Semiempirical Molecular Orbital Models based on the Neglect of Diatomic Differential Overlap Approximation

Tamara Husch, Alain C. Vaucher, and Markus Reiher*

ETH Zürich, Laboratorium für Physikalische Chemie, Vladimir-Prelog-Weg 2, 8093 Zürich, Switzerland.

July 30, 2018

Abstract

Semiempirical molecular orbital (SEMO) models based on the neglect of diatomic differential overlap (NDDO) approximation efficiently solve the self-consistent field equations by rather drastic approximations. The computational efficiency comes at the cost of an error in the electron-electron repulsion integrals. The error may be compensated by the introduction of parametric expressions to evaluate the electron-electron repulsion integrals, the one-electron integrals, and the core-core repulsion. We review the resulting formalisms of popular NDDO-SEMO models (such as the MNDO(/d), AM1, PMx, and OMx models) in a concise and self-contained manner. We discuss the approaches to implicitly and explicitly describe electron correlation effects within NDDO-SEMO models and we dissect strengths and weaknesses of the different approaches in a detailed analysis. For this purpose, we consider the results of recent benchmark studies. Furthermore, we apply bootstrapping to perform a sensitivity analysis for a selection of parameters in the MNDO model. We also identify systematic limitations of NDDO-SEMO models by drawing on an analogy to Kohn–Sham density functional theory.

*corresponding author: markus.reiher@phys.chem.ethz.ch.
1 Introduction

The driving force for the development of semiempirical molecular orbital (SEMO) models has always been the desire to accelerate quantum chemical calculations. At the outset of the development of SEMO models in the middle of the last century, the goal was to carry out electronic structure calculations for small molecules, which was not routinely possible with ab initio electronic structure methods at that time. Since then, theoretical chemistry has seen a remarkable development in terms of computational resources, but also in terms of ab initio methodology. One must not forget that most electronic structure methods which we apply routinely today, such as Kohn–Sham density functional theory (KS-DFT) and coupled cluster theory, were developed concurrently with today’s SEMO models. As a consequence of algorithmic and methodological developments, accurate ab initio electronic structure methods have long replaced SEMO models in their original areas of application (electronic structure calculations for small molecules). Nevertheless, SEMO models did not become extinct. Instead, they opened up different areas of application which can broadly be divided into three categories (see also Ref. [14] for a recent review): (i) simulations of very large systems such as proteins and those with thousands of small molecules, (ii) calculations for a large number of isolated and unrelated medium-sized molecules, e.g., in virtual high-throughput screening schemes for materials discovery and docking-and-scoring of potential drug candidates, and (iii) entirely new applications such as real-time quantum chemistry where ultra-fast SEMO models allow the perception of visual and haptic feedback in real time when manipulating medium-sized molecular structures.

In this work, we review, dissect, and analyze SEMO models which apply the neglect of diatomic differential overlap (NDDO) approximation (NDDO-SEMO models). These models are currently among the most popular SEMO models and all members of this class of SEMO models share the same conceptual framework. Another class of semiempirical models, which is under continuous development, are tight-binding versions of KS-DFT. We will not discuss these density-functional tight-binding models as the focus of this work is on NDDO-SEMO models.

The central NDDO approximation drastically reduces the computational effort associated with the calculation of electron-electron repulsion integrals (ERIs), and hence, leads to a significant speed-up. However, it took over ten years to successfully incorporate the NDDO approximation in a useful SEMO model (see Figure 1), the Modified Neglect of Diatomic Overlap (MNDO) model proposed in 1977. Since that time, small adjustments
Figure 1: Chronology of key steps in the development of NDDO-SEMO models starting with the introduction of the NDDO approximation in 1965 (highlighted in red) and reaching until today. Popular NDDO-SEMO models (MNDO, AM1, PM3, MNDO/d, RM1, PM6, PM7, OM1, OM2, OM3, and OM2-ML) are highlighted in blue, novel suggestions (NO-MNDO, PMO, OM2-ML, and hpCADD) in yellow, and semiclassical correction terms in green. The number of citations of the original publications was assessed with Google Scholar.

were made to the original MNDO model which gave rise to different closely related models such as AM1, PM3, MNDO/d, RM1, PM6, and PM7. In the nineties of the last century, the development of another family of NDDO-SEMO models began, the Orthogonalization-Corrected Models OM1, OM2, OM3.

Although the development efforts were consistently accompanied by articles, reviews, and books, there is no single resource which contains the detailed formalisms of the NDDO-SEMO models in a form which enables their facile implementation. Currently, for example, the implementation of a modern NDDO-SEMO model such as PM7 requires the consultation of at least ten references. Some of these references contain errors or misprints (which are clarified later on in this work) and some of them may be hard to obtain because they are books and dissertations written over forty years ago, some of them in German. Furthermore, the varying notation, adoption of jargon, and introduction of acronyms may hamper an
in-depth understanding. An in-depth understanding is, however, mandatory to be able to implement NDDO-SEMO models following the original references. As a concise and self-contained presentation of the formalism of these models is appropriate to understand their central ideas, we intend to provide such an overview in this work.

In recent years, a variety of semiclassical correction terms was designed to correct for specific flaws of NDDO-SEMO models, e.g., for hydrogen-bonding interactions,\textsuperscript{87–93} dispersion interactions,\textsuperscript{87,90–92,94} halogen-bonding interactions,\textsuperscript{91,95} and other pairwise interactions.\textsuperscript{96,97} The form of these terms is documented in a self-contained manner in the respective publications. The introduction of these terms does not explicitly affect the electronic structure part of the models. Hence, we refer to a recent review by Christensen et al.\textsuperscript{98} on this topic and we instead focus on the question of how NDDO-SEMO models attempt to approximate the electronic structure problem in this work.

This review is organized as follows: We first briefly introduce the notation and the quantum chemical foundations necessary to discuss NDDO-SEMO models which makes this review self-contained (Section 2). NDDO-SEMO models share, apart from the NDDO approximation, the application of a minimal basis set and the restriction of the number of explicitly considered electrons. We outline the formalism essential to all NDDO-SEMO models (Section 3) before moving on to specific NDDO-SEMO models (Sections 4–6). After discussing how NDDO-SEMO models are assembled, we discuss how (static and dynamical) electron correlation effects are captured. Generally, there are two strategies to tackle this problem: The calibration of parameters incorporated in NDDO-SEMO models against accurate reference data and the explicit description of electron correlation effects (Sections 7 and 8). We summarize the current state of knowledge with respect to both of these aspects and draw some more general conclusions for the prospects of NDDO-SEMO models (Section 9).

2 Setting the Stage

Electronic structure methods aim at the solution of the electronic Schrödinger equation,

\[ \mathcal{H}_{el} \Psi_{el}^{\{R_i\}} (\{r_i\}) = E_{el}^{\{R_i\}} \Psi_{el}^{\{R_i\}} (\{r_i\}), \]  

(1)

which asserts that we can calculate the electronic energy \( E_{el}^{\{R_i\}} \) from the electronic wave function \( \Psi_{el}^{\{R_i\}} (\{r_i\}) \) by applying the electronic Hamilto-
nian operator $\mathcal{H}_{\text{el}}$. In the Born–Oppenheimer approximation, $E_{\text{el}}^{(\mathbf{R}_i)}$ and $\Psi_{\text{el}}^{(\mathbf{R}_i)}(\{r_i\})$ depend parametrically on the fixed (indicated by the tilde and by giving them as superscripts) coordinates of the $N$ atomic nuclei ($\{\mathbf{R}_i\}$) of a system. The electronic wave function $\Psi_{\text{el}}^{(\mathbf{R}_i)}(\{r_i\})$ depends on the coordinates of $n$ electrons ($\{r_i\}$). The Hamiltonian operator $\mathcal{H}_{\text{el}}$ contains operators for the $n$ kinetic energy contributions of the electrons and for the electrostatic pair interaction energies of electrons and the $N$ atomic nuclei (in Hartree atomic units (a.u.)),

$$\mathcal{H}_{\text{el}} = \sum_{i=1}^{n} \left( -\frac{1}{2} \nabla_i^2 - \sum_{I=1}^{N} \frac{Z_I}{|r_i - \mathbf{R}_I|} \right) + \sum_{I=1}^{N} \sum_{J>I}^{N} \frac{Z_I Z_J}{|\mathbf{R}_I - \mathbf{R}_J|}. $$

The gradient defined for the coordinates of electron $i$ is denoted as $\nabla_i$ and $Z_I$ denotes the nuclear charge of the $I$-th nucleus (note that capital letters denote quantities defined for atomic nuclei). It is convenient to collect the first two terms in a one-electron operator $h_i$, the third term in a two-electron operator $g_{(i,j)}$, and the nuclear interaction energy in $V$,

$$\mathcal{H}_{\text{el}} = \sum_{i=1}^{n} h_i + \sum_{i=1}^{n} \sum_{j>i}^{n} g_{(i,j)} + V. $$

(3)

Since the early days of quantum mechanics, various approximations were developed to solve Eq. (1). In the following, we focus on Hartree–Fock (HF) theory to lay the foundation for the discussion of NDDO-SEMO models (see, e.g., Ref. 99 for a detailed presentation of HF theory in a one-electron basis set). The exact wave function $\Psi_{\text{el}}^{(\mathbf{R}_i)}(\{r_i\})$ is approximated by the HF wave function $\Phi_{\text{HF}}^{(\mathbf{R}_i)}(\{r_i\})$, which is constructed as the antisymmetrized product of one-particle functions $\psi_i^{(\mathbf{R}_i)}(r_i)$ (i.e., molecular spin orbitals),

$$\Psi_{\text{el}}^{(\mathbf{R}_i)}(\{r_i\}) \approx \Phi_{\text{HF}}^{(\mathbf{R}_i)}(\{r_i\}) = A \prod_{i=1}^{n} \psi_i^{(\mathbf{R}_i)}(r_i). $$

(4)

Antisymmetrization of the product of one particle-states by means of the antisymmetrization operator $A$ implements the Pauli principle. We approximate the spatial orbitals that enter the spin orbitals as linear combinations of $M$ atom-centered basis functions $\chi_{\mu}^{I} = \chi_{\mu}^{\mathbf{R}_i}(r)$ (not-th basis function of type $\chi$ centered on atom $I$) weighted with the expansion coefficients
\( \chi C \{ \hat{R}_I \} = \{ \chi C_{\mu I} \} \),

\[
\psi_i(\hat{R}_I)(r) = \sum_{\mu=1}^{M} \chi C_{\mu I} \chi_{\mu I}(r).
\] (5)

For the sake of brevity, we drop the superscripts \( \{ \hat{R}_I \} \) and \( \hat{R}_I \) in the following. We require the \( \chi \)-basis to be locally orthogonal, i.e., the overlap \( \chi S_{\mu \nu} = \langle \chi_{\mu I} | \chi_{\nu I} \rangle \) of different basis functions centered on the same atom must be zero,

\[
\chi S_{\mu \nu} = \begin{cases} 
\langle \chi_{\mu I} | \chi_{\nu I} \rangle & I \neq J, \ \forall \mu, \nu \\
\delta_{\mu \nu} & I = J, \ \forall \mu, \nu 
\end{cases}
\] (6)

in order to be able to apply the NDDO approximation. This is no general requirement for HF theory, but the introduction of another basis would complicate the notation. We will discuss this requirement in detail in Section 3.1.

The following equations are given for the spin-restricted formulation for the sake of simplicity. The central step underlying a canonical HF calculation in basis-set representation is then the iterative solution of the nonlinear Roothaan–Hall equation,

\[
\chi F \chi C = \chi S \chi C \epsilon,
\] (7)

for which we first need to calculate the Fock matrix \( \chi F = \chi F(\chi C) \) and the overlap matrix \( \chi S \). We obtain the matrix of basis set expansion coefficients \( \chi C \) and the diagonal matrix of orbital energies \( \epsilon \) as the solution of this generalized eigenvalue equation. The left superscript ‘\( \chi \)’ continues to indicate that the calculations are carried out in the \( \chi \)-basis. It is necessary to explicitly specify the basis because we will operate with different bases throughout this work. The matrix of orbital energies \( \epsilon \) is invariant under unitary matrix transformations by which one basis is transformed into another one. Consequently, \( \epsilon \) does not carry a superscript.

We can transform Eq. (7) to read

\[
\chi S^{-\frac{1}{2}} \chi F \chi S^{-\frac{1}{2}} \chi S^{\frac{1}{2}} \chi C = \chi S^{\frac{1}{2}} \chi C \epsilon,
\] (8)

which may be re-written in a simpler way as

\[
\phi F \phi C = \phi C \epsilon,
\] (9)

where

\[
\phi F = \chi S^{-\frac{1}{2}} \chi F \chi S^{-\frac{1}{2}}
\] (10)
and

\[ \phi C = \chi S^{\frac{1}{2}} \chi \phi. \]  

This constitutes a transformation of the Fock and coefficient matrices to the Löwdin orthogonalized \( \phi \)-basis (indicated by a left superscript \( \phi \)). The Löwdin orthogonalized basis functions \( \phi = \{ \phi_\mu \} \) and the locally orthogonal basis functions \( \chi = \{ \chi_\mu^I \} \) are related through

\[ \phi_\nu = \sum_{\mu=1}^{M} (\chi S^{-\frac{1}{2}})_{\mu\nu} \chi_\mu^I. \]

(12)

Obviously, the basis functions \( \phi_\mu \) are not centered on a single atom and therefore do not carry a superscript \( I \). The solution of the Roothaan–Hall equations in either basis (Eqs. (7) and (9)) requires a calculation of one-electron integrals and of ERIs. In this work, we employ Dirac’s bra-ket notation for the one-electron integrals,

\[ \langle \chi_\mu^I | h | \chi_\nu^L \rangle = \int \chi_\mu^I(r_1) \left[ -\frac{1}{2} \nabla^2_{1} - \sum_{l=1}^{N} \frac{Z_l}{|r_1 - R_l|} \right] \chi_\nu^L(r_1) d^3r_1 \]  

(13)

and the ERIs,

\[ \langle \chi_\mu^I \chi_\nu^L | \chi_\lambda^K \chi_\sigma^L \rangle = \int \int \chi_\mu^I(r_1) \chi_\nu^L(r_1) \frac{1}{|r_1 - r_2|} \chi_\lambda^K(r_2) \chi_\sigma^L(r_2) d^3r_1 d^3r_2, \]

(14)

in the \( \chi \)-basis. A Fock matrix element in the \( \chi \)-basis is then evaluated as,

\[ \chi F_{\mu\nu} = \langle \chi_\mu^I | h | \chi_\nu^L \rangle + \sum_{\lambda=1}^{M} \sum_{\sigma=1}^{M} \chi P_{\mu\sigma} \left[ \langle \chi_\mu^I \chi_\nu^L | \chi_\lambda^K \chi_\sigma^L \rangle - \frac{1}{2} \langle \chi_\mu^I \chi_\sigma^L | \chi_\lambda^K \chi_\nu^L \rangle \right], \]

(15)

where the ERIs are contracted with elements of the density matrix \( \chi P \). The elements of the density matrix in closed-shell systems are given by

\[ \chi P_{\mu\nu} = 2 \sum_{i=1}^{n/2} \chi C_{\mu i} \chi C_{\nu i}, \]

(16)

(assuming real expansion coefficients). The Fock matrix elements in the \( \phi \)-basis are assembled analogously,

\[ \phi F_{\mu\nu} = \langle \phi_\mu | h | \phi_\nu \rangle + \sum_{\lambda=1}^{M} \sum_{\sigma=1}^{M} \phi P_{\mu\sigma} \left[ \langle \phi_\mu \phi_\nu | \phi_\lambda \phi_\sigma \rangle - \frac{1}{2} \langle \phi_\mu \phi_\sigma | \phi_\lambda \phi_\nu \rangle \right]. \]

(17)
Eqs. (7) and (9) must be solved iteratively because the Fock matrix elements depend on the elements of the density matrix which is why the Roothaan–Hall equations are also known as the self-consistent field (SCF) equations.

For the following discussion, it is convenient to divide the Fock matrices into one-electron matrices $H$ and two-electron matrices $G$. The two-electron matrices $G$ can be further divided into the Coulomb matrices $J$ and the exchange matrices $K$, so that

$$
\chi F = \chi H + \chi G = \chi H + \chi J + \chi K,
$$

(18)

and

$$
\phi F = \phi H + \phi G = \phi H + \phi J + \phi K.
$$

(19)

After reaching self consistency, the total electronic HF energy $E_{el}^{HF}$ is calculated from the resulting density matrices, Fock matrices, and the nucleus-nucleus repulsion energy,

$$
E_{el}^{HF} = \frac{1}{2} \sum_{\mu=1}^{M} \sum_{\nu=1}^{M} \chi P_{\nu\mu} (\chi H_{\mu\nu} + \chi F_{\mu\nu}) + V,
$$

(20)

and

$$
E_{el}^{HF} = \frac{1}{2} \sum_{\mu=1}^{M} \sum_{\nu=1}^{M} \phi P_{\nu\mu} (\phi H_{\mu\nu} + \phi F_{\mu\nu}) + V.
$$

(21)

3 General Considerations for the Formalism of NDDO-SEMO Models

All NDDO-SEMO models describe a way to efficiently approximate the Fock matrix. Formally, the assembly of the Fock matrix in the course of the iterative solution of the SCF equations requires the calculation, repeated processing, and (if possible) storage of $M^4$ ERIs. Consequently, a lot of effort was put into the development of strategies to reduce the computational cost associated with this step.

3.1 Neglect of Diatomic Differential Overlap

One of these strategies is the NDDO approximation which drastically reduces the number of ERIs that must be calculated explicitly to assemble $\phi G$. 

The NDDO approximation,
\[
\langle \phi_\mu \phi_\nu | \phi_\lambda \phi_\sigma \rangle \approx \delta_{IJ} \delta_{KL} \langle \chi_\mu^I \chi_\nu^J | \chi_\lambda^K \chi_\sigma^L \rangle,
\]
(22)
specifies how ERIs in the \( \phi \)-basis may be approximated based on the values of the respective ERIs in the \( \chi \)-basis. It is not immediately obvious why Eq. (22) should hold true, especially in view of Eq. (12), but numerical data suggest that there is some merit to the NDDO approximation. We additionally illustrate the NDDO approximation in Figure 2 for three examples. Eq. (22) asserts that \( \langle \phi_\mu \phi_\nu | \phi_\lambda \phi_\sigma \rangle \) is negligibly small if \( I \neq J \) or \( K \neq L \), i.e., if \( \chi_\mu^I \) and \( \chi_\nu^J \) are centered on different atoms or if \( \chi_\lambda^K \) and \( \chi_\sigma^L \) are centered on different atoms. As a consequence, the formal scaling of the ERI evaluation step is reduced from \( \mathcal{O}(M^4) \) to \( \mathcal{O}(M^2) \). We see that this statement holds true for the three examples in Figure 2 because all blue crosses are located close to the horizontal dashed lines. Furthermore, Eq. (22) states that \( \langle \phi_\mu \phi_\nu | \phi_\lambda \phi_\sigma \rangle \) is approximately equal to \( \langle \chi_\mu^I \chi_\nu^J | \chi_\lambda^K \chi_\sigma^L \rangle \) when \( I = J \) and \( K = L \), e.g., the value of \( \langle \phi_1 \phi_1 | \phi_1 \phi_1 \rangle \) is approximately equal to the value of \( \langle \chi_1^I \chi_1^J | \chi_1^K \chi_1^L \rangle \). We see that this part of the NDDO approximation also holds true for the three examples that we considered for calculations in Figure 2 because the red circles are located close to the diagonal dashed lines. Evidently, the NDDO approximation emulates a basis transformation for the ERIs, and hence, also for the two-electron matrix, i.e.,
\[
\chi G^{\text{NDDO}} \approx \phi G
\]
(23)
This means that the matrix elements $\phi^G_{\mu\nu}$ can be approximately determined based on ERIs in the $\chi$-basis by

$$
\phi^G_{\mu\nu} \approx \chi^G_{\mu\nu} NDDO = \sum_{\lambda=1}^{M} \sum_{\sigma=1}^{M} \chi^{P NDDO}_{\lambda\sigma} \left( \delta_{IJ} \delta_{KL} \langle \chi^I_{\mu} \chi^J_{\nu} | \chi^K_{\lambda} \chi^L_{\sigma} \rangle ight) - \delta_{IK} \delta_{JL} \frac{1}{2} \langle \chi^I_{\mu} \chi^K_{\lambda} | \chi^J_{\nu} \chi^L_{\sigma} \rangle.
$$

(25)

The NDDO approximation may also be formulated in the $\phi$-basis,

$$
\langle \phi_\mu \phi_\nu | \phi_\lambda \phi_\sigma \rangle \approx \delta_{IJ} \delta_{KL} \langle \phi_\mu \phi_\nu | \phi_\lambda \phi_\sigma \rangle,
$$

(26)

which is, however, not very illuminating in the context of NDDO-SEMO models as it is not obvious what the meaning of $I$, $J$, $K$, and $L$ in Eq. (26) actually is for the basis functions $\phi$ that are not centered on single atoms (see Eq. (6)). Additionally, it is important to note that all NDDO-SEMO models calculate the ERIs in the $\chi$-basis and do not carry out an explicit basis transformation to the $\phi$-basis. In order to understand how NDDO-SEMO models work, we need to understand how this implicit basis transformation occurs, i.e., Eq. (22).

We emphasize that the NDDO statement is not that the ERIs in the $\chi$-basis are close to zero when $I \neq J$ or $K \neq L$. Only their corresponding ERIs in the $\phi$-basis are approximately zero. Figure 2 also illustrates this statement as several ERIs for which $I \neq J$ or $K \neq L$ are as large as 0.3 a.u. in the $\chi$-basis. It is therefore misleading to formulate the NDDO approximation in the $\chi$-basis,

$$
\langle \chi^I_{\mu} \chi^J_{\nu} | \chi^K_{\lambda} \chi^L_{\sigma} \rangle \approx \delta_{IJ} \delta_{KL} \langle \chi^I_{\mu} \chi^J_{\nu} | \chi^K_{\lambda} \chi^L_{\sigma} \rangle.
$$

(27)

The NDDO approximation leads to uncontrollable errors for the ERIs in the $\phi$-basis, which propagate to all quantities based on these erroneous ERIs (most importantly to electronic energies). Most likely due to the uncontrollable errors and the lack of systematic improbability, the NDDO approximation has never found any use in $ab$ initio theories.

In NDDO-SEMO models, the NDDO approximation is coupled to multiple other approximations, not least to correct for the errors introduced by the NDDO approximation itself. These additional approximations generally concern the calculation of the nonzero ERIs in the $\chi$-basis, the elements of the one-electron matrix in the $\phi$-basis, and $V$. Hence, we must
specify for each NDDO-SEMO model which parametrized expressions were applied to evaluate these three quantities. Throughout this work, we mostly adhered to the original parameter abbreviations. We present a comparison of the parameter abbreviations which we chose to the ones in the original literature in Tables S2 and S3 in the Supporting Information.

3.2 Restriction to an Effective Valence Shell

NDDO-SEMO models further reduce the computational effort by restricting the number of explicitly considered electrons. When restricting the number of explicitly considered electrons, it is necessary to specify for each atom I which of its $n_I$ electrons are considered valence (‘$v$’ electrons) $(n_{v,I})$ and which ones are considered core (‘$c$’ electrons) $(n_{c,I} = n_I - n_{v,I})$. Accordingly, each atomic core then exhibits a core charge $Q_I$,

$$Q_I = Z_I - n_{c,I}.$$  

(28)

Note that a rigorous distinction between core charge $Q_I$ and nuclear charge $Z_I$ is crucial for NDDO-SEMO models. Both quantities are required in parametric expressions in the formalism of some NDDO-SEMO models (see, e.g., Eq. (62) in Section 4.5.3). No rigorous method exists to justify a specific choice for $n_{v} = \sum_{I} n_{v,I}$ and $n_{c} = \sum_{I} n_{c,I}$. Within the NDDO-SEMO models, $n_{v,I}$ is restricted drastically so that $n_{v,I} = 1$, $n_{v,I} = 4$, $n_{v,I} = 5$, and $n_{v,I} = 6$ for hydrogen, carbon, nitrogen, and oxygen atoms, respectively.\[50,56,57,60\] In fact, $n_{v,I} \leq 12$ for all elements, and in principle, no more than 2 \(s\)-electrons, 6 \(p\)-electrons, and 10 \(d\)-electrons could be considered per atom within the current formalism of the popular NDDO-SEMO models.\[50,56,57,60\] We specify $n_{v,I}$ for all elements up to $Z_I = 83$ in Table S1 in the Supporting Information. We note that usually it is considered better to avoid too large $n_{c,I}$ in \textit{ab initio} calculations (see, e.g., Ref. [124] in which it was shown that a choice of $n_{v,I} = 4$ for lead and for titanium (as in PM6 and PM7) is not adequate to yield accurate electronic energies).

In \textit{ab initio} theories, one may approximate the effects of the core electrons by an effective core potential (ECP)\[124,125\] so that for $n_{v}$ valence electrons, the full Hamiltonian (Eq. (2)) is replaced by an approximate valence-only
Hamiltonian $H_{el,v}$,

$$
H_{el,v} = \sum_{i=1}^{n_v} \left( -\frac{1}{2} \nabla_i^2 + \sum_{I} \left( -\frac{Q_I}{|r_i - R_I|} + \text{ECP}_I \right) \right) + \sum_{i=1}^{n_v} \sum_{j>i}^{n_v} \frac{1}{|r_i - r_j|} + \sum_{I=1}^{N} \sum_{J>I}^{N} \frac{Q_I Q_J}{|R_I - R_J|},
$$

or in short notation,

$$
H_{el,v} = \sum_{i=1}^{n_v} h_{v,i} + \sum_{i=1}^{n_v} \sum_{j>i}^{n_v} g_{(i,j)} + V_v.
$$

The core-core repulsion energy is denoted as $V_v$ and the effective one-electron operator is denoted as $h_{v,i}$. The effective one-electron operator $h_{v,i}$ incorporates an effective core potential describing the interaction with the core electrons of atom $I$, $\text{ECP}_I$. We will not specify the different functional forms for the ECPs which are applied in $ab$ initio theories here (see, e.g., Refs. 124,125 for recent reviews) because they differ strongly from the ones which are applied in NDDO-SEMO models today.126–129 We will discuss the form of $\text{ECP}_I$ for each individual NDDO-SEMO model below.

### 3.3 Restriction of the Basis Set Expansion

Another way of reducing the computational effort is the restriction of the number of basis functions $M$ in Eq. (5). Generally, the number of basis function which are activated for an atom $I$, $M_I$, is less than or equal to nine. For each atom, at most, one $s$-type, three $p$-type, and five $d$-type basis functions may be considered, i.e., the number of basis functions activated for a molecule is

$$
M = \sum_{I=1}^{N} M_I \leq 9N.
$$

All NDDO-SEMO models apply only one $s$-type basis function for hydrogen. For carbon, nitrogen, and oxygen, the NDDO-SEMO models activate one $s$-type basis function and three $p$-type basis function. The basis functions which are considered for an atom of a certain element type are given in Table S1 in the Supporting Information.

