cases shown so far, the nuclei (especially ^{13}C) closest to the HA were affected most by spin-orbit effects. However, the effect is able to propagate[48] and in rare cases, it can even be much larger for atoms further away. This seems to be the case in bismabenzene, where the Bi$-^{13}$C effect is largest in the *para* position. We therefore studied a bismabenzene derivative that could be synthesized[201] in order to test the Δ_{SO}-ML method for a representative extreme case with only little similar data available for training. As before, a conformer ensemble was generated and refined at the r^{2}SCAN-3c level of theory and solvation was included applying CPCM (chloroform). The reference SO contributions are depicted in Figure D.10 and the results are listed in Table D.4.

Most importantly, the extreme $\Delta_{SO}\delta$ value of the *para*-^{13}C needs to be included in order to achieve a qualitative agreement with the experiment (i.e., $\delta(^{13}\text{C}, \textit{meta}) < \delta(^{13}\text{C}, \textit{para})$). Furthermore, despite being visibly too low for *para*-^{13}C, the predicted $\Delta_{SO}\delta$ values are in qualitative agreement with the reference method. Including both ML contributions (SO and corr) does not recover the correct ordering of the chemical shifts, but significantly approaches the experimental results. A similar behavior is observed for ^{1}H NMR with the *meta*-^{1}H being affected most. Since a satisfying agreeing is not achieved even with including both the true SO contribution and the ML-predicted correlation correction, we attribute the major part of the remaining error to solvation and dynamic effects. Nevertheless, the example of the bismabenzene compound shows that the Δ_{SO}-ML method provides reasonable approximations to the SO contribution to NMR chemical shifts even in potentially

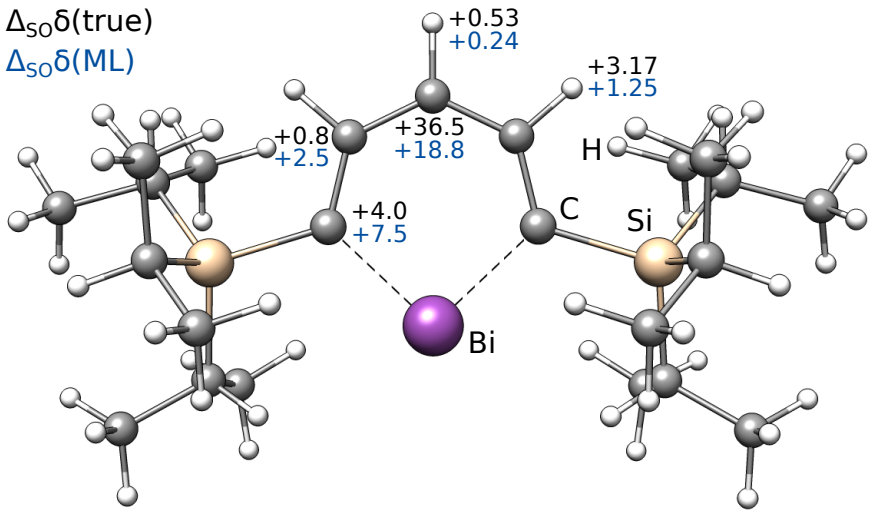


Figure D.10: Lowest conformer of the investigated bismabenzene derivative and $\Delta_{SO}\delta$ values given in ppm for the aromatic ^{13}C and ^{1}H nuclei as calculated at the SO-ZORA-PBE0/TZ2P level of theory (true, black) or predicted via the Δ_{SO}-ML method (blue) and Boltzmann-weighted over all conformers.

unexpected cases.

For the lowest-energy conformer of the bismabenzene derivative, timing evaluations were performed at different theory levels (see Figure D.11). While several hours are required using the all-electron SR *low-level* methods (e.g., PBE0/ZORA-def2-TZVP), the SO reference calculation takes multiple days. In contrast, the training of the Δ_{SO}-ML model lasts a few minutes (for ten training runs for statistical averaging) and the prediction of $\Delta_{SO}\delta$ is done in

Table D.4: Chemical shift data for the investigated bismabenzene derivative calculated with the *low-level* method (SR-PBE0/ZORA-def2-TZVP) and the spin-orbit (SO) and correlation (corr) contributions resulting in total values with both contributions predicted with ML (ML/ML) or using the true SO contribution (true/ML). All chemical shift values are Boltzmann-averaged and include a solvation contribution from CPCM.

	^{13}C			^{1}H	
	ortho	*meta*	*para*	*meta*	*para*
Low-level	241.5	147.3	121.9	8.93	7.57
$\Delta_{SO}\delta$ (ML)	7.5	2.5	18.8	1.25	0.24
$\Delta_{SO}\delta$ (true)	4.0	0.8	36.5	3.17	0.53
$\Delta_{corr}\delta$ (ML)	−13.7	−2.9	−4.4	−0.25	−0.15
Total (ML/ML)	235.2	147.0	136.3	9.93	7.66
Total (true/ML)	231.8	145.2	153.9	11.85	7.95
Experiment	222.4	136.5	153.5	11.62	7.68

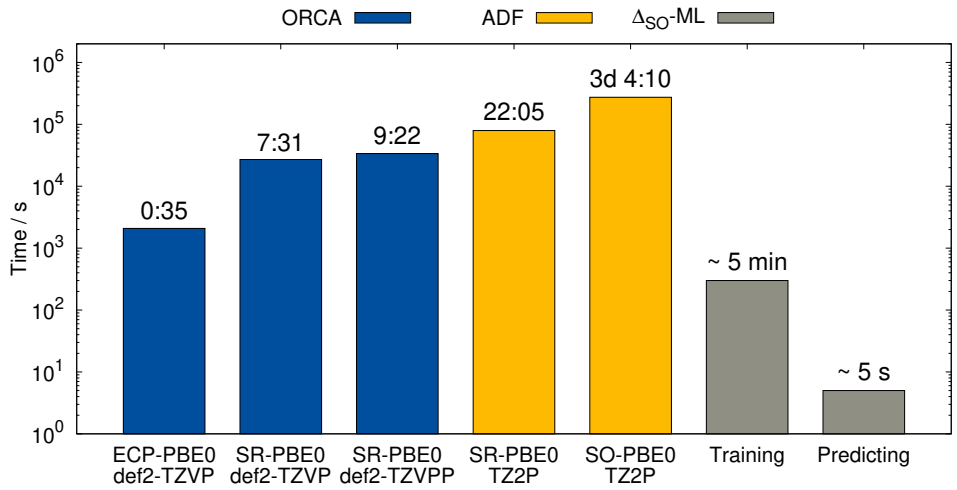


Figure D.11: Timings for NMR shielding calculations of the bismabenzene derivative at different DFT levels (ORCA) compared with SR and SO calculations (ADF) and the time required by the Δ_{SO}-ML model. Calculations were performed in parallel on four cores of an Intel® Xeon® CPU E3-1270 v5 @ 3.60 GHz. Values in hours:min; the ML evaluations were performed on a usual desktop PC.

seconds on a usual desktop computer. The speed advantage of the presented method is thus evident and it is emphasized that using an ECP-based *low-level* method has the potential of an even larger speedup with a comparable performance of the Δ_{SO}-ML method.

D.4 Conclusion

The consideration of relativistic spin-orbit effects for molecules containing heavy atoms is vital for the reliable simulation of their ^{13}C and ^{1}H NMR spectra. This treatment is computationally much more demanding than a non-relativistic calculation and requires non-standard procedures and software, making it less accessible for non-expert users and unsuitable for screening purposes. We presented a machine learning regression-based approach called Δ_{SO}-ML to approximate the SO contribution in ^{13}C and ^{1}H NMR chemical shifts. The underlying data set contains 6388 structures with 17 of the most important heavy elements from group 12 to 17 and NMR calculations at a hybrid DFT level including SO-relativistic treatment via the ZORA technique for 38740 ^{13}C and 64436 ^{1}H NMR chemical shifts. Moreover, the data set can easily be extended by including more diverse structures, e.g., with total charges or further heavy atoms. We showed that the method recovers about 85% of the SO contribution for ^{13}C (70% for ^{1}H) on the test data subset. It is further transferable for use with other density functionals than the base method PBE0 and other approaches for including scalar-relativistic effects, such as ECPs. Since the SO contribution $\Delta_{SO}\delta$ depends only slightly on the DFA and basis set as shown for heavy atom methyl compounds, the generalizability of the Δ_{SO}-ML method renders it broadly applicable for a wide range of DFT methods with a fairly good accuracy even without retraining the model

for other *low-level* methods. In principle, it is not even limited to DFT, but only needs a method that can provide the required input features. This might be correlated methods such as coupled cluster or semiempirical approaches. Predicting $\Delta_{SO}\delta$ is done in a few seconds and only requires the converged *low-level* NMR shielding calculation and the pre-trained ML model, making it superior to other low-cost methods, such as linear regression techniques, that are only applicable to special problems.

Off its training and test data set, the method proved powerful for the prediction of SO contributions caused by nearby Sn and Pb atoms in realistic systems. Moreover, a workflow that systematically addresses all main error sources in NMR parameter computation significantly reduces the deviations to experimental data throughout all 17 HAs resulting in an average error reduction by about 50% when both the Δ_{SO}- and Δ_{corr}-ML corrections are applied. If the computational resources allow an explicit treatment of the SO-relativistic effects, the Δ_{SO}-ML method can function as a diagnostic prescreening tool to identify systems with potentially large SO contributions that are subsequently treated on a higher level of theory only if necessary. The potential fields of application of the Δ_{SO}-ML method lie in high-throughput workflows such as structure assignment methods that can be improved when a higher level of theory is used[428] and as an additional ingredient in low-cost composite method approaches that rarely include any relativistic treatment.[220] To conclude, the new Δ_{SO}-ML method is able to robustly predict SO contributions to NMR chemical shifts for large systems and delivers its full potential when used together with the Δ_{corr}-ML correction in low-cost NMR prediction schemes.

Supporting Information

Electronic supplementary information available. See DOI: 10.1039/d3cp05556f.

- Additional information and analyses regarding some special aspects that were discussed here (PDF)
- Archive of all relevant chemical structures in Cartesian coordinate (.xyz) format (ZIP)
- Raw data of all evaluations presented here and in the Supporting Information (XLSX)

Acknowledgements

J.B.K.B. is most grateful to the "Fonds der Chemischen Industrie (FCI)" for financial support. S.G. is thankful to the German Research Foundation (DFG) for funding through the SPP 2363: "Utilization and Development of Machine Learning for Molecular Applications – Molecular Machine Learning". Further, S.G. and M.B. gratefully acknowledge financial support of the Max Planck Society through the Max Planck fellow program. Open Access funding provided by the Max Planck Society.

Conflicts of Interest

There are no conflicts to declare.

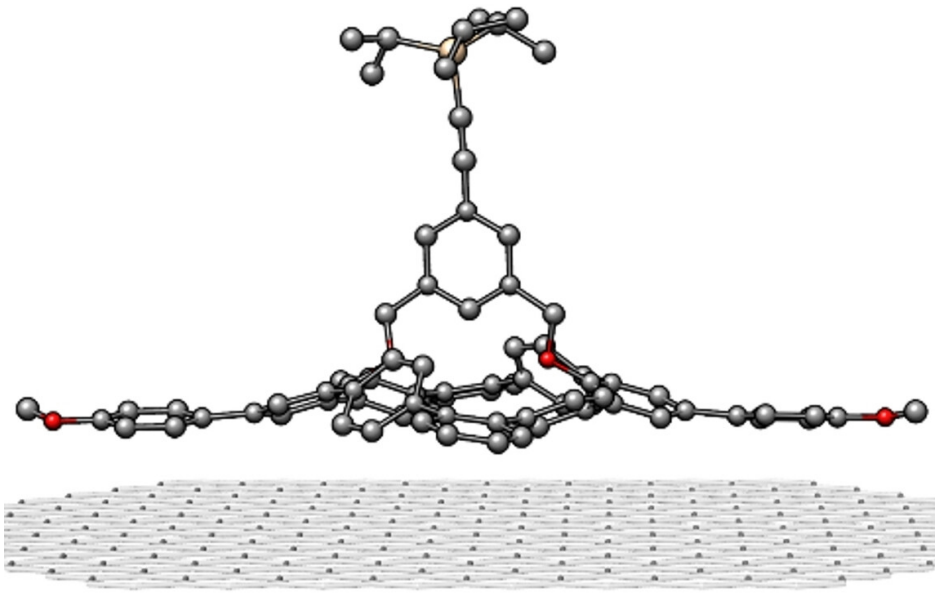
Data Availability Statement

The implementation of the Δ_{SO}-ML and Δ_{corr}-ML methods as well as the associated data sets can be found at https://github.com/grimme-lab/ml4nmr.

APPENDIX E

Appendix: Modular Bicyclophane-Based Molecular Platforms

D. Kalle, J. Bahr, T. J. Keller, J. B. Kleine Büning, S. Grimme, M. Bursch, S.-S. Jester, and S. Höger, *Chem. Eur. J.* **2023**, *29*, e202302662. DOI: 10.1002/chem.202302662.



Bicyclophane-based molecular platforms form extended 2D crystalline monolayers after adsorption from solution onto HOPG and contain orthogonal pillars that point towards the supernatant solution phase.

Abstract

The modular synthesis of a series of nanoscale phenylene bicyclophanes with an intraannular orthogonal pillar is described. The compounds are obtained by a Suzuki cross-coupling condensation and are characterized by mass spectrometry and NMR spectroscopy as well as *in situ* scanning tunneling microscopy at the solid/liquid interface of highly ordered pyrolytic graphite. In addition, their structures and conformations are supported by quantum chemical calculations, also after adsorption to the substrate. A set of two alkyl chain substitution patterns as well as a combination of both were investigated with respect to their ability to form extended 2D-crystalline superstructures on graphite. It shows that not the most densely packed surface coverage gives the most stable structure, but the largest number of alkyl chains per molecule determines the structural robustness to alterations at the pillar functionality.

E.1 Introduction

The molecular platform approach allows the decoration of surfaces with specific functionalities at a specific distance to the floor.[202–206] To do so, a pillar is mounted to the platform, with the latter having the tendency to adsorb flat on the surface. The pillar then points upwards, perpendicular to the surface, and carries at its end the functional group or molecular functionality. Their elevation from the surface is given by the length of the pillar and their spatial separation from each other by the lateral extension of the platform. Both values are to a certain extent freely adjustable. In addition, coadsorption with platform molecules that contain no or even other functional groups, either in form of a regular cocrystal or on random lattice sites, allows a further dilution of the specific functionalities in the nanopattern.[207] Method of choice to investigate these structures is scanning tunneling microscopy (STM) which can be applied in situ at the solid/liquid interface between highly oriented pyrolytic graphite (HOPG) and a low volatile solvent in which the platform molecules are dissolved. For HOPG, it is rather easy to prepare an area with atomically flat layers having extensions of several hundred nanometers.

In order to keep the large platform molecules processable, generally flexible side chains are attached to the aromatic backbone. The decoration with long alkyl chains increases the compound solubility and lowers their melting point, where branched alkyl chains are usually superior to linear chains.[208,209] Contrary, linear chains play an active role in the pattern formation when the molecules are adsorbed from solution onto HOPG. They usually align along one of the HOPG main axes and tend to interact with each other by forming close van der Waals contacts.[429–433]

Recently, we described a molecular platform based on a spoked wheel structure in which the central hub carries additionally a pillar with a molecular functionality at its end.[210] With a flexible connection, e.g., an alkyl chain, the functionality can tumble around the pillar or even adsorb at the graphite, as shown for a phenyl-C61-butyric acid ester. Contrary, if the functionality is mounted to the pillar *via* a rigid connector, as shown for an aryelene dye at the end of a phenylene-ethynylene pillar, a decoupling from the surface is achieved and the dye points about 3 nm out of the graphite plane. The spoked wheel forms in that case such a robust pattern on the surface that the 2D-lattice is not influenced by the kind of functionality at the end of the pillar.

E.2 Results and Discussion

In our present work, we describe a systematic study on phenylene-based bicyclophanes as platforms for the attachment of functional groups.[211,212] They contain an intraannular 1,3-arylene dibenzylether moiety to which a pillar is mounted at its 5-position.[434] Due to steric constraints, this points into the solvent reservoir after the molecules self-assemble from solution onto the HOPG. At the end of the pillar, we investigated a TIPS-protected acetylene, a benzyl triazole, and a phenyl terpyridyl as model systems of different sizes with different

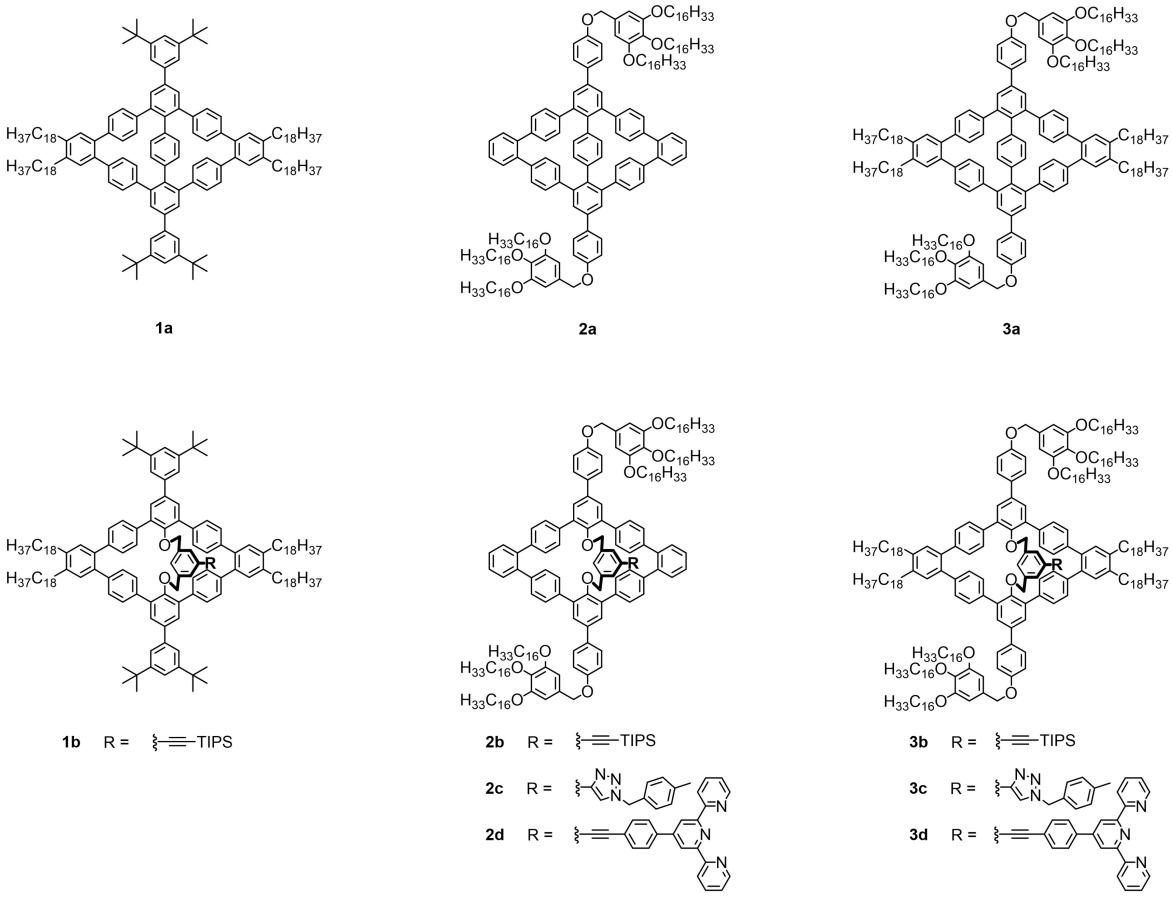


Figure E.1: Structures of the reference bicyclophanes (**1a**, **2a**, **3a**) and the bicyclophanes with intraannular pillars that are oriented orthogonally to the molecular planes (**1b**, **2b–d**, **3b–d**).

ways of their introduction (click chemistry and Sonogashira coupling). Bicyclophanes without the central functional groups serve as reference compounds to investigate whether the pillar and its functionalization influence the adsorption behavior, the lattice geometry and parameters. The structures of the compounds described here are shown in Figure E.1 and stand for a systematic study on the number and position of long alkyl/alkoxy side chains attached to a backbone of certain size and shape. Specifically, we investigate individual alkyl side chains attached to the bicyclophane backbone at the "west/east" positions and 3,4,5-trialkoxybenzyl substituents at the "north/south" positions,[435–438] as well as a combination of both substitution patterns.

The final steps of the syntheses of the bicyclophanes **1a** and **1b** are shown in Figure E.2; for all other products and intermediates experimental details can be found in the Supporting Information (SI). Fourfold Suzuki coupling of the respective tetraiodides (**4**) or bromides (**6**) with two equivalents of diboronate **5** gave the bicyclophanes in 49% (**1a**) and 11% (**1b**) yield, respectively. The 5-position of the central aromatics can be widely varied allowing

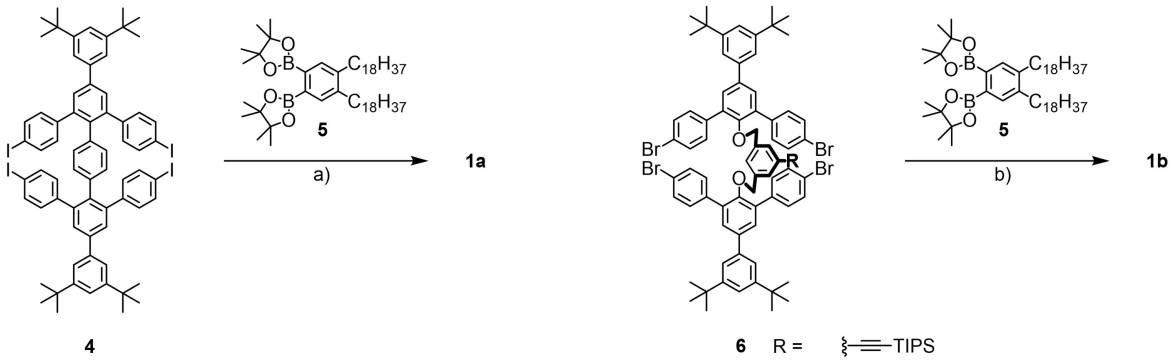


Figure E.2: Synthesis of **1a** and **1b**: a) $Pd(PPh_3)_4$, Na_2CO_3, Aliquat 336, water, toluene (49%); b) $Pd(PPh_3)_4$, Na_2CO_3, Aliquat 336, water, toluene (11%).

the preparation of different functionalities to point into the volume phase after adsorption of the molecules onto graphite. After the formation of **1b**, the ^{1}H NMR signals for the central aromatic that contains the pillar and the signal for the *m*-terphenylene unit are strongly shifted compared to the open precursor **6**. Such a shift is not observed while forming **1a** from **4**. Therefore, model compounds **7** and **8** were prepared where the same NMR signal shift upon bicyclophane formation was observed (Figure E.3), as similar also for the bicyclophanes **2b** and **3b** (see SI). After bicyclophane formation, the signal (s) at the central *m*-terphenylene aromatics shifts by 0.23 ppm downfield (from 7.51 ppm to 7.74 ppm). Even more pronounced is the signal shift of the central phenylene unit with 0.38 ppm for d (from 6.29 ppm to 6.67 ppm) and remarkable 1.59 ppm for t (from 5.87 ppm to 7.46 ppm). Consequently, compounds **7** and **8** were further investigated by quantum chemical calculations[22,247,307,308,387] to elucidate the dominant structural motif and the origin of the extraordinary chemical shifts of the discussed ^{1}H NMR signals.