The application of such a minimal valence-shell basis set has the practical advantage that the basis functions are inherently locally orthogonal (Eq. (6)),

12
i.e., the NDDO approximation is straightforwardly applicable.\textsuperscript{110,114,121,122} A minimal basis set is, however, generally unsuitable for the description of atoms in molecules. Conceptually, molecules are characterized by interacting atoms which polarize each other through additional external fields exerted by electrons and nuclei of the other atoms that distort the spherical symmetry of an atom. The description of the polarization of the electron density requires basis functions with a higher angular momentum. Consequently, it is common knowledge that calculations with a minimal basis sets do not yield reliable relative energies, force constants, electric dipole moments, static dipole polarizabilities, and other properties.\textsuperscript{130–134} More specifically, it would, for example, not be possible to obtain satisfactory results for polarizabilities or for the description of non-covalent interactions at the full-configuration-interaction (FCI) limit when applying a minimal basis set.\textsuperscript{100} A minimal basis set may, however, be sufficient to predict reasonable molecular equilibrium structures\textsuperscript{130–132,135} which led to a re-consideration of minimal-basis-set HF as a quick preliminary structure optimization method in recent years.\textsuperscript{136,137} Hence, the application of minimal basis sets may be adequate if the area of application of the NDDO-SEMO model is restricted accordingly.

4 The Modified Neglect of Diatomic Overlap (MNDO) Model

The MNDO model is the first successful SEMO model which is based on the NDDO approximation.\textsuperscript{49,50,138} The MNDO model activates one \(s\)- and three \(p\)-type basis functions (\(\chi\)-basis) for carbon, nitrogen, oxygen, and fluorine.\textsuperscript{50} Only one \(s\)-type basis function is retained for hydrogen.\textsuperscript{50} The MNDO model was later on extended to a larger part of the periodic table (see, e.g., Ref.\textsuperscript{139} and also Section 4.5.1). The Slater-type basis functions incorporate a parameter, the exponent \(\zeta_\mu\), which depends on the element type (here indicated by the superscript atomic number \(Z_I\)) of the atom \(I\) on which the basis function \(\chi\) is centered. The same \(\zeta_\mu\) is applied for \(s\)- and \(p\)-type basis functions, i.e., \(\zeta_s = \zeta_p\). We explicitly indicate for each parameter its dependencies (e.g., \(Z_I\) of the atom \(I\) on which a \(\chi\)-basis function is centered or the orbital-type of the basis function \(\chi\), i.e., \(s\), \(p\), or more generally, the angular momentum of the \(\mu\)-th \(\chi\)-basis function \(l(\mu)\)).

4.1 Parametrization of One-Center ERIs

At most, one \(s\)-type basis function and three \(p\)-type basis functions are centered on one atom. Altogether, six unique nonzero one-center ERIs
\[ \langle \chi^I_\mu \chi^J_\nu | \chi^I_\lambda \chi^J_\sigma \rangle \] may arise. Five of the six ERIs are substituted for element- and orbital-type-dependent parameters,

\[ \gamma^{ZI}_{ss} = \langle s^I s^I | s^I s^I \rangle, \tag{32} \]
\[ \gamma^{ZI}_{pp} = \langle p^I p^I | p^I p^I \rangle, \tag{33} \]
\[ \gamma^{ZI}_{sp} = \langle s^I s^I | p^I p^I \rangle, \tag{34} \]
\[ \gamma^{ZI}_{pp'} = \langle p^I p^I | p^{I'} p^{I'} \rangle, \tag{35} \]
and
\[ \gamma^{ZI}_{sp} = \langle s^I s^I | p^I p^{I'} \rangle, \tag{36} \]
where we only indicate the orbital type (s or p) of the basis functions when specifying the ERIs. The magnetic quantum numbers do not need to be explicitly specified because otherwise rotational invariance could not be guaranteed.\[140\] Hence, it is only necessary to indicate for the p-type basis functions whether the magnetic quantum number is the same (no prime, e.g., \( \langle p^I p^I | p^I p^I \rangle = \langle p^I p^I | p^I p^I \rangle \)) or different (prime, e.g., \( \langle p^I p^I | p^I p^I \rangle = \langle p^I p^I | p^I p^I \rangle \)).

The sixth one-center ERI is calculated from \( \gamma^{ZI}_{pp} \) and \( \gamma^{ZI}_{pp'} \),

\[ \langle p^I p^I | p^I p^I \rangle = \frac{1}{2} (\gamma^{ZI}_{pp} - \gamma^{ZI}_{pp'}). \tag{37} \]

If \( \gamma^{ZI}_{pp} < \gamma^{ZI}_{pp'} \), \( \langle p^I p^I | p^I p^I \rangle \) will be negative which leads to issues in the determination of the two-center ERIs (see Section 2 in the Supporting Information). In the original MNDO model, \( \langle p^I p^I | p^I p^I \rangle \) is always positive so that no issues arise, but \( \gamma^{ZI}_{pp} \) may be smaller than \( \gamma^{ZI}_{pp'} \) for extensions of the MNDO model\[56,57,139\]. All other one-center ERIs in the minimal valence-shell basis (e.g., \( \langle s^I s^I | s^I s^I \rangle \), \( \langle s^I s^I | p^I p^I \rangle \), \( \langle s^I s^I | p^I p^I \rangle \), or \( \langle p^I p^I | p^I p^I \rangle \)) are exactly zero.

### 4.2 Approximating the Two-Center ERIs

The two-center ERIs \( \langle \chi^I_\mu \chi^J_\nu | \chi^I_\lambda \chi^J_\sigma \rangle \), \( I \neq J \), can be interpreted as the electrostatic interaction between a charge distribution \( \chi^I_\mu \chi^J_\nu \) centered on atom I and a charge distribution \( \chi^I_\lambda \chi^J_\sigma \) centered on atom J. Each possible charge distribution \( \chi^I_\mu \chi^J_\nu \) in the s, p minimal valence-shell basis is approximately represented as a truncated classical multipole expansion\[49,60,63,64\] (see Ref. 49 or the Supporting Information for explicit formulae). Dewar and Thiel decided to specify four individual arrangements of discrete point charges for this purpose; a monopole \( q \), a dipole \( \mu_{xy,z} \), a linear quadrupole \( Q_{xz,yy,zz} \), and a square quadrupole \( Q_{xy,xz,yz} \) (see Figure 3). The distances \( D_\mu \) and \( D_Q \)
between the point charges in each configuration were chosen such that the multipole moment of each point charge configuration approximated the one of the corresponding charge distribution which is ensured by calculating $D_\mu$ and $D_Q$ based on $\zeta^{2t}$ (see Ref. 49 or the Supporting Information). As soon as the positions of the point charges in space have been specified (by defining $D_\mu$ and $D_Q$ and the local arrangements around the atomic nuclei), one can readily calculate the electrostatic potential energy as a sum over all possible electrostatic interactions of the point charges. The approximation of a two-center ERI as the electrostatic interaction of discrete point charges will, however, break down when the distance between the atomic nuclei,

\begin{equation}
\tilde{R}_{IJ} = |\tilde{R}_I - \tilde{R}_J|,
\end{equation}

becomes small. We illustrate this at the example of the two-center ERI $\langle s^I s^I | s^J s^J \rangle$ which is approximated by the electrostatic interaction of a point charge located on $I$ and one located on $J$. If $\tilde{R}_{IJ} = 0$, a singularity would arise for the electrostatic interaction between the two point charges. This would not be the case if we considered the electrostatic interaction of a charge distribution centered on $I$ and one centered on $J$. As a consequence, the expression to calculate the electrostatic interaction between point charges is modified in an empirical manner in such a way that it yields the one-center ERIs (Eqs. (32)–(104)) in the limit $\tilde{R}_{IJ} = 0$, (‘Klopman formula’, Eq. (14) in Ref. 141),

\begin{equation}
\langle s^I s^I | s^J s^J \rangle \approx \left[ s^I s^I | s^J s^J \right] = \left[ \tilde{R}_{IJ}^2 + \left( \frac{1}{2Z_{ss}^I} + \frac{1}{2Z_{ss}^J} \right)^2 \right]^{-\frac{1}{2}} .
\end{equation}
In the following, we denote an ERI calculated in the Klopman approximation by square brackets to easily distinguish them from analytically calculated ERIs (denoted in angle brackets). From Eq. (39) and Figure 4, we see that for large $\tilde{R}_{IJ}$ (and constant $\gamma^Z_{ss}$ and $\gamma^Z_{ss}$), the term approaches the regular Coulomb interaction of two elementary point charges, $[s^I s^J | s^I s^J] \approx 1/\tilde{R}_{IJ}$. In the limit $\tilde{R}_{IJ} = 0$ (where $\gamma^Z_{ss} = \gamma^Z_{ss}$), the expression reduces to $[s^I s^J | s^I s^J] = \gamma^Z_{ss}$. The value of $\gamma^Z_{ss}$ determines how closely the approximate ERI follows the analytical one (see also Figure 4). Usually $\gamma^Z_{ss}$ is chosen such that the semiempirical two-center ERIs are smaller than the analytical values. E.g., Pariser and Parr, Dewar and Klopman, and Voigt argued that in this way dynamic electron correlation effects can be emulated. We will analyze this claim in detail in Section 7.

$$\langle \chi^I \mu | \chi^I \mu | \chi^J \nu | \chi^J \nu \rangle$$ [a.u.]

Figure 4: Dependence of $\langle \chi^I \mu | \chi^I \mu | \chi^J \nu | \chi^J \nu \rangle$ (black line) and $[\chi^I \mu | \chi^I \mu | \chi^J \nu | \chi^J \nu]$ on $\tilde{R}_{IJ} = |\tilde{R}_{H1} - \tilde{R}_{H2}|$ in an H$_2$ molecule described by an MNDO-3G basis. We applied a Gaussian-type basis set (denoted MNDO-3G) to calculate the analytical ERI values instead of the Slater-type basis set inherent to the MNDO model. We generated the MNDO-3G basis based on $\zeta_{ZI}$. We chose different values for $\gamma^1_{ss}$ in Eq. (39): $\gamma^1_{ss} = 0.83$ a.u. (analytical one-center ERI, red line), $\gamma^1_{ss} = 0.65$ a.u. (green line), and $\gamma^1_{ss} = 0.45$ a.u. (MNDO value, blue line).

For the other two-center ERIs, similar formulae as the one in Eq. (39) can be derived which yield the respective one-center ERI in the limit $\tilde{R}_{IJ} = 0$ (see Ref. 49 or the Supporting Information). For the calculation of two-center ERIs, which involve at least one $p$-type basis function, a local coordinate system is adopted. This local coordinate system is defined based on $\tilde{R}_I$ and $\tilde{R}_J$ (see Figure 5 and Section 5 in the Supporting Information). The results
Figure 5: Illustration of the difference of the local (red axes) and global (black axes) coordinate systems when calculating $\langle s^I p_x^I | p_x^J p_y^J \rangle$ in aspirin. The positions of the point charges with which the charge distributions $s^I p_x^I$ and $p_x^J p_y^J$ are approximated are indicated by blue spheres.

obtained in this local coordinate system have to be transformed to yield the ERIs in the global coordinate system. The necessary transformations can be formulated in terms of rotation matrices which are outlined in the Supporting Information.

Issues with the presented approach were detected years after the introduction of the MNDO model. It was remarked that rotational invariance was not satisfied for the ERIs $[p_x^I p_y^I | p_x^J p_y^J]$ due to the chosen point charge configurations. It was then suggested to impose rotational invariance by setting

$$[p_x^I p_y^I | p_x^J p_y^J] = 0.5 \left( [p_x^I p_y^I | p_x^I p_y^J] - [p_x^I p_y^J | p_x^J p_y^J] \right).$$

(40)

Moreover, the Klopman approximation causes distinct errors in the ERIs which culminates in an infinite error in periodic electronic structure calculations. To be able to apply the MNDO in periodic electronic structure calculations, an additional scaling factor has to be introduced to yield the exact limit for large $\tilde{R}_{IJ}$.

4.3 Assembling the Symmetrically Orthogonalized One-Electron Matrix

The MNDO model does not provide an explicit way to account for the change from the $\chi$-basis to the $\phi$-basis for the one-electron matrix $\mathbf{H}$. It is assumed that the elements of $\phi \mathbf{H}$ are approximately equal to $\chi \mathbf{H}$ when an appropriate parametrization is chosen.

$$\phi \mathbf{H} \approx \chi \mathbf{H}^{\text{param}}.$$  

(41)

An example discussed in Refs. explains why this is not generally possible (see Figure 6). We may consider the dihydrogen molecule ($H_1 - H_2$) in a minimal basis set consisting of two 1s orbitals $\chi = \{ \chi_1^H, \chi_2^H \}$.  

17
For the moment, we neglect electron-electron interactions, i.e., we solve the eigenvalue equations,

\[ \phi^H \phi C = \phi \epsilon, \quad (42) \]

in the \( \phi \)-basis, or analogously in the \( \chi \)-basis,

\[ \chi^H \chi C = \chi S \chi \epsilon. \quad (43) \]

We now want to know whether it is possible to obtain the same \( \epsilon = \epsilon_{\text{param}} \) when solving an eigenvalue equation of the type

\[ \chi^H \chi_{\text{param}} C = \chi \epsilon_{\text{param}}, \quad (44) \]

where we neglect \( \chi S \). When solving Eqs. (42) and (43), we see that \(|\phi H_{11} - \epsilon_{11}|\) and \(|\phi H_{11} - \epsilon_{22}|\) are equally large while \(|\chi H_{11} - \epsilon_{11}|\) and \(|\chi H_{11} - \epsilon_{22}|\) differ. Independently of the parametrization, it is not possible to obtain different \(|\chi H_{11}^\text{param} - \epsilon_{11}^\text{param}|\) and \(|\chi H_{22}^\text{param} - \epsilon_{22}^\text{param}|\) when solving Eq. (44) because \( \chi S \) is neglected in Eq. (44) (see Figure 6).

Figure 6: Illustration of the effect of parametrization when orthogonalization effects are not considered at the example of dihydrogen (\( \chi = \{ \chi^H_1, \chi^H_2 \} \)).

a) Values of \( \epsilon_{11} \) and \( \epsilon_{22} \) when solving the eigenvalue equation \( \chi^H \chi C = \chi \epsilon \) without considering \( \chi S \). b) Value of \( \epsilon_{11} \) and \( \epsilon_{22} \) obtained when solving the eigenvalue equation \( \phi^H \phi C = \phi \epsilon \). The eigenvalues are split asymmetrically with respect to the \( \chi H_{11} = \chi H_{22} \) reference. c) Effect of parametrization to reproduce the lowest energy eigenvalue when solving the equation \( \chi^H_{\text{param}} \chi C = \chi \epsilon_{\text{param}} \).

Apart from this example, it is evident that orthogonalization effects,

\[ \phi H_{\mu \nu} = \sum_{\lambda=1}^{M} \sum_{\sigma=1}^{M} \left( \chi S^{-\frac{1}{2}} \right)_{\mu \lambda} \chi H_{\lambda \sigma} \left( \chi S^{-\frac{1}{2}} \right)_{\sigma \nu}, \quad (45) \]
cannot be captured by introducing element-dependent parameters which was, for instance, pointed out in Refs. 121,153–155. The matrix element $\chi^H_{\mu\nu}$ depends on contributions from all matrix elements of $\chi^H$. Consequently, the parametrization would need to depend on the chemical environment of each atom in some manner. The MNDO parameters, however, are only element-dependent and do not depend on the chemical environment. Despite this inherent limitation, MNDO has been a very successful model and we will continue to discuss how the contributions to $\phi^H_{\mu\nu}$ are evaluated in the following Sections (Sections 4.3.1 and 4.3.2). We assume that the inclusion of empirical parameters accounts for orthogonalization effects in some average manner, and hence, retain the superscript $\phi$ for $\phi^H_{\mu\nu}$.

The parametric expressions applied for the evaluation of the matrix elements $\phi^H_{\mu\nu}$ differ depending on the number of atoms on which the corresponding basis functions $\chi^I_{\mu}$ and $\chi^J_{\nu}$ are centered: (i) $\chi^I_{\mu}$ and $\chi^I_{\nu}$ are centered on a single atom $I$ (one-center one-electron matrix elements) and (ii) $\chi^I_{\mu}$ and $\chi^J_{\nu}$ are centered on different atomic nuclei $I \neq J$ (two-center one-electron matrix elements).

4.3.1 One-Center One-Electron Matrix Elements

In the case that $\chi^I_{\mu}$ and $\chi^I_{\nu}$ are centered on the same atom, the analytical matrix elements in the $\chi$-basis are given by

$$\chi^H_{\mu\nu} = \left\langle \chi^I_{\mu} \left| -\frac{1}{2} \nabla^2 \right| \chi^I_{\nu} \right\rangle - \left\langle \chi^I_{\mu} \left| \frac{Q_I}{|r_i - R_I|} \right| \chi^I_{\nu} \right\rangle - \sum_{J=1, J \neq I}^N \left\langle \chi^I_{\mu} \left| \frac{Q_J}{|r_i - R_J|} \right| \chi^J_{\nu} \right\rangle.$$  (46)

These first two terms in the right hand side of Eq. (46) only refer to the atom $I$ (‘one-center’ one-electron contributions to $\chi^H_{\mu\nu}$). The remaining ‘two-center’ one-electron contributions to $\chi^H_{\mu\nu}$ describe the electrostatic attraction between the charge distribution $\chi^I_{\mu} \chi^I_{\nu}$ and the atomic cores $J \neq I$.

In view of Eq. (46), it is apparent why Dewar and Thiel suggested\[20\] to calculate $\phi^H_{\mu\nu}$ by means of

$$\phi^H_{\mu\nu} = U^{Z_I}_{(\mu)(\nu)} - \sum_{J=1, J \neq I}^N Q^J \left[ \chi^J_{\mu} \chi^J_{\nu} \right] s^J s^J.$$  (47)

The element- and orbital-type-dependent parameter $U^{Z_I}_{(\mu)(\nu)}$ comprises all one-center one-electron contributions in Eq. (46). The one-center one-electron
terms are exactly zero in a locally orthogonal basis when \( \mu \neq \nu \),

\[
U^Z_{l(\mu)l(\nu)} = \begin{cases} 
\text{const.} & \mu = \nu \\
0 & \mu \neq \nu 
\end{cases}.
\] (48)

The parameter \( U^Z_{l(\mu)l(\nu)} \) may not depend on the magnetic quantum number to ensure rotational invariance and hence, at most two parameters, \( U^{Zss}_{ll} \) and \( U^{Zpp}_{ll} \), arise per element. Within a given \( \chi \)-basis, \( U^{Zss}_{ll} \) and \( U^{Zpp}_{ll} \) can be calculated exactly and are transferable between molecules. The MNDO model, however, attempts to approximate the matrix element in the \( \phi \)-basis,

\[
\phi H_{\mu\nu} = \left< \phi_\mu \left| -\frac{1}{2} \nabla^2 \right| \phi_\nu \right> - \left< \phi_\mu \left| \frac{Q_I}{|r_1 - R_l|} \right| \phi_\nu \right> - \sum_{J=1 \atop J \neq I}^N \left< \phi_\mu \left| \frac{Q_J}{|r_1 - R_J|} \right| \phi_\nu \right>.
\] (49)

In the \( \phi \)-basis, the first two terms obviously depend on the chemical environment of the atom \( I \) which is neglected by introducing constant \( U^{Zss}_{ll} \) and \( U^{Zpp}_{ll} \). It is assumed that the calibration of \( U^{Zss}_{ss} \) and \( U^{Zpp}_{pp} \) will implicitly lead to a modeling of average orthogonalization effects and will also absorb the effects from the core electrons.

A two-center contribution to Eq. (46) is approximated by the negative electrostatic interaction energy of \( \chi^I_\mu \chi^I_\nu \) with a model charge distributions \( s^I_s, J \neq I \) scaled with \( Q_J \) in Eq. (47). Pople and Segal proposed to apply this so-called Goeppert-Mayer–Sklar approximation (named for its relation to an equation proposed in Ref. [157]) after observing that the application of the analytical expression in Eq. (46) led to far too short bond lengths for several diatomic molecules. A decade after their proposal, Coffey analyzed the Goeppert-Mayer–Sklar approximation in more detail and concluded that a fortunate error cancellation occurs, so that

\[
\left< \phi_\mu \left| \frac{Q_J}{|r_1 - R_J|} \right| \phi_\nu \right> \approx Q_J \left< \chi^I_\mu \chi^I_\nu \right>.
\] (50)

The first two terms in Eq. (50) are the so-called ‘penetration integrals’,

\[
\left< \chi^I_\mu \chi^I_\nu, J \right> \approx \left< \chi^I_\mu \left| \frac{Q_J}{|r_1 - R_J|} \right| \chi^I_\nu \right> - Q_J \left< \chi^I_\mu \chi^I_\nu | s^I_s \right>.
\] (51)
approximately cancel the orthogonalization effects, 
\[
\langle \phi_{\mu} \left| \frac{Q_J}{r_i - \tilde{R}_J} \right| \phi_{\nu} \rangle - \langle \chi_{\mu}^I \left| \frac{Q_J}{r_i - \tilde{R}_J} \right| \chi_{\nu}^I \rangle
\]
which are required to transform $\chi_{I\mu}$ from the $\chi$- into the $\phi$-basis (see Figure 1 in Ref. 158). Hence, the Goeppert-Mayer–Sklar approximation entails an implicit basis set transformation from the $\chi$- to the $\phi$-basis. The orthogonalization effects, which Coffey considered\textsuperscript{158} also included orthogonalization of the core orbitals to the valence orbitals. Hence, the application of Eq. (50) may also be interpreted as the emulation of the application of an approximate effective core potential. Unfortunately, Coffey’s analysis was restricted to the C$_2$ molecule and included several additional approximations (such as an averaging of one- and two-electron integrals)\textsuperscript{158} It is not evident whether (and appears improbable\textsuperscript{155,159} that) Coffey’s analysis can easily be generalized to arbitrary polyatomic molecules. The success of the MNDO model indicates, however, that — at least in the context of all other invoked approximations — the Goeppert-Mayer–Sklar approximation is a satisfactory one.

### 4.3.2 Two-Center One-Electron Matrix Elements

In the case that $\chi_{I\mu}$ and $\chi_{J\nu}$ are centered on different atomic nuclei ($I \neq J$), the matrix elements $\phi_{\mu\nu}^{\text{MNDO}}$ are taken to be proportional to $\chi_{S\mu\nu}$,

\[
\phi_{\mu\nu}^{\text{MNDO}} = \frac{\beta Z_{I}^{\mu} + \beta Z_{J}^{\nu}}{2} \chi_{S_{\mu\nu}}.
\]

The mean of two element- and orbital-type-dependent parameters $\beta Z_{I}^{\mu}$ and $\beta Z_{J}^{\nu}$ yields the proportionality factor. In analogy to the parameters $U_{I}^{Z_{I}}$ and $U_{J}^{Z_{J}}$, at most two parameters arise per element, $\beta Z_{I}$ and $\beta Z_{J}$. Taking $\phi_{\mu\nu}^{\text{MNDO}}$ to be proportional to $\chi_{S_{\mu\nu}}$ has a long history\textsuperscript{8,50} and the initial idea is ascribed to Mulliken.\textsuperscript{160} Generally, Eq. (52) was, however, found to be a poor approximation to the analytical value of $\phi_{\mu\nu}$, irrespective of the chosen values for $\beta_{l(\mu)}$.\textsuperscript{155,161–163} This can be attributed to the fact that $\phi_{\mu\nu}$ is not necessarily proportional to $\chi_{S_{\mu\nu}}$ (for an example, see Figure 7). Hence, not even the nodal structure of $\phi_{\mu\nu} = \phi_{\mu\nu}(\tilde{R}_{IJ})$ is captured correctly. This finding appears puzzling in view of the success of the MNDO model, apparently Eq. (52) suffices to obtain satisfactory results, e.g., for heats of formation in this context.
Figure 7: Dependence of $\phi_{H_{\mu\nu}}$ (solid lines) and $\phi_{H_{\mu\nu}^{\text{MNDO}}}$ (dashed lines) on the distance $\tilde{R}_{IJ}$ in a $C_2$ molecule described in an MNDO-3G basis ($\chi = \{ \chi_1^{C_1}, \chi_2^{C_1}, \chi_3^{C_1}, \chi_4^{C_1}, \chi_5^{C_2}, \chi_6^{C_2}, \chi_7^{C_2}, \chi_8^{C_2} \}$ where the first and fifth basis functions are 2$s$, the second and sixth are 2$p_x$, the third and seventh are 2$p_y$, and the fourth and eighth are 2$p_z$ basis functions). a) $\phi_{H_{15}^{\text{MNDO}}}$ with the best-fit parameter $\beta_s^6 = 0.1$ a.u. (black line) and $\chi_{S_{15}}$ (gray dotted line). b) $\phi_{H_{15}^{\text{MNDO}}}$ with $\beta_s^6 = 0.1$ a.u. (black lines), $\phi_{H_{18}^{\text{MNDO}}}$ with $\beta_s^6 = \beta_p^6 = -0.3$ a.u. (red lines), $\phi_{H_{48}^{\text{MNDO}}}$ with $\beta_s^6 = 1.3$ a.u. (blue lines), and $\phi_{H_{26}^{\text{MNDO}}}$ with $\beta_s^6 = -0.2$ a.u. (green lines).

4.4 Empirical Modification of Core-Core Repulsion Energy

The core-core repulsion energy in the MNDO model, $V_{v}^{\text{MNDO}}$, is also determined from a parametric expression. The substitution of the analytical expression,

$$V_{v} = \sum_{I=1}^{N} \sum_{J>I}^{N} \frac{Q_I Q_J}{\tilde{R}_{IJ}},$$

for a parametric one cannot be physically motivated. Empirically, it was determined that a parametric expression needs to be introduced to decrease the average core-core repulsion energy to define a useful NDDO-SEMO model.
The parametric expression to evaluate \( V_{\text{MNDO}} \),

\[
V_{\text{MNDO}} = \sum_{I=1}^{N} \sum_{J>I}^{N} Q_I Q_J \left[ s^I s^I | s^J s^J \right] \cdot f_{IJ}^{\text{MNDO}}
\]

\[
= \sum_{I=1}^{N} \sum_{J>I}^{N} Q_I Q_J \left[ \tilde{R}_{IJ}^2 + \left( \frac{1}{2z_I^{ss}} + \frac{1}{2z_J^{ss}} \right) f_{IJ}^{\text{MNDO}} \right]^{-\frac{1}{2}} \cdot f_{IJ}^{\text{MNDO}},
\]

features two key modifications with respect to Eq. (53). Firstly, the pairwise point-charge interaction is substituted by a scaled interaction of the charge distributions \( s^I s^I \) and \( s^J s^J \) which is evaluated in the Klopman approximation (cf. Eq. (39)). Secondly, each core-core interaction energy is scaled by \( f_{IJ}^{\text{MNDO}} \),

\[
f_{IJ}^{\text{MNDO}} = 1 + \exp \left( -\alpha z_I \tilde{R}_{IJ} \right) + \exp \left( -\alpha z_J \tilde{R}_{IJ} \right),
\]

where \( \alpha z_I \) is an element-dependent parameter. The introduction of these modifications provides a large flexibility for the MNDO model, but this flexibility comes at a high price. Most strikingly, \( V_{\text{MNDO}} \) is finite for \( \tilde{R}_{IJ} = 0 \).

More specifically, \( f_{IJ}^{\text{MNDO}} = 3.0 \) and \( V_{\text{MNDO}} = 3Q_I^2 \alpha z_I \) in the limit \( \tilde{R}_{IJ} = 0 \) for a homonuclear diatomic system (see also Figure 8). Obviously, this limit is entirely artificial. For intermediate values of \( \tilde{R}_{IJ} \), the parameter \( \alpha z_I \) determines how fast \( f_{IJ}^{\text{MNDO}} \) declines from three (\( \tilde{R}_{IJ} = 0 \)) to one (\( \tilde{R}_{IJ} \to \infty \)) if \( \alpha z_I \) is not negative. This appears to be a constraint invoked during the calibration of \( \alpha z_I \). Hence, in the limit of large \( \tilde{R}_{IJ} \), \( V_{\text{MNDO}} \) will tend toward \( V_v \) as it should. Depending on the choice of \( \alpha z_I \), \( V_{\text{MNDO}} \) may be larger or smaller than \( V_v \) for a given \( \tilde{R}_{IJ} \).

The scaling factor is not calculated according to Eq. (55) when the element pair is N–H (\( z_I = 7 \) and \( z_J = 1 \)) or O–H (\( z_I = 8 \) and \( z_J = 1 \)). Dewar and Thiel found that they could achieve a better agreement with experimental data when they instead applied the scaling factor \( f_{IJ}^{\prime,\text{MNDO}} \) for these element pairs,

\[
f_{IJ}^{\prime,\text{MNDO}} = 1 + \tilde{R}_{IJ} \exp \left( -\alpha z_I = 7,8 \tilde{R}_{IJ} \right) + \exp \left( -\alpha z_J = 1 \tilde{R}_{IJ} \right).
\]

Because there is no theoretical foundation for the introduction of Eqs. (54) and (55), it remains unclear why the application of Eq. (56) yields a better agreement with experimental data. We note that \( f_{IJ}^{\prime,\text{MNDO}} = 2 \) in the limit \( \tilde{R}_{IJ} = 0 \). The modification does, hence, not rectify the theoretically unsatisfactory situation of finite core-core repulsion energies in the limit \( \tilde{R}_{IJ} = 0 \). The scaling factor \( f_{IJ}^{\prime,\text{MNDO}} \) also tends to one for large \( \tilde{R}_{IJ} \). For given \( \alpha z_I \), \( f_{IJ}^{\prime,\text{MNDO}} < f_{IJ}^{\text{MNDO}} \) for all \( \tilde{R}_{IJ} \).
We calculated the MNDO scaling factors $f_{IJ}^{\text{MNDO}}$ and $V_{v}^{\text{MNDO}}$ with $\alpha^1 = 1.0$ Å$^{-1}$ (green lines), $\alpha^1 = 2.5$ Å$^{-1}$ (red lines), $\alpha^1 = 10.0$ Å$^{-1}$ (blue lines), and $\alpha^1 = \infty$ (i.e., $f_{IJ}^{\text{MNDO}} = 1$, gray lines). We compare $V_{v}^{\text{MNDO}}$ to $V_{v}$ (black line).

### 4.5 Direct Descendants of the MNDO Model

#### 4.5.1 Extension to $d$ Orbitals: The MNDO/$d$ Model

The acronym ‘MNDO/$d$’ denotes the extension of the MNDO model from an $s, p$ basis to an $s, p, d$ basis.\textsuperscript{53,54} The consideration of $d$-type basis functions requires, on the one hand, the specification of additional orbital-type-dependent parameters per element, and on the other hand, an adjustment of the parametric expressions themselves.

The number of unique nonzero one-center ERIs increases from six (see Eqs. (32)–(104)) for an $s, p$ basis to 58 for an $s, p, d$ basis. These 58 one-center ERIs are determined analytically\textsuperscript{24} from a set of auxiliary orbital exponents $\zeta^{2t}_s$, $\zeta^{2t}_p$, and $\zeta^{2t}_d$.\textsuperscript{53} The auxiliary orbital exponents are derived from the fitted parameters $\gamma^{2t}_s$, $\gamma^{2t}_p$, and $\gamma^{2t}_d$ which are, e.g., applied to calculate the overlap integrals. The formulae for calculating the one-center ERIs from the auxiliary Slater exponents $\zeta^{2t}_s$, $\zeta^{2t}_p$, and $\zeta^{2t}_d$ are given in Refs. \textsuperscript{84,85}. It appears that several of the formulae presented in Ref. \textsuperscript{84} (Eqs. (17), (51), (53), (54), (56), and (57)) contain typographical mistakes which we correct in the Supporting Information (Eqs. (S19)–(S76)). These errors affect the one-center ERIs of the types $\langle p'_z d_{xz} | p'_z d_{xz} \rangle$, $\langle p'_z d_{xz} | p'_z d_{yz} \rangle$, $\langle s' d_{xz} | p'_z p'_z \rangle$, $\langle s' d_{xz} | p'_y p'_y \rangle$, $\langle p'_y p'_y | s' d_{xz} \rangle$, $\langle p'_z p'_z | s' d_{xz} \rangle$, $\langle p'_z p'_y | s' d_{yz} \rangle$, $\langle p'_z p'_y | s' d_{yz} \rangle$.\textsuperscript{24}
\[ \langle p_x^I p_z^J | s^I d^J_{x-z} \rangle, \text{ and } \langle p_x^I p_z^J | s^I d^J_{xy} \rangle. \]

Thiel and Voityuk also extended the formalism to approximate two-center ERIs in a point-charge model to charge distributions including \(d\)-type orbitals.\(^{53,54}\) For this purpose, they introduced a new quadrupole point charge configuration \(Q_{x,y,x,y}\) (see also Figure S1 in the Supporting Information). One can then straightforwardly apply the concepts of the multipole expansion introduced in Section 4.2 and derive the necessary formulae for all possible combinations of arising multipoles (see Refs. 53,54 or the Supporting Information). Thiel and Voityuk claimed\(^{53,54}\) that the interaction energy of the quadrupoles \(Q_{xz}\) and \(Q_{xy}\) is zero which is not true. This mistake results in an inconsistent evaluation of all two-center ERIs which involve a \(d\)-type orbital and one of the following charge distributions:

\[ \langle p_x^I p_z^J | s^I d^J_{x,y} \rangle, \langle p_x^I p_z^J | s^I d^J_{x,y} \rangle, \langle p_{x-y}^I | s^I d^J_{x,y} \rangle, \langle p_x^I p_z^J | s^I d^J_{x,y} \rangle, \langle p_{x-y}^I | s^I d^J_{x,y} \rangle, \langle p_{x^2-y^2}^I | s^I d^J_{x-y} \rangle. \]

The corrected formulae are presented in Refs. 164,165 and in the Supporting Information.

The two-center ERIs are also applied to calculate the core-core repulsion energy (Eq. (54)) and contributions to the one-electron matrix (Eq. (47)). In these equations, the atomic core \(I\) was described by a charge distribution \(s^I\) which involves the parameter \(\gamma_{ss}^I\). The MNDO/d formalism makes these expressions independent from the parameter \(\gamma_{ss}^I\). Therefore, the atomic core \(I\) is described by a spherical charge distribution \(\rho_{\text{core}}^I\). The electrostatic interaction energy with this charge distribution is computed within the Klopman–Ohno approximation, so that Eq. (54) is substituted for

\[
V_{\text{MNDO/d}}^I = \sum_{l=1}^N \sum_{J>I}^N Q_J Q_J \left[ \rho_{\text{core}}^I \rho_{\text{core}}^J \right] \cdot f_{IJ}^{\text{MNDO}},
\]

\[ (57)\]

where \(\rho_{\text{core}}^I \rho_{\text{core}}^J = 1/(2\gamma_{ss}^I)\) for elements which do not activate \(d\)-type orbitals and \(\rho_{\text{core}}^I \rho_{\text{core}}^J \neq 1/(2\gamma_{ss}^I)\) for elements which activate \(d\)-type orbitals. Similar adjustments are necessary for the formula to calculate the one-electron matrix (Eq. (47)) which is now evaluated as

\[
\phi H_{\mu\nu}^{\text{MNDO}} = U_{\mu\nu}^{Z_I} - \sum_{J=1}^N Q_J \left[ \chi_{\mu}^I \chi_{\nu}^I \rho_{\text{core}}^J \right] \cdot f_{IJ}^{\text{MNDO}},
\]

\[ (58)\]

Additionally, we have to specify the parameters \(U_{dd}^{Z_I}, \beta_{d}^{Z_I}, \text{ and } \zeta_{d}^{Z_I}\) to assemble the one-electron matrix.
4.5.2 The Austin Models (AM)

The Austin model 1 (AM1)\textsuperscript{51} differs from the MNDO model in the way in which the scaling factor for the pairwise core-core repulsion energies is determined,

\[
V_v^{AM1} = \sum_{i=1}^{N} \sum_{j>i}^{N} Q_i Q_j \left[ s^i s^j \right] \cdot f_{ij}^{AM1}.
\] (59)

The scaling factor \( f_{ij}^{AM1} \) is defined as the sum of the scaling factor \( f_{ij}^{MNDO} \) (Eq. (55)) and an element-specific number of additional Gaussian functions \( A^{Z_i} \textsuperscript{51,166} \).

\[
f_{ij}^{AM1} = f_{ij}^{MNDO} + \sum_{a=1}^{A^{Z_i}} K_a^{Z_i} \exp \left( -L_a^{Z_i} \left( \bar{R}_{ij} - M_a^{Z_i} \right)^2 \right) + \sum_{a=1}^{A^{Z_j}} K_a^{Z_j} \exp \left( -L_a^{Z_j} \left( \bar{R}_{ij} - M_a^{Z_j} \right)^2 \right). \] (60)

The shape of the \( a \)-th Gaussian function is characterized by the element-dependent parameters \( K_a^{Z_i}, L_a^{Z_i}, \) and \( M_a^{Z_i} \). The sign of \( K_a^{Z_i} \) determines whether the \( a \)-th Gaussian increases (positive sign) or decreases (negative sign) \( f_{ij}^{AM1} \) at a given \( \bar{R}_{ij} \). \( L_a^{Z_i} \) determines the width of the Gaussian; it must be positive because \( f_{ij}^{AM1} \) would otherwise tend to infinity for large \( \bar{R}_{ij} \). \( M_a^{Z_i} \) specifies where the \( a \)-th Gaussian is centered and, hence, for which \( \bar{R}_{ij} \) additional repulsive or attractive interactions are added. If more than one Gaussian is added to the scaling factor \( (A^{Z_i} > 1) \), \( f_{ij}^{AM1} \) becomes a quite involved function. It can decrease the core-core repulsion energy at certain distances (e.g., at \( \bar{R}_{ij} = 1.6 \text{ Å} \) in Figure 9) and increase it at other distances (e.g., at \( \bar{R}_{ij} = 2.1 \text{ Å} \) in Figure 9). Hence, \( f_{ij}^{AM1} \) offers a tremendous flexibility and allows for tightly focused fine-tuning to achieve a better agreement with reference data. Simultaneously, the addition of Gaussian functions introduces a high degree of arbitrariness, which has already been noted by Dewar and co-workers when they introduced this modification\textsuperscript{51}

A popular reparameterization of AM1\textsuperscript{51} was presented by Rocha et al. under the name Recife model 1 (RM1)\textsuperscript{55} Its formalism is identical to that of AM1\textsuperscript{55}

The AM1 model was also generalized to include \( d \) orbitals (AM1/d\textsuperscript{167}) in the same way in which MNDO was generalized to MNDO/d. Additionally,
Figure 9: Dependence of a) $f_{IJ}^{AM1}$ (red line) and b) $V_v^{AM1}$ (red line) on $\tilde{R}_{IJ}$ in H$_2$. Diagram a) illustrates the contributions of the 3 Gaussian functions (dashed lines) with $K_0^1 = 0.10$, $L_0^1 = 5.0$ Å$^{-2}$, $M_0^1 = 0.80$ Å (blue dashed line), $K_1^1 = -0.15$, $L_1^1 = 20.0$ Å$^{-2}$, $M_1^1 = 1.6$ Å (orange dashed line), and $K_2^1 = 0.15$, $L_2^1 = 20.0$ Å$^{-2}$, $M_2^1 = 2.1$ Å (green dashed line). b) Comparison of $V_v^{AM1}$ (red line) with $V_v$ (black line).

The scaling factor $f_{IJ}^{AM1}$ is usually slightly modified for heavier elements,

$$f_{IJ}^{AM1} = 1 + 2x_{Z_I, Z_J} \exp \left( -\alpha_{Z_I, Z_J} \tilde{R}_{IJ} \right)$$

$$+ \sum_{a=1}^{A_{Z_I}} K_a^{Z_I} \exp \left( -L_a^{Z_I} \left( \tilde{R}_{IJ} - M_a^{Z_I} \right)^2 \right)$$

$$+ \sum_{a=1}^{A_{Z_J}} K_a^{Z_J} \exp \left( -L_a^{Z_J} \left( \tilde{R}_{IJ} - M_a^{Z_J} \right)^2 \right),$$

so that it contains element-pair-dependent parameters $x_{Z_I, Z_J}$ (denoted as $\delta$ in Ref. [167]) and $\alpha_{Z_I, Z_J}$ [167].

The acronym AM1* denotes a popular reparameterization of the AM1/d model [168–176] which is implemented in the EMPIRE suite of programs [150].

4.5.3 The Parametrized Models (PMx)

Stewart introduced three popular NDDO-SEMO models, the parametrized models (PMx, $x = 3, 6, 7$) [53, 55, 57]. The parametrized models regard all element-dependent parameters as independent [52] and the element-dependent parameters are calibrated individually, hence the name. The MNDO/d model specifies a set of auxiliary orbital exponents $\zeta_{Z_I}^{s}$, $\zeta_{Z_I}^{p}$, and $\zeta_{Z_I}^{d}$ which are
deduced from the parameters $\gamma_{ss}^{Zt}$, $\gamma_{pp}^{Zt}$, and $\gamma_{dd}^{Zt}$, respectively. In the PMx models, the conceptual relation of the auxiliary orbital exponents to $\gamma_{ss}^{Zt}$, $\gamma_{pp}^{Zt}$, and $\gamma_{dd}^{Zt}$ is ignored for main-group elements. For several transition metals, the one-center ERIs $\langle s^I s^J | d_{xx}^I d_{xx}^J \rangle = \langle s^I s^J | d_{yz}^I d_{xy}^J \rangle$ is analogous to Eq. (57) (MNDO/d core-core repulsion energy), but applies a different scaling factor, $f_{IJ}^{\text{PM6}}$, and adds an additional term to each pairwise interaction. This additional term re-introduces a singularity for $\tilde{R}_{IJ} = 0$ (see also Figure 10). One could therefore conclude that the expression is physically more consistent. The term was designed to resemble the repulsive part of the Lennard-Jones potential. The prefactor of $10^{-8}$ electron volt (eV) is an empirical choice which is not further commented on in Ref. 56. It appears to be chosen such that $V_v^{\text{PM6}}$ is only affected by the Lennard-Jones-like term for very small $\tilde{R}_{IJ}$.

The PM6 model (s, p, d basis), by contrast, features several modifications with respect to PM3 and, hence, AM1. Most prominently, the parametric expression which is applied to calculate $V_v$, was modified even further. In general, the PM6 core-core repulsion energy, $V_v^{\text{PM6}}$, is given by

$$V_v^{\text{PM6}} = \sum_{I=1}^{N} \sum_{J> I}^{N} \left( Q_I Q_J \left[ \varrho_{\text{core}}^I \varrho_{\text{core}}^J \right] \cdot f_{IJ}^{\text{PM6}} + 10^{-8} \text{eV} \left( \frac{Z_i^{1/3} Z_j^{1/3}}{\tilde{R}_{IJ}} \right)^{12} \right).$$

Eq. 62 resembles Eq. 57 (MNDO/d core-core repulsion energy), but applies a different scaling factor, $f_{IJ}^{\text{PM6}}$, and adds an additional term to each pairwise interaction. This additional term re-introduces a singularity for $\tilde{R}_{IJ} = 0$ (see also Figure 10). One could therefore conclude that the expression is physically more consistent. The term was designed to resemble the repulsive part of the Lennard-Jones potential. The prefactor of $10^{-8}$ electron volt (eV) is an empirical choice which is not further commented on in Ref. 56. It appears to be chosen such that $V_v^{\text{PM6}}$ is only affected by the Lennard-Jones-like term for very small $\tilde{R}_{IJ}$.

The scaling factor $f_{IJ}^{\text{PM6}}$ was constructed in analogy to the scaling factor which was proposed for heavier elements for the AM1/d model (Eq. 61),

$$f_{IJ}^{\text{PM6}} = 1 + 2 \alpha_i \alpha_j x^{Zt_i Zj} \exp \left( -\alpha_i^{Zt_i} \left( \tilde{R}_{IJ} + 0.0003 \text{Å}^{-5} \tilde{R}_{IJ}^6 \right) \right) + K^{Zt_i} \exp \left( -L^{Zt_i} \left( \tilde{R}_{IJ} - M^{Zt_i} \right)^2 \right) + K^{Zj} \exp \left( -L^{Zj} \left( \tilde{R}_{IJ} - M^{Zj} \right)^2 \right).$$

It also contains element-pair-dependent parameters $x^{Zt_i Zj}$ and $\alpha_i^{Zt_i}$. In comparison to Eq. (6) in Ref. 56, we replaced $x^{Zt_i Zj}$ for $2 \alpha_i^{Zt_i}$; this is necessary to achieve an agreement with the implementation in MOPAC with
Figure 10: \( V_{PM6}^v \) (red line) and \( V_v \) (black line) in a.u. for \( H_2 \) for various distances \( \tilde{R}_{IJ} \) in Å. \( V_{PM6}^v \) is decomposed in the contribution of the first term in Eq. (61) (blue dashed line) and that of the second term in Eq. (61) (green dashed line).

the parameter values reported for \( x_{Z_I,Z_J} \) in Ref. 56. Stewart restricted the number of additional Gaussian functions to one per element, so that we do not have to specify an index anymore for the parameters characterizing the Gaussian functions (i.e., \( K_{Z_I}, L_{Z_I}, \) and \( M_{Z_I} \) instead of \( K_{a}, L_{a}, \) and \( M_{a} \), respectively). In comparison to \( f_{IJ}^{AM1} \), a term \( 0.0003 \text{Å}^{-5} \tilde{R}_{IJ}^6 \) was added to the exponential scaling function. Apparently, this modification enabled a better agreement with reference data for rare-gas compounds.\(^{56}\) The PM6 model defines additional special expressions which are only applied for certain compound classes or for the evaluation of the scaling factors for certain atom pairs (i.e., C–H, N–H, O–H, C–C, and Si–O). We discuss these minor modifications in the Supporting Information.

Disturbingly, the PM6 model contains special corrections to the heats of formations at 298 K, \( \Delta H_{298K}^f \), for several compound classes. In the PM6 model, the predicted \( \Delta H_{298K}^f \) in kcal mol\(^{-1}\) is empirically modified depending on a measure for the non-planarity of the amine nitrogen atom, \( \phi \).\(^{56}\)

\[
\Delta H_{298K}^f = \Delta H_{298K,PM6}^f - 0.5 \text{ kcal mol}^{-1} \exp (-10\phi). \tag{64}
\]

The measure for the non-planarity of the amine nitrogen atom is determined as \( 2\pi \) minus the sum of the three bond angles involving the amine nitrogen atom. For a perfectly planar amine, \( \Delta H_{298K}^f \) is reduced by 0.5 kcal mol\(^{-1}\). With an increasing pyramidalization of the amine, \( \Delta H_{298K}^f \) is reduced by a smaller amount. Additionally, the PM6 model (as implemented in MOPAC)
includes an undocumented modification to \( \Delta H_f^{298K} \) when the computed bond order for a carbon–carbon bond exceeds 2.5, as for example, in acetylenic bonds. A contribution of 12.0 kcal mol\(^{-1}\) is added to \( \Delta H_f^{298K} \) for every detected acetylenic bond, e.g., MOPAC outputs \( \Delta H_f^{298K} = 57.4 \) kcal mol\(^{-1}\) for acetylene. If one applies the formulae specified in Ref. 56 instead, one would obtain \( \Delta H_f^{298K} = 45.4 \) kcal mol\(^{-1}\) for this molecule.

The PM7 model was introduced as the successor of the PM6 model in 2013.\(^{57}\) The largest changes were again made to the core-core repulsion energy. It became evident that it is essential that the two-center ERIs decrease to the exact value at large distances when applying a SEMO model in periodic calculations.\(^{149}\) Hence, Eq. (39) was modified so that the ERI \( [s^I s^I | s^J s^J] \) is approximated as,

\[
[s^I s^I | s^J s^J]_{PM7} = \frac{1}{\tilde{R}_{IJ}} \exp \left( -0.22 \AA^{-2} (\tilde{R}_{IJ} - 7.0 \AA)^2 \right) + \left( 1 - \exp \left( -0.22 \AA^{-2} (\tilde{R}_{IJ} - 7.0 \AA)^2 \right) \right) \times \left[ \tilde{R}_{IJ}^2 + \left( \frac{1}{2\gamma_{ss}} + \frac{1}{2\gamma_{ss}} \right)^2 \right]^{-\frac{1}{2}}. 
\]

The value 7.0 Å was apparently chosen as some random distance which is far larger than usual bond lengths.\(^{57}\) This equation is also consulted to evaluate \( V_v^{PM7} \),

\[
V_v^{PM7} = \sum_{I=1}^{N} \sum_{J>I}^{N} Q_I Q_J \left[ g^I_{core} | g^J_{core} \right]_{PM7} \cdot f^{PM6}_{IJ} + 10^{-8} \text{eV} \left( \frac{Q_I^{1/3} Q_J^{1/3}}{\tilde{R}_{IJ}} \right)^{12}. 
\]

In Eq. (66), we did not include the additional empirical corrections for hydrogen bonding and dispersion interactions which are inherent to the PM7 model and described in Refs. 57,89. Note that the description of dispersion interactions in PM7 creates a conceptual problem as pointed out by Grimme et al.\(^{179}\) Ref. 57 states that the dispersion energy is damped down and truncated at longer distances. This is obviously not sensible for dispersion interactions which are long-range interactions and was also shown to cause significant errors for larger systems.\(^{180,181}\)

5 The Orthogonalization-Corrected Models (OM\(x\))

The OM\(x\) \((x = 1, 2, 3)\) models activate one \( s\)-type basis functions for hydrogen and one \( s\)- and three \( p\)-type basis functions for carbon, nitrogen, oxygen,
and fluorine.\(^{58,60,151,152}\) Each of these basis function consists of three primitive Gaussian functions, \(^{58,60,117}\) and hence, we denote the basis sets for the OM1, OM2, and OM3 models with OM1-3G, OM2-3G, and OM3-3G, respectively. The OM\(x\)-3G basis sets are based on the ECP-3G basis set.\(^{117}\) The exponents of the primitive Gaussian functions of the ECP-3G basis are scaled with \((\zeta Z_I)^2\) to yield the OM\(x\)-3G basis sets. The factor \(\zeta Z_I\) is a parameter of the respective OM\(x\) model.\(^{60}\) The OM\(x\) models currently do only provide parameters for hydrogen, carbon, nitrogen, oxygen, and fluorine.\(^{60}\)

5.1 Approximation of Electron-Electron Repulsion Integrals

In analogy to MNDO-type models, the five one-center ERIs \(\langle \chi_I^\mu \chi_I^\nu | \chi_J^\lambda \chi_J^\sigma \rangle\) arising in the minimal \(s,p\) basis are substituted for the parameters \(\gamma^{ss}_{\alpha \alpha}, \gamma^{pp}_{\beta \beta}, \gamma^{sp}_{\alpha \beta}, \gamma^{sp}_{\beta \alpha}\) (see also Eqs. (32)–(36)).

Within the OM\(x\) models, the value of the two-center ERIs \(\langle \chi_I^\mu \chi_I^\nu | \chi_J^\lambda \chi_J^\sigma \rangle, I \neq J\) is determined analytically. The analytical values of the two-center ERIs are then scaled with the so-called Klopman–Ohno factor \(f_{IJ}^{KO}\) when assembling \(\phi^G \approx \chi^G\) NDDO, \(\langle \chi_I^\mu \chi_I^\nu | \chi_J^\lambda \chi_J^\sigma \rangle = f_{IJ}^{KO} \cdot \langle \chi_I^\mu \chi_I^\nu | \chi_J^\lambda \chi_J^\sigma \rangle\) \(\langle s^I s^J | s^I s^J \rangle\).\(^{68}\)

\[
f_{IJ}^{KO} = \left[ \frac{s^I s^J | s^I s^J}{\langle s^I s^J | s^I s^J \rangle} \right] \left[ \frac{\tilde{R}_{IJ}^2 + \left( \frac{1}{2\gamma_{ss}} + \frac{1}{2\gamma_{pp}} \right) \gamma_{sp}^2}{\langle s^I s^J | s^I s^J \rangle} \right]^{-\frac{1}{2}}. \tag{68}\]

When \(\chi_I^\mu, \chi_I^\nu, \chi_J^\lambda, \) and \(\chi_J^\sigma\) are \(s\)-type basis functions, the product of \(f_{IJ}^{KO}\) and \(\langle s^I s^J | s^I s^J \rangle\) reduces to the MNDO-type model,

\[
\langle s^I s^J | s^I s^J \rangle^{OMx} = \left[ \frac{s^I s^J | s^I s^J}{\langle s^I s^J | s^I s^J \rangle} \right] \langle s^I s^J | s^I s^J \rangle = \left[ \frac{s^I s^J | s^I s^J}{\langle s^I s^J | s^I s^J \rangle} \right]. \tag{69}\]

In the case of one basis function, \(\chi_I^\mu, \chi_I^\nu, \chi_J^\lambda, \) or \(\chi_J^\sigma,\) not being an \(s\)-type basis function, the two-center ERI is still scaled with \(f_{IJ}^{KO}\).

Let us now examine how this Klopman–Ohno scaling affects the value which enters the two-electron matrices. Generally, \(f_{IJ}^{KO}\) tends to one in the limit \(\tilde{R}_{IJ} \rightarrow \infty\), i.e., the unscaled analytical value for \(\langle \chi_I^\mu \chi_I^\nu | \chi_J^\lambda \chi_J^\sigma \rangle\) is applied. It would be theoretically satisfactory if we had a smooth transition from the
one-center ERIs \( \langle \chi^I_{\mu} \chi^I_{\sigma} | \chi^J_{\lambda} \chi^J_{\sigma} \rangle = \gamma_{ss}^I, \gamma_{pp}^I, \gamma_{sp}^I, \gamma_{sp}'^I \) to the two-center ERIs \( f_{I,J}^{KO} \cdot \langle \chi^I_{\mu} \chi^J_{\nu} | \chi^I_{\lambda} \chi^J_{\sigma} \rangle \). This situation would occur if the analytical one-center limits were chosen for \( \gamma_{ss}^I, \gamma_{pp}^I, \gamma_{sp}^I, \gamma_{sp}'^I \), and \( \gamma_{sp}^I \). If \( \gamma_{ss}^I = \langle s^I s^I | s^I s^I \rangle \), \( f_{I,J}^{KO} \) would be one in the limit \( \tilde{R}_{IJ} = 0 \),

\[
f_{I,J}^{KO} = \frac{\gamma_{ss}^I}{\langle s^I s^I | s^I s^I \rangle} = \frac{\langle s^I s^I | s^I s^I \rangle}{\langle s^I s^I | s^I s^I \rangle} = 1.
\]

(70)

Hence, all two-center ERIs are scaled with a factor of one in the limit \( \tilde{R}_{IJ} = 0 \) and the analytical one-center limit is recovered. In this case, the scaled two-center ERIs differ negligibly from the analytical two-center ERIs (see Figure 11).

**Figure 11:** Dependence of a) the Klopman–Ohno factor \( f_{I,J}^{KO} \) and b) the value of \( f_{I,J}^{KO} \) multiplied by \( \langle s^I p^I_z | s^I p^I_z \rangle \) on \( \tilde{R}_{IJ} \) in a \( C_2 \) molecule described by an OM2-3G basis. We show the values for \( f_{I,J}^{KO} = 1 \) (black lines), Eq. (67) with \( \gamma_{ss}^I = 0.81 \) a.u. (analytical one-center limit, red lines), Eq. (67) with \( \gamma_{ss}^I = 0.65 \) a.u. (green lines), and Eq. (67) with \( \gamma_{ss}^I = 0.45 \) a.u. (OMx value, blue lines). The analytical one-center limit \( \langle s^I s^I | s^I s^I \rangle = 0 \) a.u. is highlighted by a black arrow and the OMx one-center limit \( \gamma_{sp}^I = 0.42 \) a.u. is not shown here.