Conformational sampling was conducted at the GFN2-xTB[ALPB][27,118] level using the CREST[25] algorithm. The obtained conformer ensembles were further refined at the r^{2}SCAN-3c+G_{solv}(COSMO-RS)+G_{mRRHO}(SPH)//r^{2}SCAN-3c[COSMO][114,115,172,421] level using the CENSO[29] program. For the open compound **7**, the most favorable conformers are found to have a dominant structural motif of the central phenyl ring being sandwiched between phenyl rings of the *m*-terphenylene unit. This results in close contact of the protons s, d, and t with the ring current anisotropy regions of the *m*-terphenylene subunits explaining the pronounced shielding effects in the NMR spectrum (Figure E.3). The origin of this shielding is further supported by nucleus independent chemical shielding (NICS)[439] calculations at the r^{2}SCAN0/def2-TZVPP[112,156,198] level. The NICS calculations were performed for a model system of the lowest conformer with removed central phenyl unit to allow additional NICS evaluation at the actual position of protons d and t (Figure E.4, for further information see the SI). The NICS values amount to –3.1 ppm at the position of t and –1.2 ppm averaged at d being in qualitative agreement with the observed shifts compared to the closed compound **8**. In compound **8**, the ring current anisotropy effect on the respective

protons vanishes as no close proximity of them to the now-fused *m*-terphenylene units is observed. Even though, the relative shifts of the strongly shifted protons going from **7** to **8** is explained qualitatively, NMR chemical shift calculations of the Boltzmann averaged conformer ensembles of both compounds indicate a significant overestimation of this effect, specifically for t. To verify the r^{2}SCAN0[CPCM[112,156,198]]/def2-TZVPP chemical shifts, a novel machine learning (ML) correction[197] to reproduce CCSD(T) quality NMR chemical shifts was applied. The ML correction yields only small corrections to the original relative r^{2}SCAN0 shifts. As the discrepancy between experimental and computed chemical shifts is not solved by inclusion of Boltzmann averaged conformer ensembles, molecular dynamics simulations at the GFN2-xTB[ALPB] were performed starting from the lowest conformer to investigate the dynamical effects that are expected to be important for the non-covalent proton$\cdots$phenyl contacts. Accordingly, averaged ML-corrected NMR chemical shifts were calculated for 50 equidistant structure snapshots from the MD simulation. The resulting relative shifts of +0.35 (s), +0.39 (d), and +1.72 ppm (t) (+0.38, +0.47, and +1.68 ppm, respectively, if the ML correction is applied) are in excellent agreement with the experimental results even though the absolute chemical shifts deviate in an expected error range of 0.1-0.3 ppm.[33] The QM calculations explain not only the observed chemical shift changes upon cyclization, they also support the proposed chemical structures of **7** and **8**, and therefore indirectly also the other structures described in this paper. Moreover, they substantiate the orthogonal arrangement of the pillar with respect to the bicyclophane backbone in solution.

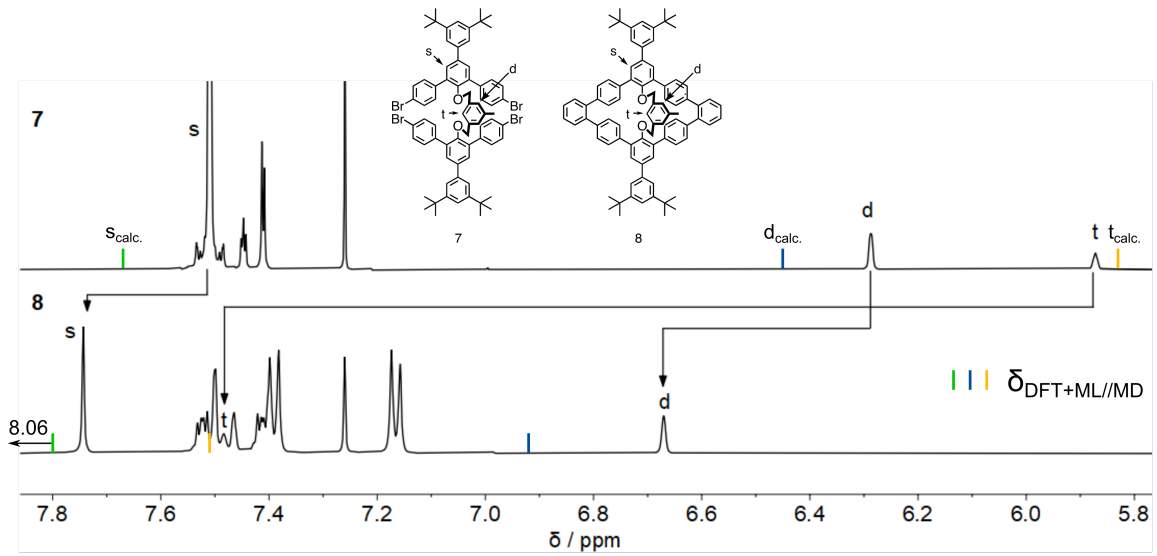


Figure E.3: Model compounds **7** and **8** and shifts of the ^{1}H NMR signals of the respective hydrogen atoms upon cyclization. Calculated, MD simulation averaged, ML-corrected ^{1}H NMR chemical shifts of protons t, s, and d (Note: t and d are not resolved in the spectra).

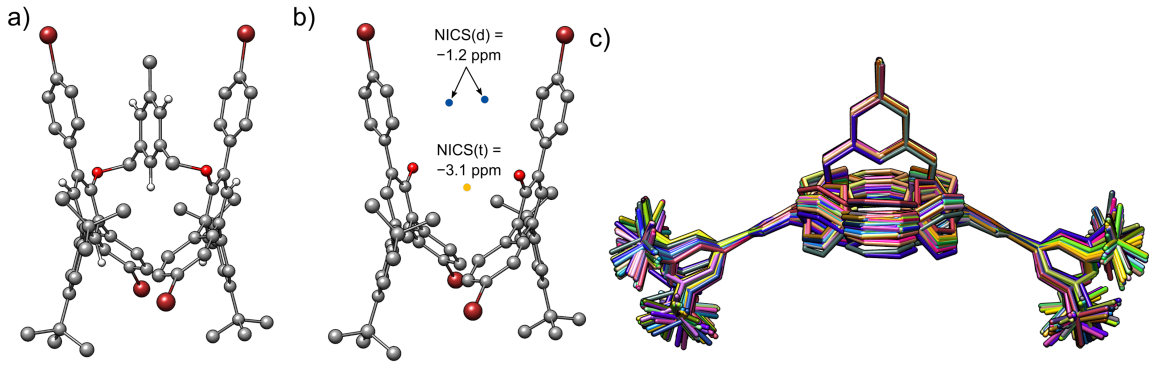


Figure E.4: a) Lowest (free energy) conformer of **7** at the level given below and b) NICS values at the positions of t and d; c) overlay of the 50 lowest (free energy) conformers of **8**. Conformer free energies computed at the r^{2}SCAN-3c+G_{solv}(COSMO-RS)+G_{mRRHO}(SPH)//r^{2}SCAN-3c[COSMO]; NICS values at the r^{2}SCAN0[CPCM]/def2-TZVPP//r^{2}SCAN-3c[COSMO] level.

As seen by STM (Figure E.5a), **1a** packs densely to form a two-dimensionally (2D) crystalline self-assembled monolayer (SAM) at the solid/liquid interface of HOPG and a 10^{-4} M solution of the compound in 1-phenyloctane (PHO).* Rows of the aromatic backbones (that assemble along unit cell vector b) appear in bright color. These are separated by aliphatic regions that appear darker and are resolved down to the level of individual methylene units, arranged in the conformer with their carbon atoms parallel to the HOPG substrate.[429–433] Two pairs of octadecyl side chains, located at the "west" and "east" positions of the bicyclophane backbones, respectively, are oriented along one of the main axis directions of the underlying HOPG substrate, d, and interdigitate with alkyl chains of the neighboring backbones, thus defining the row distance, $a/2$. The backbones of adjacent rows are oriented along c_1 and c_2 with $\gamma(b, c_1) = \gamma(b, c_2) = (37 \pm 5)°$. To this packing, a rectangular unit cell with parameters $a = (7.5 \pm 0.2)$ nm, $b = (1.8 \pm 0.2)$ nm, $\gamma(a, b) = (90 \pm 2)°$, and an orientation of $\gamma(a, d) = (4 \pm 2)°$, incorporating a total of two molecules, is indexed. Unexpectedly, the molecule design of **1a** does not allow for an exchange of the intraannular phenylene group with regard to the ability to adsorb on the HOPG surface. With a pillar-like unit (carrying a triisopropylsilyl(TIPS)-protected acetylene group) attached at the backbone interior, **1b** could not be observed forming SAMs in the same manner as shown for **1a**. Therefore, while the reference compound **1a** forms a regular 2D pattern, the platform molecule **1b** does not behave alike. We conclude that the limited number of only four alkyl chains per molecule allows the pattern formation of flat **1a**, but analog structures with a TIPS-acetylene substituted pillar that expands into the volume phase are not robust enough to be visualized

* As a routine procedure and unless otherwise noted, all samples were prepared by applying 0.2 µL of a solution of the respective compound in PHO onto HOPG at an elevated temperature of 80 °C, keeping it at this temperature for 20 s, and allowing it to cool to room temperature before performing the STM measurement with the tip immersed into the solution. This thermal annealing procedure was done to overcome possible kinetic energy barriers, avoiding amorphous regions, and promoting the annealing of possible defects by reorientation. See, e.g., Refs. [440, 441].

under STM conditions, if it is formed at all. At present, although highly speculative, we ascribed this behavior to the increased solubility of **1b** in relation to the number and length of linear alkyl chains that mediate the surface adsorption.

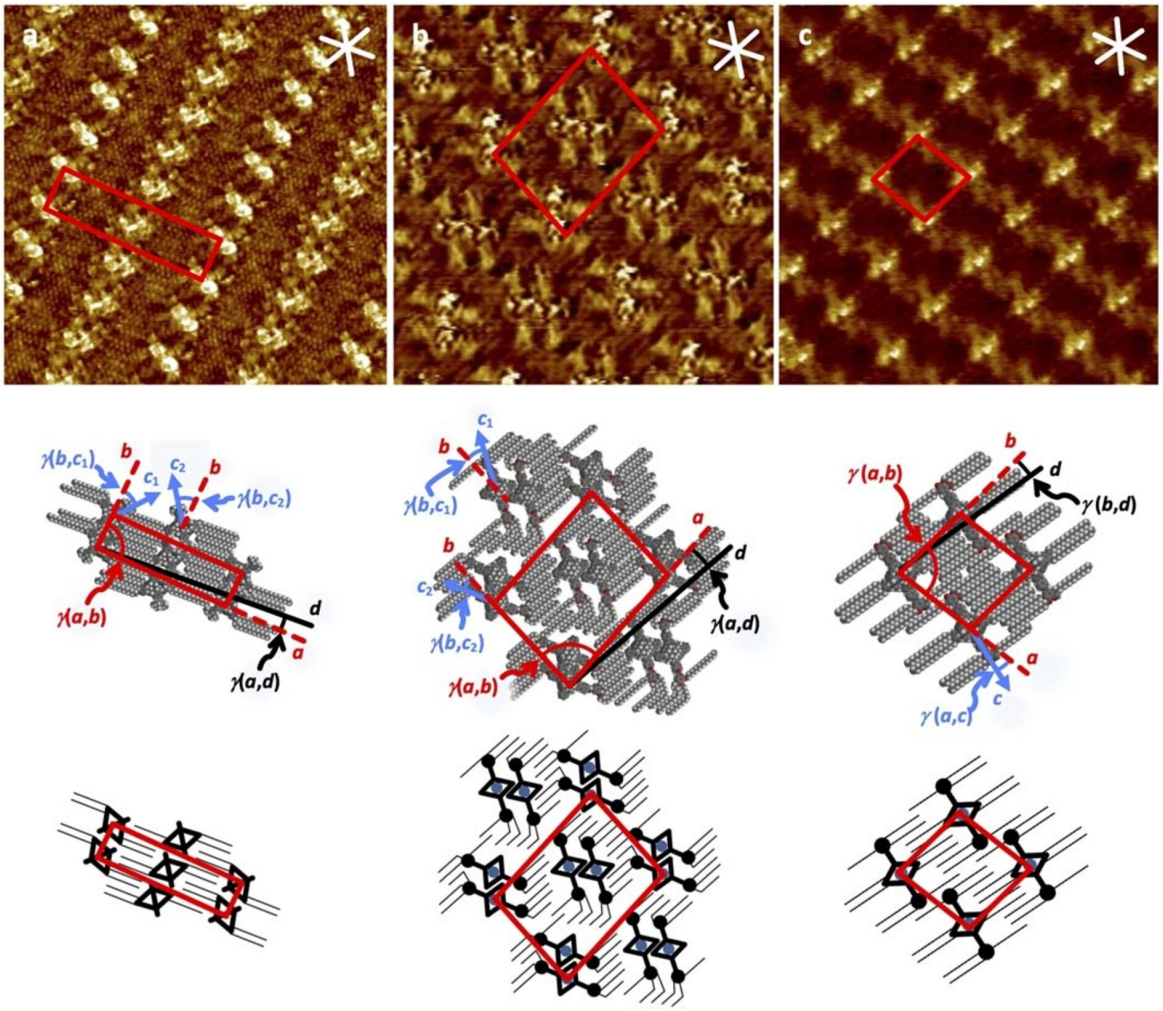


Figure E.5: STM images, proposed supramolecular model, and schematic representation of the SAM of a) **1a**, b) **2b**, and c) **3d** at the solid/liquid interface between HOPG and a solution of the respective compound in PHO. All samples thermally annealed for 20 s at 80 °C. a) $c(\mathbf{1a}) = 10^{-4}$ M, 16.6×16.6 nm^2, $V_S = -0.36$ V, $I_t = 12$ pA; unit cell parameters: $a = (7.5 \pm 0.2)$ nm, $b = (1.8 \pm 0.2)$ nm, $\gamma(a,b) = (90 \pm 2)°$, $\gamma(a,d) = (4 \pm 2)°$, $\gamma(b,c_1) = \gamma(b,c_2) = (37 \pm 5)°$. b) $c(\mathbf{2b}) = 10^{-4}$ M, 20.0×20.0 nm^2, $V_S = 0.8$ V, $I_t = 15$ pA; unit cell parameters: $a = (7.6 \pm 0.2)$ nm, $b = (5.9 \pm 0.2)$ nm, $\gamma(a,b) = (86 \pm 2)°$, $\gamma(a,d) = (7 \pm 2)°$, $\gamma(b,c_1) = \gamma(b,c_2) = (25 \pm 5)°$. c) $c(\mathbf{3d}) = 10^{-5}$ M, 20.0×20.0 nm^2, $V_S = 1.0$ V, $I_t = 11$ pA; unit cell parameters: $a = (3.7 \pm 0.2)$ nm, $b = (3.5 \pm 0.2)$ nm, $\gamma(a,b) = (82 \pm 2)°$, $\gamma(b,d) = (8 \pm 2)°$, $\gamma(a,c) = (20 \pm 5)°$. Red lines indicate the unit cell. White and black lines, respectively, indicate the HOPG main axis directions. Blue arrows indicate the orientation of the molecular backbones.

As such, an adapted architecture **2a-d** was designed, carrying six alkyl chains per molecule *via* two tris(hexadecyloxy)benzyloxy units attached at the "north" and "south" corners of the bicyclophane (instead of the four octadecyl chains at the "west" and "east" positions in **1a** and **b**). With this, the structure **2b**, containing a TIPS-protected pillar unit, adsorbs on the HOPG surface and forms a 2D periodic SAM observable *via* STM (Figure E.5b). Again, aromatic and aliphatic areas are distinguishable by their contrast. The brightest image regions are attributed to the pillar units in the centers of the aromatic backbones, and medium-bright features aligned pairwise are indexed to the "north/south" axes of the backbones. These pairs are alternately oriented along c_1 and c_2 with $\gamma(b, c_2) = \gamma(b, c_2) = (25 \pm 5)°$. To the nanopattern, an oblique unit cell containing a total number of four molecules with parameters $a = (7.6 \pm 0.2)$ nm, $b = (5.9 \pm 0.2)$ nm, $\gamma(a, b) = (86 \pm 2)°$, and a relative orientation to the HOPG main axis direction, d, of $\gamma(a, d) = (7 \pm 2)°$ is indexed. The surface periphery surrounding each backbone pair is covered with the attached alkoxy side chains, following d. More precisely, of each pair of 3,4,5-tris(hexadecyloxy)benzyl units (of both "north" and "south" corners), five chains point into one direction and one chain points into the opposite direction, although this requires a bending of some chains. Intermediate gaps are most probably covered by PHO (solvent) molecules. The pillar-pillar distance between the dimers of **2b** is with (1.6 ± 0.2) nm (Figure S34b) slightly shorter than the molecule centers within the rows of **1a** ((1.8 ± 0.2) nm), showing that steric reason for the absence of a stable 2D structure of **1b** can be excluded. The overall packing, however, is noticeably less dense than that of **1a** as not the complete surface in the unit cell is covered by alkyl chains or backbones. For **2a**, with the same perimeter but without a central pillar unit, the same pattern with an analogous unit cell is formed (see Figure S34a), showing that the adapted design based on 3,4,5-tris(hexadecyloxy)benzyl units is able to accommodate some level of structural change of the interior without changing the overall packing concept.

The molecular structure of **2b** on a graphene sheet was modelled at the GFN2-xTB level (with methoxy groups instead of 3,4,5-tris(hexadecyloxy)benzyloxy substituents, Figure E.6). The bicyclic backbone flattens upon adsorption, thus maximizing its contact with the graphene. The structure shows quite convincingly that the pillar is in an upright conformation pointing with its end to the solution phase and that no tendency of the pillar unit to bend towards the graphene sheet was observed. In the same manner, **2c** with the *p*-tolyl pillar unit adheres to the same surface pattern design as **2a** and **2b** with an alike unit cell (see Figure S34c). Thus, platform molecule **2b** as well as the reference compound **2a** form the same 2D pattern, and also a further transformation at the end substituted of the pillar is possible and does not change the pattern formation behavior, as **2c** demonstrates.

However, with a considerably larger functional unit at the end of the pillar, as in **2d**, the surface pattern changes and leads to a completely different unit cell with an arrangement of the molecules in rows (see Figure S34d). The structure model of **2d** hypothetically organized in the same way as **2a–c** (Figure S35) shows that although the substituents at the pillar are not in close contact to each other there is not enough space for solvation shells. Contrary, the new organization in rows results now in shortest pillar-pillar distances between the molecules of (2.3 ± 0.2) nm (Figure S34d). This shows that six alkoxy chains per backbone

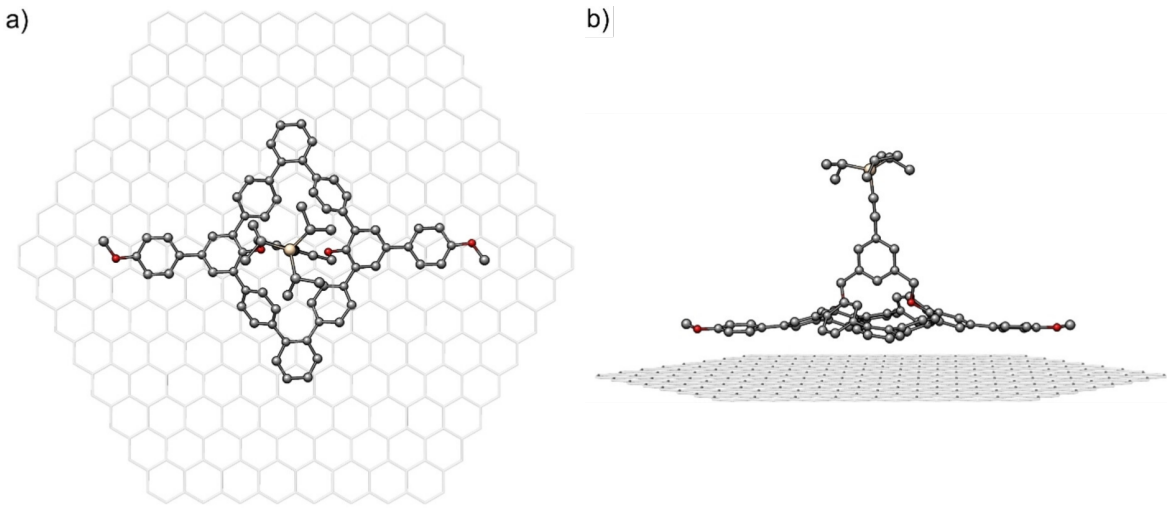


Figure E.6: Molecular structure of **2b** (3,4,5-trialkoxybenzyl substituents replaced by methyl groups) optimized on a graphene sheet at the GFN2-xTB level a) top view, b) side view.

stabilize the pattern such that minor changes at the pillar end are tolerated and keep the 2D pattern unaltered, but larger modifications with more bulky substituents are not allowed.

Based on these results, a third molecule architecture regarding the chain periphery was designed. **3d** contains the alkoxy chains of the tris(hexadecyloxy)benzyloxy units at the "north/south" positions of the bicyclophane as well as alkyl chains at the "east/west" corners of the aromatic backbone ("all four directions"). The goal of this combination of both previous structures was a stabilization of the formed surface pattern by increasing the surface adsorption energy and by keeping the pillars of the molecules so far apart that even the large terpyridyl groups do not sterically interfere with each other. The resulting SAM is shown in Figure E.5c. **3d** contains the same large substituent at the end of the pillar as **2d**, and all compounds (**3a** to **d**) organize similarly and form the same surface patterns with alike unit cells (see Figure S36). All ten alkyl/alkoxy chains per molecule are able to adsorb to the surface. The oblique unit cell contains one molecule, meaning all backbones are oriented in the same direction and are equivalent to one another. This results in a pattern of individual bicyclophanes, organized in rows and separated by their chain periphery in which all chains are aligned along one of the HOPG main axes. **3a–d** are decorated with the largest number of alkyl chains per molecule (out of the three designs presented here) and form a 2D structure with very small unoccupied surface areas, thereby resulting in the most stable pattern with regard to changes of the actual structure of the interior. This packing allows for large pillar-to-pillar distances of (3.7 ± 0.2) nm, enough to allow a functionalization of the pillars with sterically more demanding substituents and keeping the 2D structure constant. A comparison of the three alkyl chain substitution patterns ("east/west", "north/south", and all four directions) shows that a structural motif that is robust against variations at the end of the pillar is obtained when a large number of alkyl chains per molecule positioned at

all four directions of the molecule keep the pillars spatially separated and guarantees a high adsorption tendency.

E.3 Conclusions

We have synthesized a series of bicyclophane-based molecular platforms with orthogonal pillars. The structures were characterized by NMR spectroscopy and mass spectrometry. The NMR data were fully supported by state-of-the-art quantum mechanical calculations. In addition, a conformational analysis shows that in solution the pillars are orthogonal to the bicyclophane backbones. When adsorbed from solution onto HOPG, the compounds form extended 2D crystalline monolayers which were studied *in situ* at the solid/liquid interface by means of STM. The images show the orthogonal pillar orientation, which is in agreement with theoretical calculations of a model compound on a graphene sheet. The distance between the adsorbed pillars is to some extent adjustable by the number of attached alkyl groups adsorbed along the HOPG main axes, and their interdigitation is the driving force for the pattern formation. However, large functional groups at the pillar ends can influence the pattern to such an extent that either no 2D adsorbate or a different crystal structure was observed. With increasing the number of alkyl side chains per molecule at all four sites of the molecule, a stable pattern was created in which the pillars are spatially separated and which is robust enough to withstand different pillar functionalizations.

Supporting Information

The authors have cited additional references within the Supporting Information.[121,137,312,392,442–451]

Acknowledgements

The German Research Foundation (DFG) is gratefully acknowledged for financial support through the SPP 2363 "Utilization and Development of Machine Learning for Molecular Applications – Molecular Machine Learning", the RTG-2591 "TIDE – Template-Designed Organic Electronics" as well as DFG project 265513312. Further, S. G. and M. B. gratefully acknowledge financial support of the Max Planck Society through the Max Planck fellow program. Open Access funding enabled and organized by Projekt DEAL.

Conflict of Interests

The authors declare no conflict of interest.

Data Availability Statement

The data that support the findings of this study are available in the supplementary material of this article.

Bibliography

[1] P. A. M. Dirac. "Quantum Mechanics of Many-Electron Systems." *Proc. R. Soc. Lond. A* **1929**, *123*, 714–733. DOI: 10.1098/rspa.1929.0094.

[2] N. Martín and L. T. Scott. "Challenges in aromaticity: 150 years after Kekulé's benzene." *Chem. Soc. Rev.* **2015**, *44*, 6397–6400. DOI: 10.1039/C5CS90085A.

[3] X.-H. Zhu, N. Zhang, Y. Zhang, X. Zhang, K. Ugurbil, and W. Chen. "*In vivo* ^{17}O NMR approaches for brain study at high field." *NMR Biomed.* **2005**, *18*, 83–103. DOI: 10.1002/nbm.930.

[4] F. Alkan and C. Dybowski. "Calculation of chemical-shift tensors of heavy nuclei: a DFT/ZORA investigation of ^{199}Hg chemical-shift tensors in solids, and the effects of cluster size and electronic-state approximations." *Phys. Chem. Chem. Phys.* **2014**, *16*, 14298–14308. DOI: 10.1039/C4CP01682C.