Usually, \( \gamma_{ss}^I \) is, however, chosen to be significantly smaller than the analytical one-center ERI limit \( \langle s^I s^I | s^I s^I \rangle - \gamma_{ss}^I = 0.36 \) a.u. in our example in Figure 11. Consequently, a theoretically unsatisfactory situation arises which we illustrate at the example of the \( \langle s^I p^I_z | s^I p^I_z \rangle \) ERI in \( C_2 \). The parametrized one-center ERI limits of interest are \( \gamma_{ss}^6 = 0.45 \) a.u. and \( \gamma_{sp}^6 = 0.42 \) a.u. We then observe a discontinuity from the point
\[ \hat{R}_{I,J} = 0 \] where \( \langle s^I p^I_z | s^J p^J_z \rangle_{\text{OM} \times} = \gamma_{sp}^h = 0.42 \text{ a.u.} \) to the point \( \hat{R}_{I,J} \ll 1 \) where \( \langle s^I p^I_z | s^J p^J_z \rangle_{\text{OM} \times} = f_{K}^{\text{KO}} \langle s^I p^I_z | s^J p^J_z \rangle = 0.09 \text{ a.u.} \) While these discontinuities are unsatisfactory, they do not appear to lead to practical issues in the calculations.

5.2 Approximation of the Symmetrically Orthogonalized One-Electron Matrix

An exact orthogonalization of \( \chi H \) (e.g., by the transformation \( \chi S^{-\frac{1}{2}} \chi H \chi S^{-\frac{1}{2}} \)) was initially attempted, but did not provide a useful model\(^{58,59,182}\) as confirmed by other studies\(^{122,154,159,184-187}\). Kolb and Thiel therefore decided to develop approximate orthogonalization corrections to be added to \( \chi H \).\(^{58,182}\) These and other\(^{162,188-190}\) approximate orthogonalization corrections, are based on an expansion of \( \chi S^{-\frac{1}{2}} \) into a power series,

\[
\chi S^{-\frac{1}{2}} = (1 + \chi S')^{-\frac{1}{2}} = 1 - \frac{1}{2} \chi S' + \frac{3}{8} \chi S'^2 - \frac{5}{16} \chi S'^3 + O(\chi S'^4),
\]

(71)

where \( \chi S' \) is defined as

\[
\chi S' = \chi S - 1.
\]

(72)

The transformation of \( \chi H \) to \( \phi H \),

\[
\phi H = (\chi S')^{-\frac{1}{2}} \chi H (\chi S')^{-\frac{1}{2}} = (1 + \chi S')^{-\frac{1}{2}} \chi H (1 + \chi S')^{-\frac{1}{2}}
\]

(73)

is then approximated as

\[
\phi H \approx \chi H - \frac{1}{2} (\chi S' \chi H + \chi H \chi S') + \frac{3}{8} (\chi S'^2 \chi H + \chi H \chi S'^2)
\]

\[ + \frac{1}{4} \chi S' \chi H \chi S' + O(\chi S'^3).
\]

(74)

Accordingly, a matrix element \( \phi H_{\mu\nu} \) is approximated as

\[
\phi H_{\mu\nu} \approx \chi H_{\mu\nu} - \frac{1}{2} \sum_{\lambda=1}^{M} (\chi S'_{\mu\lambda} \chi H_{\lambda\nu} + \chi H_{\mu\lambda} \chi S'_{\lambda\nu})
\]

\[ + \frac{1}{8} \sum_{\lambda=1}^{M} \sum_{\sigma=1}^{M} (3 \chi S'_{\mu\lambda} \chi S'_{\lambda\sigma} \chi H_{\sigma\nu} + 3 \chi H_{\mu\lambda} \chi S'_{\lambda\sigma} \chi S'_{\sigma\nu} + 2 \chi S'_{\mu\lambda} \chi H_{\lambda\sigma} \chi S'_{\sigma\nu}).
\]

(75)

Gray and Stone showed\(^{191}\) that this power series expansion is nonconvergent in the general case. More specifically, the power series expansion fails...
to converge when the largest eigenvalue of $\chi'S'$ exceeds 1.091 which is often the case (e.g., it is 1.3 for methane and 2.1 for benzene when applying an ECP-3G basis set). Chandler and Gradee\cite{121} and Neymeyr\cite{110,111} subsequently introduced alternative convergent power series expansions. These were, however, not applied to derive approximate orthogonalization corrections for the OMx models, nor for any other NDDO-SEMO model. The nonconvergence of Eq. (74) does not appear to be a problem in practice, which may be attributed to the fact that it was only taken as a guideline to develop parametric expressions. Eq. (74) is therefore not directly applied to carry out the basis transformation in the OMx models (recall that the exact transformation of $\chi^H$ from the $\chi$- to the $\phi$-basis does not yield a useful NDDO-SEMO model).

Analogously to the MNDO model, different parametric expressions are applied for the evaluation of the matrix elements $\phi_{HOMx}^{\mu\nu}$ depending on the number of atoms on which the corresponding basis functions $\chi$ are centered (one-center one-electron and two-center one-electron matrix elements).

5.2.1 One-Center One-Electron Matrix Elements

We first discuss how $\phi_{HOMx}^{\mu\nu}$ is determined when the corresponding basis functions $\chi$ are centered on the same atom, i.e., $\chi^I_\mu$ and $\chi^I_\nu$ are both centered on the atom $I$. In this case, the matrix elements $\phi_{HOMx}^{\mu\nu}$ are given by

$$
\phi_{HOMx}^{\mu\nu} = \chi_{HOMx}^{\mu\nu} - \frac{1}{2} F_1^{Z_1} \sum_{\lambda=1}^M \chi_{S_{\mu\lambda}}' \theta(\chi^I_{\nu}, \chi^I_{\lambda}) + \theta(\chi^I_{\mu}, \chi^I_{\lambda}) \chi_{S_{\nu\lambda}}' \\
+ \frac{1}{8} F_2^{Z_2} \sum_{\lambda=1}^M \chi_{S_{\mu\lambda}}' \chi_{S_{\nu\lambda}}' (\eta(\chi^I_{\nu}, \chi^I_{\lambda}) + \eta(\chi^I_{\mu}, \chi^I_{\lambda}) - \eta(\chi^I_{\lambda}, \chi^I_{\mu})) - \eta(\chi^I_{\nu}, \chi^I_{\lambda})
$$

(76)

where $F_1^{Z_1}$ and $F_2^{Z_2}$ are element-dependent parameters. Additionally, Eq. (76) contains the functions $\theta(\chi^I_{\nu}, \chi^I_{\lambda})$ and $\eta(\chi^I_{\nu}, \chi^I_{\lambda})$ which we will specify in the following paragraphs. When comparing Eqs. (75) and (76), we notice several similarities and differences. Firstly, both equations start with the corresponding matrix element in the $\chi$-basis. The second contribution to $\phi_{HOMx}^{\mu\nu}$ is similar to the second contribution to $\phi_{HOMx}^{\mu\nu}$. The analytical expression (Eq. (75)) does not contain the parameter $F_1^{Z_1}$ which implies that $F_1^{Z_1}$ should be close to one when the analytical expression is approximated (cf. Ref. 58). The function $\theta(\chi^I_{\nu}, \chi^I_{\lambda})$ appears to model an entry in $\chi^H$. More specifically, it models the matrix elements $\chi H_{\mu\lambda}$ when $\chi^I_{\mu}$ and $\chi^I_{\lambda}$ are centered on different atoms, i.e., $I \neq J$. This is apparent when studying $\chi_{S_{\mu\lambda}}'$ which is only different from
zero when \( \chi^I_\mu \) and \( \chi^J_\lambda \) are centered on different atoms due to the condition of local orthogonality and the way in which \( \chi S' \) is constructed (Eq. (72)). In summary, the second contribution to Eq. (76) is identical to the analytical expression when \( F^Z I_1 = 1 \) and \( \theta(\chi^I_\mu, \chi^J_\lambda) = \chi^H_\mu \lambda \). The relation of the third contribution to Eqs. (76) and (75) is harder to see. It includes the approximation that all four-center contributions are neglected, see Ref. 59 for a detailed derivation. Hence, it is impossible to establish a similar relationship between \( F^Z I_2 \) and \( \eta(\chi^I_\mu, \chi^J_\lambda) \) in the parametric expression and analogs in the analytical expression. In the OM3 model, \( F^Z I_2 = 0 \) whereas in the OM1 and OM2 models both orthogonalization corrections are considered.

We now examine how the contributions to \( \phi^H_{\text{OM}} x^\mu_\nu \) are evaluated. Its first contribution, \( \chi^H_{\text{OM}} x^\mu_\nu \), on the right-hand side of Eq. (76) is given by

\[
\chi^H_{\text{OM}} x^\mu_\nu = U^Z I_{(\mu)(\nu)} + \sum_{J \neq I}^N \langle \chi^I_\mu | \text{ECP}_J | \chi^I_\nu \rangle - \sum_{J \neq I}^N \{ Q_J \left[ \chi^I_\mu | s^J s^J \right] - f_{IJ}^K \langle \chi^I_\mu \chi^I_\nu | s^J s^J \rangle \} \]

(77)

where the element- and orbital-type-dependent parameter \( U^Z I_{(\mu)(\nu)} \) replaces the calculation of \( \langle \chi^I_\mu | -\frac{1}{2} \nabla^2 | \chi^I_\nu \rangle \) (cf. the MNDO model). Kolb and Thiel explicitly stated that \( U^Z I_{(\mu)(\nu)} \) is also assumed to include all contributions from the core electrons.58

The next term is the contribution from the ECP with all other atoms \( J \). In the first OM\( x \) model, OM1, the contributions to the one-electron matrix are evaluated analytically as presented in Refs. 117,192. The analytical results were then subjected to Klopman–Ohno scaling.55 In the OM2 and OM3 models, the \textit{ab initio} effective core potential in the \( \chi \)-basis was substituted for a semiempirical one.59,60

\[
\langle \chi^I_\mu | \text{ECP}_K | \chi^I_\nu \rangle \approx - \langle \chi^I_\mu | \omega^K \rangle \theta(\chi^I_\nu, \omega^K) - \theta(\chi^I_\mu, \omega^K) \langle \omega^K | \chi^I_\nu \rangle - \langle \chi^I_\mu | \omega^K \rangle \langle \omega^K | \chi^I_\nu \rangle W^Z(K).
\]

(78)

In this expression, we introduced an auxiliary set of basis functions \( \omega = \{ \omega^I \} \) (no additional subscript index is necessary because there is at most one additional \( s \)-type basis function per atom). Each basis function \( \omega \) is
characterized by an orbital exponent $\zeta^Z_I$. Note that an orbital $\omega^I$ is generally not locally orthogonal to $\chi^I_\mu$. An additional element-dependent parameter $W^Z(K)$ enters Eq. (78).

The last contribution to Eq. (77) describes the interaction of the charge distribution $\chi^I_\mu \chi^I_\nu$ with all other atomic nuclei $J \neq I$. It is composed of three contributions: The first one is identical to the one applied to describe this interaction in the MNDO model (Eq. (50)). As stated in Section 4.3.1, this expression is the result of an error compensation between the so-called penetration integrals (Eq. (51)) and the orthogonalization corrections. Because the OM$^x$ models explicitly consider orthogonalization corrections, also the penetration integrals must be considered which make up the last two terms (subjected to Klopman–Ohno scaling). If $\chi^I_\mu$ and $\chi^I_\nu$ are $s$-type orbitals, Eq. (77) will reduce to

$$\chi H^{\text{OM}^x}_{\mu\nu} = U_{ss} - \sum_{J \neq I}^N f_{IJ}^{\text{KO}} \left\langle s^I | \frac{Q_J}{|r_i - \tilde{R}_J|} | s^J \right\rangle + \sum_{J \neq I}^N f_{IJ}^{\text{KO}} \left\langle s^I | \text{ECP}_J | s^J \right\rangle,$$  

(79)

because the first two remaining terms in Eq. (77) cancel out exactly,

$$Q_J \left[ s^I s^J | s^I s^J \right] - f_{IJ}^{\text{KO}} Q_J \left\langle s^I s^J | s^I s^J \right\rangle = Q_J \left[ s^I s^J | s^I s^J \right]$$

$$-Q_J \left( \left\langle s^I s^J | s^I s^J \right\rangle \left\langle s^I s^J | s^I s^J \right\rangle \right) = 0.$$  

(80)

If either $\chi^I_\mu$ or $\chi^I_\nu$ are not $s$-type orbitals, Eq. (80) will not hold true and we will have a contribution $Q_J \left( \left[ \chi^I_\mu \chi^I_\nu | s^I s^J \right] - f_{IJ}^{\text{KO}} \left\langle \chi^I_\mu \chi^I_\nu | s^I s^J \right\rangle \right)$ to Eq. (77).

Generally, however, the differences in the results of Eq. (77) and Eq. (79) are quite small.

The function $\theta (\chi^I_\mu, \chi^I_\nu)$ consists of the product of a phase vector, $(-1)^{l(\lambda) + m(\lambda)}$, the scaled square-root of the interatomic distance $\tilde{R}_{IJ}$, and a Gaussian contribution depending on $\tilde{R}_{IJ}^2$ (see Figure 12). The function $\theta (\chi^I_\mu, \chi^I_\nu)$ is evaluated in the same local coordinate system which is applied in MNDO-type methods (see Section 4.2) and have to be transformed accordingly. The scaling factor for the square-root of the interatomic distance $\tilde{R}_{IJ}$ is determined from element- and orbital-type-dependent parameters $b_{l(\mu)}$. The width of the Gaussian function is determined by element- and orbital-type-dependent parameters $a_{l(\mu)}$, it
determines how fast the whole function approaches to zero. The parameters $a_{l(\mu)}^Z$ must be positive to obtain a sensible expression. Within OMO models, the parameters $b_{l(\mu)}$ are unanimously negative. Hence, the phase vector determines the sign of $\theta(\chi^I_{\mu}, \chi^J_{\lambda})$. Consequently, the sign of $\theta(\chi^I_{\mu}, \chi^J_{\lambda})$ does not depend on $\tilde{R}_{IJ}$. The function $\theta(\chi^I_{\mu}, \chi^J_{\lambda})$ was designed to emulate $H_{\mu\lambda}$. The matrix element $H_{\mu\lambda}$, which the function $\theta(\chi^I_{\mu}, \chi^J_{\lambda})$ is supposed to model, may, however, have a different sign for different $\tilde{R}_{IJ}$. In fact, Kolb and Thiel included an example where $H_{\mu\lambda}$ changes its sign for different $\tilde{R}_{IJ}$ in Figure 2 of Ref. 58 (see also Figure 12 in the present work). Furthermore, it is inter-

Figure 12: Dependence of $H_{\mu\nu}$ (solid lines) and $\theta(\chi^I_{\mu}, \chi^J_{\lambda})$ (dashed lines) on the distance $\tilde{R}_{IJ}$ in a C$_2$ molecule described in an OM2-3G basis ($\chi = \{ \chi_1^{C_1}, \chi_2^{C_1}, \chi_3^{C_1}, \chi_4^{C_1}, \chi_5^{C_2}, \chi_6^{C_2}, \chi_7^{C_2}, \chi_8^{C_2} \}$ where the first and fifth basis functions are 2s, the second and sixth are 2px, the third and seventh are 2py, and the fourth and eighth are 2pz basis functions). a) $\theta(s^{C_1}, s^{C_2})$ with the best-fit parameters $\beta^6_s = -38.9$ a.u. and $a^6_s = 1.0$ a.u. The contributions to $\theta(s^{C_1}, s^{C_2})$ are divided into the scaled square-root contributions (dashed gray line) and the Gaussian contribution (dashed-dotted gray line). b) $\theta(s^{C_1}, s^{C_2})$ with $\beta^6_s = -38.9$ a.u. and $a^6_s = 1.0$ a.u. (black lines), $\theta(s^{C_1}, p_z^{C_2})$ with $\beta^6_p = \beta^6_s = -11.2$ a.u. and $a^6_s = a^6_p = 0.6$ a.u. (red lines), $\theta(p_x^{C_1}, p_z^{C_2})$ with $\beta^6_p = -0.2$ a.u. and $a^6_p = 0.1$ a.u. (blue lines), and $\theta(p_x^{C_1}, p_z^{C_2})$ with $\beta^6_p = -4.2$ a.u. and $a^6_p = 1.2$ a.u. (green lines).

esting that no ECP is explicitly considered for the two-center one-electron matrix elements which is apparently assumed to be absorbed into Eq. (81).

Finally, we need to evaluate the function $\eta(\chi^I_{\mu}, \chi^J_{\lambda})$ to assemble Eq. (76). It is evaluated similarly to the corresponding local one-electron matrix ele-
ment in the MNDO model. The function \( \eta(\chi^I, \chi^K) \) is evaluated as

\[
\eta(\chi^I, \chi^K) = U^Z_I - Q_K [\chi^I \chi^K | s^K s^K] 
\]  
(82)

(cf. Eq. (47)). Note that the order in which the basis functions are written in the function matters, i.e., \( \eta(\chi^I, \chi^K) \neq \eta(\chi^K, \chi^I) \). The function \( \eta(\chi^I, \chi^K) \) is not rotationally invariant, so that the functions have to be averaged when \( \chi^I \) is a \( p \)-type basis function,

\[
\eta(p^I, \chi^K) = \frac{1}{3} (\eta(p^I_x, \chi^K) + \eta(p^I_y, \chi^K) + \eta(p^I_z, \chi^K)) 
\]  
(83)

5.2.2 Two-Center One-Electron Matrix Elements

While Kolb and Thiel explicitly pointed out that \( \theta(\chi^I, \chi^J) \) is assumed to contain orthogonalization corrections,\(^{\text{58}}\) it turned out that it cannot accomplish this fully.\(^{\text{59}}\) As a remedy, Weber and Thiel\(^{\text{59}}\) developed an orthogonalization correction for matrix elements \( \phi_{HOMx}^{\mu\nu} \) for which \( \chi^I \) and \( \chi^J \) are centered on different atoms,

\[
\phi_{HOMx}^{\mu\nu} = \theta(\chi^I, \chi^J) 
\]

\[
- \frac{1}{2} G_1^{ZI} + \frac{G_1^{ZJ}}{2} \sum_{\lambda=1}^{M} (1 - \delta_{IK})(1 - \delta_{JK}) \left( \chi^{I\lambda}_\mu \chi^{J\lambda}_\nu \theta(\chi^I, \chi^K) + \theta(\chi^J, \chi^K) \chi^{I\lambda}_\nu \chi^{J\lambda}_\mu \right) 
\]

\[
+ \frac{1}{8} G_2^{ZI} + \frac{G_2^{ZJ}}{2} \sum_{\lambda=1}^{M} (1 - \delta_{IK})(1 - \delta_{JK}) \chi^{I\lambda}_\mu \chi^{J\lambda}_\mu \left( \eta(\chi^I, \chi^K) + \eta(\chi^J, \chi^K) - \eta(\chi^K, \chi^I) - \eta(\chi^K, \chi^J) \right) 
\]  
(84)

which contains the element-dependent parameters \( G_1^{ZI} \) and \( G_2^{ZI} \). This equation significantly differs from the analytic expression for the transformation of the matrix elements (Eq. (75)). Most importantly, the corrections do only include terms with basis functions which are centered on a third atom \( (K \neq I \neq J; \text{indicated by } (1 - \delta_{IK})(1 - \delta_{JK}) \text{ in Eq. (84)). Other than that similar considerations apply as in the one-center case. The similarity of Eqs. (77) and (84) might imply that } F_1^{ZI} \approx 2 \cdot G_1^{ZI} \text{ and } F_2^{ZI} \approx 2 \cdot G_2^{ZI} \text{ which is, however, generally not the case.}

In the earliest variant, OM1, \( G_1^{ZI} = 0 \) and \( G_2^{ZI} = 0 \), i.e., orthogonalization corrections are only considered when \( \chi^I \) and \( \chi^J \) are centered on a single atom. The latest version, OM3, sets \( F_2^{ZI} = 0 \) and \( G_2^{ZI} = 0 \). The OM2 model considers all orthogonalization corrections.
5.3 Empirical Scaling of the Core-Core Repulsion Energy

The contribution of each pairwise repulsion of atomic cores $I, J \neq I$ is scaled with the Klopman–Ohno factor $f_{KIO}^{IJO}$:

$$V_{OM}^{OMx} = \sum_{I=1}^{N} \sum_{J>I}^{N} f_{KIO}^{IJO} \frac{Q_I Q_J}{\tilde{R}_{IJ}}.$$  \hspace{1cm} (85)

It is argued\textsuperscript{58–60,152} that the core-core repulsion energy needs to be reduced for small interatomic distances to ensure a balance within the model (ERIs and contributions to the core-electron attraction terms are also scaled with $f_{KIO}^{IJO}$). In contrast to the MNDO-type core-core repulsion (Eqs. (54), (59), and (62)), Eq. (85) has a singularity for $\tilde{R}_{IJ} = 0$. When $\tilde{R}_{IJ} \rightarrow \infty$, $f_{KIO}^{IJO} \rightarrow 1$, i.e., the core-core repulsion energy approximates the point-charge model asymptotically.

6 Other NDDO-SEMO models

Several other models, which have not found widespread popularity, introduce new conceptual ideas beyond the MNDO-type and OMx models. In the following sections, several of these ideas are reviewed and discussed without providing the complete formalism for these models. This list is by no means complete. We selected models which feature conceptually large differences to the introduced models, but are still built around the NDDO approximation. For other NDDO-SEMO models see also Refs.\textsuperscript{104,146–148,185,193–198}

6.1 The Nonorthogonalized Modified Neglect of Differential Overlap (NO-MNDO) Model

An obvious weakness of the MNDO model is the lack of explicit orthogonalization corrections to $H$ (Section 4.3). Sattelmeyer et al. claimed that this problem can be addressed by introducing the overlap matrix into the SCF equations\textsuperscript{61} and coined the name Nonorthogonalized Modified Neglect of Differential Overlap (NO-MNDO) model for this procedure.\textsuperscript{61} Let us neglect the presence of parameters for the moment and assume that all integrals are calculated analytically. Their suggestion then reads, in our notation,

$$\chi (\chi H + \chi G^{NDDO}) \chi C = \chi S \chi C \epsilon,$$  \hspace{1cm} (86)
which can be reformulated to
\[
(\phi H + \chi S^{-\frac{1}{2}} \chi G^{\text{NDDO}} \chi S^{-\frac{1}{2}}) \phi C = \phi C \epsilon,
\] (87)
when applying Eq. (23). While it is obvious that the one-electron matrix is explicitly orthogonalized, a problem arises. The NDDO approximation emulates a basis transformation for \(G\) (cf. Eq. (23)). The application of a matrix transformation to \(\chi G^{\text{NDDO}}\) is therefore not sensible. When applying the matrix transformation again, we end up with a situation where
\[
\chi S^{-\frac{1}{2}} \chi G^{\text{NDDO}} \chi S^{-\frac{1}{2}} \neq \phi G
\] (88)
and
\[
\chi S^{-\frac{1}{2}} \chi G^{\text{NDDO}} \chi S^{-\frac{1}{2}} \neq \chi G
\] (89)
(see Figure 13).

Figure 13: Graphical representation of \(\chi G^{\text{NDDO}} - \chi G\) (left), \(\chi G^{\text{NDDO}} - \phi G\) (middle left), \(\chi S^{-\frac{1}{2}} \chi G^{\text{NDDO}} \chi S^{-\frac{1}{2}} - \chi G\) (middle right), and \(\chi S^{-\frac{1}{2}} \chi G^{\text{NDDO}} \chi S^{-\frac{1}{2}} - \phi G\) (right) for a benzene molecule described in an MNDO-3G basis. The entries are colored according to their values from negative (blue) to zero (white) to positive (red).

According to Sattelmeyer et al.\textsuperscript{61} the NO-MNDO model appeared to significantly improve on the MNDO model. We speculate that the good performance of the NO-MNDO model might be due to a combination of two reasons: (i) The parameters in the NO-MNDO model might provide a sufficient flexibility to remedy the conceptual shortcomings and (ii) the explicit orthogonalization of \(\chi H\) might outweigh the conceptual error to a certain extent. The improvements of NO-MNDO in areas which are typically associated with orthogonalization errors for MNDO\textsuperscript{60} (e.g., wrong barriers of rotations about single bonds) might be taken as an indicator for this statement.
In any case, it appears promising to attempt the construction of a similar (MNDO-type) model which does not share the conceptual difficulties of NO-MNDO. Without considering the parametrization, the corresponding SCF equations might read,

\[
\left( \chi S^{-\frac{1}{2}} \chi H S^{-\frac{1}{2}} + \chi G^{\text{NDDO}} \right) \phi_C = \phi_C \epsilon.
\] (90)

It might be possible that such an attempt could result in a model which is more accurate than NO-MNDO, and hence, significantly more accurate than MNDO.

### 6.2 The Polarized Molecular Orbital (PMO\(x\)) Models

The Polarized Molecular Orbital (PMO\(x\), \(x = 1, 2\)) models\(^{62–66}\) were developed in an attempt to provide a more accurate description of noncovalent interactions and polarization effects than possible with the standard NDDO-SEMO models. It is built upon the MNDO model, but features a key difference in its formalism: The PMO\(x\) models activate one \(s\)- and three \(p\)-type basis functions for hydrogen (compared to only one \(s\)-type basis function for MNDO). Truhlar and co-workers determined that the activation of diffuse \(p\)-type basis functions for hydrogen atoms is already sufficient to obtain a significant improvement in the description of polarization effects in \textit{ab initio} studies\(^{62–64}\) Similar results were also published before in a different context\(^{199–201}\). Such a basis set, nevertheless, fulfills the condition of local orthogonality. The addition of \(p\)-type basis functions for hydrogen atoms was, furthermore, accompanied by changes to the parametric expressions applied to evaluate the one-electron matrix elements and the core-core repulsion energy\(^{65}\).

### 6.3 The Machine Learning OM2 (ML-OM2) Model

Dral, von Lilienfeld, and Thiel suggested\(^{67}\) to combine machine learning techniques with NDDO-SEMO models which resulted in the machine learning OM2 (ML-OM2) model. The formalism of the ML-OM2 model is identical to that of the OM2 model. It differs from the OM2 model only in the value of the parameter \((\zeta^6)^2\) with which the exponents of the primitive Gaussian functions of the ECP-3G basis functions for carbon are scaled.\(^{67}\) Dral \textit{et al.} applied\(^{67}\) kernel ridge regression to predict \((\zeta^6)^2\) for individual molecules, i.e., \((\zeta^6)^2\) was not assumed to be a constant element-dependent parameter in ML-OM2. As a consequence, the resulting model offers a much greater flexibility. The mean absolute error in predicted atomization enthalpies could
be reduced from 26.4 kJ mol\(^{-1}\) with OM2 to 7.1 kJ mol\(^{-1}\) with ML-OM2 for a test set of organic molecules.\(^{67}\)

### 6.4 The High-Performance Computer-Aided Drug Design (hpCADD) Model

Very recently, Thomas et al. introduced the High-Performance Computer-Aided Drug Design (hpCADD) model\(^{68}\) which differs from an MNDO-type model in the dependence of the parameters. In the MNDO model, all parameters are element-dependent. Thomas et al. proposed to adopt the concept of ‘atom types’ (well-known for force fields) into an MNDO-type model, i.e., they proposed to make the parameters in the MNDO model environment-dependent\(^{68}\). E.g., hpCADD does not only comprise one parameter set for sulfur, but separate sets of parameters for a sulfur atom which is part of a \(\pi\)-system (such as the one in thiophene) and for a sulfur atom which is part of a thiol group.\(^{68}\) Hence, this conceptually follows the introduction of valence states of atoms in molecules, which is known to advance parametrized concepts such as electronegativity.\(^{202}\) We will come back to the advantages and disadvantages which are associated with such an approach in Section 7.3.