[5] H. W. Spiess. "*50th Anniversary Perspective*: The Importance of NMR Spectroscopy to Macromolecular Science." *Macromolecules* **2017**, *50*, 1761–1777. DOI: 10.1021/acs.macromol.6b02736.

[6] D. Hahn, A. Masoudi Alavi, V. Hopp, and P. Quirmbach. "Phase development of phosphate-bonded Al_2O_3-$MgAl_2O_4$ high-temperature ceramics: XRD and solid-state NMR investigations." *J. Am. Ceram. Soc.* **2021**, *104*, 6625–6642. DOI: 10.1111/jace.18022.

[7] I. Speciale, A. Notaro, P. Garcia-Vello, F. Di Lorenzo, S. Armiento, A. Molinaro, R. Marchetti, A. Silipo, and C. De Castro. "Liquid-state NMR spectroscopy for complex carbohydrate structural analysis: A hitchhiker's guide." *Carbohydr. Polym.* **2022**, *277*, 118885. DOI: 10.1016/j.carbpol.2021.118885.

[8] S. Krause, F. S. Reuter, S. Ehrling, V. Bon, I. Senkovska, S. Kaskel, and E. Brunner. "Impact of Defects and Crystal Size on Negative Gas Adsorption in DUT-49 Analyzed by *In Situ* ^{129}Xe NMR Spectroscopy." *Chem. Mater.* **2020**, *32*, 4641–4650. DOI: 10.1021/acs.chemmater.0c01059.

[9] M. Bornemann-Pfeiffer, J. Wolf, K. Meyer, S. Kern, D. Angelone, A. Leonov, L. Cronin, and F. Emmerling. "Standardization and Control of Grignard Reactions in a Universal Chemical Synthesis Machine using online NMR." *Angew. Chem. Int. Ed.* **2021**, *60*, 23202–23206. DOI: 10.1002/anie.202106323.

[10] E. Bartalucci, C. Schumacher, L. Hendrickx, F. Puccetti, I. d'Anciães Almeida Silva, R. Dervişoğlu, R. Puttreddy, C. Bolm, and T. Wiegand. "Disentangling the Effect of Pressure and Mixing on a Mechanochemical Bromination Reaction by Solid-State NMR Spectroscopy." *Chem. Eur. J.* **2023**, *29*, e202203466. DOI: 10.1002/chem.202203466.

[11] M. Pellecchia, D. S. Sem, and K. Wüthrich. "NMR in drug discovery." *Nat. Rev. Drug Discov.* **2002**, *1*, 211–219. DOI: 10.1038/nrd748.

[12] H.-Y.-Y. Yao, J.-Q. Wang, J.-Y. Yin, S.-P. Nie, and M.-Y. Xie. "A review of NMR analysis in polysaccharide structure and conformation: Progress, challenge and perspective." *Food Res. Int.* **2021**, *143*, 110290. DOI: 10.1016/j.foodres.2021.110290.

[13] Y. Chen, S. R. Smock, A. H. Flintgruber, F. A. Perras, R. L. Brutchey, and A. J. Rossini. "Surface Termination of $CsPbBr_3$ Perovskite Quantum Dots Determined by Solid-State NMR Spectroscopy." *J. Am. Chem. Soc.* **2020**, *142*, 6117–6127. DOI: 10.1021/jacs.9b13396.

[14] D. J. Kubicki, D. Prochowicz, E. Salager, A. Rakhmatullin, C. P. Grey, L. Emsley, and S. D. Stranks. "Local Structure and Dynamics in Methylammonium, Formamidinium, and Cesium Tin(II) Mixed-Halide Perovskites from ^{119}Sn Solid-State NMR." *J. Am. Chem. Soc.* **2020**, *142*, 7813–7826. DOI: 10.1021/jacs.0c00647.

[15] I. Ojima. "Exploration of Fluorine Chemistry at the Multidisciplinary Interface of Chemistry and Biology." *J. Org. Chem.* **2013**, *78*, 6358–6383. DOI: 10.1021/jo400301u.

[16] G.-N. W. Gomes, M. Krzeminski, A. Namini, E. W. Martin, T. Mittag, T. Head-Gordon, J. D. Forman-Kay, and C. C. Gradinaru. "Conformational Ensembles of an Intrinsically Disordered Protein Consistent with NMR, SAXS, and Single-Molecule FRET." *J. Am. Chem. Soc.* **2020**, *142*, 15697–15710. DOI: 10.1021/jacs.0c02088.

[17] Y. Hu, K. Cheng, L. He, X. Zhang, B. Jiang, L. Jiang, C. Li, G. Wang, Y. Yang, and M. Liu. "NMR-Based Methods for Protein Analysis." *Anal. Chem.* **2021**, *93*, 1866–1879. DOI: 10.1021/acs.analchem.0c03830.

[18] Y. Xiang, G. Zheng, Z. Liang, Y. Jin, X. Liu, S. Chen, K. Zhou, J. Zhu, M. Lin, H. He, J. Wan, S. Yu, G. Zhong, R. Fu, Y. Li, and Y. Yang. "Visualizing the growth process of sodium microstructures in sodium batteries by in-situ ^{23}Na MRI and NMR spectroscopy." *Nat. Nanotechnol.* **2020**, *15*, 883–890. DOI: 10.1038/s41565-020-0749-7.

[19] P. Sormanni, D. Piovesan, G. T. Heller, M. Bonomi, P. Kukic, C. Camilloni, M. Fuxreiter, Z. Dosztanyi, R. V. Pappu, M. M. Babu, S. Longhi, P. Tompa, A. K. Dunker, V. N. Uversky, S. C. E. Tosatto, and M. Vendruscolo. "Simultaneous quantification of protein order and disorder." *Nat. Chem. Biol.* **2017**, *13*, 339–342. DOI: 10.1038/nchembio.2331.

[20] W. Zhao, L. D. Fernando, A. Kirui, F. Deligey, and T. Wang. "Solid-state NMR of plant and fungal cell walls: A critical review." *Solid State Nucl. Magn. Reson.* **2020**, *107*, 101660. DOI: 10.1016/j.ssnmr.2020.101660.

[21] A. Wacker, J. E. Weigand et al. "Secondary structure determination of conserved SARS-CoV-2 RNA elements by NMR spectroscopy." *Nucleic Acids Res.* **2020**, *48*, 12415–12435. DOI: 10.1093/nar/gkaa1013.

[22] M. Bursch, J.-M. Mewes, A. Hansen, and S. Grimme. "Best-Practice DFT Protocols for Basic Molecular Computational Chemistry." *Angew. Chem. Int. Ed.* **2022**, *61*, e202205735. DOI: 10.1002/anie.202205735.

[23] P. Pyykkö. "Perspective on Norman Ramsey's theories of NMR chemical shifts and nuclear spin–spin coupling." *Theor. Chem. Acc.* **2000**, *103*, 214–216. DOI: 10.1007/s002149900011.

[24] M. W. Lodewyk, M. R. Siebert, and D. J. Tantillo. "Computational Prediction of ^{1}H and ^{13}C Chemical Shifts: A Useful Tool for Natural Product, Mechanistic, and Synthetic Organic Chemistry." *Chem. Rev.* **2012**, *112*, 1839–1862. DOI: 10.1021/cr200106v.

[25] P. Pracht, F. Bohle, and S. Grimme. "Automated exploration of the low-energy chemical space with fast quantum chemical methods." *Phys. Chem. Chem. Phys.* **2020**, *22*, 7169–7192. DOI: 10.1039/c9cp06869d.

[26] S. Grimme, C. Bannwarth, and P. Shushkov. "A Robust and Accurate Tight-Binding Quantum Chemical Method for Structures, Vibrational Frequencies, and Noncovalent Interactions of Large Molecular Systems Parametrized for All spd-Block Elements (Z = 1–86)." *J. Chem. Theory Comput.* **2017**, *13*, 1989–2009. DOI: 10.1021/acs.jctc.7b00118.

[27] C. Bannwarth, S. Ehlert, and S. Grimme. "GFN2-xTB – An Accurate and Broadly Parametrized Self-Consistent Tight-Binding Quantum Chemical Method with Multipole Electrostatics and Density-Dependent Dispersion Contributions." *J. Chem. Theory Comput.* **2019**, *15*, 1652–1671. DOI: 10.1021/acs.jctc.8b01176.

[28] S. Spicher and S. Grimme. "Robust Atomistic Modeling of Materials, Organometallic, and Biochemical Systems." *Angew. Chem. Int. Ed.* **2020**, *59*, 15665–15673. DOI: 10.1002/anie.202004239.

[29] S. Grimme, F. Bohle, A. Hansen, P. Pracht, S. Spicher, and M. Stahn. "Efficient Quantum Chemical Calculation of Structure Ensembles and Free Energies for Nonrigid Molecules." *J. Phys. Chem. A* **2021**, *125*, 4039–4054. DOI: 10.1021/acs.jpca.1c00971.

[30] S. Grimme, C. Bannwarth, S. Dohm, A. Hansen, J. Pisarek, P. Pracht, J. Seibert, and F. Neese. "Fully Automated Quantum-Chemistry-Based Computation of Spin–Spin-Coupled Nuclear Magnetic Resonance Spectra." *Angew. Chem. Int. Ed.* **2017**, *56*, 14763–14769. DOI: 10.1002/anie.201708266.

[31] F. Bohle, J. Seibert, and S. Grimme. "Automated Quantum Chemistry-Based Calculation of Optical Rotation for Large Flexible Molecules." *J. Org. Chem.* **2021**, *86*, 15522–15531. DOI: 10.1021/acs.joc.1c02008.

[32] R. Jain, T. Bally, and P. R. Rablen. "Calculating Accurate Proton Chemical Shifts of Organic Molecules with Density Functional Methods and Modest Basis Sets." *J. Org. Chem.* **2009**, *74*, 4017–4023. DOI: 10.1021/jo900482q.

[33] D. Flaig, M. Maurer, M. Hanni, K. Braunger, L. Kick, M. Thubauville, and C. Ochsenfeld. "Benchmarking Hydrogen and Carbon NMR Chemical Shifts at HF, DFT, and MP2 Levels." *J. Chem. Theory Comput.* **2014**, *10*, 572–578. DOI: 10.1021/ct400780f.

[34] M. T. de Oliveira, J. M. A. Alves, A. A. C. Braga, D. J. D. Wilson, and C. A. Barboza. "Do Double-Hybrid Exchange–Correlation Functionals Provide Accurate Chemical Shifts? A Benchmark Assessment for Proton NMR." *J. Chem. Theory Comput.* **2021**, *17*, 6876–6885. DOI: 10.1021/acs.jctc.1c00604.

[35] M. A. Iron. "Evaluation of the Factors Impacting the Accuracy of ^{13}C NMR Chemical Shift Predictions using Density Functional Theory – The Advantage of Long-Range Corrected Functionals." *J. Chem. Theory Comput.* **2017**, *13*, 5798–5819. DOI: 10.1021/acs.jctc.7b00772.

[36] P. Gao, J. Zhang, and H. Chen. "A systematic benchmarking of ^{31}P and ^{19}F NMR chemical shift predictions using different DFT/GIAO methods and applying linear regression to improve the prediction accuracy." *Int. J. Quantum Chem.* **2021**, *121*, e26482. DOI: 10.1002/qua.26482.

[37] S. K. Latypov, F. M. Polyancev, D. G. Yakhvarov, and O. G. Sinyashin. "Quantum chemical calculations of ^{31}P NMR chemical shifts: scopes and limitations." *Phys. Chem. Chem. Phys.* **2015**, *17*, 6976–6987. DOI: 10.1039/C5CP00240K.

[38] L. B. Krivdin. "Recent advances in computational ^{31}P NMR: Part 1. Chemical shifts." *Magn. Reson. Chem.* **2020**, *58*, 478–499. DOI: 10.1002/mrc.4965.

[39] G. L. Stoychev, A. A. Auer, and F. Neese. "Efficient and Accurate Prediction of Nuclear Magnetic Resonance Shielding Tensors with Double-Hybrid Density Functional Theory." *J. Chem. Theory Comput.* **2018**, *14*, 4756–4771. DOI: 10.1021/acs.jctc.8b00624.

[40] D. Sethio, G. Raggi, R. Lindh, and M. Erdélyi. "Halogen Bond of Halonium Ions: Benchmarking DFT Methods for the Description of NMR Chemical Shifts." *J. Chem. Theory Comput.* **2020**, *16*, 7690–7701. DOI: 10.1021/acs.jctc.0c00860.

[41] C. J. Schattenberg and M. Kaupp. "Extended Benchmark Set of Main-Group Nuclear Shielding Constants and NMR Chemical Shifts and Its Use to Evaluate Modern DFT Methods." *J. Chem. Theory Comput.* **2021**, *17*, 7602–7621. DOI: 10.1021/acs.jctc.1c00919.

[42] C. J. Schattenberg, M. Lehmann, M. Bühl, and M. Kaupp. "Systematic Evaluation of Modern Density Functional Methods for the Computation of NMR Shifts of 3d Transition-Metal Nuclei." *J. Chem. Theory Comput.* **2022**, *18*, 273–292. DOI: 10.1021/acs.jctc.1c00964.

[43] J. Gauss. "Analytic second derivatives for the full coupled-cluster singles, doubles, and triples model: Nuclear magnetic shielding constants for BH, HF, CO, N_2, N_2O, and O_3." *J. Chem. Phys.* **2002**, *116*, 4773–4776. DOI: 10.1063/1.1462612.

[44] A. A. Auer, J. Gauss, and J. F. Stanton. "Quantitative prediction of gas-phase ^{13}C nuclear magnetic shielding constants." *J. Chem. Phys.* **2003**, *118*, 10407–10417. DOI: 10.1063/1.1574314.

[45] M. E. Harding, J. Gauss, and P. von Ragué Schleyer. "Why Benchmark-Quality Computations Are Needed To Reproduce 1-Adamantyl Cation NMR Chemical Shifts Accurately." *J. Phys. Chem. A* **2011**, *115*, 2340–2344. DOI: 10.1021/jp1103356.

[46] M. Jaszuński, S. P. A. Sauer, R. Faber, and D. J. D. Wilson. "NMR parameters of FNNF as a test for coupled-cluster methods: CCSDT shielding and CC3 spin–spin coupling." *Phys. Chem. Chem. Phys.* **2020**, *22*, 21350–21359. DOI: 10.1039/d0cp02730h.

[47] J. Autschbach. "Calculation of Heavy-Nucleus Chemical Shifts. Relativistic All-Electron Methods." In *Calculation of NMR and EPR Parameters*. Ed. by M. Kaupp, M. Bühl, and V. G. Malkin. John Wiley & Sons, Ltd., **2004**. Chap. 14, pp. 227–247. DOI: 10.1002/3527601678.ch14.

[48] J. Vícha, J. Novotný, S. Komorovsky, M. Straka, M. Kaupp, and R. Marek. "Relativistic Heavy-Neighbor-Atom Effects on NMR Shifts: Concepts and Trends across the Periodic Table." *Chem. Rev.* **2020**, *120*, 7065–7103. DOI: 10.1021/acs.chemrev.9b00785.

[49] A. Rodriguez-Fortea, P. Alemany, and T. Ziegler. "Density Functional Calculations of NMR Chemical Shifts with the Inclusion of Spin–Orbit Coupling in Tungsten and Lead Compounds." *J. Phys. Chem. A* **1999**, *103*, 8288–8294. DOI: 10.1021/jp9912004.

[50] D. Paschoal, C. Fonseca Guerra, M. A. L. de Oliveira, T. C. Ramalho, and H. F. Dos Santos. "Predicting Pt-195 NMR Chemical Shift Using New Relativistic All-Electron Basis Set." *J. Comput. Chem.* **2016**, *37*, 2360–2373. DOI: 10.1002/jcc.24461.

[51] J. Autschbach and M. Sterzel. "Molecular Dynamics Computational Study of the ^{199}Hg–^{199}Hg NMR Spin–Spin Coupling Constants of $[Hg–Hg–Hg]^{2+}$ in SO_2 Solution." *J. Am. Chem. Soc.* **2007**, *129*, 11093–11099. DOI: 10.1021/ja073166+.

[52] K. Jakubowska and M. Pecul. "Nuclear magnetic resonance parameters in Zn_2, Cd_2 and Hg_2 dimers: relativistic calculations." *Theor. Chem. Acc.* **2021**, *140*, 26. DOI: 10.1007/s00214-021-02720-5.

[53] T. B. Demissie. "Relativistic effects on the NMR parameters of Si, Ge, Sn, and Pb alkynyl compounds: Scalar versus spin-orbit effects." *J. Chem. Phys.* **2017**, *147*, 174301. DOI: 10.1063/1.4996712.

[54] Y. Y. Rusakov, I. L. Rusakova, and L. B. Krivdin. "On the HALA effect in the NMR carbon shielding constants of the compounds containing heavy p-elements." *Int. J. Quantum Chem.* **2016**, *116*, 1404–1412. DOI: 10.1002/qua.25196.

[55] W. S. McCulloch and W. Pitts. "A logical calculus of the ideas immanent in nervous activity." *Bull. Math. Biophys.* **1943**, *5*, 115–133. DOI: 10.1007/BF02478259.

[56] J. Gasteiger and J. Zupan. "Neural Networks in Chemistry." *Angew. Chem. Int. Ed.* **1993**, *32*, 503–527. DOI: 10.1002/anie.199305031.

[57] N. Artrith, K. T. Butler, F.-X. Coudert, S. Han, O. Isayev, A. Jain, and A. Walsh. "Best practices in machine learning for chemistry." *Nat. Chem.* **2021**, *13*, 505–508. DOI: 10.1038/s41557-021-00716-z.

[58] K. T. Butler, D. W. Davies, H. Cartwright, O. Isayev, and A. Walsh. "Machine learning for molecular and materials science." *Nature* **2018**, *559*, 547–555. DOI: 10.1038/s41586-018-0337-2.

[59] F. Strieth-Kalthoff, F. Sandfort, M. H. S. Segler, and F. Glorius. "Machine learning the ropes: principles, applications and directions in synthetic chemistry." *Chem. Soc. Rev.* **2020**, *49*, 6154–6168. DOI: 10.1039/c9cs00786e.

[60] Web of Science: Search for *machine learning* and *chemistry* as topics. https://www.webofscience.com (accessed: January 22, 2024).

[61] K. Yang, K. Swanson, W. Jin, C. Coley, P. Eiden, H. Gao, A. Guzman-Perez, T. Hopper, B. Kelley, M. Mathea, A. Palmer, V. Settels, T. Jaakkola, K. Jensen, and R. Barzilay. "Analyzing Learned Molecular Representations for Property Prediction." *J. Chem. Inf. Model.* **2019**, *59*, 3370–3388. DOI: 10.1021/acs.jcim.9b00237.

[62] K. Hansen, G. Montavon, F. Biegler, S. Fazli, M. Rupp, M. Scheffler, O. A. von Lilienfeld, A. Tkatchenko, and K.-R. Müller. "Assessment and Validation of Machine Learning Methods for Predicting Molecular Atomization Energies." *J. Chem. Theory Comput.* **2013**, *9*, 3404–3419. DOI: 10.1021/ct400195d.

[63] D. Folmsbee and G. Hutchison. "Assessing conformer energies using electronic structure and machine learning methods." *Int. J. Quantum Chem.* **2021**, *121*, e26381. DOI: 10.1002/qua.26381.

[64] J. Timoshenko, D. Lu, Y. Lin, and A. I. Frenkel. "Supervised Machine-Learning-Based Determination of Three-Dimensional Structure of Metallic Nanoparticles." *J. Phys. Chem. Lett.* **2017**, *8*, 5091–5098. DOI: 10.1021/acs.jpclett.7b02364.

[65] K. Kaufmann, C. Zhu, A. S. Rosengarten, D. Maryanovsky, T. J. Harrington, E. Marin, and K. S. Vecchio. "Crystal symmetry determination in electron diffraction using machine learning." *Science* **2020**, *367*, 564–568. DOI: 10.1126/science.aay3062.

[66] R. Ramakrishnan, M. Hartmann, E. Tapavicza, and O. A. von Lilienfeld. "Electronic spectra from TDDFT and machine learning in chemical space." *J. Chem. Phys.* **2015**, *143*, 084111. DOI: 10.1063/1.4928757.

[67] M. Gastegger, J. Behler, and P. Marquetand. "Machine learning molecular dynamics for the simulation of infrared spectra." *Chem. Sci.* **2017**, *8*, 6924–6935. DOI: 10.1039/C7SC02267K.

[68] E. Putin, A. Asadulaev, Y. Ivanenkov, V. Aladinskiy, B. Sanchez-Lengeling, A. Aspuru-Guzik, and A. Zhavoronkov. "Reinforced Adversarial Neural Computer for *de Novo* Molecular Design." *J. Chem. Inf. Model.* **2018**, *58*, 1194–1204. DOI: 10.1021/acs.jcim.7b00690.

[69] V. Bagal, R. Aggarwal, P. K. Vinod, and U. D. Priyakumar. "MolGPT: Molecular Generation Using a Transformer-Decoder Model." *J. Chem. Inf. Model.* **2022**, *62*, 2064–2076. DOI: 10.1021/acs.jcim.1c00600.

[70] A. Chandrasekaran, D. Kamal, R. Batra, C. Kim, L. Chen, and R. Ramprasad. "Solving the electronic structure problem with machine learning." *Npj Comput. Mater.* **2019**, *5*, 22. DOI: 10.1038/s41524-019-0162-7.

[71] J. Kirkpatrick, B. McMorrow, D. H. P. Turban, A. L. Gaunt, J. S. Spencer, A. G. D. G. Matthews, A. Obika, L. Thiry, M. Fortunato, D. Pfau, L. Román Castellanos, S. Petersen, A. W. R. Nelson, P. Kohli, P. Mori-Sánchez, D. Hassabis, and A. J. Cohen. "Pushing the frontiers of density functionals by solving the fractional electron problem." *Science* **2021**, *374*, 1385–1389. DOI: 10.1126/science.abj6511.

[72] D. M. Anstine, R. Zubatyuk, and O. Isayev. "AIMNet2: A Neural Network Potential to Meet your Neutral, Charged, Organic, and Elemental-Organic Needs." *ChemRxiv preprint* **2023**, chemrxiv-2023–296ch. DOI: 10.26434/chemrxiv-2023-296ch.

[73] E. Kocer, T. W. Ko, and J. Behler. "Neural Network Potentials: A Concise Overview of Methods." *Annu. Rev. Phys. Chem.* **2022**, *73*, 163–186. DOI: 10.1146/annurev-physchem-082720-034254.

[74] J. Daru, H. Forbert, J. Behler, and D. Marx. "Coupled Cluster Molecular Dynamics of Condensed Phase Systems Enabled by Machine Learning Potentials: Liquid Water Benchmark." *Phys. Rev. Lett.* **2022**, *129*, 226001. DOI: 10.1103/PhysRevLett.129.226001.

[75] F. M. Paruzzo, A. Hofstetter, F. Musil, S. De, M. Ceriotti, and L. Emsley. "Chemical shifts in molecular solids by machine learning." *Nat. Commun.* **2018**, *9*, 4501. DOI: 10.1038/s41467-018-06972-x.

[76] S. Liu, J. Li, K. C. Bennett, B. Ganoe, T. Stauch, M. Head-Gordon, A. Hexemer, D. Ushizima, and T. Head-Gordon. "Multiresolution 3D-DenseNet for Chemical Shift Prediction in NMR Crystallography." *J. Phys. Chem. Lett.* **2019**, *10*, 4558–4565. DOI: 10.1021/acs.jpclett.9b01570.

[77] Z. Yang, M. Chakraborty, and A. D. White. "Predicting chemical shifts with graph neural networks." *Chem. Sci.* **2021**, *12*, 10802–10809. DOI: 10.1039/d1sc01895g.

[78] Y. Guan, S. V. Shree Sowndarya, L. C. Gallegos, P. C. St. John, and R. S. Paton. "Real-time prediction of ^{1}H and ^{13}C chemical shifts with DFT accuracy using a 3D graph neural network." *Chem. Sci.* **2021**, *12*, 12012–12026. DOI: 10.1039/d1sc03343c.

[79] W. Gerrard, L. A. Bratholm, M. J. Packer, A. J. Mulholland, D. R. Glowacki, and C. P. Butts. "IMPRESSION – prediction of NMR parameters for 3-dimensional chemical structures using machine learning with near quantum chemical accuracy." *Chem. Sci.* **2020**, *11*, 508–515. DOI: 10.1039/c9sc03854j.