### 7 Implicit Description of Electron Correlation Effects through Parametrization

So far, we have discussed how NDDO-SEMO models approximate the SCF equations in the \(\phi\)-basis (Eq. \(9\)). Historically, NDDO-SEMO models were developed to reproduce experimental data rather than, e.g., HF data.\(^{9,50,51,127}\) Consequently, NDDO-SEMO models have to be able to capture electron correlation effects in some manner. The most popular way to describe electron correlation effects is implicit, i.e., through the calibration of the parameters incorporated in the NDDO-SEMO model.

#### 7.1 Parallels to Kohn–Sham Density-Functional Theory

A comparison to correlation functional derivations of KS-DFT is likely to highlight insufficiencies in the description of electron correlation in a parametrized single-determinant approach. The comparison of the elements of the Fock matrix for KS-DFT in the \(\chi\)-basis,

\[
\chi F_{\text{KS-DFT}} = \chi H + \chi J + (1 - \Lambda) \chi K + \Lambda V^x + V^c
\]  \( (91) \)
highlights the connection between KS-DFT and HF (through $\Lambda$). The parameter $\Lambda$ quantifies the amount of exact (HF) exchange $\chi_K$. For $0 < \Lambda \leq 1$, we have a contribution of a (approximate) DFT contribution to the exchange potential, $V^x$, to $\chi_F^{KS-DFT}$. Additionally, $V^c$ is the DFT description of the correlation potential. HF theory does not consider a correlation potential so that when $\Lambda = 0$ and no $V^c$ is considered then $\chi_F^{KS-DFT} = \chi_F^{HF}$.

Yang and co-workers classified the most severe drawbacks in KS-DFT at the example of fractional electrons and of fractional spins for the prototypical molecules H$_2$ and H$_2^+$. The energy for a system with a fractional number of electrons (or a fractional spin) is given by the straight line connecting the energies for the system with integer electron numbers (or integer spins). Approximate density functionals and HF are not able to correctly reproduce this behavior which may be interpreted as the source of many failures of approximate KS-DFT and HF models (such as delocalization and static correlation errors).

We may also study NDDO-SEMO models in this respect for which we choose as an example an H$_8$ cube, whose structure is described in Ref. 206 (see Figure 14). Not surprisingly, the NDDO-SEMO models are not able to describe the discontinuities in the energy for integer electron numbers per hydrogen atom. The NDDO-SEMO models can also not be reparametrized to yield such a behavior. We emphasize that the significance of this result is not the inability of NDDO-SEMO models to accurately describe H$_8$ with different electron numbers. Rather, it shows that NDDO-SEMO models fail...
to describe the quantum mechanical interaction of electrons in the same way as HF and approximate KS-DFT fail to do this. We may therefore take this failure as an indication that NDDO-SEMO models share the same systematic errors as approximate KS-DFT and HF models. Hence, these systematic errors cannot be alleviated through parametrization of the existing NDDO-SEMO models. However, these fundamental errors may not severely affect the equilibrium structures of organic molecules, but they will have a larger effect for non-equilibrium structures of organic molecules and for molecules with a more complicated electronic structure such as transition-metal complexes.

7.2 NDDO-SEMO Models for Isolated Atoms

Historically, NDDO-SEMO models are built upon considerations for isolated atoms. Studying isolated atoms has two distinct advantages: (i) We do not have to consider orthogonalization effects ($\chi = \phi$ for an isolated atom) and (ii) the NDDO approximation is no approximation in this special case. The one-electron matrix elements $\phi H_{\mu\nu} = \chi H_{\mu\nu}$ are equal to the corresponding one-center parameters $U_{Z_{I}\ell(l)(\mu)(\nu)}$. The one-center ERIs ($\gamma_{Z_{I}ss}$, $\gamma_{Z_{I}pp}$, $\gamma_{Z_{I}sp}$, $\gamma_{Z_{I}pp}$', and $\tilde{\gamma}_{Z_{I}sp}$), and if $d$-type orbitals are activated, the additional parameters specified in Section 4.5.1 will enter the two-electron matrix. For each element, the one-center parameters originally are calibrated with respect to reference electronic energies $E_{el}^{ref}$ for isolated atoms and monatomic ions (e.g., the one-center parameters for carbon (C) are calibrated with respect to $E_{el}^{ref}$ for C$^{3+}$ ($n_{v} = 1$), C$^{2+}$ ($n_{v} = 2$), C$^{+}$ ($n_{v} = 3$), C ($n_{v} = 4$), and C$^{-}$ ($n_{v} = 5$) in Ref. 183). $E_{el}^{ref}$ may be approximately determined from atomic ionization energies, or, quite recently, from coupled cluster data Refs. 183 and 209 showed that it is not possible to achieve a good agreement between $E_{el}^{ref}$ and $E_{el}^{NDDO-SEMO}$ with a single element-dependent parameter set for a range of monatomic ions. Margraf and co-workers, however, achieved a good agreement with the first ionization potentials and electron affinities for the neutral atoms. Oleari et al. 183 and subsequent studies 207,208 found that the one-center parameters vary in a remarkably regular manner with respect to $Z_{I}$ (quadratic dependence of $U_{Z_{I}ss}$, $U_{Z_{I}pp}$, and $U_{Z_{I}sp}$, and linear dependence for the one-center ERI parameters on $Z_{I}$; see also Figure 15). In NDDO-SEMO models, the one-center parameters are, however, not determined with respect to data for atoms, but with respect to data for molecules. Interestingly, the regularity of the one-center parameters with respect to $Z_{I}$ disappears for heavier elements ($Z_{I} > 23$) when taking molecular data as reference data.
(compare, e.g., the red squares (atomic data as reference data) and blue circles (molecular data as reference data) in Figure 15). We may take this as a direct and method-inherent hint that the description of transition-metal complexes will be more challenging than the description of organic compounds with NDDO-SEMO models.

Figure 15: $U_{pp}^{Z_I}$ determined in a fit to atomic data in Refs. 183,208 (red squares) and Ref. 209 (green dashes) and $U_{pp}^{Z_I}$ determined in a fit to molecular data for MNDO50,139 (black crosses), PM656 (light blue triangles), and PM757 (dark blue circles) for $Z_I = 2, 3, ..., 54$. We highlight the transition metal blocks ($21 \leq Z_I \leq 30$, and $39 \leq Z_I \leq 48$) by a yellow background and we indicate $Z_I$ of noble gases by vertical gray dashed lines.

7.3 General Parametrization Procedure

In general, parameters $p = (p_1, p_2, ..., p_p)^T$ are calibrated against a reference data set $D$ which comprises $D$ data triples,

$$D = \{(y_d, x_d, w_d)\}, \text{ with } d = (1, 2, ..., D),$$

consisting of (i) target observables $y_d$, (ii) input variables $x_d$ (e.g., atomic coordinates, charge, and spin multiplicity of a molecule), and (iii) weights $w_d$. Traditionally, the target observables are measured heats of formation $\Delta H_f^{298K}$ at 298 K, structural variables (bond distances or bond angles), dipole moments, and first vertical ionization potentials for a variety of molecules,50,56,57 or calculated electronic energy differences. The prediction of an observable by an NDDO-SEMO model, $f(x_d, p)$, is determined
by $x_d$ and $p$. The parameter set $p$ is then calibrated through minimization of an error function $E$ which is evaluated from the sum of weighted square differences between $y_d$ and $f(x_d, p)$:

$$E = E_D(p) = \sum_{d=1}^{D} w_d [y_d - f(x_d, p)]^2,$$  \hspace{1cm} (93)

where $w_d$ are the weights. The minimization of $E$ with respect to $p$ in a nonlinear ordinary least squares fit,

$$\frac{\partial E_D(p)}{\partial p} = 0, \quad \forall p \in p,$$  \hspace{1cm} (94)

yields an optimal parameter set $p_D$ with respect to the reference data set $D$. Different optimization algorithms, e.g., the Levenberg–Marquardt algorithm,\textsuperscript{152} gradient-based methods,\textsuperscript{52,56,57} genetic algorithms,\textsuperscript{63,66,212,214} and line-search algorithms\textsuperscript{50} can be straightforwardly applied for this task.

### 7.3.1 Applying Molecular Data Including Nuclear Effects as Reference Data

As we already noted, $D$ traditionally incorporates measured $\Delta H_{f}^{298\text{K}}$ for a variety of molecules.\textsuperscript{50,51,56,57} When applying an NDDO-SEMO model, $\Delta H_{f}^{298\text{K}}$ is usually predicted based on the electronic energies of the molecule and the constituent atoms, and the heats of formation of the atoms at 298 K, $\Delta H_{f,I}^{298\text{K}}$ \textsuperscript{50,80,210}

$$\Delta H_{f}^{298\text{K}} \approx E_{\text{el}}^{\{R_{i}\}} + \sum_{I=1}^{N} \left( \Delta H_{f,I}^{298\text{K}} - E_{\text{el}}^{R_{i}} \right).$$  \hspace{1cm} (95)

The heat of formation of the atom at 298 K is taken from experimental data (for instance from Ref. \textsuperscript{215}). We can examine which approximations are included in Eq. (95) by comparing it with the standard expression to calculate $\Delta H_{f}^{298\text{K}}$ from first principles,\textsuperscript{216}

$$\Delta H_{f}^{298\text{K}} = E_{\text{el}}^{\{\tilde{R}_{i}\}} + \text{ZPE} + H_{\text{rest}}(T) + \sum_{I=1}^{N} \left( \Delta H_{f,I}^{298\text{K}} - E_{\text{el}}^{\tilde{R}_{i}} \right).$$  \hspace{1cm} (96)

Compared to Eq. (95), Eq. (96) incorporates the zero-point energy (ZPE) and the temperature-dependent translational, rotational, and vibrational contributions (if coupling of degrees of freedom is neglected), $H_{\text{rest}}(T)$. Hence, the
parameters of an NDDO-SEMO model must account for the neglect of ZPE and $H_{\text{rest}}(T)$ when calculating $\Delta H_f^{298\text{K}}$ according to Eq. (95). Consequently, an NDDO-SEMO electronic energy in a traditional parameterization cannot be considered a pure electronic energy. This is a very unsatisfactory situation from a theoretical point of view (as, e.g., also noted in Refs. [72,211,217]). We would like to emphasize that, in principle, the standard protocol (Eq. (96)) and specialized approaches tailored toward SEMO models [217,218] could be readily applied instead of Eq. (95).

Hicks and Thiel studied [211] the severity of this conceptual inconsistency by reparametrizing MNDO with respect to electronic atomization energies ($\Delta E_{\text{el}}^{\text{at}}$),

$$\Delta E_{\text{el}}^{\text{at}} = E_{\text{el}}(\tilde{R}_i) - \sum_{l=1}^{N} E_{\text{el}}(\tilde{R}_l).$$  (97)

Hicks and Thiel found [211] that the errors between reference and predicted $\Delta H_f^{298\text{K}}$ and the errors between reference and predicted $\Delta E_{\text{el}}^{\text{at}}$ are similarly large. They therefore concluded [211] that the errors are dominated by the error in the MNDO electronic energies rather than by the error caused by applying Eq. (95). Their study was, however, limited to 36 medium-sized hydrocarbon compounds. Later, it was found that the application of Eq. (95) in the parametrization process is the reason for poor results for very small (e.g., diatomic) and large compounds in comparison to medium-sized compounds [96,219] (see, e.g., Figure 2 in Ref. [219]). This is not surprising because $D$ is dominated by medium-sized organic compounds. The opinion that the most severe errors stem from the NDDO-SEMO model itself and not from the application of Eq. (95), however, persisted in the literature [219].

7.3.2 Dependence of $p_D$ on $D$

For the prediction of properties for molecules not included in $D$, one needs to estimate the uncertainties of $p$. We recently demonstrated [220,221] how to apply nonparametric bootstrapping [222,223] in order to calibrate physicochemical property models with a limited amount of data and to determine the uncertainties of the incorporated parameters. Here, we re-optimize $p^m = \{\beta_1, \beta_6, \beta_6\}$ for the MNDO model with respect to a model data set $D^m$ containing twelve measured $\Delta H_f^{298\text{K}}$ of hydrocarbon compounds (see Figure 16a). Starting from the MNDO values for $p^m$, we determine the optimal parameter set $p_D^{m*}$ by minimizing $\mathcal{E}_{D^m}(p^m)$ with the Nelder–Mead simplex algorithm. The application of nonparametric bootstrap sampling now enables the quantification of the dependence $p^{m*}$ on the choice of $D^m$. We generate $B = 1000$ bootstrap samples $\{D^b\}$, $b = (1,2,...,B)$ by drawing $D$ elements
with replacement at random from $\mathcal{D}^m$. For each bootstrap sample $\mathcal{D}^m_b$, we determine the optimal parameter set $\mathbf{p}_{\mathcal{D}^m_b}$ by minimizing $\mathcal{E}_{\mathcal{D}^m_b}(\mathbf{p})$. We then determine the mean of the parameters $\bar{\mathbf{p}}_{\mathcal{D}^m}$ from all bootstrap samples,

$$\bar{\mathbf{p}}_{\mathcal{D}^m} = \frac{1}{B} \sum_{b=1}^{B} \mathbf{p}_{\mathcal{D}^m_b}. \quad (98)$$

Overall, $\bar{\mathbf{p}}_{\mathcal{D}^m}$ coincides nicely with $\mathbf{p}_{\mathcal{D}^m}$ and with the MNDO values. This means that we arrive at a very similar final parameter set, but we have gained significantly more knowledge from the parametrization procedure than from a fit to $\mathcal{D}^m$ alone. Figure 16b shows the distribution of $\mathbf{p}_{\mathcal{D}^m_b}$ which we obtained for the $B$ bootstrap samples $\{\mathcal{D}^m_b\}$. The parameters $\beta^6_s$ and $\beta^6_p$ differ significantly when they are calibrated with respect to different $\mathcal{D}^m_b$ ($-0.87$ a.u. < $\beta^6_s$ < −0.33 a.u. and $-0.51$ a.u. < $\beta^6_p$ < −0.17 a.u.). The parameter $\beta^1_s$, by contrast, hardly varies for different $\mathcal{D}^m_b$ ($-0.29$ a.u. < $\beta^1_s$ < −0.24 a.u.). Simply put, this means that we were not able to identify a single value for $\beta^6_s$ and $\beta^6_p$ which minimizes all different $\mathcal{E}_{\mathcal{D}^m_b}$. Rather, very different values for $\beta^6_s$ and $\beta^6_p$ are ideal to describe different $\mathcal{D}^m_b$. From the bootstrap samples, we can then sample the model prediction uncertainty for the target property, $\Delta H^{298\text{K}}$ which yields very large 95% confidence intervals for all molecules in $\mathcal{D}^m$ (> 20.0 kJ mol$^{-1}$, see the Supporting Information).

Figure 16: a) Model reference data set $\mathcal{D}^m$ consisting of dihydrogen (1), methane (2), ethane (3), ethene (4), ethyne (5), cyclopropane (6), cyclobutane (7), benzene (8), neopentane (9), n-butane (10), adamantane (11), and 1,3-butadine (12). b) MNDO parameter values (red lines), the $\mathbf{p}_{\mathcal{D}^m_b}$ values (blue lines), the $B \mathbf{p}_{\mathcal{D}^m_b}$ values (gray lines), and the $\bar{\mathbf{p}}_{\mathcal{D}^m}$ values (black lines) in a.u.
The fact that there is no single transferable parameter set has been noted before. Scholten remarked that different parameter values are well-suited to describe different properties for the same set of reference molecules. The parametrization of the HpCADD model demonstrated that the parameters for hydrogen atoms vary by 406% when considering different environments. Very recently, Oreluk et al. systematically assessed the variability of the PM7 parameters for a set of linear alkanes and came to the conclusion that no single set of parameters is consistent with the entire data set. Oreluk et al. propagated the uncertainties for the PM7 parameters to the prediction of heats of formations which then enables the attachment of an error bar to it.

7.3.3 Insights from Benchmark Studies

Not surprisingly, NDDO-SEMO models are unable to describe systems with strong electron correlation. Such systems are, however, present in $\mathcal{D}$ for some NDDO-SEMO models (see, e.g., the chromium dimer and $\text{CrO}_3$ which both exhibit a very strong multiconfigurational character) are contained in the PM6 and PM7 reference data set. The inclusion of systems with strong electron correlation in $\mathcal{D}$ may lead to a bias in $p$ which would at least partially explain the generally poor accuracy for transition-metal complexes. Despite significant efforts, it was not yet possible to create an NDDO-SEMO model which achieves a similar accuracy with respect to the reference data for transition-metal complexes as for organic compounds.

Figure 17: Deviation of electronic ligand dissociation energies $\Delta E$ in kJ mol$^{-1}$ calculated with PM6 and PM7 from DLPNO-CCSD(T) energies for the ten reactions in the WCCR10 set. We did not include $\Delta\Delta E = 786.8$ kJ mol$^{-1}$ for reaction 4 for PM6 in this Figure.

In this respect, we assess the performance of PM6 and PM7 for the WCCR10 set. The WCCR10 set contains ten ligand dissociation energies of large transition-metal complexes which feature different transition
metals (Au, Ag, Pt, Ru, Cu, Pd) and a diverse selection of ligand environments. The PM6 and PM7 ligand dissociation energies deviate significantly (on average 130.6 kJ mol$^{-1}$ and 114.1 kJ mol$^{-1}$, respectively) from reference DLPNO-CCSD(T) ligand dissociation energies$^{232}$ (see Figure 17). While a deviation of PM6 and PM7 energies from DLPNO-CCSD(T) data is not particularly surprising, the severeness of the failure of PM6 and PM7 might be. The PM6 ligand dissociation energy for reaction 4, for instance, is strongly negative ($\Delta E = -579.8$ kJ mol$^{-1}$).

Figure 18 shows that the PM7 structure of the charged product of reaction 1 is strongly distorted compared to the BP86/def2-QZVPP reference structure taken from Ref. 233 even though the deviation of the PM7 ligand dissociation energy from the DLPNO-CCSD(T) energy is only 29.4 kJ mol$^{-1}$ for this reaction. In fact, the structures may be so severely distorted that a re-optimization with BP86/def2-QZVPP starting from the PM6 or PM7 optimized structures does not yield the original BP86/def2-QZVPP minimum-energy structures from which the PM6 and PM7 optimizations were started (e.g., reactant of reaction 9 in Figure 18). Great caution is therefore in order when applying NDDO-SEMO models to transition-metal complexes in general.

![Figure 18: Overlay of the BP86/def2-QZVPP structure taken from Ref. 233 (solid) and the PM7 structure (translucent) of the charged product of reaction 1 (left) and of the reactant of reaction 9 (right) Element color code: carbon, gray; nitrogen, blue; chlorine, green; palladium, teal; platinum, silver. Hydrogen atoms are omitted for clarity.](image)

NDDO-SEMO models are mostly applied to study organic compounds which do not exhibit strong electron correlation. Recent benchmark studies show$^{234,235}$ that OMx models with dispersion corrections are slightly superior to MNDO-type models. The performance of NDDO-SEMO models in extensive benchmark sets such as the GMTKN24 database$^{234}$ is quite impressive considering their high computational efficiency (mean absolute deviation...
< 33 kJ mol\(^{-1}\) for OM3 at the GMTKN24 database\(^{[234]}\). Nevertheless, it is insightful to take a closer look at the distribution of the individual errors: It is not rare that a given NDDO-SEMO model either over- or underestimates relative electronic energies by over 80 kJ mol\(^{-1}\) (see, e.g., Figures 1–4 in Ref. \(^{[235]}\)). In special cases, the errors can be attributed to the insufficiency of the basis set (e.g., to explain the failure to describe nitro compounds\(^{219}\)) or to the absence of orthogonalization corrections in MNDO-type models (which is, e.g., assumed to be responsible for wrong rotation barriers\(^{219}\)). However, it is basically impossible to rationalize why some error occurs in general due to the diversity of the approximations invoked in an NDDO-SEMO model.

### 7.3.4 Focused Reparameterization

A way to reduce the error is the restriction of the domain of applicability. Rossi and Truhlar proposed to adjust the parameters to describe specific reactions yielding specific reaction parameters\(^{[212]}\). This approach has become more popular in recent years and parameters were adjusted to study specific compound classes and specific reactions (see, e.g., Refs. \(^{[67, 236-244]}\) and other references citing Ref. \(^{[212]}\)).

A focused reparameterization is, however, plagued by problems. Obviously, it cannot resolve systematic errors (cf. Section 7.1). Additionally, it may be difficult to curtail the domain of applicability adequately, i.e., to decide for a specific system whether it is similar enough to the ones for which it was parametrized. To define such a structure-based metric has been a long-standing goal in machine learning applied to chemistry\(^{[245]}\) and is related to the present problem. Strictly speaking, parameters are only valid for one arrangement of the atomic nuclei because they implicitly encode orthogonalization effects which, obviously, depend on the atomic nuclei. This statement is valid for all NDDO-SEMO models (even for the OM\(_{\text{fr}}\) models) because the parametric expressions are tuned to compensate for errors in \(\phi G\). Additionally, we might need different parameters for two atoms of the same element type in the same molecule (e.g., when they are encountered in different local environments represented by different valence states). In the worst case, this means that we need to have separate parameters for every valence state of an atom in a molecule. It may be possible to define ‘atom types’ as proposed in Ref. \(^{[68]}\) However, as a matter of principle it remains challenging to divide the atoms in a molecule into different atom types in a meaningful way. Moreover, when defining atom types, one obviously would inherit all of the problems associated with the definition of atom types from force-field development which does not appear particularly appealing for a method rooted in the first principles of quantum mechanics. The practical
consequence of these considerations is that the number of parameters which we have to determine increases dramatically, e.g., in the case of MNDO with its six adjustable parameters per element to $6N$ parameters per molecule in the worst case which would bring SEMO models very close to machine-learning approaches. For standard static benchmark approaches that apply a fixed amount of pre-defined reference data, it is hardly imaginable how one could achieve a similar increase in the number of reference data so that we can determine $p$ in a well-defined manner (i.e., so that $D \gg p$).

### 7.4 Improving Parametric Functions

Contemporary NDDO-SEMO models have limitations which cannot be addressed by reparameterization and overcoming these limitations requires the adoption of novel parametric functions. Unfortunately, the NDDO approximation causes large and uncontrollable errors in the ERIs in the $\phi$-basis and these errors propagate to all quantities calculated on the basis of the ERIs. Contemporary NDDO-SEMO models counteract the errors by introducing parametric expressions to evaluate the one-electron matrix and the core-core repulsion energy, i.e., they rely on error cancellation. This raises the question why one does not directly correct $\phi G$ or the ERIs in the $\phi$-basis. A reason might be that the ERIs in the $\phi$-basis, $\phi G$, and also $\chi G$ encode information on the whole molecule (cf. Eqs. (12), (15), and (17), respectively). By contrast, the contributions to $\chi H$ and $V_v$ are straightforwardly transferable from molecule to molecule. It is therefore comparatively easy to develop transferable parametric expressions to model $\chi H$ and $V_v$ on the examples of simple model systems (e.g., by considering diatomic systems). Unfortunately, approximating $\chi H$ and $V_v$ well is not sufficient for the design of a reliable NDDO-SEMO model. Instead, the parametric expressions applied to approximate $\chi H$ and $V_v$ need to be flexible enough to compensate for the errors in $\phi G$. Hence, we may anticipate that the improvement of the parametric expressions is as complicated as the direct correction for the error in $\phi G$.

We recently introduced\textsuperscript{116} a strategy to directly correct for the error caused by the NDDO approximation in $\phi G$ which we call the correction inheritance for semiempirics (CISE) approach. We drew inspiration from the work carried out by Roby and Sinanoğlu who suggested\textsuperscript{104} to scale $\chi G^{NDDO}$ with a scaling matrix $\Gamma$ to obtain a better estimate for $\phi G$:

$$\phi G \approx \Gamma \chi G^{NDDO}.$$  

The goal of Roby and Sinanoğlu in 1969 was to speed up single-point HF calculations for a diverse set of small molecules, and hence, they attempted
to define universal rules to assemble $\Gamma$ which, not surprisingly, turned out to be impossible.\cite{footnote:0} It is, however, possible, to exactly determine $\Gamma$ \cite{footnote:1} for a given structure $\{\tilde{R}_I^n\}$ from a reference self-consistent field (i.e., HF, KS-DFT, or general multi-configurational SCF) calculation (yielding the exact $\phi G(\{\tilde{R}_I^n\})$,)

$$
\Gamma(\{\tilde{R}_I^n\}) = \phi G(\{\tilde{R}_I^n\}) \cdot \left(\chi G^{\text{NDDO}}(\{\tilde{R}_I^n\})\right)^{-1}.
$$

(100)

Obviously, we will then not achieve a speed-up with respect to the reference calculation. We found\cite{footnote:1} that $\Gamma(\{\tilde{R}_I^n\})$ is transferable to a certain degree in a sequence of related structures, i.e., for two similar structures $\{\tilde{R}_I^n\}$ and $\{\tilde{R}_I^{(n+1)}\}$,

$$
\phi G(\{\tilde{R}_I^{(n+1)}\}) \approx \Gamma(\{\tilde{R}_I^n\}) \cdot \chi G^{\text{NDDO}}(\{\tilde{R}_I^{(n+1)}\}),
$$

(101)

for which we achieved a speed-up at a negligible loss of accuracy. We also showed\cite{footnote:1} that a correction to $\chi G^{\text{NDDO}}(\{\tilde{R}_I^{(n+1)}\})$ can be constructed in different ways, departing from a Roby–Sinanoğlu-type approach. We proposed\cite{footnote:1} to construct additive corrections $\Gamma_J$ and $\Gamma_K$ to the matrices $\chi J^{\text{NDDO}}$ and to $\chi K^{\text{NDDO}}$, respectively,

$$
\phi G(\{\tilde{R}_I^{(n+1)}\}) \approx \Gamma_J(\{\tilde{R}_I^n\}) + \chi J^{\text{NDDO}}(\{\tilde{R}_I^{(n+1)}\}) + \Gamma_K(\{\tilde{R}_I^n\}) + \chi K^{\text{NDDO}}(\{\tilde{R}_I^{(n+1)}\}).
$$

(102)

The CISE approach has a potential for application whenever we are interested in obtaining electronic energies for sequences of related structures, e.g., in the context of kinetic modeling,\cite{footnote:2,footnote:3} in real-time\cite{footnote:4,footnote:5,footnote:6} and automated\cite{footnote:7,footnote:8} reaction-mechanism explorations, or in reaction and first-principles\cite{footnote:9,footnote:10} molecular dynamics simulations. The CISE approach differs conceptually from the existing NDDO-SEMO models insofar as that no determination of parameters in a statistical calibration is required. Instead, we maintain complete error control on the resulting model because we could straightforwardly determine $\Gamma(\{\tilde{R}_I^{(n+1)}\})$ for a given molecule with nuclear coordinates $\{\tilde{R}_I^{(n+1)}\}$. By contrast, we cannot straightforwardly determine the best $p(\{\tilde{R}_I^{(n+1)}\})$ for the molecule $n + 1$ for contemporary NDDO-SEMO models.

53
8 Explicit Description of Electron Correlation Effects

The last question which we address in this work is whether one could, in principle, obtain FCI quality results in a given one-electron basis when applying an NDDO-SEMO reference wave function. It was suggested that all of the developed wave function methods can be (and many have been) straightforwardly applied after carrying out an NDDO-SEMO calculation which then essentially substitutes the HF calculation. It is, however, important to recognize that the NDDO approximation affects the ERIs over the molecular orbitals,

$$\langle \psi_i \psi_j | \psi_k \psi_l \rangle = \sum_{\mu=1}^{M} \sum_{\nu=1}^{M} \sum_{\lambda=1}^{M} \sum_{\sigma=1}^{M} \phi C_{\mu i} \phi C_{\nu j} \langle \phi_{\mu} \phi_{\nu} | \phi_{\lambda} \phi_{\sigma} \rangle \phi C_{\lambda k} \phi C_{\sigma l}$$

$$\approx \sum_{\mu=1}^{M} \sum_{\nu=1}^{M} \sum_{\lambda=1}^{M} \sum_{\sigma=1}^{M} \phi C_{\mu i} \phi C_{\nu j} \delta_{IJ} \delta_{KL} \langle \chi_I^{J} | \chi_{L}^{K} \rangle \phi C_{\lambda k} \phi C_{\sigma l}.$$  

Previous results by Thiel and co-workers and Clark and co-workers showed that the correlation energy calculated with single-reference perturbation theories evaluated for an NDDO-SEMO reference is about one order of magnitude too small, which was also corroborated by our recent results. If we do not explicitly correct for the errors caused by the NDDO approximation, we will not be able to adequately capture dynamic electron correlation effects and we must rely on the proper calibration of the parameters to achieve this.

It is no surprise that single-determinantal NDDO-SEMO models do not adequately capture static electron correlation effects and that static electron correlation effects have to be considered explicitly. Static electron correlation effects may be described through a multi-reference configuration interaction (MR-CI) procedure (including single and double excitations) using the graphical unitary group approach (GUGA). Another approach is the application of an unrestricted natural orbital complete active space (UNO-CAS) or configuration interaction (UNO-CI) ansatz. Such methods are usually applied to describe excited states and the dynamics of excited states. Note that all NDDO-SEMO models apply a valence-shell minimal basis sets which prevents a description of Rydberg states. Usually, MNDO-type models tend to underestimate excited-state energies due to the symmetric splitting of bonding and antibonding orbitals (see also Section 4.3), whereas OMx models showed an overall good performance.
An open question in the NDDO-SEMO/MRCI approaches is whether contributions to the correlation energy may be doubly counted. We may draw the parallel to KS-DFT as it was combined with MR-CI where a similar issue arises. The main problem of KS-DFT/MR-CI is the double counting of the correlation energy which can be alleviated through the introduction of empirical parameters. Similar measures have apparently not been taken when combining NDDO-SEMO models with MR-CI approaches. Recent benchmarks show, however, that double counting and the error caused by the NDDO approximation in the ERIs over molecular orbitals appear not to be an issue in practice.

9 The Future of NDDO-SEMO Models

The success of NDDO-SEMO models is largely based on the effectiveness with which they allow one to solve the SCF equations. Contemporary NDDO-SEMO models yield electronic energies about three orders of magnitude faster than HF or KS-DFT models. The acceleration is largely due to the NDDO approximation which drastically reduces the number of ERIs to be computed and processed in the course of a calculation. The price to pay for the acceleration are significant errors in the ERIs in the Löwdin orthogonalized basis. As a consequence, the NDDO approximation must be tied to many other approximations in the one-electron matrix and in the core-core repulsion energy to define a meaningful NDDO-SEMO model. In this work, we presented a comprehensive overview of the parametric expressions applied in the MNDO-type (MNDO, MNDO/d, AM1, PM3, PM6, PM7, and RM1) and OM models.

We outlined the systematic limitations which NDDO-SEMO model face. First, severe limitations are caused by the application of a small basis set. The drastic restriction of the basis set size will, in general, prohibit the determination of accurate relative electronic energies, force constants, and polarizabilities. The increase of the basis set size is, however, challenging within the framework of contemporary NDDO-SEMO models for practical reasons. Second, systematic errors are caused by the adoption of a parametrized mean-field framework. The examination of the parallels to KS-DFT revealed that NDDO-SEMO models fail to describe the general behavior of electronic energy as a function of the electron number, in the same way as in HF and KS-DFT models. We may therefore anticipate that NDDO-SEMO models will be plagued by the same difficulties in describing electron correlation, irrespective of their specific parametrization.

In general, the parameters of NDDO-SEMO models are calibrated with
respect to experimental reference data. When calibrating the parameters, one first encounters difficulties associated with the current practice of calculating heats of formation which leads to a contamination of the parameters with nuclear-motion contributions. Consequently, the electronic energy calculated with an NDDO-SEMO model cannot be considered a pure electronic energy which, however, does not appear to have severe practical consequences. This conceptual inconsistency could simply be alleviated by adjusting the parametrization procedure. Recent benchmark studies\textsuperscript{224,225} showed that NDDO-SEMO models are notoriously unreliable. Large errors are observed\textsuperscript{224,225} for molecules which do not show any apparent strong electron correlation. This may be partially explained by the fact that it is highly unlikely that there is a single parameter set which is suited to describe all molecules. We believe that bootstrap sampling\textsuperscript{220,222,249,276} offers an interesting insight into the parameterization of NDDO-SEMO models. When recalibrating a selection of parameters of the MNDO model in this work, we discovered that the parameters have to adopt significantly different values to describe different molecules well.

We briefly reviewed our recent proposal for system-focused NDDO-SEMO models that yield accurate results for structures related to a reference structure. Our CISE approach has the advantage that we are able to determine the parametrization of a corrective matrix directly for a given structure from a reference calculation. We, hence, do not have to apply a statistical procedure to calibrate parameters. This convenience obviously comes at the cost that the approach is restricted to the investigation of sequences of related structures which, however, are key areas of application for NDDO-SEMO models (e.g., structure optimization, Born–Oppenheim molecular dynamics, and real-time reactivity exploration).

To conclude this overview, we would like to stress that the age of NDDO-SEMO models is far from being over. Although we pointed out several (conceptual and practical) difficulties, we want to highlight again that contemporary NDDO-SEMO models achieve, overall, a remarkably high accuracy with respect to experimental data. To make NDDO-SEMO models useful for predictive work, we, however, have to know when, and why, they fail. This may, for instance, be achieved through statistical learning models\textsuperscript{277,278}

Acknowledgements

This work was supported by the Schweizerischer Nationalfonds. We are grateful to Professor Tim Clark for drawing our attention to the undocumented 6 kcal/mol correction in MOPAC and for providing information on AM1* that
allowed us to implement this approach.

10 Appendix

10.1 Computational Methodology

We implemented the MNDO(/d), AM1(/d), PM3, PM6, OM1, OM2, and OM3 models in our cross-platform quantum chemistry package SCINE. This new module of SCINE, SCINESEMO, will be made available on our Web page and can be applied as a stand-alone SEMO program or within the SCINE framework.

We evaluated ERIs in the \( \chi \)-basis with PYSCF (version 1.4). The ERIs in the \( \chi \)-basis were transformed to the corresponding ERIs in the \( \phi \)-basis with the AO2MO integral transformation module of PYSCF.

Lastly, we evaluated heat of formations at 298 K with Mopac 2016 in the course of the re-optimization of the parameters. We specified non-standard parameters through the keyword ‘EXTERNAL’ and we provide the corresponding parameter files as Supplementary Material.

The coordinates of all molecules considered in this work are given as Supplementary Material.

10.2 Basic Specifications

The MNDO-type models, MNDO,\(^5\) MNDO/d,\(^5\) AM1,\(^5\) PM3,\(^5\) PM6,\(^5\) PM7 and PM7 are freely available in the MOPAC2016 program. Throughout this work, MOPAC2016 served as our reference implementation for these NDDO-SEMO models because the parameters for the PM\(_x\)\(^5\) and RM1\(^5\) models and for many elements for the MNDO(/d) and AM1 models\(^5\) were determined with MOPAC2016.

In this work, we uncovered mistakes in the equations which we found implemented in many programs, also in MOPAC2016. Note that it is not easily possible to correct these mistake by altering the implementation because the parameterization of the NDDO-SEMO models was carried out with the erroneous implementation. Instead, one would have to determine a new set of parameters when implementing the correct equations. Implementations of MNDO-type models are also available in other programs. If one wishes to check if an MNDO model is implemented in the same way as in MOPAC2016, one can compare the parameter values and the values of the one- and two-center ERIs to the ones provided by MOPAC2016 when invoking the keyword Hcore.
The parameters for the O\(M_x\) models were determined with the (not freely available) MNDO2005 program\(^{252}\). We verified our implementation of the O\(M_x\) models by comparison to numerical data provided in Refs.\(^{58,60,151}\),\(^{152,235}\).

A calculation with an NDDO-SEMO model requires the specification of the number of explicitly considered electrons \(n_v\), of the basis functions which are activated, and of a set of parameters for every element in the system of interest. We specify these quantities in Table 1 for the MNDO,\(^{50}\) MNDO/\(d\),\(^{53,54}\) AM1,\(^{51}\) AM1*,\(^{51,167–169,171–176,283–285}\) PM3,\(^{52}\) RM1,\(^{55}\) PM6,\(^{56}\) PM7,\(^{57}\) OM1,\(^{58}\) OM2,\(^{59}\) and OM3\(^{60}\) models.

| Element | Availability of Parameters | \(Z\) | \(n_v\) | Basis Functions |
|---------|---------------------------|------|-------|----------------|
| H       | MNDO(/d), AM1, AM1*, PM3, PM6, PM7, RM1, OM1, OM2, OM3 | 1    | 1     | 1s            |
| He      | MNDO(/d), AM1, PM3, PM6, PM7 | 2    | 2     | 1s,2p         |
| Li      | MNDO(/d), AM1, PM3, PM6, PM7 | 3    | 1     | 2s,2p         |
| Be      | MNDO(/d), AM1, PM3, PM6, PM7 | 4    | 2     | 2s,2p         |
| B       | MNDO(/d), AM1*, PM3, PM6, PM7 | 5    | 3     | 2s,2p         |
| C       | MNDO(/d), AM1, AM1*, PM3, PM6, PM7, RM1, OM1, OM2, OM3 | 6    | 4     | 2s,2p         |
| N       | MNDO(/d), AM1, AM1*, PM3, PM6, PM7, RM1, OM1, OM2, OM3 | 7    | 5     | 2s,2p         |
| O       | MNDO(/d), AM1, AM1*, PM3, PM6, PM7, RM1, OM1, OM2, OM3 | 8    | 6     | 2s,2p         |
| F       | MNDO(/d), AM1, AM1*, PM3, PM6, PM7, RM1, OM1, OM2, OM3 | 9    | 7     | 2s,2p         |
| Ne      | MNDO(/d), AM1, PM3, PM6, PM7 | 10   | 6     | 2p,3s         |
| Na      | MNDO(/d), AM1, PM3, PM6, PM7 | 11   | 1     | 3s,3p         |
| Mg      | MNDO(/d), AM1, PM3, PM6, PM7 | 12   | 2     | 3s,3p         |
| Al      | MNDO, AM1, PM3 | 13   | 3     | 3s,3p,3d      |
| Si      | MNDO, AM1, PM3 | 14   | 4     | 3s,3p         |
| P       | MNDO, AM1, PM3, RM1 | 15   | 5     | 3s,3p         |
| S       | MNDO, AM1, PM3, RM1 | 16   | 6     | 3s,3p         |
| Element | Availability of Parameters | Z  | \( n_v \) | Basis Functions |
|---------|---------------------------|----|----------|----------------|
| Cl      | MNDO, AM1, PM3, RM1       | 16 | 6        | 3s, 3p, 3d    |
|         | MNDO/d, AM1*, PM6, PM7    | 17 | 7        | 3s, 3p        |
| Ar      | MNDO, AM1, PM3, PM6, PM7  | 18 | 6        | 3p, 4s        |
| K       | MNDO, AM1, PM3, PM6, PM7  | 19 | 1        | 4s, 4p        |
| Ca      | MNDO, AM1, PM3, PM6, PM7  | 20 | 2        | 4s, 4p        |
| Sc      | PM6, PM7                  | 21 | 3        | 3d, 4s, 4p    |
| Ti      | AM1*, PM6, PM7            | 22 | 4        | 3d, 4s, 4p    |
| V       | AM1*, PM6, PM7            | 23 | 5        | 3d, 4s, 4p    |
| Cr      | AM1*, PM6, PM7            | 24 | 6        | 3d, 4s, 4p    |
| Mn      | AM1*, PM6, PM7            | 25 | 7        | 3d, 4s, 4p    |
| Fe      | AM1*, PM6, PM7            | 26 | 8        | 3d, 4s, 4p    |
| Co      | AM1*, PM6, PM7            | 27 | 9        | 3d, 4s, 4p    |
| Ni      | AM1*, PM6, PM7            | 28 | 10       | 3d, 4s, 4p    |
| Cu      | AM1*, PM6, PM7            | 29 | 11       | 3d, 4s, 4p    |
| Zn      | MNDO, AM1, PM3, PM6, PM7  | 30 | 2        | 4s, 4p        |
|         | AM1*                     | 30 | 12       | 3d, 4s, 4p    |
| Ga      | MNDO, AM1, PM3, PM6, PM7  | 31 | 3        | 4s, 4p        |
| Ge      | MNDO, AM1, PM3, PM6, PM7  | 32 | 4        | 4s, 4p        |
| As      | MNDO, AM1, PM3            | 33 | 5        | 4s, 4p        |
|         | PM6, PM7                  | 33 | 5        | 4s, 4p, 4d    |
| Se      | MNDO, AM1, PM3, PM6, PM7  | 34 | 6        | 4s, 4p, 4d    |
| Br      | MNDO, AM1, PM3, RM1       | 35 | 7        | 4s, 4p        |
|         | MNDO/d, AM1*, PM6, PM7    | 35 | 7        | 4s, 4p, 4d    |
| Kr      | MNDO, AM1, PM3, PM6, PM7  | 36 | 6        | 4p, 5s        |
| Rb      | MNDO, AM1, PM3, PM6, PM7  | 37 | 1        | 5s, 5p        |
| Sr      | MNDO, AM1, PM3, PM6, PM7  | 38 | 2        | 5s, 5p        |
| Y       | PM6, PM7                  | 39 | 3        | 4d, 5s, 5p    |
| Zr      | AM1*, PM6, PM7            | 40 | 4        | 4d, 5s, 5p    |
| Nb      | PM6, PM7                  | 41 | 5        | 4d, 5s, 5p    |
| Mo      | AM1*, PM6, PM7, AM1       | 42 | 6        | 4d, 5s, 5p    |
| Tc      | PM6, PM7                  | 43 | 7        | 4d, 5s, 5p    |
| Ru      | PM6, PM7                  | 44 | 8        | 4d, 5s, 5p    |
| Rh      | PM6, PM7                  | 45 | 9        | 4d, 5s, 5p    |
| Pd      | AM1*, PM6, PM7            | 46 | 10       | 4d, 5s, 5p    |
| Ag      | AM1*, PM6, PM7            | 47 | 11       | 4d, 5s, 5p    |
| Cd      | MNDO/d, PM3, PM6, PM7     | 48 | 2        | 5s, 5p        |
| In      | MNDO, AM1, PM3, PM6, PM7  | 49 | 3        | 5s, 5p        |
| Sn      | MNDO, AM1, PM3, PM6, PM7  | 50 | 4        | 5s, 5p        |
| Element | Availability of Parameters | Z | \(n_v\) | Basis Functions |
|--------|---------------------------|---|--------|----------------|
| Sb     | MNDO, AM1, PM3, PM6, PM7  | 51| 5      | 5s, 5p         |
| Te     | MNDO, AM1, PM3, PM6, PM7  | 52| 6      | 5s, 5p, 5d     |
| I      | MNDO, AM1, PM3, RM1       | 53| 7      | 5s, 5p         |
|        | MNDO/d, AM1*, PM6, PM7    | 53| 7      | 5s, 5p, 5d     |
| Xe     | MNDO, AM1, PM6, PM7       | 54| 6      | 5p, 6s         |
| Cs     | MNDO, AM1, PM3, PM6, PM7  | 55| 1      | 6s, 6p         |
| Ba     | MNDO, AM1, PM3, PM6, PM7  | 56| 2      | 6s, 6p         |
| La     | PM6, PM7                  | 57| 3      | 5d, 6s, 6p     |
| Lu     | PM6, PM7                  | 71| 3      | 5d, 6s, 6p     |
| Hf     | PM6, PM7                  | 72| 4      | 5d, 6s, 6p     |
| Ta     | PM6, PM7                  | 73| 5      | 5d, 6s, 6p     |
| W      | PM6, PM7                  | 74| 6      | 5d, 6s, 6p     |
| Re     | PM6, PM7                  | 75| 7      | 5d, 6s, 6p     |
| Os     | PM6, PM7                  | 76| 8      | 5d, 6s, 6p     |
| Ir     | PM6, PM7                  | 77| 9      | 5d, 6s, 6p     |
| Pt     | PM6, PM7                  | 78| 10     | 5d, 6s, 6p     |
| Au     | AM1*, PM6, PM7            | 79| 11     | 5d, 6s, 6p     |
| Hg     | MNDO(/d), AM1, PM3, PM6, PM7 | 80| 2      | 6s, 6p         |
| Tl     | MNDO, AM1, PM3, PM6       | 81| 3      | 6s, 6p         |
|        | PM7                       | 81| 3      | 6s, 6p, 6d     |
| Pb     | MNDO, AM1, PM3, PM6, PM7  | 82| 4      | 6s, 6p         |
| Bi     | MNDO, AM1, PM3, PM6       | 83| 5      | 6s, 6p         |
|        | PM7                       | 83| 5      | 6s, 6p, 6d     |

While we mostly adhered to the original parameter abbreviations, we chose to re-name several parameters to avoid confusion with other quantities. The NDDO-SEMO models were developed independently of each other, and hence, they also sometimes apply different parameter names. We indicate in Tables 2 and 3 how the parameter abbreviations introduced in the main text relate to the ones chosen in several popular publications.
Table 2: Relation of the parameter abbreviations introduced in the main text for MNDO-type models to the ones chosen in several popular publications.

| Parameter Abbreviations | MOPAC | MNDO | AM1 | AM1* | MNDO/d | PM6 | PM7 |
|-------------------------|-------|------|-----|------|--------|-----|-----|
| U_{ss}                  | USS   | U_{ss} | U_{ss} | U_{ss} | U_{ss} | U_{ss} | U_{ss} |
| U_{pp}                  | UPP   | U_{pp} | U_{pp} | U_{pp} | U_{pp} | U_{pp} | U_{pp} |
| U_{dd}                  | UDD   | —     | —     | U_{dd} | U_{dd} | —     | —     |
| ζ_{s}                   | ZS    | ζ_{s} | ζ_{s} | ζ_{s} | ζ_{s} | ζ_{s} | ζ_{s} |
| ζ_{p}                   | ZP    | ζ_{p} | ζ_{p} | ζ_{p} | ζ_{p} | ζ_{p} | ζ_{p} |
| ζ_{d}                   | ZD    | —     | —     | ζ_{d} | ζ_{d} | ζ_{d} | ζ_{d} |
| β_{s}                   | BETAS | β_{s} | β_{s} | β_{s} | β_{s} | β_{s} | β_{s} |
| β_{p}                   | BETAP | β_{p} | β_{p} | β_{p} | β_{p} | β_{p} | β_{p} |
| β_{d}                   | BETAD | —     | —     | β_{d} | β_{d} | β_{d} | β_{d} |
| γ_{ss}                  | GSS   | g_{ss} | g_{ss} | g_{ss} | g_{ss} | g_{ss} | g_{ss} |
| γ_{pp}                  | GPP   | g_{pp} | g_{pp} | g_{pp} | g_{pp} | g_{pp} | g_{pp} |
| γ_{sp}                  | GSP   | g_{sp} | g_{sp} | g_{sp} | —     | g_{sp} | —     |
| γ_{pp'}                 | GP2   | g_{pp'} | g_{pp'} | g_{pp'} | —     | —     | g_{pp'} |
| γ_{sp'}                 | HSP   | h_{sp} | h_{sp} | h_{sp} | —     | —     | h_{sp} |
| ζ'_{s}                  | ZSN   | —     | —     | z_{sn} | ζ'_{s} | z_{sn} | ζ'_{sn} |
| ζ'_{p}                  | ZPN   | —     | —     | z_{pn} | ζ'_{p} | z_{pn} | ζ'_{pn} |
| ζ'_{d}                  | ZDN   | —     | —     | z_{dn} | ζ'_{d} | z_{dn} | ζ'_{dn} |
| ρ                      | P09   | —     | —     | ρ(core) | g_{core} | ρ(core) | —     |
| K_a                    | FN1a  | —     | K_a  | —     | —     | a     | —     |
| L_a                    | FN2a  | —     | L_a  | —     | —     | b     | —     |
| M_a                    | FN3a  | —     | M_a  | —     | —     | c     | —     |
| α                      | ALP   | —     | α    | —     | α    | —     | —     |
| α'                     | ALPB  | —     | α'   | α'/ij | —     | α    | —     |
| x                      | XFAC  | —     | —     | δ_{ij} | —     | x    | —     |

Table 3: Relation of the parameter abbreviations introduced in the main text for OMx models to the ones chosen in several popular publications.

| Parameter Abbreviations | OM1 | OM2 | OM3 |
|-------------------------|-----|-----|-----|
| U_{ss}                  | U_{ss} | U_{ss} | U_{ss} |
| U_{pp}                  | U_{pp} | U_{pp} | U_{pp} |
| Scaling factor          | ζ    | ζ    | ζ    |
| Scaling factor for core orbitals | ζ_{a} | ζ_{a} | —     |
Furthermore, Thiel and co-workers denote \( \theta(\chi^\dagger_I, \chi^\dagger_J) \) with \( \beta_{\mu \lambda} \) in Refs. 58–60, and \( \eta(\chi^\dagger_I, \chi^\dagger_J) \) is denoted as \( H^{IJ}_{\mu \mu} \) or as \( H^{loc}_{\mu \mu, J} \).  

### 10.3 Parametrization of One-Center ERIs

The one-center ERI \( \langle p^I p^I|p^I p^I \rangle \) is calculated from \( \gamma_{\mu \mu}^{Z_I} \) and \( \gamma_{\mu \mu}^{Z_I} \),

\[
\langle p^I p^I|p^I p^I \rangle = \frac{1}{2} \left( \gamma_{\mu \mu}^{Z_I} - \gamma_{\mu \mu}^{Z_I} \right). \tag{104}
\]

A practical issue with Eq. (104) arises when \( \gamma_{\mu \mu}^{Z_I} < \gamma_{\mu \mu}^{Z_I} \) so that \( \langle p^I p^I|p^I p^I \rangle \) is negative. In this case, it will not be possible to determine a parameter necessary to calculate the distance \( D_Q \) in the quadrupole moments according to Eq. (115) because the applied numerical procedure will not converge. The search for the parameter value was terminated after 5 iterations in MOPAC2016 even if it had not converged yet which is likely the reason why this failure has not been detected yet.

This is not a practical issue in the original MNDO model which was parameterized for hydrogen, carbon, nitrogen, and oxygen, nor for the RM1 model or the OM\( x \) models.

When specifying the MNDO, AM1, and PM3 model, \( \langle p^I p^I|p^I p^I \rangle \) is evaluated strictly according to Eq. (104) in MOPAC2016. This leads to negative \( \langle p^I p^I|p^I p^I \rangle \) for Ga, Sr, Xe, Ba, and Tl for the MNDO model, for Li, Be, Mg, Ga, Sr, Mo, Sb, Xe, Ba, and Tl for the AM1 model, and for Be, Mg, Rb, Sr, Xe, Cs, Ba, Hg, and Tl for the PM3 model.

Curiously, \( \langle p^I p^I|p^I p^I \rangle \) appears to be evaluated according to

\[
\langle p^I p^I|p^I p^I \rangle = -\frac{1}{2} \left( \gamma_{\mu \mu}^{Z_I} - \gamma_{\mu \mu}^{Z_I} \right) \tag{105}
\]

for several elements when specifying the MNDO/d, PM6, and PM7 models. Eq. (105) is applied for Al, Si, P, S, Cl, Br, and I for the MNDO/d model,
for Al, Si, P, S, Cl, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, As, Br, Y, Zr, Nb, Mo, 
Te, Ru, Rh, Pd, Ag, Sb, I, La, Lu, Hf, Ta, W, Re, Os, Ir, Pt, and Au for the 
PM6 model, and for Al, Si, P, S, Cl, Sc, Ti, V, Cr, Mn, Fe, Co, Ni, Cu, As, 
Br, Y, Zr, Nb, Mo, Te, Ru, Rh, Pd, Ag, Sb, I, La, Lu, Hf, Ta, W, Re, Os, 
Ir, Pt, Au, Tl, and Bi for the PM7 model. In MNDO/d, all \( \langle p_I' p_J' \mid p_I' p_J' \rangle \) are 
then positive. For several elements, PM6 and PM7 apply Eq. (104) which 
then yields negative \( \langle p_I' p_J' \mid p_I' p_J' \rangle \): He, Be, F, Ne, Na, Mg, Ar, K, Ca, Ga, 
Kr, Sr, In, Xe, and Hg (both models), Tl (PM6), and B and Pb (PM7).

10.4 Evaluation of Two-Center ERIs in MNDO-type Models with an \( s, p \) basis set

The two-center ERIs \( \langle \chi_I^I \chi_J^J \mid \chi_I^I \chi_J^I \rangle \), \( I \neq J \) can be interpreted as the 
electrostatic interaction between a charge distribution \( \chi_I^I \chi_J^J \) centered on atom \( I \) 
and a charge distribution \( \chi_I^I \chi_J^I \) centered on atom \( J \). The different possible 
charge distributions \( \chi_I^I \chi_J^J \) in the \( s, p \) minimal valence-shell basis are listed 
in Table 4. Each charge distribution is approximately represented as a trun-
cated classical multipole expansion of \( T_{\mu \nu} \) multipoles \( \Theta_{\mu \nu}^t \), \( t = (1, 2, \ldots, T_{\mu \nu}) \).[99] 
The two-center ERI is then approximated as the electrostatic interaction energy \( U(\Theta_{\mu \nu}^t, \Theta_{\lambda \sigma}^s) \) of the \( T_{\mu \nu} \) multipoles \( \Theta_{\mu \nu}^t \) specified for \( \chi_I^I \chi_J^J \) with the \( T_{\lambda \sigma} \) multipoles \( \Theta_{\lambda \sigma}^s \) specified for \( \chi_J^J \chi_J^I \),

\[
\langle \chi_I^I \chi_J^J \mid \chi_I^I \chi_J^I \rangle \approx \sum_{t=1}^{T_{\mu \nu}} \sum_{s=1}^{T_{\lambda \sigma}} U(\Theta_{\mu \nu}^t, \Theta_{\lambda \sigma}^s). 
\] (106)

The multipoles \( \Theta_{\mu \nu}^t \) may be a monopole \( q_I^I \), a dipole \( \mu_I^I, \mu_J^J \), a linear quadrupole 
\( Q_{xx}^{I}, Q_{yy}^{I} \), and a square quadrupole \( Q_{xy}^{I}, Q_{yz}^{I} \). (see also Figure 3 in the main 
text). Table 4 indicates which multipoles \( \Theta_{\mu \nu}^t \) appear in the multipole ex-

Table 4: Number of multipoles \( T_{\mu \nu} \) and types of multipoles applied to repre-
sent the charge distribution \( \chi_I^I \chi_J^J \) in an \( s, p \) basis.

| Charge Distribution | \( T_{\mu \nu} \) | Multipoles |
|--------------------|----------------|-----------|
| \( s^I s^I \)     | 1              | \( q_I^I \) |
| \( p_x^I p_x^I \) | 2              | \( q_I^I, Q_{xx}^{I} \) |
| \( p_y^I p_y^I \) | 2              | \( q_I^I, Q_{yy}^{I} \) |
| \( p_x^I p_y^I \) | 2              | \( q_I^I, Q_{xy}^{I} \) |
| \( s^I p_x^I \)   | 1              | \( \mu_x^I \) |
| \( s^I p_y^I \)   | 1              | \( \mu_y^I \) |

63
The specification of the positions of the point charges which make up the dipoles and the quadrupoles necessitates the specification of $D_{\mu,sp}$ and $D_{Q,pp}$ (see also Figure 3 in the main text). In the main text, we denoted $D_{\mu,sp}$ and $D_{Q,pp}$ as $D_\mu$ and $D_Q$, respectively, to keep the notation uncluttered. When an $s,p,d$ basis set is considered (see Section 10.5), we need to specify additional subscript identifiers indicating which kind of charge distribution is approximated. The distances $D_{\mu,sp}$ and $D_{Q,pp}$ between the point charges are chosen such that the multipole moment of the point charge configuration approximates the one of the corresponding charge distribution.