[80] W. Gerrard, C. Yiu, and C. P. Butts. "Prediction of ^{15}N chemical shifts by machine learning." *Magn. Reson. Chem.* **2022**, *60*, 1087–1092. DOI: 10.1002/mrc.5208.

[81] S. K. Chandy and K. Raghavachari. "MIM-ML: A Novel Quantum Chemical Fragment-Based Random Forest Model for Accurate Prediction of NMR Chemical Shifts of Nucleic Acids." *J. Chem. Theory Comput.* **2023**, *19*, 6632–6642. DOI: 10.1021/acs.jctc.3c00563.

[82] M. C. Venetos, M. Wen, and K. A. Persson. "Machine Learning Full NMR Chemical Shift Tensors of Silicon Oxides with Equivariant Graph Neural Networks." *J. Phys. Chem. A* **2023**, *127*, 2388–2398. DOI: 10.1021/acs.jpca.2c07530.

[83] B. Meyer, T. Hansen, D. Nute, P. Albersheim, A. Darvill, W. York, and J. Sellers. "Identification of the ^{1}H-NMR Spectra of Complex Oligosaccharides with Artificial Neural Networks." *Science* **1991**, *251*, 542–544. DOI: 10.1126/science.1990429.

[84] A. M. Sarotti. "Successful combination of computationally inexpensive GIAO ^{13}C NMR calculations and artificial neural network pattern recognition: a new strategy for simple and rapid detection of structural misassignments." *Org. Biomol. Chem.* **2013**, *11*, 4847–4859. DOI: 10.1039/c3ob40843d.

[85] A. Howarth, K. Ermanis, and J. M. Goodman. "DP4-AI automated NMR data analysis: straight from spectrometer to structure." *Chem. Sci.* **2020**, *11*, 4351–4359. DOI: 10.1039/D0SC00442A.

[86] I. M. Novitskiy and A. G. Kutateladze. "DU8ML: Machine Learning-Augmented Density Functional Theory Nuclear Magnetic Resonance Computations for High-Throughput In Silico Solution Structure Validation and Revision of Complex Alkaloids." *J. Org. Chem.* **2022**, *87*, 4818–4828. DOI: 10.1021/acs.joc.2c00169.

[87] M. M. Zanardi and A. M. Sarotti. "GIAO C–H COSY Simulations Merged with Artificial Neural Networks Pattern Recognition Analysis. Pushing the Structural Validation a Step Forward." *J. Org. Chem.* **2015**, *80*, 9371–9378. DOI: 10.1021/acs.joc.5b01663.

[88] P. A. Unzueta, C. S. Greenwell, and G. J. O. Beran. "Predicting Density Functional Theory-Quality Nuclear Magnetic Resonance Chemical Shifts via Δ-Machine Learning." *J. Chem. Theory Comput.* **2021**, *17*, 826–840. DOI: 10.1021/acs.jctc.0c00979.

[89] J. Li, J. Liang, Z. Wang, A. L. Ptaszek, X. Liu, B. Ganoe, M. Head-Gordon, and T. Head-Gordon. "Highly Accurate Prediction of NMR Chemical Shifts from Low-Level Quantum Mechanics Calculations Using Machine Learning." *arXiv preprint* **2023**, arXiv:2306.08269. DOI: 10.48550/arXiv.2306.08269.

[90] P. Gao, J. Zhang, Q. Peng, J. Zhang, and V.-A. Glezakou. "General Protocol for the Accurate Prediction of Molecular ^{13}C/^{1}H NMR Chemical Shifts via Machine Learning Augmented DFT." *J. Chem. Inf. Model.* **2020**, *60*, 3746–3754. DOI: 10.1021/acs.jcim.0c00388.

[91] A. Szabo and N. S. Ostlund. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*. Dover Publications, Inc., **1996**. ISBN: 987-0-486-69186-2.

[92] F. Jensen. *Introduction to Computational Chemistry. Third Edition*. John Wiley & Sons, Ltd., **2017**. ISBN: 978-1-118-82599-0.

[93] E. Schrödinger. "An Undulatory Theory of the Mechanics of Atoms and Molecules." *Phys. Rev.* **1926**, *28*, 1049–1070. DOI: 10.1103/PhysRev.28.1049.

[94] R. L. Workman, V. D. Burkert, V. Crede, E. Klempt, U. Thoma, L. Tiator et al. (Particle Data Group). "Review of Particle Physics." *Prog. Theor. Exp. Phys.* **2022**, *2022*, 083C01. DOI: 10.1093/ptep/ptac097.

[95] M. Born and R. Oppenheimer. "Zur Quantentheorie der Molekeln." *Ann. Phys.* **1927**, *389*, 457–484. DOI: 10.1002/andp.19273892002.

[96] J. C. Slater. "Atomic Shielding Constants." *Phys. Rev.* **1930**, *36*, 57–64. DOI: 10.1103/PhysRev.36.57.

[97] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon. "A fifth-order perturbation comparison of electron correlation theories." *Chem. Phys. Lett.* **1989**, *157*, 479–483. DOI: 10.1016/S0009-2614(89)87395-6.

[98] R. J. Bartlett and M. Musiał. "Coupled-cluster theory in quantum chemistry." *Rev. Mod. Phys.* **2007**, *79*, 291–352. DOI: 10.1103/RevModPhys.79.291.

[99] P. Hohenberg and W. Kohn. "Inhomogeneous Electron Gas." *Phys. Rev.* **1964**, *136*, B864–B871. DOI: 10.1103/PhysRev.136.B864.

[100] J. P. Perdew and K. Schmidt. "Jacob's Ladder of Density Functional Approximations for the Exchange-Correlation Energy." *AIP Conf. Proc.* **2001**, *577*, 1–20. DOI: 10.1063/1.1390175.

[101] J. P. Perdew, K. Burke, and M. Ernzerhof. "Generalized Gradient Approximation Made Simple." *Phys. Rev. Lett.* **1996**, *77*, 3865–3868. DOI: 10.1103/PhysRevLett.77.3865.

[102] A. D. Becke. "Density-functional exchange-energy approximation with correct asymptotic behavior." *Phys. Rev. A* **1988**, *38*, 3098–3100. DOI: 10.1103/PhysRevA.38.3098.

[103] J. P. Perdew. "Density-functional approximation for the correlation energy of the inhomogeneous electron gas." *Phys. Rev. B* **1986**, *33*, 8822–8824. DOI: 10.1103/PhysRevB.33.8822.

[104] J. Tao, J. P. Perdew, V. N. Staroverov, and G. E. Scuseria. "Climbing the Density Functional Ladder: Nonempirical Meta–Generalized Gradient Approximation Designed for Molecules and Solids." *Phys. Rev. Lett.* **2003**, *91*, 146401. DOI: 10.1103/PhysRevLett.91.146401.

[105] J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, and J. Sun. "Accurate and Numerically Efficient r^{2}SCAN Meta-Generalized Gradient Approximation." *J. Phys. Chem. Lett.* **2020**, *11*, 8208–8215. DOI: 10.1021/acs.jpclett.0c02405.

[106] C. Lee, W. Yang, and R. G. Parr. "Development of the Colle-Salvetti correlation-energy formula into a functional of the electron density." *Phys. Rev. B* **1988**, *37*, 785–789. DOI: 10.1103/PhysRevB.37.785.

[107] A. D. Becke. "Density-functional thermochemistry. III. The role of exact exchange." *J. Chem. Phys.* **1993**, *98*, 5648–5652. DOI: 10.1063/1.464913.

[108] P. J. Stephens, F. J. Devlin, C. F. Chabalowski, and M. J. Frisch. "*Ab Initio* Calculation of Vibrational Absorption and Circular Dichroism Spectra Using Density Functional Force Fields." *J. Phys. Chem.* **1994**, *98*, 11623–11627. DOI: 10.1021/j100096a001.

[109] C. Adamo and V. Barone. "Toward reliable density functional methods without adjustable parameters: The PBE0 model." *J. Chem. Phys.* **1999**, *110*, 6158–6170. DOI: 10.1063/1.478522.

[110] S. Grimme. "Semiempirical hybrid density functional with perturbative second-order correlation." *J. Chem. Phys.* **2006**, *124*, 034108. DOI: 10.1063/1.2148954.

[111] L. Goerigk and S. Grimme. "Efficient and Accurate Double-Hybrid-Meta-GGA Density Functionals–Evaluation with the Extended GMTKN30 Database for General Main Group Thermochemistry, Kinetics, and Noncovalent Interactions." *J. Chem. Theory Comput.* **2011**, *7*, 291–309. DOI: 10.1021/ct100466k.

[112] V. Barone and M. Cossi. "Quantum Calculation of Molecular Energies and Energy Gradients in Solution by a Conductor Solvent Model." *J. Phys. Chem. A* **1998**, *102*, 1995–2001. DOI: 10.1021/jp9716997.

[113] M. Cossi, N. Rega, G. Scalmani, and V. Barone. "Energies, Structures, and Electronic Properties of Molecules in Solution with the C-PCM Solvation Model." *J. Comput. Chem.* **2003**, *24*, 669–681. DOI: 10.1002/jcc.10189.

[114] A. Klamt and G. Schüürmann. "COSMO: A New Approach to Dielectric Screening in Solvents with Explicit Expressions for the Screening Energy and its Gradient." *J. Chem. Soc., Perkin Trans. 2* **1993**, 799–805. DOI: 10.1039/P29930000799.

[115] A. Klamt. "Conductor-like Screening Model for Real Solvents: A New Approach to the Quantitative Calculation of Solvation Phenomena." *J. Phys. Chem.* **1995**, *99*, 2224–2235. DOI: 10.1021/j100007a062.

[116] A. Klamt, V. Jonas, T. Bürger, and J. C. W. Lohrenz. "Refinement and Parametrization of COSMO-RS." *J. Phys. Chem. A* **1998**, *102*, 5074–5085. DOI: 10.1021/jp980017s.

[117] G. Sigalov, A. Fenley, and A. Onufriev. "Analytical electrostatics for biomolecules: Beyond the generalized Born approximation." *J. Chem. Phys.* **2006**, *124*, 124902. DOI: 10.1063/1.2177251.

[118] S. Ehlert, M. Stahn, S. Spicher, and S. Grimme. "Robust and Efficient Implicit Solvation Model for Fast Semiempirical Methods." *J. Chem. Theory Comput.* **2021**, *17*, 4250–4261. DOI: 10.1021/acs.jctc.1c00471.

[119] S. Grimme. "Supramolecular Binding Thermodynamics by Dispersion-Corrected Density Functional Theory." *Chem. Eur. J.* **2012**, *18*, 9955–9964. DOI: 10.1002/chem.201200497.

[120] S. Spicher and S. Grimme. "Efficient Computation of Free Energy Contributions for Association Reactions of Large Molecules." *J. Phys. Chem. Lett.* **2020**, *11*, 6606–6611. DOI: 10.1021/acs.jpclett.0c01930.

[121] S. Grimme. "Exploration of Chemical Compound, Conformer, and Reaction Space with Meta-Dynamics Simulations Based on Tight-Binding Quantum Chemical Calculations." *J. Chem. Theory Comput.* **2019**, *15*, 2847–2862. DOI: 10.1021/acs.jctc.9b00143.

[122] F. Calvo, E. Pahl, M. Wormit, and P. Schwerdtfeger. "Evidence for Low-Temperature Melting of Mercury owing to Relativity." *Angew. Chem. Int. Ed.* **2013**, *52*, 7583–7585. DOI: 10.1002/anie.201302742.

[123] S. Löffelsender, P. Schwerdtfeger, S. Grimme, and J.-M. Mewes. "It's Complicated: On Relativistic Effects and Periodic Trends in the Melting and Boiling Points of the Group 11 Coinage Metals." *J. Am. Chem. Soc.* **2022**, *144*, 485–494. DOI: 10.1021/jacs.1c10881.

[124] P. Pyykkö and J. P. Desclaux. "Relativity and the Periodic System of Elements." *Acc. Chem. Res.* **1979**, *12*, 276–281. DOI: 10.1021/ar50140a002.

[125] P. Schwerdtfeger. "Relativistic Effects in Gold Chemistry. 2. The Stability of Complex Halides of Gold(III)." *J. Am. Chem. Soc.* **1989**, *111*, 7261–7262. DOI: 10.1021/ja00200a056.

[126] B. A. Hess. "Relativistic electronic-structure calculations employing a two-component no-pair formalism with external-field projection operators." *Phys. Rev. A* **1986**, *33*, 3742–3748. DOI: 10.1103/PhysRevA.33.3742.

[127] G. Jansen and B. A. Hess. "Revision of the Douglas-Kroll transformation." *Phys. Rev. A* **1989**, *39*, 6016–6017. DOI: 10.1103/PhysRevA.39.6016.

[128] E. van Lenthe, E. J. Baerends, and J. G. Snijders. "Relativistic regular two-component Hamiltonians." *J. Chem. Phys.* **1993**, *99*, 4597–4610. DOI: 10.1063/1.466059.

[129] E. van Lenthe, J. G. Snijders, and E. J. Baerends. "The zero-order regular approximation for relativistic effects: The effect of spin–orbit coupling in closed shell molecules." *J. Chem. Phys.* **1996**, *105*, 6505–6516. DOI: 10.1063/1.472460.

[130] W. Kutzelnigg and W. Liu. "Quasirelativistic theory equivalent to fully relativistic theory." *J. Chem. Phys.* **2005**, *123*, 241102. DOI: 10.1063/1.2137315.

[131] M. Iliaš and T. Saue. "An infinite-order two-component relativistic Hamiltonian by a simple one-step transformation." *J. Chem. Phys.* **2007**, *126*, 064102. DOI: 10.1063/1.2436882.

[132] D. Peng, W. Liu, Y. Xiao, and L. Cheng.
"Making four- and two-component relativistic density functional methods fully equivalent based on the idea of "from atoms to molecule"."
J. Chem. Phys. **2007**, *127*, 104106. DOI: 10.1063/1.2772856.

[133] G. Katti, S. A. Ara, and A. Shireen.
"Magnetic Resonance Imaging (MRI) – A Review." *Int. J. Dent. Clin.* **2011**, *3*, 65–70.

[134] M. Hesse, H. Meier, and B. Zeeh.
Spektroskopische Methoden in der organischen Chemie. 7., überarbeitete Auflage.
Georg Thieme Verlag, **2005**. ISBN: 978-3-135-76107-7.

[135] M. Kaupp, M. Bühl, and V. G. Malkin. *Calculation of NMR and EPR Parameters*.
John Wiley & Sons, Ltd., **2004**. ISBN: 987-3-527-30779-6. DOI: 10.1002/3527601678.

[136] F. Jensen.
"Segmented Contracted Basis Sets Optimized for Nuclear Magnetic Shielding."
J. Chem. Theory Comput. **2015**, *11*, 132–138. DOI: 10.1021/ct5009526.

[137] R. Ditchfield. "Self-consistent perturbation theory of diamagnetism I. A gauge-invariant LCAO method for N.M.R. chemical shifts."
Mol. Phys. **1974**, *27*, 789–807. DOI: 10.1080/00268977400100711.

[138] K. Wolinski, J. F. Hinton, and P. Pulay. "Efficient Implementation of the Gauge-Independent Atomic Orbital Method for NMR Chemical Shift Calculations."
J. Am. Chem. Soc. **1990**, *112*, 8251–8260. DOI: 10.1021/ja00179a005.

[139] G. Schreckenbach and T. Ziegler. "Calculation of NMR Shielding Tensors Using Gauge-Including Atomic Orbitals and Modern Density Functional Theory."
J. Phys. Chem. **1995**, *99*, 606–611. DOI: 10.1021/j100002a024.

[140] W. Kutzelnigg. "Theory of Magnetic Susceptibilities and NMR Chemical Shifts in Terms of Localized Quantities." *Isr. J. Chem.* **1980**, *19*, 193–200.
DOI: 10.1002/ijch.198000020.

[141] M. Schindler and W. Kutzelnigg.
"Theory of magnetic susceptibilities and NMR chemical shifts in terms of localized quantities. II. Application to some simple molecules."
J. Chem. Phys. **1982**, *76*, 1919–1933. DOI: 10.1063/1.443165.

[142] C. van Wüllen and W. Kutzelnigg.
"Calculation of nuclear magnetic resonance shieldings and magnetic susceptibilities using multiconfiguration Hartree–Fock wave functions and local gauge origins."
J. Chem. Phys. **1996**, *104*, 2330–2340. DOI: 10.1063/1.470928.

[143] E. Bartalucci, A. A. Malär, A. Mehnert, J. B. Kleine Büning, L. Günzel, M. Icker, M. Börner, C. Wiebeler, B. H. Meier, S. Grimme, B. Kersting, and T. Wiegand.
"Probing a Hydrogen-π Interaction Involving a Trapped Water Molecule in the Solid State." *Angew. Chem. Int. Ed.* **2023**, *62*, e202217725. DOI: 10.1002/anie.202217725.

[144] J. Vícha, P. Švec, Z. Růžičková, M. A. Samsonov, K. Bártová, A. Růžička, M. Straka, and M. Dračínský.
"Experimental and Theoretical Evidence of Spin-Orbit Heavy Atom on the Light Atom ^{1}H NMR Chemical Shifts Induced through H $\cdots$ I^{-} Hydrogen Bond."
Chem. Eur. J. **2020**, *26*, 8698–8702. DOI: 10.1002/chem.202001532.

[145] A. Géron.
Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow. 2nd Edition.
O'Reilly Media, Inc., **2019**. ISBN: 978-1-492-03264-9.

[146] Q. Zhang, J. Lu, and Y. Jin. "Artificial intelligence in recommender systems."
Complex Intell. Syst. **2021**, *7*, 439–457. DOI: 10.1007/s40747-020-00212-w.

[147] O. I. Abiodun, A. Jantan, A. E. Omolara, K. V. Dada, A. M. Umar, O. U. Linus, H. Arshad, A. A. Kazaure, U. Gana, and M. U. Kiru. "Comprehensive Review of Artificial Neural Network Applications to Pattern Recognition."
IEEE Access **2019**, *7*, 158820–158846. DOI: 10.1109/ACCESS.2019.2945545.

[148] G. Cooper. "Examining Science Education in ChatGPT: An Exploratory Study of Generative Artificial Intelligence." *J. Sci. Educ. Technol.* **2023**, *32*, 444–452.
DOI: 10.1007/s10956-023-10039-y.

[149] E. Adamopoulou and L. Moussiades. "An Overview of Chatbot Technology."
In *Artificial Intelligence Applications and Innovations*.
Ed. by I. Maglogiannis, L. Iliadis, and E. Pimenidis.
Springer Nature Switzerland AG, **2020**. pp. 373–383.
DOI: 10.1007/978-3-030-49186-4_31.

[150] E. Heid, C. J. McGill, F. H. Vermeire, and W. H. Green.
"Characterizing Uncertainty in Machine Learning for Chemistry."
J. Chem. Inf. Model. **2023**, *63*, 4012–4029. DOI: 10.1021/acs.jcim.3c00373.

[151] J. A. Tossell and P. Lazzeretti. "*Ab initio* calculations of ^{29}Si NMR chemical shifts for some gas phase and solid state silicon fluorides and oxides."
J. Chem. Phys. **1986**, *84*, 369–374. DOI: 10.1063/1.450146.

[152] J. Casanovas, F. Illas, and G. Pacchioni. "Ab initio calculations of ^{29}Si solid state NMR chemical shifts of silane and silanol groups in silica."
Chem. Phys. Lett. **2000**, *326*, 523–529. DOI: 10.1016/S0009-2614(00)00818-6.

[153] A. I. Poblador-Bahamonde, R. Poteau, C. Raynaud, and O. Eisenstein.
"DFT calculations of ^{29}Si-NMR chemical shifts in Ru(II) silyl complexes: Searching for trends and accurate values." *Dalton Trans.* **2011**, *40*, 11321–11326.
DOI: 10.1039/c1dt11135c.

[154] J. Autschbach, K. Sutter, L. A. Truflandier, E. Brendler, and J. Wagler.
"Atomic Contributions from Spin-Orbit Coupling to ^{29}Si NMR Chemical Shifts in Metallasilatrane Complexes." *Chem. Eur. J.* **2012**, *18*, 12803–12813.
DOI: 10.1002/chem.201200746.

[155] S. Grimme, J. G. Brandenburg, C. Bannwarth, and A. Hansen. "Consistent structures and interactions by density functional theory with small atomic orbital basis sets." *J. Chem. Phys.* **2015**, *143*, 054107. DOI: 10.1063/1.4927476.

[156] F. Weigend and R. Ahlrichs. "Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy." *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305. DOI: 10.1039/b508541a.

[157] S. Grimme. "Semiempirical GGA-Type Density Functional Constructed with a Long-Range Dispersion Correction." *J. Comput. Chem.* **2006**, *27*, 1787–1799. DOI: 10.1002/jcc.20495.

[158] V. N. Staroverov, G. E. Scuseria, J. Tao, and J. P. Perdew. "Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes." *J. Chem. Phys.* **2003**, *119*, 12129–12137. DOI: 10.1063/1.1626543.

[159] V. N. Staroverov, G. E. Scuseria, J. Tao, and J. P. Perdew. "Erratum: "Comparative assessment of a new nonempirical density functional: Molecules and hydrogen-bonded complexes" [J. Chem. Phys. 119, 12129 (2003)]." *J. Chem. Phys.* **2004**, *121*, 11507. DOI: 10.1063/1.1795692.

[160] T. W. Keal and D. J. Tozer. "The exchange-correlation potential in Kohn–Sham nuclear magnetic resonance shielding calculations." *J. Chem. Phys.* **2003**, *119*, 3015–3024. DOI: 10.1063/1.1590634.

[161] Y. Zhao and D. G. Truhlar. "The M06 suite of density functionals for main group thermochemistry, thermochemical kinetics, noncovalent interactions, excited states, and transition elements: two new functionals and systematic testing of four M06-class functionals and 12 other functionals." *Theor. Chem. Acc.* **2008**, *120*, 215–241. DOI: 10.1007/s00214-007-0310-x.

[162] F. Jensen. "Polarization consistent basis sets: Principles." *J. Chem. Phys.* **2001**, *115*, 9113–9125. DOI: 10.1063/1.1413524.

[163] T. H. Dunning. "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen." *J. Chem. Phys.* **1989**, *90*, 1007–1023. DOI: 10.1063/1.456153.

[164] K. A. Peterson. "Systematically convergent basis sets with relativistic pseudopotentials. I. Correlation consistent basis sets for the post-*d* group 13–15 elements." *J. Chem. Phys.* **2003**, *119*, 11099–11112. DOI: 10.1063/1.1622923.

[165] E. Van Lenthe and E. J. Baerends. "Optimized Slater-Type Basis Sets for the Elements 1–118." *J. Comput. Chem.* **2003**, *24*, 1142–1156. DOI: 10.1002/jcc.10255.

[166] J. Holeček, M. Nádvorník, K. Handlíř, and A. Lyčka.
"^{13}C and ^{119}Sn NMR spectra of Di-n-butyltin(IV) compounds."
J. Organomet. Chem. **1986**, *315*, 299–308. DOI: 10.1016/0022-328X(86)80450-8.

[167] L. Pellerito and L. Nagy. "Organotin(IV)$^{n+}$ complexes formed with biologically active ligands: equilibrium and structural studies, and some biological aspects."
Coord. Chem. Rev. **2002**, *224*, 111–150. DOI: 10.1016/S0010-8545(01)00399-X.

[168] L. Ronconi and P. J. Sadler. "Applications of heteronuclear NMR spectroscopy in biological and medicinal inorganic chemistry."
Coord. Chem. Rev. **2008**, *252*, 2239–2277. DOI: 10.1016/j.ccr.2008.01.016.

[169] A. Bagno, G. Casella, and G. Saielli. "Relativistic DFT Calculation of ^{119}Sn Chemical Shifts and Coupling Constants in Tin Compounds."
J. Chem. Theory Comput. **2006**, *2*, 37–46. DOI: 10.1021/ct050173k.

[170] C. Zeppek, J. Pichler, A. Torvisco, M. Flock, and F. Uhlig. "Aryltin chlorides and hydrides: Preparation, detailed NMR studies and DFT calculations."
J. Organomet. Chem. **2013**, *740*, 41–49. DOI: 10.1016/j.jorganchem.2013.03.012.

[171] K. K. Pandey. "Relativistic DFT calculations of structure and ^{119}Sn NMR chemical shifts for bent M–Sn–C bonding in Power's metallostannylenes of chromium, molybdenum, tungsten and iron and diaryl stannylenes."
J. Organomet. Chem. **2016**, *815–816*, 23–34. DOI: 10.1016/j.jorganchem.2016.05.006.

[172] S. Grimme, A. Hansen, S. Ehlert, and J.-M. Mewes.
"r^{2}SCAN-3c: A "Swiss army knife" composite electronic-structure method."
J. Chem. Phys. **2021**, *154*, 064103. DOI: 10.1063/5.0040021.

[173] F. Alkan, S. T. Holmes, R. J. Iuliucci, K. T. Mueller, and C. Dybowski.
"Spin–orbit effects on the ^{119}Sn NMR magnetic-shielding tensor in solids: a ZORA/DFT investigation." *Phys. Chem. Chem. Phys.* **2016**, *18*, 18914–18922.
DOI: 10.1039/C6CP03807G.

[174] P. Pollak and F. Weigend.
"Segmented Contracted Error-Consistent Basis Sets of Double- and Triple-ζ Valence Quality for One- and Two-Component Relativistic All-Electron Calculations."
J. Chem. Theory Comput. **2017**, *13*, 3696–3705. DOI: 10.1021/acs.jctc.7b00593.

[175] Y. J. Franzke, R. Treß, T. M. Pazdera, and F. Weigend.
"Error-consistent segmented contracted all-electron relativistic basis sets of double- and triple-zeta quality for NMR shielding constants."
Phys. Chem. Chem. Phys. **2019**, *21*, 16658–16664. DOI: 10.1039/C9CP02382H.

[176] Y. Zhao and D. G. Truhlar.
"A new local density functional for main-group thermochemistry, transition metal bonding, thermochemical kinetics, and noncovalent interactions."
J. Chem. Phys. **2006**, *125*, 194101. DOI: 10.1063/1.2370993.

[177] Y. Zhang and W. Yang.
"Comment on "Generalized Gradient Approximation Made Simple"."
Phys. Rev. Lett. **1998**, *80*, 890. DOI: 10.1103/PhysRevLett.80.890.

[178] A. M. Teale, O. B. Lutnæs, T. Helgaker, D. J. Tozer, and J. Gauss.
"Benchmarking density-functional theory calculations of NMR shielding constants and spin–rotation constants using accurate coupled-cluster calculations."
J. Chem. Phys. **2013**, *138*, 024111. DOI: 10.1063/1.4773016.

[179] J. Gauss and J. F. Stanton. "Perturbative treatment of triple excitations in coupled-cluster calculations of nuclear magnetic shielding constants."
J. Chem. Phys. **1996**, *104*, 2574–2583. DOI: 10.1063/1.471005.

[180] K. Ito, Y. Obuchi, E. Chikayama, Y. Date, and J. Kikuchi.
"Exploratory machine-learned theoretical chemical shifts can closely predict metabolic mixture signals." *Chem. Sci.* **2018**, *9*, 8213–8220. DOI: 10.1039/c8sc03628d.

[181] P. Gao, J. Zhang, Y. Sun, and J. Yu.
"Toward Accurate Predictions of Atomic Properties via Quantum Mechanics Descriptors Augmented Graph Convolutional Neural Network: Application of This Novel Approach in NMR Chemical Shifts Predictions."
J. Phys. Chem. Lett. **2020**, *11*, 9812–9818. DOI: 10.1021/acs.jpclett.0c02654.

[182] Y. Kwon, D. Lee, Y.-S. Choi, M. Kang, and S. Kang.
"Neural Message Passing for NMR Chemical Shift Prediction."
J. Chem. Inf. Model. **2020**, *60*, 2024–2030. DOI: 10.1021/acs.jcim.0c00195.

[183] Z. Zhang, J. Zhu, S. Chen, W. Sun, and D. Wang. "Liquid Fluxional Ga Single Atom Catalysts for Efficient Electrochemical CO_2 Reduction."
Angew. Chem. Int. Ed. **2023**, *62*, e202215136. DOI: 10.1002/anie.202215136.

[184] K. Chen, Y. Zhang, J. Xiang, X. Zhao, X. Li, and K. Chu. "p-Block Antimony Single-Atom Catalysts for Nitric Oxide Electroreduction to Ammonia."
ACS Energy Lett. **2023**, *8*, 1281–1288. DOI: 10.1021/acsenergylett.2c02882.

[185] H. W. Moon and J. Cornella. "Bismuth Redox Catalysis: An Emerging Main-Group Platform for Organic Synthesis." *ACS Catal.* **2022**, *12*, 1382–1393.
DOI: 10.1021/acscatal.1c04897.

[186] K. Guan, L. Tao, R. Yang, H. Zhang, N. Wang, H. Wan, J. Cui, J. Zhang, H. Wang, and H. Wang. "Anti-Corrosion for Reversible Zinc Anode via a Hydrophobic Interface in Aqueous Zinc Batteries." *Adv. Energy Mater.* **2022**, *12*, 2103557.
DOI: 10.1002/aenm.202103557.

[187] C. Xiao, W. Song, J. Liang, J. Zhang, Z. Huang, J. Zhang, H. Wang, C. Zhong, J. Ding, and W. Hu. "P-block tin single atom catalyst for improved electrochemistry in a lithium–sulfur battery: a theoretical and experimental study."
J. Mater. Chem. A **2022**, *10*, 3667–3677. DOI: 10.1039/D1TA09422J.

[188] H. Dong, C. Ran, W. Gao, N. Sun, X. Liu, Y. Xia, Y. Chen, and W. Huang. "Crystallization Dynamics of Sn-Based Perovskite Thin Films: Toward Efficient and Stable Photovoltaic Devices." *Adv. Energy Mater.* **2022**, *12*, 2102213. DOI: 10.1002/aenm.202102213.

[189] J. S. Kim, J.-M. Heo, G.-S. Park, S.-J. Woo, C. Cho, H. J. Yun, D.-H. Kim, J. Park, S.-C. Lee, S.-H. Park, E. Yoon, N. C. Greenham, and T.-W. Lee. "Ultra-bright, efficient and stable perovskite light-emitting diodes." *Nature* **2022**, *611*, 688–694. DOI: 10.1038/s41586-022-05304-w.

[190] W. Koch, M. Czop, K. Iłowiecka, A. Nawrocka, and D. Wiącek. "Dietary Intake of Toxic Heavy Metals with Major Groups of Food Products—Results of Analytical Determinations." *Nutrients* **2022**, *14*, 1626. DOI: 10.3390/nu14081626.

[191] N. Natasha, M. Shahid, I. Bibi, J. Iqbal, S. Khalid, B. Murtaza, H. F. Bakhat, A. B. U. Farooq, M. Amjad, H. M. Hammad, N. K. Niazi, and M. Arshad. "Zinc in soil-plant-human system: A data-analysis review." *Sci. Total Environ.* **2022**, *808*, 152024. DOI: 10.1016/j.scitotenv.2021.152024.

[192] S. Ruangritchankul, C. Sumananusorn, J. Sirivarasai, W. Monsuwan, and P. Sritara. "Association between Dietary Zinc Intake, Serum Zinc Level and Multiple Comorbidities in Older Adults." *Nutrients* **2023**, *15*, 322. DOI: 10.3390/nu15020322.

[193] E. Scarpellini, L. M. Balsiger, V. Maurizi, E. Rinninella, A. Gasbarrini, N. Giostra, P. Santori, L. Abenavoli, and C. Rasetti. "Zinc and gut microbiota in health and gastrointestinal disease under the COVID-19 suggestion." *BioFactors* **2022**, *48*, 294–306. DOI: 10.1002/biof.1829.

[194] X. Zheng, B. Ren, X. Li, H. Yan, Q. Xie, H. Liu, J. Zhou, J. Tian, and K. Huang. "Selenoprotein F knockout leads to glucose and lipid metabolism disorders in mice." *J. Biol. Inorg. Chem.* **2020**, *25*, 1009–1022. DOI: 10.1007/s00775-020-01821-z.

[195] M. Schwarz, C. E. Meyer, A. Löser, K. Lossow, J. Hackler, C. Ott, S. Jäger, I. Mohr, E. A. Eklund, A. A. H. Patel, N. Gul, S. Alvarez, I. Altinonder, C. Wiel, M. Maares, H. Haase, A. Härtlova, T. Grune, M. B. Schulze, T. Schwerdtle, U. Merle, H. Zischka, V. I. Sayin, L. Schomburg, and A. P. Kipp. "Excessive copper impairs intrahepatocyte trafficking and secretion of selenoprotein P." *Nat. Commun.* **2023**, *14*, 3479. DOI: 10.1038/s41467-023-39245-3.

[196] I. A. Gallardo, D. A. Todd, S. T. Lima, J. R. Chekan, N. H. Chiu, and E. W. Taylor. "SARS-CoV-2 Main Protease Targets Host Selenoproteins and Glutathione Biosynthesis for Knockdown via Proteolysis, Potentially Disrupting the Thioredoxin and Glutaredoxin Redox Cycles." *Antioxidants* **2023**, *12*, 559. DOI: 10.3390/antiox12030559.

[197] J. B. Kleine Büning and S. Grimme. "Computation of CCSD(T)-Quality NMR Chemical Shifts via Δ-Machine Learning from DFT." *J. Chem. Theory Comput.* **2023**, *19*, 3601–3615. DOI: 10.1021/acs.jctc.3c00165.

[198] M. Bursch, H. Neugebauer, S. Ehlert, and S. Grimme. "Dispersion corrected r^2SCAN based global hybrid functionals: r^2SCANh, r^2SCAN0, and r^2SCAN50." *J. Chem. Phys.* **2022**, *156*, 134105. DOI: 10.1063/5.0086040.

[199] J. B. Stückrath, T. Gasevic, M. Bursch, and S. Grimme. "Benchmark Study on the Calculation of ^{119}Sn NMR Chemical Shifts." *Inorg. Chem.* **2022**, *61*, 3903–3917. DOI: 10.1021/acs.inorgchem.1c03453.

[200] T. Gasevic, J. B. Kleine Büning, S. Grimme, and M. Bursch. "Benchmark Study on the Calculation of ^{207}Pb NMR Chemical Shifts." *Inorg. Chem.*, *Manuscript in revision.*

[201] T. Ishii, K. Suzuki, T. Nakamura, and M. Yamashita. "An Isolable Bismabenzene: Synthesis, Structure, and Reactivity." *J. Am. Chem. Soc.* **2016**, *138*, 12787–12790. DOI: 10.1021/jacs.6b08714.

[202] B. Baisch, D. Raffa, U. Jung, O. M. Magnussen, C. Nicolas, J. Lacour, J. Kubitschke, and R. Herges. "Mounting Freestanding Molecular Functions onto Surfaces: The Platform Approach." *J. Am. Chem. Soc.* **2009**, *131*, 442–443. DOI: 10.1021/ja807923f.

[203] J. A. Mann, J. Rodríguez-López, H. D. Abruña, and W. R. Dichtel. "Multivalent Binding Motifs for the Noncovalent Functionalization of Graphene." *J. Am. Chem. Soc.* **2011**, *133*, 17614–17617. DOI: 10.1021/ja208239v.

[204] L. Gerhard, K. Edelmann, J. Homberg, M. Valášek, S. G. Bahoosh, M. Lukas, F. Pauly, M. Mayor, and W. Wulfhekel. "An electrically actuated molecular toggle switch." *Nat. Commun.* **2017**, *8*, 14672. DOI: 10.1038/ncomms14672.

[205] J. Homberg, M. Lindner, L. Gerhard, K. Edelmann, T. Frauhammer, Y. Nahas, M. Valášek, M. Mayor, and W. Wulfhekel. "Six state molecular revolver mounted on a rigid platform." *Nanoscale* **2019**, *11*, 9015–9022. DOI: 10.1039/C9NR00259F.

[206] L. Sosa-Vargas, E. Kim, and A.-J. Attias. "Beyond "decorative" 2D supramolecular self-assembly: strategies towards functional surfaces for Nanotechnology." *Mater. Horiz.* **2017**, *4*, 570–583. DOI: 10.1039/C7MH00127D.

[207] T. R. Rusch, M. Hammerich, R. Herges, and O. M. Magnussen. "Molecular platforms as versatile building blocks for multifunctional photoswitchable surfaces." *Chem. Commun.* **2019**, *55*, 9511–9514. DOI: 10.1039/C9CC04528G.

[208] M. J. Hollamby and T. Nakanishi. "The power of branched chains: optimising functional molecular materials." *J. Mater. Chem. C* **2013**, *1*, 6178–6183. DOI: 10.1039/C3TC31381F.

[209] T. Lei, J.-Y. Wang, and J. Pei. "Roles of Flexible Chains in Organic Semiconducting Materials." *Chem. Mater.* **2014**, *26*, 594–603. DOI: 10.1021/cm4018776.

[210] G. Poluektov, T. J. Keller, A. Jochemich, A. Krönert, U. Müller, S. Spicher, S. Grimme, S.-S. Jester, and S. Höger. "Supramolecular Nanopatterns of Molecular Spoked Wheels with Orthogonal Pillars: The Observation of a Fullerene Haze." *Angew. Chem. Int. Ed.* **2021**, *60*, 27264–27270. DOI: 10.1002/anie.202111869.

[211] S. Henzel, S. Becker, D. Hennen, T. J. Keller, J. Bahr, S.-S. Jester, and S. Höger. "Highly Strained Nanoscale Bicyclophane Monolayers Entering the Third Dimension: A Combined Synthetic and Scanning Tunneling Microscopy Investigation." *ChemPlusChem* **2021**, *86*, 803–811. DOI: 10.1002/cplu.202000711.

[212] G. Ohlendorf, C. W. Mahler, S.-S. Jester, G. Schnakenburg, S. Grimme, and S. Höger. "Highly Strained Phenylene Bicyclophanes." *Angew. Chem. Int. Ed.* **2013**, *52*, 12086–12090. DOI: 10.1002/anie.201306299.

[213] M. Bursch, T. Gasevic, J. B. Stückrath, and S. Grimme. "Comprehensive Benchmark Study on the Calculation of ^{29}Si NMR Chemical Shifts." *Inorg. Chem.* **2021**, *60*, 272–285. DOI: 10.1021/acs.inorgchem.0c02907.

[214] C. Riplinger and F. Neese. "An efficient and near linear scaling pair natural orbital based local coupled cluster method." *J. Chem. Phys.* **2013**, *138*, 034106. DOI: 10.1063/1.4773581.

[215] C. Riplinger, B. Sandhoefer, A. Hansen, and F. Neese. "Natural triple excitations in local coupled cluster calculations with pair natural orbitals." *J. Chem. Phys.* **2013**, *139*, 134101. DOI: 10.1063/1.4821834.

[216] P. Pinski, C. Riplinger, E. F. Valeev, and F. Neese. "Sparse maps—A systematic infrastructure for reduced-scaling electronic structure methods. I. An efficient and simple linear scaling local MP2 method that uses an intermediate basis of pair natural orbitals." *J. Chem. Phys.* **2015**, *143*, 034108. DOI: 10.1063/1.4926879.

[217] C. Riplinger, P. Pinski, U. Becker, E. F. Valeev, and F. Neese. "Sparse maps—A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory." *J. Chem. Phys.* **2016**, *144*, 024109. DOI: 10.1063/1.4939030.

[218] H. Neugebauer, P. Pinski, S. Grimme, F. Neese, and M. Bursch. "Assessment of DLPNO-MP2 Approximations in Double-Hybrid DFT." *J. Chem. Theory Comput.* **2023**, *19*, 7695–7703. DOI: 10.1021/acs.jctc.3c00896.

[219] G. L. Stoychev, A. A. Auer, J. Gauss, and F. Neese. "DLPNO-MP2 second derivatives for the computation of polarizabilities and NMR shieldings." *J. Chem. Phys.* **2021**, *154*, 164110. DOI: 10.1063/5.0047125.

[220] J. Liang, Z. Wang, J. Li, J. Wong, X. Liu, B. Ganoe, T. Head-Gordon, and M. Head-Gordon. "Efficient Calculation of NMR Shielding Constants Using Composite Method Approximations and Locally Dense Basis Sets." *J. Chem. Theory Comput.* **2023**, *19*, 514–523. DOI: 10.1021/acs.jctc.2c00933.

[221] S. Zheng and J. Autschbach.
"Modeling of Heavy-Atom–Ligand NMR Spin–Spin Coupling in Solution: Molecular Dynamics Study and Natural Bond Orbital Analysis of Hg–C Coupling Constants."
Chem. Eur. J. **2011**, *17*, 161–173. DOI: 10.1002/chem.201001343.

[222] L. B. Krivdin.
"Recent advances in computational ^{31}P NMR: Part 2. Spin–spin coupling constants."
Magn. Reson. Chem. **2020**, *58*, 500–511. DOI: 10.1002/mrc.4973.

[223] X.-X. Zhang, T. Deng, and G.-Z. Jia. "Nuclear spin-spin coupling constants prediction based on XGBoost and LightGBM algorithms."
Mol. Phys. **2020**, *118*, e1696478. DOI: 10.1080/00268976.2019.1696478.

[224] M. Dračínský and P. Bouř.
"Computational Analysis of Solvent Effects in NMR Spectroscopy."
J. Chem. Theory Comput. **2010**, *6*, 288–299. DOI: 10.1021/ct900498b.

[225] P. Lafiosca, F. Rossi, F. Egidi, T. Giovannini, and C. Cappelli.
"Multiscale Frozen Density Embedding/Molecular Mechanics Approach for Simulating Magnetic Response Properties of Solvated Systems."
J. Chem. Theory Comput. **2024**, *20*, 266–279. DOI: 10.1021/acs.jctc.3c00850.

[226] J. R. Cheeseman, G. W. Trucks, T. A. Keith, and M. J. Frisch. "A comparison of models for calculating nuclear magnetic resonance shielding tensors."
J. Chem. Phys. **1996**, *104*, 5497–5509. DOI: 10.1063/1.471789.

[227] C. van Wüllen. "A comparison of density functional methods for the calculation of phosphorus-31 NMR chemical shifts." *Phys. Chem. Chem. Phys.* **2000**, *2*, 2137–2144. DOI: 10.1039/b000461h.

[228] G. L. Stoychev, A. A. Auer, R. Izsák, and F. Neese. "Self-Consistent Field Calculation of Nuclear Magnetic Resonance Chemical Shielding Constants Using Gauge-Including Atomic Orbitals and Approximate Two-Electron Integrals."
J. Chem. Theory Comput. **2018**, *14*, 619–637. DOI: 10.1021/acs.jctc.7b01006.

[229] Y. Zhao and D. G. Truhlar. "Improved Description of Nuclear Magnetic Resonance Chemical Shielding Constants Using the M06-L Meta-Generalized-Gradient-Approximation Density Functional."
J. Phys. Chem. A **2008**, *112*, 6794–6799. DOI: 10.1021/jp804583d.

[230] T. Gregor, F. Mauri, and R. Car.
"A comparison of methods for the calculation of NMR chemical shifts."
J. Chem. Phys. **1999**, *111*, 1815–1822. DOI: 10.1063/1.479451.

[231] T. Heine, A. Goursot, G. Seifert, and J. Weber.
"Performance of DFT for ^{29}Si NMR Chemical Shifts of Silanes."
J. Phys. Chem. A **2001**, *105*, 620–626. DOI: 10.1021/jp002495k.

[232] C. Corminbœuf, T. Heine, and J. Weber.
"^{29}Si NMR chemical shifts of silane derivatives." *Chem. Phys. Lett.* **2002**, *357*, 1–7. DOI: 10.1016/S0009-2614(02)00372-X.

[233] T. M. Alam and M. Henry.
"Empirical calculations of ^{29}Si NMR chemical shielding tensors: A partial charge model investigation of hydrolysis in organically modified alkoxy silanes." *Phys. Chem. Chem. Phys.* **2000**, *2*, 23–28. DOI: 10.1039/a906445a.

[234] D. Auer, M. Kaupp, and C. Strohmann. ""Unexpected" ^{29}Si NMR Chemical Shifts in Heteroatom-Substituted Silyllithium Compounds: A Quantum-Chemical Analysis." *Organometallics* **2004**, *23*, 3647–3655. DOI: 10.1021/om049812k.

[235] L. J. P. van der Boon, J. H. Hendriks, D. Roolvink, S. J. O'Kennedy, M. Lutz, J. C. Slootweg, A. W. Ehlers, and K. Lammertsma.
"Dynamic Conformational Behavior in Stable Pentaorganosilicates." *Eur. J. Inorg. Chem.* **2019**, *2019*, 3318–3328. DOI: 10.1002/ejic.201900641.

[236] M. I. Arz, D. Geiß, M. Straßmann, G. Schnakenburg, and A. C. Filippou.
"Silicon(I) chemistry: the NHC-stabilised silicon(I) halides $Si_2X_2(Idipp)_2$ (X = Br, I) and the disilicon(I)-iodido cation $[Si_2(I)(Idipp)_2]^+$." *Chem. Sci.* **2015**, *6*, 6515–6524. DOI: 10.1039/c5sc02681d.

[237] L. A. Truflandier, E. Brendler, J. Wagler, and J. Autschbach.
"^{29}Si DFT/NMR Observation of Spin–Orbit Effect in Metallasilatrane Sheds Some Light on the Strength of the Metal→Silicon Interaction." *Angew. Chem. Int. Ed.* **2011**, *50*, 255–259. DOI: 10.1002/anie.201005431.

[238] M. Karni, Y. Apeloig, N. Takagi, and S. Nagase.
"Ab Initio and DFT Study of the ^{29}Si NMR Chemical Shifts in RSi≡SiR." *Organometallics* **2005**, *24*, 6319–6330. DOI: 10.1021/om058033g.

[239] V. A. Du, G. N. Stipicic, and U. Schubert.
"^{29}Si NMR Shielding Calculations Employing Density Functional Theory, Focussing on Hypervalent Silicon Compounds." *Eur. J. Inorg. Chem.* **2011**, *2011*, 3365–3373. DOI: 10.1002/ejic.201100286.

[240] A. D. Boese and N. C. Handy. "A new parametrization of exchange–correlation generalized gradient approximation functionals." *J. Chem. Phys.* **2001**, *114*, 5497–5503. DOI: 10.1063/1.1347371.

[241] C. Zhang, P. Patschinski, D. S. Stephenson, R. Panisch, J. H. Wender, M. C. Holthausen, and H. Zipse. "The calculation of ^{29}Si NMR chemical shifts of tetracoordinated silicon compounds in the gas phase and in solution." *Phys. Chem. Chem. Phys.* **2014**, *16*, 16642–16650. DOI: 10.1039/C4CP01736F.

[242] S. V. Fedorov, Y. Y. Rusakov, and L. B. Krivdin.
"Quantum-Chemical Calculations of NMR Chemical Shifts of Organic Molecules: XV. Relativistic Calculations of ^{29}Si NMR Chemical Shifts of Silanes."
Russ. J. Org. Chem **2017**, *53*, 643–651. DOI: 10.1134/S1070428017050013.

[243] F. Neese. "The ORCA program system." *WIREs Comput. Mol. Sci.* **2012**, *2*, 73–78.
DOI: 10.1002/wcms.81.

[244] ORCA – *An ab initio, DFT and semiempirical SCF-MO package*, Versions 4.2.1, 5.0.3, and 5.0.4, Max-Planck-Institut für Kohlenforschung, Mülheim a. d. Ruhr, Germany **2019**, **2022**, https://www.faccts.de/orca.

[245] G. te Velde, F. M. Bickelhaupt, E. J. Baerends, C. Fonseca Guerra, S. J. A. van Gisbergen, J. G. Snijders, and T. Ziegler. "Chemistry with ADF."
J. Comput. Chem. **2001**, *22*, 931–967. DOI: 10.1002/jcc.1056.

[246] AMS – *Amsterdam Modeling Suite*, Versions 2019.303, 2020.102, and 2022.103, SCM, Theoretical Chemistry, Vrije Universiteit, Amsterdam, The Netherlands, **2019**, **2020**, **2022**, http://www.scm.com.

[247] xTB – *Semiempirical Extended Tight-Binding Program Package*, Versions 6.3.3, 6.5.0, and 6.6.0, Universität Bonn, Mulliken Center for Theoretical Chemistry, Bonn, Germany **2020**, **2022**, **2023**, https://github.com/grimme-lab/xtb/releases.