We first need to introduce the function $A(\chi_{I\mu}^f, \chi_{I\nu}^f, a)$:

$$A(\chi_{I\mu}^f, \chi_{I\nu}^f, a) = (2\zeta_{\mu})^{n_{\mu}+\frac{a}{2}}(2\zeta_{\nu})^{n_{\nu}+\frac{a}{2}}(\zeta_{\mu} + \zeta_{\nu})^{-n_{\mu} - n_{\nu} - a - 1} \times [((2n_{\mu})!(2n_{\nu})!)^{-\frac{1}{2}}(n_{\mu} + n_{\nu} + a)!],$$

(107)

where $\zeta_{\mu}$ is the orbital exponent of $\chi_{I\mu}^f$, $n_{\mu}$ is the principal quantum number associated with $\chi_{I\mu}^f$, and $a$ characterizes the angular momentum of the multipole ($a \in \{0, 1, 2, \ldots\}$). The distances $D_{\mu,sp}$ and $D_{Q,pp}$ are then given by:

$$D_{\mu,sp} = 3^{-\frac{1}{2}}A(\chi_{I\mu}^f, \chi_{I\nu}^f, 1)$$

(108)

and

$$D_{Q,pp} = 5^{-\frac{1}{2}}\sqrt{A(\chi_{I\mu}^f, \chi_{I\nu}^f, 2)},$$

(109)

respectively. The implementation of the formulae to calculate $D_{\mu,sp}$ and $D_{Q,pp}$ can be verified by comparison of $D_{\mu,sp}$ and $D_{Q,pp}$ to DD2 and DD3, respectively, which MOPAC2016 provides when specifying the keyword HCORE.

When comparing our implementation to MOPAC2016, we noticed that the MNDO-type models, in which $s$- and $p$-type basis functions with different principal quantum numbers ($n_s \neq n_p$) are activated for an atom (i.e., for He, Ne, Ar, Kr, and Xe; see Table 1), do not actually apply the different principal quantum numbers to evaluate $D_{\mu,sp}$ and $D_{Q,pp}$. Instead, only the lower principal quantum number is applied.

With the help of $D_{\mu,sp}$ and $D_{Q,pp}$, we can specify the positions and charges of the $C$ individual point charges $q_c$, $c = (1, 2, \ldots, C)$ of the multipoles relative to the atom $I$ on which the charge distribution $\chi_{I\mu}^f\chi_{I\nu}^f$ is centered (see Table 5).
Table 5: Specification of the position \( r_c = (r_{c,x}, r_{c,y}, r_{c,z}) \) of the \( C \) point charges \( q_c, c = (1, 2, ..., C) \) chosen to represent a specific multipole (Table 4). The total number of point charges \( C \) is given in brackets after the specification of the multipole. The positions \( r_c \) are given in relation to the origin of the local coordinate system (Section 10.6). The charges \( q_c \) are given in atomic units, i.e., as multiples of the elementary charge.

| Multipole | \( r_{c,x} \) | \( r_{c,y} \) | \( r_{c,z} \) | \( q_c \) |
|-----------|----------------|----------------|----------------|---------|
| \( q (C = 1) \) | 0.0 | 0.0 | 0.0 | +1.00 |
| \( \mu_x (C = 2) \) | \(-D_{\mu,sp}\) | 0.0 | 0.0 | \(-0.50\) |
| \( +D_{\mu,sp}\) | 0.0 | 0.0 | +0.50 |
| \( \mu_y (C = 2) \) | 0.0 | \(-D_{\mu,sp}\) | 0.0 | \(-0.50\) |
| 0.0 | \(+D_{\mu,sp}\) | 0.0 | +0.50 |
| \( \mu_z (C = 2) \) | 0.0 | 0.0 | \(-D_{\mu,sp}\) | \(-0.50\) |
| 0.0 | 0.0 | \(+D_{\mu,sp}\) | +0.50 |
| \( Q_{xx} (C = 3) \) | \(-2D_{Q,pp}\) | 0.0 | 0.0 | +0.25 |
| \(+2D_{Q,pp}\) | 0.0 | 0.0 | +0.25 |
| 0.0 | 0.0 | 0.0 | \(-0.50\) |
| \( Q_{yy} (C = 3) \) | 0.0 | \(-2D_{Q,pp}\) | 0.0 | +0.25 |
| 0.0 | \(+2D_{Q,pp}\) | 0.0 | +0.25 |
| 0.0 | 0.0 | 0.0 | \(-0.50\) |
| \( Q_{zz} (C = 3) \) | 0.0 | 0.0 | \(-2D_{Q,pp}\) | +0.25 |
| 0.0 | 0.0 | \(+2D_{Q,pp}\) | +0.25 |
| 0.0 | 0.0 | 0.0 | \(-0.50\) |
| \( Q_{xy} (C = 4) \) | \(+D_{Q,pp}\) | \(+D_{Q,pp}\) | 0.0 | +0.25 |
| \(-D_{Q,pp}\) | \(-D_{Q,pp}\) | 0.0 | +0.25 |
| \(+D_{Q,pp}\) | \(-D_{Q,pp}\) | 0.0 | \(-0.25\) |
| \(-D_{Q,pp}\) | \(+D_{Q,pp}\) | 0.0 | \(-0.25\) |
| \( Q_{xz} (C = 4) \) | \(+D_{Q,pp}\) | 0.0 | \(+D_{Q,pp}\) | +0.25 |
| \(-D_{Q,pp}\) | 0.0 | \(-D_{Q,pp}\) | +0.25 |
| \(+D_{Q,pp}\) | 0.0 | \(-D_{Q,pp}\) | \(-0.25\) |
| \(-D_{Q,pp}\) | 0.0 | \(+D_{Q,pp}\) | \(-0.25\) |
| \( Q_{yz} (C = 4) \) | 0.0 | \(+D_{Q,pp}\) | \(+D_{Q,pp}\) | +0.25 |
| 0.0 | \(-D_{Q,pp}\) | \(-D_{Q,pp}\) | +0.25 |
| 0.0 | \(+D_{Q,pp}\) | \(-D_{Q,pp}\) | \(-0.25\) |
| 0.0 | \(-D_{Q,pp}\) | \(+D_{Q,pp}\) | \(-0.25\) |

After the specification of the position and value of the point charges, we can straightforwardly assess the electrostatic potential energy \( U(\Theta^{\alpha\mu}, \Theta^{\beta\sigma}) \).
of the interaction of the $C_t$ and $C_s$ point charges making up $\Theta_{\mu\nu}^t$ and $\Theta_{\lambda\sigma}^s$, respectively (in atomic units),

$$U(\Theta_{\mu\nu}^t, \Theta_{\lambda\sigma}^s) = \sum_{c=1}^{C_t} \sum_{d=1}^{C_s} \frac{q_c^t q_d^s}{|r_c^t - r_d^s|^2}. \quad (110)$$

As stated in the main text, this interaction is not calculated analytically, but within the empirical Klopman approximation\cite{49,50,141}. The Klopman approximation is introduced to be able to recover the respective one-center ERIs $\langle \chi_{\mu}^I | \chi_{\nu}^J \rangle$ in the limit $\tilde{R}_{IJ} = 0$. In general, this means that the denominator in Eq. (110) is modified in such a way that we obtain the correct one-center limit,

$$U(\Theta_{\mu\nu}^t, \Theta_{\lambda\sigma}^s) = \sum_{c=1}^{C_t} \sum_{d=1}^{C_s} \frac{q_c^t q_d^s}{\sqrt{|r_c^t - r_d^s|^2} + (\vartheta_{I}^t(\chi_{\mu}^I) + \vartheta_{J}^t(\chi_{\nu}^J))^2}. \quad (111)$$

The term $\vartheta_{I}^t$ depends on the multipole to which the point charge $q_c$ belongs (see Table 5). If $q_c$ is part of a monopole $q^I$ representing the charge distribution $s^I s^J$, $\vartheta_{q,ss}$ is applied in Eq. (111). The four terms $\vartheta_{q,ss}^t$, $\vartheta_{q,pp}^t$, $\vartheta_{\mu,sp}^t$, and $\vartheta_{Q,pp}^t$ are calculated with reference to the one-center ERIs $\gamma_{ss}^I$, $\gamma_{sp}^I$, $\gamma_{pp}^I$, and $\gamma_{2p}^I$\cite{49},

$$\vartheta_{q,ss}^t = \frac{1}{2\gamma_{ss}^I}, \quad (112)$$

$$\vartheta_{q,pp}^t = \vartheta_{q,ss}^t, \quad (113)$$

$$(\vartheta_{q,ss}^t)^{-1} - \left[ (\vartheta_{q,ss}^t)^2 + (D_{q,sp})^2 \right]^{-\frac{1}{2}} = \frac{4}{3\gamma_{sp}^I}, \quad (114)$$

and

$$(\vartheta_{q,pp}^t)^{-1} - 2 \left[ (\vartheta_{q,pp}^t)^2 + (D_{q,pp})^2 \right]^{-\frac{1}{2}} + \left[ (\vartheta_{q,pp}^t)^2 + 2(D_{q,pp})^2 \right]^{-\frac{1}{2}} = \frac{24}{25} \langle p^I p^J | p^I p^J \rangle, \quad (115)$$

The parameters cannot be calculated analytically, but have to be determined in an iterative numerical procedure. The values for $\vartheta_{q,ss}^t$, $\vartheta_{q,pp}^t$, $\vartheta_{\mu,sp}^t$, and $\vartheta_{Q,pp}^t$ can be compared to those provided for P01, P07, P02, and P03, respectively, when specifying the keyword Hcore in MOPAC2016.

The implementation of the procedure to calculate the two-center ERIs can be compared to the implementation in MOPAC2016 when specifying the keyword Hcore for the calculation of the electronic energy of a diatomic...
molecule which is aligned along the z-axis. The values of the two-center ERIs are then listed under \textit{Two-Electron Matrix in HCORE} (note that these values really are the values for the ERIs and not the two-electron matrix entries). The first one hundred entries are the one-center ERIs for the first atom (i.e., \(\gamma_{ss}, \gamma_{sp}, \gamma_{pp}, \tilde{\gamma}_{pp}'\), and zeros). The next one hundred entries are the two-center ERIs which arise between the first and the second atom. The order in which the two-center ERIs are given is described in Ref. \footnote{287} or in the file `mopac\_eris\_sp.dat'.

10.5 MNDO-type Models with an \(s, p, d\) basis set

10.5.1 Evaluation of One-Center ERIs

The one-center ERIs \(\langle \chi_{\mu[I]}^{I} | \chi_{\nu[I]}^{I} | \chi_{\lambda[I]}^{I} | \chi_{\sigma[I]}^{I} \rangle\) will be calculated analytically if \(s-, p-,\) and \(d\)-type basis functions are activated for the atom \(I\). For this purpose, the one-center ERIs are re-written as:\footnote{85}

\[
\langle \chi_{\mu[I]}^{I} | \chi_{\nu[I]}^{I} | \chi_{\lambda[I]}^{I} | \chi_{\sigma[I]}^{I} \rangle = \sum_{k=0}^{C(k(l(m)(\mu), l(\nu)(m(\nu)))) \cdot R(k)(\chi_{\mu[I]}^{I}, \chi_{\nu[I]}^{I}, \chi_{\lambda[I]}^{I}, \chi_{\sigma[I]}^{I})}
\]

where \(C(k(l(m)(\mu), l(\nu)(m(\nu))))\) denotes the so-called angular coefficients and \(R(k)(\chi_{\mu[I]}^{I}, \chi_{\nu[I]}^{I}, \chi_{\lambda[I]}^{I}, \chi_{\sigma[I]}^{I})\) the radial integrals. The radial integrals \(R(k)(\chi_{\mu[I]}^{I}, \chi_{\nu[I]}^{I}, \chi_{\lambda[I]}^{I}, \chi_{\sigma[I]}^{I})\) are calculated as follows (Eq. (3) in Ref. \footnote{85}),

\[
\begin{align*}
R(k)(\chi_{\mu[I]}^{I}, \chi_{\nu[I]}^{I}, \chi_{\lambda[I]}^{I}, \chi_{\sigma[I]}^{I}) &= \frac{(2\zeta_{\mu[I]}^{I})^{n_{\sigma}+1/2}(2\zeta_{\nu[I]}^{I})^{n_{\nu}+1/2}}{(2\zeta_{\mu[I]}^{I})^{n_{\sigma}+1/2}(2\zeta_{\nu[I]}^{I})^{n_{\nu}+1/2}} \times \frac{(n_{\sigma} + n_{\lambda} + k)!}{(\zeta_{\sigma[I]}^{I} + \zeta_{\lambda[I]}^{I})^{n_{\sigma}+n_{\lambda}+k+1}} \left\{ \begin{array}{l}
(\zeta_{\mu[I]}^{I} + \zeta_{\nu[I]}^{I})^{n_{\mu}+n_{\nu}-k} \\
(\zeta_{\mu[I]}^{I} + \zeta_{\nu[I]}^{I})^{n_{\mu}+n_{\nu}+k} \\
(\zeta_{\mu[I]}^{I} + \zeta_{\nu[I]}^{I})^{n_{\mu}+n_{\nu}+k} \\
\end{array} \right. \\
&- \sum_{k'=1}^{n_{\lambda}+n_{\sigma} - k} \left( \zeta_{\sigma[I]}^{I} \right)^{n_{\lambda}+n_{\sigma} + k-k'+1} \left( \zeta_{\mu[I]}^{I} + \zeta_{\nu[I]}^{I} \right)^{n_{\mu}+n_{\nu}+n_{\lambda}+n_{\sigma}-k'+1} \\
&+ \sum_{k'=1}^{n_{\lambda}+n_{\sigma} - k} \left( \zeta_{\sigma[I]}^{I} \right)^{n_{\lambda}+n_{\sigma} + k-k'+1} \left( \zeta_{\mu[I]}^{I} + \zeta_{\nu[I]}^{I} \right)^{n_{\mu}+n_{\nu}+n_{\lambda}+n_{\sigma}-k'+1} \right),
\end{align*}
\]

where \(\zeta_{\mu[I]}^{I}\) is the auxiliary orbital exponent and \(n_{\mu}\) the principal quantum number associated with the basis function \(\chi_{\mu[I]}^{I}\).
Pelikán and Nagy determined\textsuperscript{83} the values of the 58 unique nonzero one-center ERIs $\langle \chi^i_{\mu}, \chi^j_{\nu} | \chi^k_{\lambda}, \chi^l_{\sigma} \rangle$ in terms of $R^k(\chi^i_{\mu}, \chi^j_{\nu}, \chi^k_{\lambda}, \chi^l_{\sigma})$ by explicitly evaluating Eq. (116). For this purpose, they used the angular coefficients $C^k(l(\mu)m(\mu), l(\nu)m(\nu))$ which are presented in Table 1 on pp. 178–179 in Ref. \textsuperscript{83}. In their work, the term $R^k(\chi^i_{\mu}, \chi^j_{\nu}, \chi^k_{\lambda}, \chi^l_{\sigma})$ is also denoted as $R^k_{l(\mu)l(\nu)(\lambda)(\sigma)}$. It is customary\textsuperscript{83,84} to introduce the quantities $F^k_{l(\mu)l(\nu)}$ and $G^k_{l(\mu)l(\nu)}$ to simplify the notation. The formulae for the one-center ERIs are presented in Table 2 of Ref. \textsuperscript{84} but some contain typographical mistakes which we clarify here (corrected formulae are indicated by an asterisk appended to the equation number); for the 58 one-center ERIs, they read in our notation:

$$F^k_{l(\mu)l(\nu)} \equiv R^k_{l(\mu)l(\nu)}$$

and

$$G^k_{l(\mu)l(\nu)} \equiv R^k_{l(\mu)l(\nu)}$$

(118)

(119)

to simplify the notation. The formulae for the one-center ERIs are presented in Table 2 of Ref. \textsuperscript{84} but some contain typographical mistakes which we clarify here (corrected formulae are indicated by an asterisk appended to the equation number); for the 58 one-center ERIs, they read in our notation:

$$\langle s^I s^I | s^I s^I \rangle = F^0_{ss}$$

$$\langle s^I s^I | p^I_y p^I_y \rangle = \langle s^I s^I | p^I_y p^I_y \rangle = \langle s^I s^I | p^I_y p^I_y \rangle = F^0_{sp}$$

$$\langle s^I p^I_x | s^I p^I_x \rangle = \langle s^I p^I_x | s^I p^I_x \rangle = \langle s^I p^I_x | s^I p^I_x \rangle = \frac{1}{3} G^1_{sp}$$

(120)

(121)

(122)

$$\langle p^I_x p^I_x | p^I_y p^I_y \rangle = \langle p^I_x p^I_x | p^I_y p^I_y \rangle = \langle p^I_x p^I_x | p^I_y p^I_y \rangle = F^0_{pp} + \frac{4}{25} F^2_{pp}$$

$$\langle p^I_y p^I_y | p^I_y p^I_y \rangle = \langle p^I_y p^I_y | p^I_y p^I_y \rangle = \langle p^I_y p^I_y | p^I_y p^I_y \rangle = F^0_{pp} - \frac{2}{25} F^2_{pp}$$

(123)

(124)

$$\langle p^I_x p^I_y | p^I_x p^I_y \rangle = \langle p^I_x p^I_y | p^I_x p^I_y \rangle = \langle p^I_x p^I_y | p^I_x p^I_y \rangle = \frac{3}{25} F^2_{pp}$$

(125)

$$\langle p^I_x d^I_{xz} | p^I_y d^I_{xz} \rangle = \langle p^I_x d^I_{xz} | p^I_y d^I_{xz} \rangle = \frac{1}{15} G^1_{pd} + \frac{18}{245} G^3_{pd}$$

(126)

$$\langle p^I_x d^I_{xz} | p^I_y d^I_{xz} \rangle = \langle p^I_x d^I_{xz} | p^I_y d^I_{xz} \rangle = \langle p^I_x d^I_{xz} | p^I_y d^I_{xz} \rangle = -\frac{\sqrt{3}}{15} G^1_{pd} - \frac{\sqrt{27}}{245} G^3_{pd}$$

(127)

$$\langle p^I_x d^I_{xz} | p^I_y d^I_{xz} \rangle = \frac{1}{5} G^1_{pd} - \frac{21}{245} G^3_{pd}$$

(128)

$$\langle p^I_y d^I_{xy} | p^I_y d^I_{xy} \rangle = -\frac{1}{5} G^1_{pd} + \frac{21}{245} G^3_{pd}$$

(129)

$$\langle p^I_y d^I_{xy} | d^I_{xz} d^I_{xz} \rangle = \langle p^I_y d^I_{xy} | d^I_{xz} d^I_{xz} \rangle = F^0_{pd} - \frac{2}{35} F^2_{pd}$$

(130)

$$\langle p^I_y d^I_{xy} | d^I_{xz} d^I_{xz} \rangle = \langle p^I_y d^I_{xy} | d^I_{xz} d^I_{xz} \rangle = -\frac{\sqrt{12}}{35} F^2_{pd}$$

(131)
\[
\langle p'_y d'_{xx} | p'_y d'_{x2-y'_2} \rangle = \frac{\sqrt{3}}{15} G^1_{pd} + \frac{\sqrt{27}}{245} G^3_{pd} \\
\langle p'_z d'_{xx} | p'_z d'_{zz} \rangle = \frac{4}{15} G^1_{pd} + \frac{27}{245} G^3_{pd} \\
\langle p'_z d'_{xx} | p'_z d'_{z2-y'_2} \rangle = \frac{3}{49} G^3_{pd} \\
\langle p'_z d'_{xz} | p'_z d'_{xz} \rangle = \frac{\sqrt{12}}{15} G^1_{pd} - \frac{\sqrt{243}}{245} G^3_{pd} \\
\langle p'_z d'_{xy} | p'_z d'_{xy} \rangle = \frac{1}{5} G^1_{pd} - \frac{6}{245} G^3_{pd} \\
\langle p'_z d'_{xz} | p'_y d'_{yz} \rangle = -\frac{3}{49} G^3_{pd} \\
\langle p'_z d'_{xz} | p'_z d'_{y'_2} \rangle = -\frac{\sqrt{3}}{15} G^1_{pd} + \frac{\sqrt{332}}{245} G^3_{pd} \\
\langle p'_z d'_{yz} | p'_y d'_{y'_2} \rangle = -\frac{1}{5} G^1_{pd} + \frac{6}{245} G^3_{pd} \\
\langle p'_z p'_y | d'_{xx} d'_{zz} \rangle = F^0_{pd} + \frac{4}{35} F^2_{pd} \\
\langle p'_z p'_y | d'_{xz} d'_{xz} \rangle = F^0_{pd} + \frac{4}{35} F^2_{pd} \\
\langle p'_x | p'_y | d'_{xx} d'_{yz} \rangle = \frac{3}{35} F^2_{pd} \\
\langle p'_x | p'_y | d'_{x2-y'_2} d'_{yz} \rangle = \frac{3}{35} F^2_{pd} \\
\langle p'_x | p'_y | d'_{xz} d'_{yz} \rangle = \frac{3}{35} F^2_{pd} \\
\langle p'_x | p'_y | d'_{xz} d'_{zz} \rangle = -\frac{3}{35} F^2_{pd} \\
\langle p'_x | p'_y | d'_{x2-y'_2} d'_{xz} \rangle = -\frac{3}{35} F^2_{pd} \\
\langle p'_x | p'_y | d'_{x2-y'_2} d'_{yz} \rangle = -\frac{3}{35} F^2_{pd} 
\]

(132) (133) (134) (135) (136) (137) (138) (139) (140) (141) (142) (143) (144) (145)
\[ \langle p_{l}^t p_{l}^t | d_{lx}^t d_{lx}^t \rangle = \langle p_{l}^t p_{l}^t | d_{ly}^t d_{ly}^t \rangle = \langle p_{l}^t p_{l}^t | d_{lxy}^t d_{lxy}^t \rangle = \langle p_{l}^t p_{l}^t | d_{l}^t d_{l}^t \rangle = F_{pd}^0 + \frac{2}{35} F_{pd}^2 \]  
\[ \langle p_{l}^t p_{l}^t | d_{l}^t d_{l}^t \rangle = \frac{\sqrt{12}}{35} F_{pd}^2 \]  
\[ \langle s^t l^t | s^t l^t \rangle = \langle s^t d_{l}^t d_{l}^t \rangle = \frac{1}{5} C_s \]  
\[ \langle s^t s^t | d_{l}^t d_{l}^t \rangle = \langle s^t s^t | d_{l}^t d_{l}^t \rangle = F_{sd}^0 \]  
\[ \langle d_{l}^t d_{l}^t | d_{l}^t d_{l}^t \rangle = \langle d_{l}^t d_{l}^t | d_{l}^t d_{l}^t \rangle = F_{dd}^0 + \frac{4}{49} F_{dd}^2 + \frac{36}{441} F_{dd}^4 \]  
\[ \langle d_{l}^t d_{l}^t | d_{l}^t d_{l}^t \rangle = \frac{4}{49} F_{dd}^2 + \frac{15}{441} F_{dd}^4 \]  
\[ \langle d_{l}^t d_{l}^t | d_{l}^t d_{l}^t \rangle = \frac{1}{49} F_{dd}^2 + \frac{30}{441} F_{dd}^4 \]  
\[ \langle d_{l}^t d_{l}^t | d_{l}^t d_{l}^t \rangle = F_{dd}^0 - \frac{4}{49} F_{dd}^2 + \frac{6}{441} F_{dd}^4 \]  
\[ \langle d_{l}^t d_{l}^t | d_{l}^t d_{l}^t \rangle = \frac{\sqrt{3}}{49} F_{dd}^2 - \frac{\sqrt{75}}{441} F_{dd}^4 \]  
\[ \langle d_{l}^t d_{l}^t | d_{l}^t d_{l}^t \rangle = -\frac{\sqrt{3}}{49} F_{dd}^2 + \frac{\sqrt{75}}{441} F_{dd}^4 \]  
\[ \langle d_{l}^t d_{l}^t | d_{l}^t d_{l}^t \rangle = \frac{2}{49} F_{dd}^2 - \frac{24}{441} F_{dd}^4 \]  
\[ \langle d_{l}^t d_{l}^t | d_{l}^t d_{l}^t \rangle = -\frac{\sqrt{12}}{49} F_{dd}^2 + \frac{300}{441} F_{dd}^4 \]  
\[ \langle d_{l}^t d_{l}^t | d_{l}^t d_{l}^t \rangle = \frac{\sqrt{12}}{49} F_{dd}^2 - \frac{300}{441} F_{dd}^4 \]  
\[ \langle d_{l}^t d_{l}^t | d_{l}^t d_{l}^t \rangle = \frac{35}{441} F_{dd}^4 \]  
\[ \langle d_{l}^t d_{l}^t | d_{l}^t d_{l}^t \rangle = \frac{3}{49} F_{dd}^2 + \frac{20}{441} F_{dd}^4 \]  