[248] C. Bannwarth, E. Caldeweyher, S. Ehlert, A. Hansen, P. Pracht, J. Seibert, S. Spicher, and S. Grimme. "Extended tight-binding quantum chemistry methods."
WIREs Comput. Mol. Sci. **2021**, *11*, e1493. DOI: 10.1002/wcms.1493.

[249] E. Caldeweyher, C. Bannwarth, and S. Grimme.
"Extension of the D3 dispersion coefficient model." *J. Chem. Phys.* **2017**, *147*, 034112.
DOI: 10.1063/1.4993215.

[250] E. Caldeweyher, S. Ehlert, A. Hansen, H. Neugebauer, S. Spicher, C. Bannwarth, and S. Grimme.
"A generally applicable atomic-charge dependent London dispersion correction."
J. Chem. Phys. **2019**, *150*, 154122. DOI: 10.1063/1.5090222.

[251] M. Bursch, E. Caldeweyher, A. Hansen, H. Neugebauer, S. Ehlert, and S. Grimme.
"Understanding and Quantifying London Dispersion Effects in Organometallic Complexes." *Acc. Chem. Res.* **2019**, *52*, 258–266.
DOI: 10.1021/acs.accounts.8b00505.

[252] H. Tatewaki and S. Huzinaga. "A Systematic Preparation of New Contracted Gaussian-Type Orbital Sets. III. Second-Row Atoms from Li through Ne."
J. Comput. Chem. **1980**, *1*, 205–228. DOI: 10.1002/jcc.540010302.

[253] R. Sure and S. Grimme.
"Corrected Small Basis Set Hartree-Fock Method for Large Systems."
J. Comput. Chem. **2013**, *34*, 1672–1685. DOI: 10.1002/jcc.23317.

[254] K. A. Peterson, D. Figgen, E. Goll, H. Stoll, and M. Dolg. "Systematically convergent basis sets with relativistic pseudopotentials. II. Small-core pseudopotentials and correlation consistent basis sets for the post-*d* group 16–18 elements." *J. Chem. Phys.* **2003**, *119*, 11113–11123. DOI: 10.1063/1.1622924.

[255] K. Eichkorn, O. Treutler, H. Öhm, M. Häser, and R. Ahlrichs. "Auxiliary basis sets to approximate Coulomb potentials." *Chem. Phys. Lett.* **1995**, *240*, 283–290. DOI: 10.1016/0009-2614(95)00621-A.

[256] A. Hellweg, C. Hättig, S. Höfener, and W. Klopper. "Optimized accurate auxiliary basis sets for RI-MP2 and RI-CC2 calculations for the atoms Rb to Rn." *Theor. Chem. Acc.* **2007**, *117*, 587–597. DOI: 10.1007/s00214-007-0250-5.

[257] D. Andrae, U. Häußermann, M. Dolg, H. Stoll, and H. Preuß. "Energy-adjusted *ab initio* pseudopotentials for the second and third row transition elements." *Theor. Chim. Acta* **1990**, *77*, 123–141. DOI: 10.1007/BF01114537.

[258] S. K. Wolff, T. Ziegler, E. van Lenthe, and E. J. Baerends. "Density functional calculations of nuclear magnetic shieldings using the zeroth-order regular approximation (ZORA) for relativistic effects: ZORA nuclear magnetic resonance." *J. Chem. Phys.* **1999**, *110*, 7689–7698. DOI: 10.1063/1.478680.

[259] D. A. Pantazis, X.-Y. Chen, C. R. Landis, and F. Neese. "All-Electron Scalar Relativistic Basis Sets for Third-Row Transition metal Atoms." *J. Chem. Theory. Comput.* **2008**, *4*, 908–919. DOI: 10.1021/ct800047t.

[260] D. A. Pantazis and F. Neese. "All-electron scalar relativistic basis sets for the 6*p* elements." *Theor. Chem. Acc.* **2012**, *131*, 1292. DOI: 10.1007/s00214-012-1292-x.

[261] C. C. Pye and T. Ziegler. "An implementation of the conductor-like screening model of solvation within the Amsterdam density functional package." *Theor. Chem. Acc.* **1999**, *101*, 396–408. DOI: 10.1007/s002140050457.

[262] N. C. Handy and A. J. Cohen. "Left-right correlation energy." *Mol. Phys.* **2001**, *99*, 403–412. DOI: 10.1080/00268970010018431.

[263] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, L. A. Constantin, and J. Sun. "Workhorse Semilocal Density Functional for Condensed Matter Physics and Quantum Chemistry." *Phys. Rev. Lett.* **2009**, *103*, 026403. DOI: 10.1103/PhysRevLett.103.026403.

[264] J. P. Perdew, A. Ruzsinszky, G. I. Csonka, L. A. Constantin, and J. Sun. "Erratum: Workhorse Semilocal Density Functional for Condensed Matter Physics and Quantum Chemistry [Phys. Rev. Lett. 103, 026403 (2009)]." *Phys. Rev. Lett.* **2011**, *106*, 179902. DOI: 10.1103/PhysRevLett.106.179902.

[265] J. Sun, A. Ruzsinszky, and J. P. Perdew. "Strongly Constrained and Appropriately Normed Semilocal Density Functional." *Phys. Rev. Lett.* **2015**, *115*, 036402. DOI: 10.1103/PhysRevLett.115.036402.

[266] T. Yanai, D. P. Tew, and N. C. Handy. "A new hybrid exchange–correlation functional using the Coulomb-attenuating method (CAM-B3LYP)." *Chem. Phys. Lett.* **2004**, *393*, 51–57. DOI: 10.1016/j.cplett.2004.06.011.

[267] A. D. Becke. "A new mixing of Hartree–Fock and local density-functional theories." *J. Chem. Phys.* **1993**, *98*, 1372–1377. DOI: 10.1063/1.464304.

[268] A. D. Boese and J. M. L. Martin. "Development of density functional for thermochemical kinetics." *J. Chem. Phys.* **2004**, *121*, 3405–3416. DOI: 10.1063/1.1774975.

[269] J.-D. Chai and M. Head-Gordon. "Systematic optimization of long-range corrected hybrid density functionals." *J. Chem. Phys.* **2008**, *128*, 084106. DOI: 10.1063/1.2834918.

[270] J.-D. Chai and M. Head-Gordon. "Long-range corrected hybrid density functionals with damped atom–atom dispersion corrections." *Phys. Chem. Chem. Phys.* **2008**, *10*, 6615–6620. DOI: 10.1039/b810189b.

[271] Y.-S. Lin, G.-D. Li, S.-P. Mao, and J.-D. Chai. "Long-Range Corrected Hybrid Density Functionals with Improved Dispersion Corrections." *J. Chem. Theory Comput.* **2013**, *9*, 263–272. DOI: 10.1021/ct300715s.

[272] Y. Zhao and D. G. Truhlar. "Design of Density Functionals That Are Broadly Accurate for Thermochemistry, Thermochemical Kinetics, and Nonbonded Interactions." *J. Phys. Chem. A* **2005**, *109*, 5656–5667. DOI: 10.1021/jp050536c.

[273] S. Kozuch and J. M. L. Martin. "DSD-PBEP86: in search of the best double-hybrid DFT with spin-component scaled MP2 and dispersion corrections." *Phys. Chem. Chem. Phys.* **2011**, *13*, 20104–20107. DOI: 10.1039/c1cp22592h.

[274] S. Kozuch and J. M. L. Martin. "Spin-Component-Scaled Double Hybrids: An Extensive Search for the Best Fifth-Rung Functionals Blending DFT and Perturbation Theory." *J. Comput. Chem.* **2013**, *34*, 2327–2344. DOI: 10.1002/jcc.23391.

[275] G. Santra, N. Sylvetsky, and J. M. L. Martin. "Minimally Empirical Double-Hybrid Functionals Trained against the GMTKN55 Database: revDSD-PBEP86-D4, revDOD-PBE-D4, and DOD-SCAN-D4." *J. Phys. Chem. A* **2019**, *123*, 5129–5143. DOI: 10.1021/acs.jpca.9b03157.

[276] S. Kozuch, D. Gruzman, and J. M. L. Martin. "DSD-BLYP: A General Purpose Double Hybrid Density Functional Including Spin Component Scaling and Dispersion Correction." *J. Phys. Chem. C* **2010**, *114*, 20801–20808. DOI: 10.1021/jp1070852.

[277] A. Karton, A. Tarnopolsky, J.-F. Lamère, G. C. Schatz, and J. M. L. Martin. "Highly Accurate First-Principles Benchmark Data Sets for the Parametrization and Validation of Density Functional and Other Approximate Methods. Derivation of a Robust, Generally Applicable, Double-Hybrid Functional for Thermochemistry and Thermochemical Kinetics." *J. Phys. Chem. A* **2008**, *112*, 12868–12886. DOI: 10.1021/jp801805p.

[278] T. Schwabe and S. Grimme. "Towards chemical accuracy for the thermodynamics of large molecules: new hybrid density functionals including non-local correlation effects." *Phys. Chem. Chem. Phys.* **2006**, *8*, 4398–4401. DOI: 10.1039/b608478h.

[279] J. G. Brandenburg, C. Bannwarth, A. Hansen, and S. Grimme. "B97-3c: A revised low-cost variant of the B97-D density functional method." *J. Chem. Phys.* **2018**, *148*, 064104. DOI: 10.1063/1.5012601.

[280] J. Tomasi and M. Persico. "Molecular Interactions in Solution: An Overview of Methods Based on Continuous Distributions of the Solvent." *Chem. Rev.* **1994**, *94*, 2027–2094. DOI: 10.1021/cr00031a013.

[281] J. P. Perdew, A. Ruzsinszky, J. Tao, V. N. Staroverov, G. E. Scuseria, and G. I. Csonka. "Prescription for the design and selection of density functional approximations: More constraint satisfaction with fewer fits." *J. Chem. Phys.* **2005**, *123*, 062201. DOI: 10.1063/1.1904565.

[282] M. J. Allen, T. W. Keal, and D. J. Tozer. "Improved NMR chemical shifts in density functional theory." *Chem. Phys. Lett.* **2003**, *380*, 70–77. DOI: 10.1016/j.cplett.2003.08.101.

[283] M. I. Arz, M. Straßmann, D. Geiß, G. Schnakenburg, and A. C. Filippou. "Addition of Small Electrophiles to N-Heterocyclic-Carbene-Stabilized Disilicon(0): A Revisit of the Isolobal Concept in Low-Valent Silicon Chemistry." *J. Am. Chem. Soc.* **2016**, *138*, 4589–4600. DOI: 10.1021/jacs.6b01018.

[284] M. Kira, S. Ishida, T. Iwamoto, and C. Kabuto. "The First Isolable Dialkylsilylene." *J. Am. Chem. Soc.* **1999**, *121*, 9722–9723. DOI: 10.1021/ja9925305.

[285] P. R. Rablen, S. A. Pearlman, and J. Finkbiner. "A Comparison of Density Functional Methods for the Estimation of Proton Chemical Shifts with Chemical Accuracy." *J. Phys. Chem. A* **1999**, *103*, 7357–7363. DOI: 10.1021/jp9916889.

[286] A. E. Aliev, D. Courtier-Murias, and S. Zhou. "Scaling factors for carbon NMR chemical shifts obtained from DFT B3LYP calculations." *J. Mol. Struct.: THEOCHEM* **2009**, *893*, 1–5. DOI: 10.1016/j.theochem.2008.09.021.

[287] I. A. Konstantinov and L. J. Broadbelt. "Regression Formulas for Density Functional Theory Calculated ^{1}H and ^{13}C NMR Chemical Shifts in Toluene-d_8." *J. Phys. Chem. A* **2011**, *115*, 12364–12372. DOI: 10.1021/jp2060975.

[288] J. Vícha, J. Novotný, M. Straka, M. Repisky, K. Ruud, S. Komorovsky, and R. Marek. "Structure, solvent, and relativistic effects on the NMR chemical shifts in square-planar transition-metal complexes: assessment of DFT approaches." *Phys. Chem. Chem. Phys.* **2015**, *17*, 24944–24955. DOI: 10.1039/C5CP04214C.

[289] J. Vaara, P. Manninen, and P. Lantto. "Perturbational and ECP Calculation of Relativistic Effects in NMR Shielding and Spin–Spin Coupling." In *Calculation of NMR and EPR Parameters*. Ed. by M. Kaupp, M. Bühl, and V. G. Malkin. John Wiley & Sons, Ltd., **2004**. Chap. 13, pp. 209–226. DOI: 10.1002/3527601678.ch13.

[290] J. Autschbach. "Calculating NMR Chemical Shifts and *J*-Couplings for Heavy Element Compounds." *Encyclopedia of Analytical Chemistry* **2006**, pp. 1–14. DOI: 10.1002/9780470027318.a9173.

[291] P. Hrobárik, V. Hrobáriková, A. H. Greif, and M. Kaupp. "Giant Spin-Orbit Effects on NMR Shifts in Diamagnetic Actinide Complexes: Guiding the Search of Uranium(VI) Hydride Complexes in the Correct Spectral Range." *Angew. Chem. Int. Ed.* **2012**, *51*, 10884–10888. DOI: 10.1002/anie.201204634.

[292] T. Saue. "Relativistic Hamiltonians for Chemistry: A Primer." *ChemPhysChem* **2011**, *12*, 3077–3094. DOI: 10.1002/cphc.201100682.

[293] A. C. Filippou, B. Baars, O. Chernov, Y. N. Lebedev, and G. Schnakenburg. "Silicon–Oxygen Double Bonds: A Stable Silanone with a Trigonal-Planar Coordinated Silicon Center." *Angew. Chem. Int. Ed.* **2014**, *53*, 565–570. DOI: 10.1002/anie.201308433.

[294] E. Sola, A. García-Camprubí, J. L. Andrés, M. Martín, and P. Plou. "Iridium Compounds with *κ-P,P,Si* (biPSi) Pincer Ligands: Favoring Reactive Structures in Unsaturated Complexes." *J. Am. Chem. Soc.* **2010**, *132*, 9111–9121. DOI: 10.1021/ja102479h.

[295] M. Taoufik, K. C. Szeto, N. Merle, I. D. Rosal, L. Maron, J. Trébosc, G. Tricot, R. M. Gauvin, and L. Delevoye. "Heteronuclear NMR Spectroscopy as a Surface-Selective Technique: A Unique Look at the Hydroxyl Groups of γ-Alumina." *Chem. Eur. J.* **2014**, *20*, 4038–4046. DOI: 10.1002/chem.201304883.

[296] M. Kaupp, O. L. Malkina, V. G. Malkin, and P. Pyykkö. "How Do Spin–Orbit-Induced Heavy-Atom Effects on NMR Chemical Shifts Function? Validation of a Simple Analogy to Spin–Spin Coupling by Density Functional Theory (DFT) Calculations on Some Iodo Compounds." *Chem. Eur. J.* **1998**, *4*, 118–126. DOI: 10.1002/(SICI)1521-3765(199801)4:1<118::AID-CHEM118>3.0.CO;2-6.

[297] T. P. Lockhart and W. F. Manders.
"Structure Determination by NMR Spectroscopy. Correlation of $|{}^2J(^{119}\mathrm{Sn},^1\mathrm{H})|$ on the Me–Sn–Me Angle in Methyltin(IV) Compounds." *Inorg. Chem.* **1986**, *25*, 892–895. DOI: 10.1021/ic00227a002.

[298] C. van Wüllen.
"Chemical Shifts with Hartree–Fock and Density Functional Methods."
In *Calculation of NMR and EPR Parameters*.
Ed. by M. Kaupp, M. Bühl, and V. G. Malkin. John Wiley & Sons, Ltd., **2004**.
Chap. 6, pp. 83–100. DOI: 10.1002/3527601678.ch6.

[299] P. Avalle, R. K. Harris, and R. D. Fischer.
"DFT calculations of ^{119}Sn chemical shifts for organometallic cyanides."
Phys. Chem. Chem. Phys. **2002**, *4*, 3558–3561. DOI: 10.1039/B202909J.

[300] L. Wang, C. E. Kefalidis, T. Roisnel, S. Sinbandhit, L. Maron, J.-F. Carpentier, and Y. Sarazin. "Structure vs ^{119}Sn NMR Chemical Shift in Three-Coordinated Tin(II) Complexes: Experimental Data and Predictive DFT Computations."
Organometallics **2015**, *34*, 2139–2150. DOI: 10.1021/om5007566.

[301] E. Y. Pankratyev, A. R. Tulyabaev, and L. M. Khalilov.
"How Reliable are GIAO Calculations of ^{1}H and ^{13}C NMR Chemical Shifts? A Statistical Analysis and Empirical Corrections at DFT (PBE/3z) Level."
J. Comput. Chem. **2011**, *32*, 1993–1997. DOI: 10.1002/jcc.21786.

[302] W. Liu and D. Peng. "Infinite-order quasirelativistic density functional method based on the exact matrix quasirelativistic theory." *J. Chem. Phys.* **2006**, *125*, 044102. DOI: 10.1063/1.2222365.

[303] D. Peng, N. Middendorf, F. Weigend, and M. Reiher. "An efficient implementation of two-component relativistic exact-decoupling methods for large molecules."
J. Chem. Phys. **2013**, *138*, 184105. DOI: 10.1063/1.4803693.

[304] H. Kaneko, M. Hada, T. Nakajima, and H. Nakatsuji. "Spin-orbit effect on the magnetic shielding constant using the ab initio UHF method: tin tetrahalides."
Chem. Phys. Lett. **1996**, *261*, 1–6. DOI: 10.1016/0009-2614(96)00906-2.

[305] R. Vivas-Reyes, F. De Proft, M. Biesemans, R. Willem, and P. Geerlings.
"DFT Calculations of ^{119}Sn Chemical Shifts Using Gauge-Including Atomic Orbitals and Their Interpretation via Group Properties." *J. Phys. Chem. A* **2002**, *106*, 2753–2759. DOI: 10.1021/jp0145917.

[306] P. Avalle, R. K. Harris, P. B. Karadakov, and P. J. Wilson.
"Calculations of magnetic shielding for the tin nucleus in a series of tetra-organotin compounds using density functional theory."
Phys. Chem. Chem. Phys. **2002**, *4*, 5925–5932. DOI: 10.1039/B208435J.

[307] TURBOMOLE, Versions 7.5.1, 7.5.3, and 7.7.1, Universität Karlsruhe & Forschungszentrum Karlsruhe GmbH, Karlsruhe, Germany **2020**, **2023**, https://www.turbomole.org.

[308] F. Furche, R. Ahlrichs, C. Hättig, W. Klopper, M. Sierka, and F. Weigend. "Turbomole." *WIREs Comput. Mol. Sci.* **2014**, *4*, 91–100. DOI: 10.1002/wcms.1162.

[309] S. G. Balasubramani, G. P. Chen, S. Coriani, M. Diedenhofen, M. S. Frank, Y. J. Franzke, F. Furche, R. Grotjahn, M. E. Harding, C. Hättig, A. Hellweg, B. Helmich-Paris, C. Holzer, U. Huniar, M. Kaupp, A. Marefat Khah, S. Karbalaei Khani, T. Müller, F. Mack, B. D. Nguyen, S. M. Parker, E. Perlt, D. Rappoport, K. Reiter, S. Roy, M. Rückert, G. Schmitz, M. Sierka, E. Tapavicza, D. P. Tew, C. van Wüllen, V. K. Voora, F. Weigend, A. Wodyński, and J. M. Yu. "TURBOMOLE: Modular program suite for *ab initio* quantum-chemical and condensed-matter simulations." *J. Chem. Phys.* **2020**, *152*, 184107. DOI: 10.1063/5.0004635.

[310] CREST – *Conformer-Rotamer Ensemble Sampling Tool*, Versions 2.11, 2.11.2, and 2.12, Universität Bonn, Mulliken Center for Theoretical Chemistry, Bonn, Germany **2021**, **2022**, https://github.com/crest-lab/crest/releases.

[311] CENSO – *Commandline Energetic Sorting of Conformer-Rotamer Ensembles*, Versions 1.0.0, 1.1.2, and 1.2.0, Universität Bonn, Mulliken Center for Theoretical Chemistry, Bonn, Germany **2021**, **2022**, https://github.com/grimme-lab/censo/releases.

[312] F. Eckert and A. Klamt. "Fast Solvent Screening via Quantum Chemistry: COSMO-RS Approach." *AIChE J.* **2002**, *48*, 369–385. DOI: 10.1002/aic.690480220.

[313] K. Eichkorn, F. Weigend, O. Treutler, and R. Ahlrichs. "Auxiliary basis sets for main row atoms and transition metals and their use to approximate Coulomb potentials." *Theor. Chem. Acc.* **1997**, *97*, 119–124. DOI: 10.1007/s002140050244.

[314] P. Deglmann, K. May, F. Furche, and R. Ahlrichs. "Nuclear second analytical derivative calculations using auxiliary basis set expansions." *Chem. Phys. Lett.* **2004**, *384*, 103–107. DOI: 10.1016/j.cplett.2003.11.080.

[315] M. Häser, R. Ahlrichs, H. P. Baron, P. Weis, and H. Horn. "Direct computation of second-order SCF properties of large molecules on workstation computers with an application to large carbon clusters." *Theor. Chim. Acta* **1992**, *83*, 455–470. DOI: 10.1007/BF01113068.

[316] K. Reiter, F. Mack, and F. Weigend. "Calculation of Magnetic Shielding Constants with meta-GGA Functionals Employing the Multipole-Accelerated Resolution of the Identity: Implementation and Assessment of Accuracy and Efficiency." *J. Chem. Theory Comput.* **2018**, *14*, 191–197. DOI: 10.1021/acs.jctc.7b01115.

[317] Y. J. Franzke and F. Weigend. "NMR Shielding Tensors and Chemical Shifts in Scalar-Relativistic Local Exact Two-Component Theory." *J. Chem. Theory Comput.* **2019**, *15*, 1028–1043. DOI: 10.1021/acs.jctc.8b01084.

[318] E. van Lenthe, E. J. Baerends, and J. G. Snijders. "Relativistic total energy using regular approximations." *J. Chem. Phys.* **1994**, *101*, 9783–9792. DOI: 10.1063/1.467943.

[319] M. Krykunov, T. Ziegler, and E. van Lenthe. "Hybrid Density Functional Calculations of Nuclear Magnetic Shieldings Using Slater-Type Orbitals and the Zeroth-Order Regular Approximation." *Int. J. Quantum Chem.* **2009**, *109*, 1676–1683. DOI: 10.1002/qua.21985.

[320] J. W. Furness, A. D. Kaplan, J. Ning, J. P. Perdew, and J. Sun. "Correction to "Accurate and Numerically Efficient r^{2}SCAN Meta-Generalized Gradient Approximation"." *J. Phys. Chem. Lett.* **2020**, *11*, 9248. DOI: 10.1021/acs.jpclett.0c03077.

[321] Y. J. Franzke, N. Middendorf, and F. Weigend. "Efficient implementation of one- and two-component analytical energy gradients in exact two-component theory." *J. Chem. Phys.* **2018**, *148*, 104110. DOI: 10.1063/1.5022153.

[322] E. van Lenthe, A. Ehlers, and E.-J. Baerends. "Geometry optimizations in the zero order regular approximation for relativistic effects." *J. Chem. Phys.* **1999**, *110*, 8943–8953. DOI: 10.1063/1.478813.

[323] J. D. Queen, A. C. Phung, C. A. Caputo, J. C. Fettinger, and P. P. Power. "Metathetical Exchange between Metal–Metal Triple Bonds." *J. Am. Chem. Soc.* **2020**, *142*, 2233–2237. DOI: 10.1021/jacs.9b13604.

[324] T. Munguia, I. S. Pavel, R. N. Kapoor, F. Cervantes-Lee, L. Párkányi, and K. H. Pannell. "Lewis acidity of group 14 elements toward intramolecular sulfur in ortho-aryl-thioanisoles." *Can. J. Chem.* **2003**, *81*, 1388–1397. DOI: 10.1139/v03-163.

[325] M. Kira, R. Yauchibara, R. Hirano, C. Kabuto, and H. Sakurai. "Synthesis and X-ray Structure of the First Dicoordinate Dialkylstannylene That Is Monomeric in the Solid State." *J. Am. Chem. Soc.* **1991**, *113*, 7785–7787. DOI: 10.1021/ja00020a064.

[326] C. Eaborn, M. S. Hill, P. B. Hitchcock, D. Patel, J. D. Smith, and S. Zhang. "Oxidative Addition to a Monomeric Stannylene To Give Four-Coordinate Tin Compounds Containing the Bulky Bidentate Ligand $C(SiMe_3)_2SiMe_2CH_2CH_2Me_2Si(Me_3Si)_2C$. Crystal Structures of $CH_2Me_2Si(Me_3Si)_2CSnC(SiMe_3)_2SiMe_2CH_2$, $CH_2Me_2Si(Me_3Si)_2CSnMe(OCOCF_3)C(SiMe_3)_2SiMe_2CH_2$, and $(CF_3COO)_2MeSnC(SiMe_3)_2SiMe_2CH_2CH_2Me_2Si(Me_3Si)_2CSnMe(OCOCF_3)_2$." *Organometallics* **2000**, *19*, 49–53. DOI: 10.1021/om990779p.

[327] G. H. Spikes, Y. Peng, J. C. Fettinger, and P. P. Power.
"Synthesis and Characterization of the Monomeric Sterically Encumbered Diaryls E{C_6H_3-2,6-(C_6H_3-2,6-Pr^i_2)$_2$}$_2$ (E = Ge, Sn, or Pb)."
Z. Anorg. Allg. Chem. **2006**, *632*, 1005–1010. DOI: 10.1002/zaac.200500532.

[328] J. D. Erickson, R. D. Riparetti, J. C. Fettinger, and P. P. Power.
"Molecular Zinc Species with Ge–Zn and Sn–Zn Bonds: A Reversible Insertion of a Stannylene into a Zinc–Carbon Bond." *Organometallics* **2016**, *35*, 2124–2128.
DOI: 10.1021/acs.organomet.6b00344.

[329] T. Kuwabara, J. D. Guo, S. Nagase, and M. Saito.
"Diversity of the Structures in a Distannene Complex and its Reduction to Generate a Six-Membered Ti_2Sn_4 Ring Complex." *Angew. Chem. Int. Ed.* **2014**, *53*, 434–438.
DOI: 10.1002/anie.201308565.

[330] S. Ito, T. Kuwabara, and Y. Ishii. "A Tin Analogue of the Cycloheptatrienyl Anion: Synthesis, Structure, and Further Reduction to Form a Dianionic Species."
Organometallics **2020**, *39*, 640–644. DOI: 10.1021/acs.organomet.0c00042.

[331] B. D. Rekken, T. M. Brown, J. C. Fettinger, F. Lips, H. M. Tuononen, R. H. Herber, and P. P. Power. "Dispersion Forces and Counterintuitive Steric Effects in Main Group Molecules: Heavier Group 14 (Si–Pb) Dichalcogenolate Carbene Analogues with Sub-90° Interligand Bond Angles." *J. Am. Chem. Soc.* **2013**, *135*, 10134–10148.
DOI: 10.1021/ja403802a.

[332] I. Objartel, H. Ott, and D. Stalke. "Low-Temperature NMR and Crystal Structure Analyses of a Hemilabile Tin Complex." *Z. Anorg. Allg. Chem.* **2008**, *634*, 2373–2379.
DOI: 10.1002/zaac.200800294.

[333] A. C. Filippou, D. Hoffmann, and G. Schnakenburg.
"Triple bonds of niobium with silicon, germaniun and tin: the tetrylidyne complexes [(κ^3-tmps)$(CO)_2$Nb≡E–R] (E = Si, Ge, Sn; tmps = $MeSi(CH_2PMe_2)_3$; R = aryl)."
Chem. Sci. **2017**, *8*, 6290–6299. DOI: 10.1039/c7sc02708g.

[334] V. G. Kumar Das, L. Kong Mun, C. Wei, and T. C. W. Mak. "Synthesis, Spectroscopic Study, and X-ray Crystal Structure of Bis[3-(2-pyridyl)-2-thienyl-*C*,*N*]diphenyltin(IV): The First Example of a Six-Coordinate Tetraorganotin Compound."
Organometallics **1987**, *6*, 10–14. DOI: 10.1021/om00144a003.

[335] H. Grützmacher, H. Pritzkow, and F. T. Edelmann.
"Synthesis and Structure of a Monomeric Diarylstannylene."
Organometallics **1991**, *10*, 23–25. DOI: 10.1021/om00047a013.

[336] M. Saito, S. Imaizumi, T. Tajima, K. Ishimura, and S. Nagase.
"Synthesis and Structure of Pentaorganostannate Having Five Carbon Substituents."
J. Am. Chem. Soc. **2007**, *129*, 10974–10975. DOI: 10.1021/ja072478+.

[337] A. C. Filippou, P. Ghana, U. Chakraborty, and G. Schnakenburg.
"Manganese–Tin Triple Bonds: A New Synthetic Route to the Manganese Stannylidyne Complex Cation *trans*-[H(dmpe)$_2$Mn≡Sn(C$_6$H$_3$-2,6-Mes$_2$)]$^+$ (dmpe = Me$_2$PCH$_2$CH$_2$PMe$_2$, Mes = 2,4,6-Trimethylphenyl)."
J. Am. Chem. Soc. **2013**, *135*, 11525–11528. DOI: 10.1021/ja406290t.

[338] N. Chiorean, C. Coza, A. Pop, and A. Silvestru.
"New hypercoordinated diorganotin(IV) compounds with dithiocarbamato ligands. Synthesis and structural characterization." *J. Organomet. Chem.* **2019**, *880*, 83–90.
DOI: 10.1016/j.jorganchem.2018.10.030.

[339] W. D. Woodul, A. F. Richards, A. Stasch, M. Driess, and C. Jones.
"N-Heterocyclic Germylidenide and Stannylidenide Anions: Group 14 Metal(II) Cyclopentadienide Analogues." *Organometallics* **2010**, *29*, 3655–3660.
DOI: 10.1021/om100595a.

[340] F. Stella, C. Marschner, and J. Baumgartner.
"Incorporating Methyl and Phenyl Substituted Stannylene Units into Oligosilanes. The Influence on Optical Absorption Properties." *Molecules* **2017**, *22*, 2212.
DOI: 10.3390/molecules22122212.

[341] C. Kaiya, K. Suzuki, and M. Yamashita.
"A Monomeric Stannabenzene: Synthesis, Structure, and Electronic Properties."
Angew. Chem. Int. Ed. **2019**, *58*, 7749–7752. DOI: 10.1002/anie.201902639.

[342] T. Kuwabara, J.-D. Guo, S. Nagase, M. Minoura, R. H. Herber, and M. Saito.
"Enhancement of Stannylene Character in Stannole Dianion Equivalents Evidenced by NMR and Mössbauer Spectroscopy and Theoretical Studies of Newly Synthesized Silyl-Substituted Dilithiostannoles." *Organometallics* **2014**, *33*, 2910–2913.
DOI: 10.1021/om5003717.

[343] A. Hinz.
"Pseudo-One-Coordinate Tetrylenium Salts Bearing a Bulky Carbazolyl Substituent."
Chem. Eur. J. **2019**, *25*, 3267–3271. DOI: 10.1002/chem.201806346.

[344] A. C. Filippou, P. Portius, A. I. Philippopoulos, and H. Rohde.
"Triple Bonding to Tin: Synthesis and Characterization of the Stannylyne Complex *trans*-[Cl(PMe$_3$)$_4$W≡Sn−C$_6$H$_3$-2,6-Mes$_2$]." *Angew. Chem. Int. Ed.* **2003**, *42*, 445–447.
DOI: 10.1002/anie.200390135.

[345] B. Glowacki, M. Lutter, D. Schollmeyer, W. Hiller, and K. Jurkschat.
"Novel Stannatrane N(CH$_2$CMe$_2$O)$_2$(CMe$_2$CH$_2$O)SnO-*t*-Bu and Related Oligonuclear Tin(IV) Oxoclusters. Two Isomers in One Crystal."
Inorg. Chem. **2016**, *55*, 10218–10228. DOI: 10.1021/acs.inorgchem.6b01429.

[346] T. Kuwabara, J.-D. Guo, S. Nagase, T. Sasamori, N. Tokitoh, and M. Saito. "Synthesis, Structures, and Electronic Properties of Triple- and Double-Decker Ruthenocenes Incorporated by a Group 14 Metallole Dianion Ligand." *J. Am. Chem. Soc.* **2014**, *136*, 13059–13064. DOI: 10.1021/ja507330p.

[347] M. Aman, L. Dostál, Z. Růžičková, S. Mebs, J. Beckmann, and R. Jambor. "Ambiguous Role of N → Sn Coordinated Stannylene: Lewis Base or Acid?" *Organometallics* **2019**, *38*, 816–828. DOI: 10.1021/acs.organomet.8b00816.

[348] C. P. Sindlinger, A. Stasch, H. F. Bettinger, and L. Wesemann. "A nitrogen-base catalyzed generation of organotin(II) hydride from an organotin trihydride under reductive dihydrogen elimination." *Chem. Sci.* **2015**, *6*, 4737–4751. DOI: 10.1039/c5sc01561h.

[349] W. A. Merrill, R. J. Wright, C. S. Stanciu, M. M. Olmstead, J. C. Fettinger, and P. P. Power. "Synthesis and Structural Characterization of a Series of Dimeric Metal(II) Imido Complexes $\{M(\mu\text{-}NAr^{\#})\}_2$ [M = Ge, Sn, Pb; $Ar^{\#} = C_6H_3$-2,6-$(C_6H_2$-2,4,6-$Me_3)_2$] and the Related Monomeric Primary Amido Derivatives $M\{N(H)Ar^{\#}\}_2$ (M = Ge, Sn, Pb): Spectroscopic Manifestations of Secondary Metal–Ligand Interactions." *Inorg. Chem.* **2010**, *49*, 7097–7105. DOI: 10.1021/ic100831c.

[350] A. Pop, L. Wang, V. Dorcet, T. Roisnel, J.-F. Carpentier, A. Silvestru, and Y. Sarazin. "On the coordination chemistry of organochalcogenolates $R^{NMe2\wedge}E^{-}$ and $R^{NMe2\wedge}E^{\wedge}O^{-}$ (E = S, Se) onto lead(II) and lighter divalent tetrel elements." *Dalton Trans.* **2014**, *43*, 16459–16474. DOI: 10.1039/c4dt02252a.

[351] M. Novák, M. Bouška, L. Dostál, A. Růžička, A. Hoffmann, S. Herres-Pawlis, and R. Jambor. "Less Is More: Three-Coordinate C,N-Chelated Distannynes and Digermynes." *Chem. Eur. J.* **2015**, *21*, 7820–7829. DOI: 10.1002/chem.201500695.

[352] C. Förster, P. M. Becker, and K. Heinze. "A Ferrocenyl Amino Substituted Stannylene as an Intramolecular Fe→Sn Lewis Adduct." *Z. Anorg. Allg. Chem.* **2018**, *644*, 1057–1063. DOI: 10.1002/zaac.201800269.

[353] A. P. Singh, H. W. Roesky, E. Carl, D. Stalke, J.-P. Demers, and A. Lange. "Lewis Base Mediated Autoionization of $GeCl_2$ and $SnCl_2$." *J. Am. Chem. Soc.* **2012**, *134*, 4998–5003. DOI: 10.1021/ja300563g.

[354] M. Veith, M. Gasthauer, M. Zimmer, and V. Huch. "Zur Bildung einer Zinn-Zinn-Bindung in Bis(hexamethyldisilazyl)-Zinn(II) unter C-H Bindungsbruch." *Z. Anorg. Allg. Chem.* **2007**, *633*, 2274–2277. DOI: 10.1002/zaac.200700156.

[355] R. K. Raut, P. Sahoo, D. Chimnapure, and M. Majumdar. "Versatile coordinating abilities of acyclic N_4 and N_2P_2 ligand frameworks in conjunction with $Sn[N(SiMe_3)_2]_2$." *Dalton Trans.* **2019**, *48*, 10953–10961. DOI: 10.1039/c9dt00617f.

[356] G. Barone, T. G. Hibbert, M. F. Mahon, K. C. Molloy, I. P. Parkin, L. S. Price, and I. Silaghi-Dumitrescu. "Structural distortions in homoleptic $(RE)_4A$ (E = O, S, Se; A = C, Si, Ge, Sn): implications for the CVD of tin sulfides." *J. Chem. Soc., Dalton Trans.* **2001**, 3435–3445. DOI: 10.1039/b010157p.

[357] J. K. West and L. Stahl. "Reactions of Germylenes and Stannylenes with Halo(hydrocarbyl)- and Chloro(amino)phosphines: Oxidative Addition versus Ligand Transfer." *Inorg. Chem.* **2017**, *56*, 12728–12738. DOI: 10.1021/acs.inorgchem.7b01275.

[358] D. Dakternieks, H. Zhu, D. Masi, and C. Mealli. "Stereochemistry of hypervalent tin(IV) compounds. NMR and crystallographic studies of organoyltin(IV) complexes with 1,1-dithiolate ligands." *Inorg. Chim. Acta* **1993**, *211*, 155–160. DOI: 10.1016/S0020-1693(00)85595-4.

[359] T. Kuwabara, M. Nakada, J. Hamada, J. D. Guo, S. Nagase, and M. Saito. "(η^4-Butadiene)Sn(0) Complexes: A New Approach for Zero-Valent p-Block Elements Utilizing a Butadiene as a 4π-Electron Donor." *J. Am. Chem. Soc.* **2016**, *138*, 11378–11382. DOI: 10.1021/jacs.6b07304.

[360] J. J. Burke and P. C. Lauterbur. "Sn^{119} Nuclear Magnetic Resonance Spectra." *J. Am. Chem. Soc.* **1961**, *83*, 326–331. DOI: 10.1021/ja01463a016.

[361] L. Wirtz, M. Jourdain, V. Huch, M. Zimmer, and A. Schäfer. "Synthesis, Structure, and Reactivity of Disiloxa[3]tetrelocenophanes." *ACS Omega* **2019**, *4*, 18355–18360. DOI: 10.1021/acsomega.9b02605.

[362] R. R. Holmes, S. Shafieezad, J. M. Holmes, and R. O. Day. "Square-Pyramidal and Trigonal-Bipyramidal Anionic (Ethanedithiolato)stannates. Binuclear Centers with Tin–Sulfur Bonding." *Inorg. Chem.* **1988**, *27*, 1232–1237. DOI: 10.1021/ic00280a028.

[363] S. P. Constantine, H. Cox, P. B. Hitchcock, and G. A. Lawless. "Parallel Metallocenes of Germanium, Tin, and Lead." *Organometallics* **2000**, *19*, 317–326. DOI: 10.1021/om990884z.

[364] D. S. Wishart and D. A. Case. "Use of Chemical Shifts in Macromolecular Structure Determination." *Meth. Enzymol.* **2002**, *338*, 3–34. DOI: 10.1016/S0076-6879(02)38214-4.

[365] C. S. Yannoni, P. P. Bernier, D. S. Bethune, G. Meijer, and J. R. Salem. "NMR Determination of the Bond Lengths in C_{60}." *J. Am. Chem. Soc.* **1991**, *113*, 3190–3192. DOI: 10.1021/ja00008a068.

[366] M. Balodis, M. Cordova, A. Hofstetter, G. M. Day, and L. Emsley. "*De Novo* Crystal Structure Determination from Machine Learned Chemical Shifts." *J. Am. Chem. Soc.* **2022**, *144*, 7215–7223. DOI: 10.1021/jacs.1c13733.

[367] P. Schienbein. "Spectroscopy from Machine Learning by Accurately Representing the Atomic Polar Tensor." *J. Chem. Theory Comput.* **2023**, *19*, 705–712. DOI: 10.1021/acs.jctc.2c00788.

[368] H. Han and S. Choi. "Transfer Learning from Simulation to Experimental Data: NMR Chemical Shift Predictions." *J. Phys. Chem. Lett.* **2021**, *12*, 3662–3668. DOI: 10.1021/acs.jpclett.1c00578.

[369] E. A. Engel, V. Kapil, and M. Ceriotti. "Importance of Nuclear Quantum Effects for NMR Crystallography." *J. Phys. Chem. Lett.* **2021**, *12*, 7701–7707. DOI: 10.1021/acs.jpclett.1c01987.

[370] R. J. Iuliucci, J. D. Hartman, and G. J. O. Beran. "Do Models beyond Hybrid Density Functionals Increase the Agreement with Experiment for Predicted NMR Chemical Shifts or Electric Field Gradient Tensors in Organic Solids?" *J. Phys. Chem. A* **2023**, *127*, 2846–2858. DOI: 10.1021/acs.jpca.2c07657.

[371] T. Kupka, M. Stachów, M. Nieradka, J. Kaminsky, T. Pluta, and S. P. A. Sauer. "From CCSD(T)/aug-cc-pVTZ-J to CCSD(T) complete basis set limit isotropic nuclear magnetic shieldings via affordable DFT/CBS calculations." *Magn. Reson. Chem.* **2011**, *49*, 231–236. DOI: 10.1002/mrc.2738.

[372] T. Kupka, B. Ruscic, and R. E. Botto. "Toward Hartree–Fock- and Density Functional Complete Basis-Set-Predicted NMR Parameters." *J. Phys. Chem. A* **2002**, *106*, 10396–10407. DOI: 10.1021/jp020987m.

[373] T. Kupka. "H_2O, H_2, HF, F_2 and F_2O nuclear magnetic shielding constants and indirect nuclear spin–spin coupling constants (SSCCs) in the BHandH/pcJ-*n* and BHandH/XZP Kohn–Sham limits." *Magn. Reson. Chem.* **2009**, *47*, 959–970. DOI: 10.1002/mrc.2500.

[374] S. Moon and D. A. Case. "A Comparison of Quantum Chemical Models for Calculating NMR Shielding Parameters in Peptides: Mixed Basis Set and ONIOM Methods Combined with a Complete Basis Set Extrapolation." *J. Comput. Chem.* **2006**, *27*, 825–836. DOI: 10.1002/jcc.20388.

[375] T. Kupka and C. Lim. "Polarization-Consistent versus Correlation-Consistent Basis Sets in Predicting Molecular and Spectroscopic Properties." *J. Phys. Chem. A* **2007**, *111*, 1927–1932. DOI: 10.1021/jp065008v.

[376] T. Kupka, M. Stachów, M. Nieradka, J. Kaminsky, and T. Pluta. "Convergence of Nuclear Magnetic Shieldings in the Kohn–Sham Limit for Several Small Molecules." *J. Chem. Theory Comput.* **2010**, *6*, 1580–1589. DOI: 10.1021/ct100109j.

[377] D. M. Reid and M. A. Collins. "Approximating CCSD(T) Nuclear Magnetic Shielding Calculations Using Composite Methods." *J. Chem. Theory Comput.* **2015**, *11*, 5177–5181. DOI: 10.1021/acs.jctc.5b00546.

[378] S. Grimme, J. Antony, S. Ehrlich, and H. Krieg.
"A consistent and accurate *ab initio* parametrization of density functional dispersion correction (DFT-D) for the 94 elements H-Pu." *J. Chem. Phys.* **2010**, *132*, 154104.
DOI: 10.1063/1.3382344.

[379] J. Behler. "Atom-centered symmetry functions for constructing high-dimensional neural network potentials." *J. Chem. Phys.* **2011**, *134*, 074106.
DOI: 10.1063/1.3553717.

[380] A. P. Bartók, R. Kondor, and G. Csányi. "On representing chemical environments." *Phys. Rev. B* **2013**, *87*, 184115. DOI: 10.1103/PhysRevB.87.184115.

[381] J. Mason. "Conventions for the reporting of nuclear magnetic shielding (or shift) tensors suggested by participants in the NATO ARW on NMR Shielding Constants at the University of Maryland, College Park, July 1992."
Solid State Nucl. Magn. Reson. **1993**, *2*, 285–288.
DOI: 10.1016/0926-2040(93)90010-K.

[382] J. C. Facelli.
"Chemical shift tensors: Theory and application to molecular structural problems."
Prog. Nucl. Magn. Reson. Spectrosc. **2011**, *58*, 176–201.
DOI: 10.1016/j.pnmrs.2010.10.003.

[383] J. Řezáč, K. E. Riley, and P. Hobza. "S66: A Well-balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures."
J. Chem. Theory Comput. **2011**, *7*, 2427–2438. DOI: 10.1021/ct2002946.

[384] A. L. Ringer, M. S. Figgs, M. O. Sinnokrot, and C. D. Sherrill. "Aliphatic C–H/π Interactions: Methane–Benzene, Methane–Phenol, and Methane–Indole Complexes."
J. Phys. Chem. A **2006**, *110*, 10822–10828. DOI: 10.1021/jp062740l.

[385] J. Gauss and J. F. Stanton. "Gauge-invariant calculation of nuclear magnetic shielding constants at the coupled–cluster singles and doubles level."
J. Chem. Phys. **1995**, *102*, 251–253. DOI: 10.1063/1.469397.

[386] J. Gauss and J. F. Stanton.
"Coupled-cluster calculations of nuclear magnetic resonance chemical shifts."
J. Chem. Phys. **1995**, *103*, 3561–3577. DOI: 10.1063/1.470240.

[387] F. Neese. "Software update: The ORCA program system–Version 5.0."
WIREs Comput. Mol. Sci. **2022**, 1–15. DOI: 10.1002/wcms.1606.

[388] Y. Tawada, T. Tsuneda, S. Yanagisawa, T. Yanai, and K. Hirao.
"A long-range-corrected time-dependent density functional theory."
J. Chem. Phys. **2004**, *120*, 8425–8433. DOI: 10.1063/1.1688752.

[389] N. Mardirossian and M. Head-Gordon. "ωB97X-V: A 10-parameter, range-separated hybrid, generalized gradient approximation density functional with nonlocal correlation, designed by a survival-of-the-fittest strategy."
Phys. Chem. Chem. Phys. **2014**, *16*, 9904–9924. DOI: 10.1039/c3cp54374a.

[390] D. E. Woon and T. H. Dunning. "Gaussian basis sets for use in correlated molecular calculations. III. The atoms aluminum through argon."
J. Chem. Phys. **1993**, *98*, 1358–1371. DOI: 10.1063/1.464303.

[391] F. Weigend, M. Kattannek, and R. Ahlrichs. "Approximated electron repulsion integrals: Cholesky decomposition versus resolution of the identity methods."
J. Chem. Phys. **2009**, *130*, 164106. DOI: 10.1063/1.3116103.

[392] F. Neese, F. Wennmohs, A. Hansen, and U. Becker.
"Efficient, approximate and parallel Hartree–Fock and hybrid DFT calculations. A 'chain-of-spheres' algorithm for the Hartree–Fock exchange."
Chem. Phys. **2009**, *356*, 98–109. DOI: 10.1016/j.chemphys.2008.10.036.

[393] S. Kossmann and F. Neese.
"Comparison of two efficient approximate Hartee–Fock approaches."
Chem. Phys. Lett. **2009**, *481*, 240–243. DOI: 10.1016/j.cplett.2009.09.073.

[394] G. L. Stoychev, A. A. Auer, and F. Neese.
"Automatic Generation of Auxiliary Basis Sets."
J. Chem. Theory Comput. **2017**, *13*, 554–562. DOI: 10.1021/acs.jctc.6b01041.

[395] M. Kállay and J. Gauss. "Analytic second derivatives for general coupled-cluster and configuration-interaction models." *J. Chem. Phys.* **2004**, *120*, 6841–6848.
DOI: 10.1063/1.1668632.

[396] M. E. Harding, T. Metzroth, J. Gauss, and A. A. Auer.
"Parallel Calculation of CCSD and CCSD(T) Analytic First and Second Derivatives."
J. Chem. Theory Comput. **2008**, *4*, 64–74. DOI: 10.1021/ct700152c.

[397] D. A. Matthews, L. Cheng, M. E. Harding, F. Lipparini, S. Stopkowicz, T.-C. Jagau, P. G. Szalay, J. Gauss, and J. F. Stanton. "Coupled-cluster techniques for computational chemistry: The CFOUR program package."
J. Chem. Phys. **2020**, *152*, 214108. DOI: 10.1063/5.0004837.

[398] CFOUR – *Coupled-Cluster techniques for Computational Chemistry*, Version 2.1, University of Florida, Gainesville, FL 32611, USA, Universität Mainz, Mainz, Germany **2019**, https://cfour.uni-mainz.de/cfour.

[399] B. P. Pritchard, D. Altarawy, B. Didier, T. D. Gibson, and T. L. Windus.
"New Basis Set Exchange: An Open, Up-to-Date Resource for the Molecular Sciences Community." *J. Chem. Inf. Model.* **2019**, *59*, 4814–4820.
DOI: 10.1021/acs.jcim.9b00725.

[400] J. Li, J.-K. Liu, and W.-X. Wang. "GIAO ^{13}C NMR Calculation with Sorted Training Sets Improves Accuracy and Reliability for Structural Assignation."
J. Org. Chem. **2020**, *85*, 11350–11358. DOI: 10.1021/acs.joc.0c01451.

[401] F. Bohle and S. Grimme.
"Hydrocarbon Macrocycle Conformer Ensembles and ^{13}C-NMR Spectra."
Angew. Chem. Int. Ed. **2022**, *61*, e202113905. DOI: 10.1002/anie.202113905.

[402] K. Jackowski. "Multinuclear NMR spectroscopy in the gas phase." *J. Mol. Struct.* **2006**, *786*, 215–219. DOI: 10.1016/j.molstruc.2005.10.018.

[403] T. Zuschneid, H. Fischer, T. Handel, K. Albert, and G. Häfelinger. "Experimental Gas Phase ^{1}H NMR Spectra and Basis Set Dependence of *ab initio* GIAO MO Calculations of ^{1}H and ^{13}C NMR Absolute Shieldings and Chemical Shifts of Small Hydrocarbons." *Z. Naturforsch. B* **2004**, *59*, 1153–1176. DOI: 10.1515/znb-2004-1012.

[404] A.-H. Emwas, R. Roy, R. T. McKay, L. Tenori, E. Saccenti, G. A. Nagana Gowda, D. Raftery, F. Alahmari, L. Jaremko, M. Jaremko, and D. S. Wishart. "NMR Spectroscopy for Metabolomics Research." *Metabolites* **2019**, *9*, 123. DOI: 10.3390/metabo9070123.

[405] D. J. Kubicki, S. D. Stranks, C. P. Grey, and L. Emsley. "NMR spectroscopy probes microstructure, dynamics and doping of metal halide perovskites." *Nat. Rev. Chem.* **2021**, *5*, 624–645. DOI: 10.1038/s41570-021-00309-x.

[406] L. B. Krivdin. "Computational ^{1}H NMR: Part 1. Theoretical background." *Magn. Reson. Chem.* **2019**, *57*, 897–914. DOI: 10.1002/mrc.4873.

[407] L. B. Krivdin. "Computational protocols for calculating ^{13}C NMR chemical shifts." *Prog. Nucl. Magn. Reson. Spectrosc.* **2019**, *112-113*, 103–156. DOI: 10.1016/j.pnmrs.2019.05.004.

[408] S. Moncho and J. Autschbach. "Relativistic Zeroth-Order Regular Approximation Combined with Nonhybrid and Hybrid Density Functional Theory: Performance for NMR Indirect Nuclear Spin–Spin Coupling in Heavy Metal Compounds." *J. Chem. Theory Comput.* **2010**, *6*, 223–234. DOI: 10.1021/ct900535d.

[409] D. L. Crittenden. "A new double-reference correction scheme for accurate and efficient computation of NMR chemical shieldings." *Phys. Chem. Chem. Phys.* **2022**, *24*, 27055–27063. DOI: 10.1039/D2CP03992C.

[410] A. G. Kutateladze and D. S. Reddy. "High-Throughput in Silico Structure Validation and Revision of Halogenated Natural Products Is Enabled by Parametric Corrections to DFT-Computed ^{13}C NMR Chemical Shifts and Spin–Spin Coupling Constants." *J. Org. Chem.* **2017**, *82*, 3368–3381. DOI: 10.1021/acs.joc.7b00188.

[411] J. S. Smith, O. Isayev, and A. E. Roitberg. "ANI-1: an extensible neural network potential with DFT accuracy at force field computational cost." *Chem. Sci.* **2017**, *8*, 3192–3203. DOI: 10.1039/C6SC05720A.

[412] T. Zubatiuk, B. Nebgen, N. Lubbers, J. S. Smith, R. Zubatyuk, G. Zhou, C. Koh, K. Barros, O. Isayev, and S. Tretiak. "Machine learned Hückel theory: Interfacing physics and deep neural networks." *J. Chem. Phys.* **2021**, *154*, 244108. DOI: 10.1063/5.0052857.

[413] D. Chen, Z. Wang, D. Guo, V. Orekhov, and X. Qu. "Review and Prospect: Deep Learning in Nuclear Magnetic Resonance Spectroscopy." *Chem. Eur. J.* **2020**, *26*, 10391–10401. DOI: 10.1002/chem.202000246.

[414] A. Gupta, S. Chakraborty, and R. Ramakrishnan. "Revving up ^{13}C NMR shielding predictions across chemical space: benchmarks for atoms-in-molecules kernel machine learning with new data for 134 kilo molecules." *Mach. Learn.: Sci. Technol.* **2021**, *2*, 035010. DOI: 10.1088/2632-2153/abe347.

[415] S. G. Smith and J. M. Goodman. "Assigning Stereochemistry to Single Diastereoisomers by GIAO NMR Calculation: The DP4 Probability." *J. Am. Chem. Soc.* **2010**, *132*, 12946–12959. DOI: 10.1021/ja105035r.

[416] M. Ruth, D. Gerbig, and P. R. Schreiner. "Machine Learning for Bridging the Gap between Density Functional Theory and Coupled Cluster Energies." *J. Chem. Theory Comput.* **2023**, *19*, 4912–4920. DOI: 10.1021/acs.jctc.3c00274.

[417] A. Bagno and G. Saielli. "Computational NMR spectroscopy: reversing the information flow." *Theor. Chem. Acc.* **2007**, *117*, 603–619. DOI: 10.1007/s00214-006-0196-z.

[418] T. E. Field-Theodore, M. Olejniczak, M. Jaszuński, and D. J. D. Wilson. "NMR shielding constants in group 15 trifluorides." *Phys. Chem. Chem. Phys.* **2018**, *20*, 23025–23033. DOI: 10.1039/C8CP04056G.

[419] A. Bagno, F. Rastrelli, and G. Saielli. "Predicting ^{13}C NMR Spectra by DFT Calculations." *J. Phys. Chem. A* **2003**, *107*, 9964–9973. DOI: 10.1021/jp0353284.

[420] C. Adamo and V. Barone. "Toward chemical accuracy in the computation of NMR shieldings: the PBE0 model." *Chem. Phys. Lett.* **1998**, *298*, 113–119. DOI: 10.1016/S0009-2614(98)01201-9.

[421] S. Spicher and S. Grimme. "Single-Point Hessian Calculations for Improved Vibrational Frequencies and Rigid-Rotor-Harmonic-Oscillator Thermodynamics." *J. Chem. Theory Comput.* **2021**, *17*, 1701–1714. DOI: 10.1021/acs.jctc.0c01306.

[422] S. Sinnecker, A. Rajendran, A. Klamt, M. Diedenhofen, and F. Neese. "Calculation of Solvent Shifts on Electronic g-Tensors with the Conductor-Like Screening Model (COSMO) and Its Self-Consistent Generalization to Real Solvents (Direct COSMO-RS)." *J. Phys. Chem. A* **2006**, *110*, 2235–2245. DOI: 10.1021/jp056016z.

[423] R. Ahlrichs and K. May. "Contracted all-electron Gaussian basis sets for atoms Rb to Xe." *Phys. Chem. Chem. Phys.* **2000**, *2*, 943–945. DOI: 10.1039/A908859H.

[424] J. D. Rolfes, F. Neese, and D. A. Pantazis. "All-electron scalar relativistic basis sets for the elements Rb–Xe." *J. Comput. Chem.* **2020**, *41*, 1842–1849. DOI: 10.1002/jcc.26355.

[425] C. Adamo and V. Barone.
"Exchange functionals with improved long-range behavior and adiabatic connection methods without adjustable parameters: The *m*PW and *m*PW1PW models."
J. Chem. Phys. **1998**, *108*, 664–675. DOI: 10.1063/1.475428.

[426] D. J. Giesen and N. Zumbulyadis. "A hybrid quantum mechanical and empirical model for the prediction of isotropic ^{13}C shielding constants of organic molecules."
Phys. Chem. Chem. Phys. **2002**, *4*, 5498–5507. DOI: 10.1039/B206245C.

[427] P. d'Antuono, E. Botek, B. Champagne, M. Spassova, and P. Denkova.
"Theoretical investigation on ^{1}H and ^{13}C NMR chemical shifts of small alkanes and chloroalkanes." *J. Chem. Phys.* **2006**, *125*, 144309. DOI: 10.1063/1.2353830.

[428] N. Grimblat, M. M. Zanardi, and A. M. Sarotti.
"Beyond DP4: an Improved Probability for the Stereochemical Assignment of Isomeric Compounds using Quantum Chemical Calculations of NMR Shifts."
J. Org. Chem. **2015**, *80*, 12526–12534. DOI: 10.1021/acs.joc.5b02396.

[429] A. J. Groszek. "Selective adsorption at graphite/hydrocarbon interfaces."
Proc. Roy. Soc. Lond. A **1970**, *314*, 473–498. DOI: 10.1098/rspa.1970.0019.

[430] G. C. McGonigal, R. H. Bernhardt, and D. J. Thomson. "Imaging alkane layers at the liquid/graphite interface with the scanning tunneling microscope."
Appl. Phys. Lett. **1990**, *57*, 28–30. DOI: 10.1063/1.104234.

[431] J. P. Rabe and S. Buchholz. "Commensurability and Mobility in Two-Dimensional Molecular Patterns on Graphite." *Science* **1991**, *253*, 424–427.
DOI: 10.1126/science.253.5018.424.

[432] T. Yang, S. Berber, J.-F. Liu, G. P. Miller, and D. Tománek.
"Self-assembly of long chain alkanes and their derivatives on graphite."
J. Chem. Phys. **2008**, *128*, 124709. DOI: 10.1063/1.2841478.

[433] B. Ilan, G. M. Florio, M. S. Hybertsen, B. J. Berne, and G. W. Flynn.
"Scanning Tunneling Microscopy Images of Alkane Derivatives on Graphite: Role of Electronic Effects." *Nano Lett.* **2008**, *8*, 3160–3165. DOI: 10.1021/nl8014186.

[434] D. Grabowski, S. Alef, S. Becker, U. Müller, G. Schnakenburg, and S. Höger.
"Condensation of pyrylium salts with mixed anhydrides: aryl ethers, aryl amines and sterically congested aromatics." *Org. Chem. Front.* **2022**, *9*, 294–298.
DOI: 10.1039/D1QO01419F.

[435] S. Höger, K. Bonrad, A. Mourran, U. Beginn, and M. Möller.
"Synthesis, Aggregation, and Adsorption Phenomena of Shape-Persistent Macrocycles with Extraannular Polyalkyl Substituents."
J. Am. Chem. Soc. **2001**, *123*, 5651–5659. DOI: 10.1021/ja003990x.

[436] P. Wu, Q. Fan, G. Deng, Q. Zeng, C. Wang, and C. Bai. "Real Space Visualization of the Disklike Assembly Structure of Dendritic Molecules on Graphite."
Langmuir **2002**, *18*, 4342–4344. DOI: 10.1021/la0116138.

[437] J.-R. Gong, S.-B. Lei, L.-J. Wan, G.-J. Deng, Q.-H. Fan, and C.-L. Bai. "Structure and Dynamic Process of Two-Dimensional Monodendron Assembly." *Chem. Mater.* **2003**, *15*, 3098–3104. DOI: 10.1021/cm030289u.

[438] M. Fritzsche, A. Bohle, D. Dudenko, U. Baumeister, D. Sebastiani, G. Richardt, H. W. Spiess, M. R. Hansen, and S. Höger. "Empty Helical Nanochannels with Adjustable Order from Low-Symmetry Macrocycles." *Angew. Chem. Int. Ed.* **2011**, *50*, 3030–3033. DOI: 10.1002/anie.201007437.

[439] J. A. Platts and K. Gkionis. "NMR shielding as a probe of intermolecular interactions: *ab initio* and density functional theory studies." *Phys. Chem. Chem. Phys.* **2009**, *11*, 10331–10339. DOI: 10.1039/B822560E.

[440] Y. Wang, H. Xu, H. Wang, S. Li, W. Gan, and Q. Yuan. "Temperature dependent 2D self-assembled motif transition of copper–phthalocyanine derivates at air/HOPG interface: an STM study." *RSC Adv.* **2014**, *4*, 20256–20261. DOI: 10.1039/C3RA46651E.

[441] C.-J. Li, Q.-D. Zeng, Y.-H. Liu, L.-J. Wan, C. Wang, C.-R. Wang, and C.-L. Bai. "Evidence of a Thermal Annealing Effect on Organic Molecular Assembly." *ChemPhysChem* **2003**, *4*, 857–859. DOI: 10.1002/cphc.200200492.

[442] J. F. Dobson. "Alternative expressions for the Fermi hole curvature." *J. Chem. Phys.* **1993**, *98*, 8870–8872. DOI: 10.1063/1.464444.

[443] T. Zimmermann and G. W. Fischer. "Pyryliumverbindungen. 37. Arylbenzene aus 2,4,6-Triaryl-pyryliumsalzen und Carbonsäureanhydriden." *J. Prakt. Chem.* **1987**, *329*, 975–984. DOI: 10.1002/prac.19873290605.

[444] C. Mahler, U. Müller, W. M. Müller, V. Enkelmann, C. Moon, G. Brunklaus, H. Zimmermann, and S. Höger. "Synthesis of highly phenylene substituted *p*-phenylene oligomers from pyrylium salts." *Chem. Commun.* **2008**, 4816–4818. DOI: 10.1039/B807382A.

[445] A. Scarpaci, A. Nantalaksakul, J. M. Hales, J. D. Matichak, S. Barlow, M. Rumi, J. W. Perry, and S. R. Marder. "Effects of Dendronization on the Linear and Third-Order Nonlinear Optical Properties of Bis(thiopyrylium) Polymethine Dyes in Solution and the Solid State." *Chem. Mater.* **2012**, *24*, 1606–1618. DOI: 10.1021/cm3002139.

[446] T. Peterle, P. Ringler, and M. Mayor. "Gold Nanoparticles Stabilized by Acetylene-Functionalized Multidentate Thioether Ligands: Building Blocks for Nanoparticle Superstructures." *Adv. Funct. Mater.* **2009**, *19*, 3497–3506. DOI: 10.1002/adfm.200901410.

[447] Q. Zhou, P. J. Carroll, and T. M. Swager. "Synthesis of Diacetylene Macrocycles Derived from 1,2-Diethynylbenzene Derivatives: Structure and Reactivity of the Strained Cyclic Dimer." *J. Org. Chem.* **1994**, *59*, 1294–1301. DOI: 10.1021/jo00085a016.

[448] V. Percec, M. Peterca, Y. Tsuda, B. M. Rosen, S. Uchida, M. R. Imam, G. Ungar, and P. A. Heiney. "Elucidating the Structure of the $Pm\bar{3}n$ Cubic Phase of Supramolecular Dendrimers through the Modification of their Aliphatic to Aromatic Volume Ratio." *Chem. Eur. J.* **2009**, *15*, 8994–9004. DOI: 10.1002/chem.200901324.

[449] C. Sterzenbach, T. J. Keller, D. Kraus, J. M. Lupton, S.-S. Jester, and S. Höger. "Expanded all-phenylene molecular spoked wheels: cutouts of graphenylene-3." *Org. Chem. Front.* **2021**, *8*, 4980–4985. DOI: 10.1039/D1QO00876E.

[450] Z.-L. Zhou, L. Zhao, S. Zhang, K. Vincent, S. Lam, and D. Henze. "Facile Synthesis of Functionalized Bis(arylethynyl)benzene Derivatives via Sila–Sonogashira Reaction." *Synth. Commun.* **2012**, *42*, 1622–1631. DOI: 10.1080/00397911.2010.542538.

[451] F. Dumur, C. R. Mayer, E. Dumas, J. Marrot, and F. Sécheresse. "Synthesis of valuable terpyridine building blocks to generate a variety of metallodendrons by the convergent approach." *Tetrahedron Lett.* **2007**, *48*, 4143–4146. DOI: 10.1016/j.tetlet.2007.03.160.

List of Figures

1.1 The five main error sources in NMR calculations . . . 2

2.1 Example simulated ^{1}H NMR spectrum of chloroethane . . . 21
2.2 Schematic structure of a multilayer perceptron . . . 25

A.1 Overview of investigated ^{29}Si NMR chemical shifts . . . 57
A.2 Results for all investigated methods for the *SiS-L* subset . . . 58
A.3 Correlation plots applying a linear scaling correction to ^{29}Si NMR chemical shifts . . . 61
A.4 Results for all investigated basis sets for the *SiS-L* subset . . . 61
A.5 Selected molecules with heavy atoms in the vicinity of the Si atom . . . 63
A.6 Results for all investigated methods for the *SiS-H* subset . . . 65
A.7 Results for different geometry optimization levels for calculating ^{29}Si NMR chemical shifts . . . 66
A.8 Linear scaling approach to account for structural deficiencies for *SiS-L* and *SiS-H* . . . 67

B.1 Schematic protocol used for the SnS51 benchmark study . . . 77
B.2 Overview of investigated ^{119}Sn NMR chemical shifts . . . 78
B.3 Overview of all compounds included in the SnS51 set . . . 79
B.4 Results for all investigated methods for the SnS51 set . . . 82
B.5 Selected molecules from the SnS51 set and chemical shifts calculated at different levels . . . 83
B.6 Correlation plots applying a linear scaling correction to ^{119}Sn NMR chemical shifts . . . 85
B.7 Comparison of ensemble and lowest-conformer approaches within the SnS51 set . . . 86
B.8 Detailed analysis of the conformational space of compound **46** . . . 87
B.9 Results for different geometry optimization levels for calculating ^{119}Sn NMR chemical shifts . . . 88
B.10 Linear scaling approach to account for structural deficiencies for SnS51 . . . 90

C.1 Example for a geometrically distorted structure of compound **014** 97
C.2 Performance of the Δ_{corr}-ML method for the ^{1}H NMR test set 102
C.3 Performance of the Δ_{corr}-ML method for the ^{13}C NMR test set 104
C.4 Comparison of different *low-level* methods for use with Δ_{corr}-ML 106
C.5 Performance of the Δ_{corr}-ML method for the benchmark set of Flaig et al. . . . 110
C.6 Performance of the Δ_{corr}-ML method for the NS372 benchmark set 111
C.7 Performance of the Δ_{corr}-ML method for the NMRH148 benchmark set 112
C.8 Simulated NMR spectrum of butadiene corrected with Δ_{corr}-ML 113
C.9 Timing comparison of different NMR shielding calculations and the application of the Δ_{corr}-ML method 114

D.1 Key specifications of the data set of the Δ_{SO}-ML method 121
D.2 Workflow for training and applying the Δ_{SO}-ML method 123
D.3 Analysis of the ^{13}C NMR data set of the Δ_{SO}-ML method 126
D.4 Performance of the Δ_{SO}-ML method for the ^{13}C NMR test set 127
D.5 Performance of the Δ_{SO}-ML method for the ^{1}H NMR test set 129
D.6 Comparison of different *low-level* methods for use with Δ_{SO}-ML 132
D.7 Correlation between selected descriptors of the ^{13}C input feature vector of the Δ_{SO}-ML method 133
D.8 Performance of the Δ_{SO}-ML method for calculating ^{13}C NMR chemical shifts in the vicinity of Sn and Pb atoms 135
D.9 Selected molecules from the 17HAC set and components of the predicted ^{13}C NMR chemical shift 136
D.10 Computed molecular structure of the bismabenzene derivative and predicted $\Delta_{SO}\delta$ values 138
D.11 Timing comparison of different NMR shielding calculations and the application of the Δ_{SO}-ML method 139

E.1 Lewis structures of the discussed bicyclophanes 146
E.2 Synthesis of bicyclophanes **1a** and **1b** 147
E.3 ^{1}H NMR spectra of model compounds **7** and **8** and predicted chemical shift changes 148
E.4 Computed molecular structures of model compounds **7** and **8** and NICS values in **7** 149
E.5 STM images and proposed supramolecular structures of **1a**, **2b**, and **3d** adsorbed on HOPG 150
E.6 Computed molecular structure of **2b** adsorbed on a graphene sheet 152

List of Tables

A.1 List of all methods tested for computing ^{29}Si NMR chemical shifts 55
A.2 Statistics and linear scaling for all investigated methods for the *SiS-L* subset . 59
A.3 Statistics for all investigated methods for the *SiS-H* subset 64
A.4 Scaling parameters for ^{29}Si NMR for geometry optimizations with GFN2-xTB and GFN-FF 68

B.1 List of all methods tested for computing ^{119}Sn NMR chemical shifts 75
B.2 List of all compounds in the SnS51 set 80
B.3 Statistics and linear scaling for all investigated methods in the SnS51 set for lowest-conformer and ensemble approaches 84

C.1 Analysis of the importance of the different descriptor categories within the Δ_{corr}-ML method 108

D.1 Statistics of the performance of the Δ_{SO}-ML method for the test data set . . . 127
D.2 Comparison of different methods for calculating $\Delta_{SO}\delta$ 130
D.3 Statistics for applying both Δ-ML methods on the 17HAC benchmark set . . 137
D.4 Results for the Δ-ML approaches applied to the chemical shift calculation of a bismabenzene derivative 138

Acknowledgements

This thesis is the result of four years of work and since research is rarely done alone, it would not have come so far without the company and help of many people.

First, I want to thank my supervisor, Prof. Dr. Stefan Grimme, for the opportunity to work with him on such a fascinating topic and the freedom to shape the projects the way I wanted. I learned a lot from it and am happy that I could contribute to the first machine learning projects in the group. I am thankful for his professional and financial support and the possibility to visit international conferences and a summer school.

I am also grateful to Prof. Dr. Thomas Bredow for reviewing my thesis as second examiner as well as to Prof. Dr. Sigurd Höger and Prof. Dr. Eva Kiermaier for serving as chairperson and fourth member of my doctoral committee, respectively. Thank you all for your time despite being asked frequently from members of our group.

Moreover, I want to thank all my cooperation partners from inside and outside our group, in particular Dr. Markus Bursch, who collaborated with me on the majority of my projects and functioned as a mentor throughout my whole PhD, and Thomas Gasevic, who is now an expert in performing NMR benchmark studies. I very much appreciate that scientific collaborations play an important role in our group, as I have made only positive experiences with them. Therefore, further special thanks for fruitful collaborations go to Prof. Dr. Sigurd Höger, Prof. Dr. Matthias Wagner, Prof. Dr. Andreas Gansäuer, Prof. Dr. Peter Vöhringer, Ettore Bartalucci, Prof. Dr. Thomas Wiegand, as well as their research groups. In addition, I am grateful for two years of funding form the "Fonds der Chemischen Industrie (FCI)" through the Kekulé scholarship.

When it comes to this work, I want to thank Dr. Hagen Neugebauer, Sarah Löffelsender, Thomas Gasevic, and Dr. Kim Greis for valuable feedback after proofreading my thesis. During my time at the MCTC, I was accompanied by many colleagues that helped me develop both professionally and personally. It would have been half as enjoyable and productive without you. In particular, I want to thank my long-term fellow Dr. Hagen Neugebauer, who became a good friend, my inspiring former colleagues Dr. Jeroen Koopman and Dr. Sebatian Ehlert, as well as my office buddies Sarah Löffelsender and Abylay aka. Albert Katbashev. Beyond that, my thanks go to the Mulliken United pub quiz team and all faithful members of the Mate Train.

A special thank you goes to all other former and current group members for all scientific

and non-scientific conversations and activities. I explicitly want to thank Benedikt Bädorf, Fabian Bohle, Dr. Eike Caldeweyher, Marvin Friede, Thomas Froitzheim, Thomas Gasevic, Johannes Gorges, Dr. Andreas Hansen, Christian Hölzer, Julia Kohn, Lukas Kunze, Dr. Joachim Laun, Dr. Jan Mewes, Marcel Müller, Christoph Plett, Dr. Philipp Pracht, Thomas Rose, Dr. Jakob Seibert, Leopold Seidler, Dr. Sebastian Spicher, Marcel Stahn, and Lukas Wittmann. I also want to thank Jens Mekelburger and Claudia Kronz for help with organizational, IT, and administration questions.

The people outside academia were not less important in supporting me to finish this thesis. My largest gratitude goes to my wife, Lena Kleine Büning, whose persistent and loving encouragement for me was tremendous. I also want to thank my parents, other family members, and my godfather, Thomas Moser, who already requested a dedication in my dissertation when I was still in primary school. I am grateful for all my friends who have been with me since my studies in Berlin, my school time in Moers, or even earlier, and still are during my time in Bonn.

Thank you to everyone that I have mentioned or perhaps missed mentioning. You all contributed – at least indirectly – to this work.