(146)
\[
\langle d_{xy}^t d_{xy}^t d_{xy}^t \rangle = F_{dd}^0 + \frac{4}{49} F_{dd}^2 - \frac{34}{441} F_{dd}^4 \tag{161}
\]
\[
\langle d_{xy}^t d_{xy}^t d_{xz}^t \rangle = \frac{3}{49} F_{dd}^2 - \frac{15}{441} F_{dd}^4 \tag{162}
\]
\[
\langle d_{xy}^t d_{yz}^t d_{xz}^t \rangle = \frac{3}{49} F_{dd}^2 + \frac{15}{441} F_{dd}^4 \tag{163}
\]
\[
\langle d_{xy}^t d_{xy}^t d_{xz}^t d_{xz}^t \rangle = \langle d_{xy}^t d_{yz}^t d_{xy}^t d_{yz}^t \rangle = \langle d_{xz}^t d_{xz}^t d_{yz}^t d_{yz}^t \rangle = \langle d_{xz}^t d_{xz}^t d_{xy}^t d_{xy}^t \rangle = \langle d_{xz}^t d_{xz}^t d_{xz}^t d_{xz}^t \rangle
\]
\[
= \langle d_{xz}^t d_{xz}^t d_{xy}^t d_{xy}^t \rangle = F_{dd}^0 - \frac{2}{49} F_{dd}^2 - \frac{4}{441} F_{dd}^4 \tag{164}
\]
\[
\langle s^t d_{xz}^t d_{xz}^t d_{xz}^t \rangle = \frac{2}{\sqrt{245}} R_{sddd}^2 \tag{165}
\]
\[
\langle s^t d_{xz}^t d_{xz}^t d_{xz}^t \rangle = \langle s^t d_{xy}^t d_{xz}^t d_{xy}^t \rangle = \langle s^t d_{xz}^t d_{xz}^t d_{xy}^t \rangle = \langle s^t d_{xz}^t d_{xz}^t d_{xz}^t \rangle
\]
\[
= \langle s^t d_{xz}^t d_{xz}^t d_{xy}^t \rangle = -\frac{2}{\sqrt{245}} R_{sddd}^2 \tag{166}
\]
\[
\langle s^t d_{xz}^t d_{xz}^t d_{xy}^t \rangle = \langle s^t d_{xz}^t d_{xz}^t d_{xz}^t \rangle = \langle s^t d_{xz}^t d_{xz}^t d_{xy}^t \rangle = \langle s^t d_{xz}^t d_{xz}^t d_{xz}^t \rangle
\]
\[
= \langle s^t d_{xz}^t d_{xz}^t d_{xz}^t \rangle = \frac{1}{\sqrt{245}} R_{sddd}^2 \tag{167}
\]
\[
\langle s^t d_{xz}^t d_{xz}^t d_{xz}^t \rangle = \langle s^t d_{xz}^t d_{xz}^t d_{xy}^t \rangle = \langle s^t d_{xz}^t d_{xz}^t d_{xz}^t \rangle = \langle s^t d_{xz}^t d_{xz}^t d_{xz}^t \rangle
\]
\[
= \langle s^t d_{xz}^t d_{xz}^t d_{xz}^t \rangle = \frac{\sqrt{3}}{\sqrt{245}} R_{sddd}^2 \tag{168}
\]
\[
\langle s^t d_{xz}^t d_{xz}^t d_{xy}^t d_{xy}^t \rangle = \langle s^t d_{xz}^t d_{xz}^t d_{xz}^t \rangle = -\frac{\sqrt{3}}{\sqrt{245}} R_{sddd}^2 \tag{169}
\]
\[
\langle s^t d_{xz}^t d_{xz}^t p_{xy}^t s_{xy}^t \rangle = \langle s^t d_{xz}^t d_{xz}^t p_{xy}^t s_{xy}^t \rangle = -\frac{1}{\sqrt{125}} R_{sdpp}^2 \tag{170*}
\]
\[
\langle s^t p_{xy}^t s_{xz}^t d_{xz}^t \rangle = -\frac{\sqrt{3}}{\sqrt{45}} R_{spdp}^1 \tag{171}
\]
\[
\langle p_{xy}^t p_{xy}^t s_{xz}^t \rangle = -\frac{\sqrt{3}}{\sqrt{125}} R_{spdy}^2 \tag{172*}
\]
\[
\langle p_{xy}^t s_{xz}^t s_{xz}^t \rangle = \frac{2}{\sqrt{45}} R_{sdpp}^1 \tag{173*}
\]
\[
\langle p_{xy}^t s_{xz}^t p_{xy}^t d_{xz}^t \rangle = \langle p_{xy}^t s_{xz}^t p_{xy}^t d_{xz}^t \rangle = \langle p_{xy}^t s_{xz}^t p_{xy}^t d_{xz}^t \rangle = \langle p_{xy}^t s_{xz}^t p_{xy}^t d_{xz}^t \rangle
\]
\[
= \langle p_{xy}^t s_{xz}^t p_{xy}^t d_{xz}^t \rangle = \langle p_{xy}^t s_{xz}^t p_{xy}^t d_{xz}^t \rangle = \langle p_{xy}^t s_{xy}^t s_{xy}^t \rangle = \frac{\sqrt{3}}{\sqrt{45}} R_{spdp}^1 \tag{174}
\]
\[ \langle p_x^l p_y^l | s^l d^l_{xz} \rangle = \frac{2}{\sqrt{125}} R_{sdp}^2 \]  
\[ (175^*) \]

\[ \langle p_x^l p_y^l | s^l d^l_{yz} \rangle = \langle p_x^l p_y^l | s^l d^l_{y^2} \rangle \]
\[ = \frac{\sqrt{3}}{\sqrt{125}} R_{sdp}^2 \]  
\[ (176^*) \]

\[ \langle p_x^l s^l | p_y^l d^l_{xz} \rangle = \langle p_y^l s^l | p_y^l d^l_{yz} \rangle = -\frac{1}{\sqrt{45}} R_{sppd}^1 \]  
\[ (177) \]

The typographical mistakes in Ref. [84] which we corrected here can be summarized as follows:

- The factor 12 has to be substituted for \( \sqrt{12} \) in Eq. (17) of Ref. [84], see Eq. (136*).
- The radial integral \( R_{sppd}^2 \) has to be replaced by \( R_{sdp}^2 \) in Eqs. (51), (53), (56), and (57) of Ref. [84], see Eqs. (170*), (172*), (175*), and (176*), respectively.
- \( R_{spdd}^1 \) has to be replaced for \( R_{sppd}^1 \) in Eq. (54) of Ref. [84], see Eq. (173*).

These errors affect all MNDO-type models which activate \( d \)-type basis functions (i.e., MNDO/d, AM1, PM6, and PM7). Specifically, the one-center ERIs \( \langle p_x^l d^l_{x^2} | p_x^l d^l_{x^2} \rangle, \langle p_x^l d^l_{x^2} | p_x^l d^l_{yz} \rangle, \langle s^l d^l_{x^2} | p_x^l d^l_{x^2} \rangle, \langle s^l d^l_{x^2} | p_y^l d^l_{y^2} \rangle, \langle p_x^l p_y^l | s^l d^l_{x^2} \rangle, \langle p_x^l p_y^l | s^l d^l_{x^2} \rangle, \langle p_x^l p_y^l | s^l d^l_{y^2} \rangle, \langle p_x^l p_y^l | s^l d^l_{x^2} \rangle, \langle p_x^l p_y^l | s^l d^l_{x^2} \rangle, \langle p_x^l p_y^l | s^l d^l_{x^2} \rangle \) are erroneous. One would have to determine a new set of parameters when implementing the corrected equations (Eqs. (136*), (170*), (172*), (173*), (175*), and (176*)).

The implementation of all uncorrected equations can be compared against MOPAC2016 when specifying the Hcore keyword for the calculation of the electronic energy for an atom for which \( d \)-type basis functions are activated. One can then find 2025 values for the one-center ERIs (their order is detailed in Ref. [287] or in the auxiliary file 'mopac_eries_spd.dat') listed under TWO-ELECTRON MATRIX IN HCORE in the output.

10.5.2 Evaluation of Two-Center ERIs

Conceptually, the calculation of the two-center ERIs in the s,p,d basis is similar to the one in the s,p basis. If at least one of the basis functions contributing to a two-center ERI is a \( d \)-type basis function, the following equations are applied instead of the ones specified in Section 10.4.
To fully describe the point charge interactions when a $d$-type orbital is involved in a two-center ERI, Thiel and Voityuk specified an additional point charge configuration $Q_{xy,yz,xz}$ (see Figure 19). Within an $s,p,d$-orbital basis,

$$q_\mu x Q_{xy}Q_{xx}$$

Figure 19: Illustration of the configuration of point charges (blue spheres) for the monopole $q$, the dipole $\mu_x$, the linear quadrupole $Q_{xx}$, the square quadrupole $Q_{xy}$, and the quadrupole $Q_{xy,xz,yz}$. The charge for each point charge is given in units of the elementary charge. The point charges in $\mu_x$ are $2D_\mu$ apart and the ones in $Q_{xx}$, $Q_{xy}$, and $Q_{xy,xz,yz}$ are $2D_Q$ apart.

many more combinations of charge distributions are now possible than in the $s,p$ basis (see Table 6). These multipoles are applied when at least one $d$-type orbital is involved.

Table 6: Specification of multipoles which are applied to represent the possible charge distributions $\chi^I_\mu \chi^J_\nu$ in an $s,p,d$ basis set in a truncated multipole expansion. We indicate the center of the multipole by a superscript atom index.

| Charge Distribution | $T_{\mu\nu}$ | Multipoles |
|---------------------|--------------|------------|
| $s^I s^I$           | 1            | $q^I$      |
| $p^I_x p^I_x$       | 3            | $q^I, -\frac{2}{3} \left( \tilde{Q}^I_{xx} - \frac{1}{2} \tilde{Q}^I_{xy} \right), \tilde{Q}^I_{xx}$ |
| $p^I_y p^I_y$       | 3            | $q^I, -\frac{2}{3} \left( \tilde{Q}^I_{xx} - \frac{1}{2} \tilde{Q}^I_{xy} \right), -\tilde{Q}^I_{xy}$ |
| $p^I_z p^I_z$       | 2            | $q^I, \frac{4}{3} \left( \tilde{Q}^I_{xx} - \frac{1}{2} \tilde{Q}^I_{xy} \right)$ |
| $s^I p^I_x$         | 1            | $\mu^I_x$  |
| $s^I p^I_y$         | 1            | $\mu^I_y$  |
| $s^I p^I_z$         | 1            | $\mu^I_z$  |
| $p^I_x p^I_y$       | 1            | $Q^I_{xy}$ |
| $p^I_x p^I_z$       | 1            | $Q^I_{xz}$ |
| $p^I_y p^I_z$       | 1            | $Q^I_{yz}$ |
| $s^I d^I_z$         | 1            | $\sqrt{\frac{4}{3}} \left( \tilde{Q}^I_{xx} - \frac{1}{2} \tilde{Q}^I_{xy} \right)$ |
| $s^I d^I_{zx}$      | 1            | $Q^I_{xx}$ |
| Charge Distribution | $T^{\mu\nu}$ | Multipoles |
|---------------------|--------------|-------------|
| $s^t d_{yz}^t$      | 1            | $Q_{yz}^t$  |
| $s^t d_{x^2-y^2}^t$| 1            | $\tilde{Q}_{xy}^t$ |
| $s^t d_{xy}^t$     | 1            | $\left(\tilde{Q}_{zx}^t - \frac{1}{2} \tilde{Q}_{xy}^t\right)$ |
| $p_x^t d_{x^2-y^2}^t$| 1          | $-\sqrt{\frac{1}{3}} \mu_x^t$ |
| $p_x^t d_{xy}^t$   | 1            | $\mu_z^t$    |
| $p_y^t d_{xy}^t$   | 0            | —            |
| $p_x^t d_{x^2-y^2}^t$| 1          | $\mu_x^t$    |
| $p_x^t d_{xy}^t$   | 1            | $\mu_y^t$    |
| $p_y^t d_{xy}^t$   | 1            | $-\sqrt{\frac{1}{3}} \mu_y^t$ |
| $p_y^t d_{xz}^t$   | 0            | —            |
| $p_y^t d_{yz}^t$   | 1            | $\mu_x^t$    |
| $p_y^t d_{x^2-y^2}^t$| 1          | $-\mu_y^t$   |
| $p_y^t d_{xy}^t$   | 1            | $\mu_x^t$    |
| $p_y^t d_{xz}^t$   | 1            | $\mu_x^t$    |
| $p_y^t d_{yz}^t$   | 1            | $\mu_y^t$    |
| $p_z^t d_{x^2-y^2}^t$| 0          | —            |
| $p_z^t d_{xy}^t$   | 0            | —            |
| $d_{xz}^t d_{xz}^t$| 2            | $q^t \frac{2}{3} \left(\tilde{Q}_{zx}^t - \frac{1}{2} \tilde{Q}_{xy}^t\right)$ |
| $d_{z^2}^t d_{xz}^t$| 1          | $\sqrt{\frac{1}{3}} Q_{zx}^t$ |
| $d_{z^2}^t d_{yz}^t$| 1          | $\sqrt{\frac{1}{3}} Q_{yz}^t$ |
| $d_{z^2}^t d_{x^2-y^2}^t$| 1       | $-\sqrt{\frac{2}{3}} Q_{xy}^t$ |
| $d_{z^2}^t d_{xy}^t$| 1          | $-\sqrt{\frac{2}{3}} Q_{xy}^t$ |
| $d_{xz}^t d_{xz}^t$| 3            | $q^t \frac{2}{3} \left(\tilde{Q}_{zx}^t - \frac{1}{2} \tilde{Q}_{xy}^t\right)$, $\tilde{Q}_{xy}^t$ |
| $d_{xz}^t d_{yz}^t$| 1            | $Q_{xy}^t$    |
| $d_{xz}^t d_{x^2-y^2}^t$| 1       | $Q_{xz}^t$    |
| $d_{xz}^t d_{xy}^t$| 1            | $Q_{xz}^t$    |
| $d_{yz}^t d_{yz}^t$| 3            | $q^t \frac{2}{3} \left(\tilde{Q}_{yz}^t - \frac{1}{2} \tilde{Q}_{xy}^t\right)$, $-\tilde{Q}_{xy}^t$ |
| $d_{yz}^t d_{x^2-y^2}^t$| 1       | $-Q_{yz}^t$   |
| $d_{yz}^t d_{xy}^t$| 1            | $Q_{yz}^t$    |
| $d_{x^2-y^2}^t d_{x^2-y^2}^t$| 3      | $q^t \frac{2}{3} \left(\tilde{Q}_{zx}^t - \frac{1}{2} \tilde{Q}_{xy}^t\right)$, $\tilde{Q}_{xy}^t$ |
| $d_{x^2-y^2}^t d_{xy}^t$| 0       | —            |
For the charge distributions $p_{I}^{I}d_{I}^{I}$, $p_{I}^{I}d_{I}^{I}$, and $d_{I}^{I}$ no multipoles (i.e., $T_{\mu\nu} = 0$ in Table 6) are specified. This is due to the fact that octopole (or higher) moments are neglected. As a consequence, all two-center ERIs which involve at least one of these charge distributions are exactly zero for all $\tilde{R}_{I,J}$.

Note how the multipoles specified for the charge distributions $p_{I}^{I}p_{I}^{I}$, $p_{I}^{I}p_{I}^{I}$, and $p_{I}^{I}p_{I}^{I}$ are different than the ones specified in Table 4 which means that the representation of $p_{I}^{I}p_{I}^{I}$ and $p_{I}^{I}p_{I}^{I}$ differs depending on whether the second charge distribution contains a $d$-type orbital or not.

Additionally, we are now in a situation where a dipole moment $\mu_{I}^{I}$ is applied to describe different charge distributions, e.g., $s_{I}^{I}p_{I}^{I}$ and $d_{I}^{I}p_{I}^{I}$.

As a consequence, we need to specify different distances for the multipoles which appear in the multipole expansions describing a charge distribution. In addition to $D_{\mu,sp}$ and $D_{Q,pp}$, we must define the distances $D_{\mu,sp}$, $D_{Q,sp}$, and $Q_{\mu,dd}$. These distances are again defined with respect to $A(\chi_{I}^{I}, \chi_{I}^{I}, a)$ (Eq. (107))

$$D_{\mu,sp} = 5^{-\frac{1}{4}}A(\chi_{I}^{I}, \chi_{I}^{I}, 1),$$

$$D_{Q,sp} = 15^{-\frac{1}{4}}\sqrt{A(\chi_{I}^{I}, \chi_{I}^{I}, 2)},$$

and

$$D_{Q,dd} = 7^{-\frac{1}{4}}\sqrt{A(\chi_{I}^{I}, \chi_{I}^{I}, 2)).}$$

The implementation of the formulae to calculate $D_{\mu,sp}$, $D_{Q,sp}$, and $D_{Q,dd}$ can be verified by comparison to the respective values supplied by MOPAC2011 when specifying the keyword HCORE. For each element, the value tabulated as DD2 corresponds to $D_{\mu,sp}$, DD3 to $D_{Q,pp}$, DD4 to $D_{Q,sp}$, DD5 to $D_{\mu,sp}$, and DD6 to $D_{Q,dd}$.

The positions of the point charges arising for a monopole $q$, a dipole $\mu_{I}, \nu_{I}$, a linear quadrupole $Q_{xx,yy,zz}$, and a square quadrupole $Q_{xy,yz,xz}$ are listed in Table 5. The positions of the point charges for the quadrupole $Q_{xy,xx,zz}$ relative to atom $I$ on which the charge distribution $\chi_{I}^{I}$ is centered are given in Table 7.
Table 7: Specification of the position \( \mathbf{r}_C = (r_{C,x}, r_{C,y}, r_{C,z}) \) of the \( C \) point charges chosen to represent a specific multipole (Table 4). The number of point charges is given in brackets after the specification of the multipole moment. The positions \( \mathbf{r}_c \) are given in relation to the origin of the local coordinate system (Section 10.6). The charges \( q_C \) are given in atomic units.

| Multipole | \( r_{C,x} \) | \( r_{C,y} \) | \( r_{C,z} \) | \( q_C \) |
|-----------|-------------|-------------|-------------|---------|
| \( Q_{xy} (C = 4) \) | \( +\sqrt{2}D_{Q,pp} \) | 0.0 | 0.0 | +0.25 |
| | \( -\sqrt{2}D_{Q,pp} \) | 0.0 | 0.0 | +0.25 |
| | 0.0 | \( -\sqrt{2}D_{Q,pp} \) | 0.0 | -0.25 |
| | 0.0 | \( +\sqrt{2}D_{Q,pp} \) | 0.0 | -0.25 |
| \( Q_{xz} (C = 4) \) | \( +\sqrt{2}D_{Q,pp} \) | 0.0 | 0.0 | +0.25 |
| | \( -\sqrt{2}D_{Q,pp} \) | 0.0 | 0.0 | +0.25 |
| | 0.0 | 0.0 | \( -\sqrt{2}D_{Q,pp} \) | -0.25 |
| | 0.0 | 0.0 | \( +\sqrt{2}D_{Q,pp} \) | -0.25 |
| \( Q_{yz} (C = 4) \) | 0.0 | \( +\sqrt{2}D_{Q,pp} \) | 0.0 | +0.25 |
| | 0.0 | \( -\sqrt{2}D_{Q,pp} \) | 0.0 | +0.25 |
| | 0.0 | 0.0 | \( -\sqrt{2}D_{Q,pp} \) | -0.25 |
| | 0.0 | 0.0 | \( +\sqrt{2}D_{Q,pp} \) | -0.25 |

Now we can again straightforwardly apply Eq. (111) after specifying \( \vartheta_{q,ss}^I, \vartheta_{q,pp}^I, \vartheta_{q,pp}^I, \vartheta_{q,pp}^I, \vartheta_{q,pp}^I, \vartheta_{q,pp}^I, \vartheta_{Q,pp}^I, \vartheta_{Q,sp}^I, \) and \( \vartheta_{Q,dd}^I \). The additive terms \( \vartheta_{q,ss}^I, \vartheta_{q,pp}^I, \vartheta_{q,pp}^I, \vartheta_{q,pp}^I, \vartheta_{q,pp}^I, \vartheta_{q,pp}^I, \vartheta_{Q,pp}^I, \vartheta_{Q,sp}^I, \) and \( \vartheta_{Q,dd}^I \) are determined numerically from Eqs. (112), (113), (114), and (115), respectively. The other terms are determined from the following equations:

\[
\vartheta_{q,dd}^I = \frac{1}{2F_{dd}^I} \quad (181)
\]

\[
(\vartheta_{q,pp}^I)^{-1} - \left[ (\vartheta_{q,pp}^I)^2 + (D_{q,pp})^2 \right]^{-\frac{1}{2}} = \frac{16}{15} G_{pd}^I \quad (182)
\]

\[
(\vartheta_{Q,dd}^I)^{-1} - 2 \left[ (\vartheta_{Q,dd}^I)^2 + (D_{Q,dd})^2 \right]^{-\frac{1}{2}} + \left[ (\vartheta_{Q,dd}^I)^2 + 2(D_{Q,dd})^2 \right]^{-\frac{1}{2}} = \frac{8}{5} G_{sd}^I \quad (183)
\]

and

\[
(\vartheta_{Q,dd}^I)^{-1} - 2 \left[ (\vartheta_{Q,dd}^I)^2 + (D_{Q,dd})^2 \right]^{-\frac{1}{2}} + \left[ (\vartheta_{Q,dd}^I)^2 + 2(D_{Q,dd})^2 \right]^{-\frac{1}{2}} = \frac{24}{49} F_{dd}^I \quad (184)
\]
The values for \( \vartheta_{q,ss}^I \), \( \vartheta_{q,pp}^I \), \( \vartheta_{q,dd}^I \), \( \vartheta_{\mu,sp}^I \), \( \vartheta_{\mu,pd}^I \), \( \vartheta_{Q,pp}^I \), \( \vartheta_{Q,sd}^I \), and \( \vartheta_{Q,dd}^I \) can be verified against the values for \( P_01 \), \( P_07 \), \( P_08 \), \( P_02 \), \( P_05 \), \( P_03 \), \( P_04 \), and \( P_06 \), respectively, when specifying the keyword \texttt{Hcore} in MOPAC2016.

Ref. \( \text{[53]} \) contains a mistake when describing the interaction of two multipoles, i.e., in the notation of Ref. \( \text{[53]} \),

\[
[M_{20}, M_{20}] = [\hat{Q}_{zx}, \hat{Q}_{zx}] - \frac{1}{4}[\hat{Q}_{xy}, \hat{Q}_{xy}]
\]

(p. 397 of Ref. \( \text{[53]} \)) should read instead

\[
[M_{20}, M_{20}] = [\hat{Q}_{zx}, \hat{Q}_{zx}] + \frac{3}{4}[\hat{Q}_{xy}, \hat{Q}_{xy}].
\]

Eq. \( \text{(185)} \) is implemented instead of Eq. \( \text{(186)} \) in MOPAC2016\(^{178}\). This mistake affects all MNDO-type models which activate \( d \)-type basis functions (i.e., MNDO/d, AM1, PM6, and PM7). The application of Eq. \( \text{(185)} \) results in erroneous values for those two-center ERIs that involve at least one of the following charge distributions: \( p_{Iz}^I p_{Iz}^I \), \( p_{Iz}^I p_{Iz}^I \), \( p_{Iz}^I p_{Iz}^I \), \( s^I d_{x}^I \), \( d_{x}^I d_{x}^I \), \( d_{x}^I d_{x}^I \), \( d_{x}^I d_{x}^I \), \( d_{x}^I d_{x}^I \), and \( d_{x}^I d_{x}^I \). Note that it is not easily possible to correct the mistake by substituting Eq. \( \text{(185)} \) for Eq. \( \text{(186)} \). We noted this already when discussing the errors affecting the one-center ERIs. The parameterization of the NDDO-SEMO models was carried out with erroneous implementation, and hence, the parameters cannot simply be transferred to a corrected implementation. Instead, one would have to determine a new set of parameters when implementing the correct equation (Eq. \( \text{(186)} \)).

The implementation of the procedure to calculate those two-center ERIs that are not affected by this mistake can be compared to the implementation in MOPAC2016 when invoking the keyword \texttt{Hcore} for a calculation of the electronic energy for a diatomic molecule which is aligned along the \( z \)-axis (for which \( s \)-, \( p \)-, and \( d \)-type basis functions are activated). The values of the two-center ERIs are then listed under \texttt{TWO-ELECTRON MATRIX IN HCORE}. The first 2025 entries are the one-center ERIs for the first atom. The next 2025 entries are the two-center ERIs which arise between the first and the second atom. The order in which the two-center ERIs are given is described in Ref. \( \text{[287]} \) or in the file `mopac_eris_spd.dat`.

### 10.6 Transformation from the Local to the Global Coordinate System

Within the MNDO-type models, we calculate the multipole-multipole interactions in a local coordinate system as visualized in Figure 5 in the main
text. A local coordinate system is defined for each pair of atoms \( I \) and \( J \neq I \) which are positioned at \( \mathbf{R}_I = (\tilde{R}_{I,x}, \tilde{R}_{I,y}, \tilde{R}_{I,z}) \) and \( \mathbf{R}_J = (\tilde{R}_{J,x}, \tilde{R}_{J,y}, \tilde{R}_{J,z}) \), respectively. The standard basis \( \mathbf{x}_{\text{loc}}, \mathbf{y}_{\text{loc}}, \mathbf{z}_{\text{loc}} \) of the local coordinate system can then be determined, e.g., as described in Ref. 86 or in Ref. 288. The first unit vector \( \mathbf{z}_{\text{loc}} \) is defined as the normalized vector connecting \( \tilde{R}_I \) and \( \tilde{R}_J \),

\[
\mathbf{z}_{\text{loc}} = \frac{\tilde{R}_I - \tilde{R}_J}{|\tilde{R}_I - \tilde{R}_J|} = (z_{\text{loc}}^x, z_{\text{loc}}^y, z_{\text{loc}}^z). \tag{187}
\]

The vector \( \mathbf{z}_{\text{loc}} \) can be applied to construct a perpendicular vector \( \mathbf{y}_{\text{loc}} \),

\[
\mathbf{y}_{\text{loc}} = \frac{1}{\sqrt{z_{\text{loc}}^x^2 + z_{\text{loc}}^y^2}}(-z_{\text{loc}}^y, z_{\text{loc}}^x, 0). \tag{188}
\]

The cross product of \( \mathbf{z}_{\text{loc}} \) and \( \mathbf{y}_{\text{loc}} \) yields \( \mathbf{x}_{\text{loc}} \),

\[
\mathbf{x}_{\text{loc}} = \mathbf{z}_{\text{loc}} \times \mathbf{y}_{\text{loc}}. \tag{189}
\]

We can then construct the rotation matrices which transform the results from the local to the global coordinate system.

The implementation of the transformation procedure can be compared to the implementation in MOPAC2016 when invoking the keyword \texttt{Hcore} the calculation of the electronic energy of a diatomic molecule which is \textit{not} aligned along the \( z \)-axis. The one hundred first to two hundredth entry listed under \texttt{TWO-ELECTRON MATRIX IN HCORE} are the transformed two-center ERIs.

### 10.7 Modification to the PM6 Core-Core Repulsion Energy

For certain element pairs, a scaling factor different from \( f_{IJ}^{\text{PM6}} \) (Eq. (62) in the main text) is applied in the PM6 model.\textsuperscript{53} The scaling function \( f_{IJ}^{\text{PM6}} \) to calculate the core-core repulsion energy between two carbon atoms is given by

\[
f_{IJ}^{\text{PM6}} = 1 + 2Z_I Z_J \exp\left( -a Z_I Z_J \left( \tilde{R}_{IJ} + 0.0003 \text{Å}^{-5} \tilde{R}_{IJ}^5 \right) \right)
+ 9.28 \exp\left( -5.98 \text{Å}^{-1} \tilde{R}_{IJ} \right)
+ K Z_I \exp\left( -L Z_I \left( \tilde{R}_{IJ} - M Z_I \right)^2 \right)
+ K Z_J \exp\left( -L Z_J \left( \tilde{R}_{IJ} - M Z_J \right)^2 \right). \tag{190}
\]
For N–H and O–H interactions, the scaling function \( f''_{IJ}^{\text{PM6}} \) reads,

\[
\begin{align*}
  f''_{IJ}^{\text{PM6}} &= 1 + 2x_{Z_I Z_J} \exp \left(-\alpha_{Z_I Z_J} \left(\tilde{R}_{IJ} - \frac{1}{A} \tilde{R}_{IJ}^2\right)\right) \\
  &\quad + K_{Z_I} \exp \left(-L_{Z_I} \left(\tilde{R}_{IJ} - M_{Z_I}\right)^2\right) \\
  &\quad + K_{Z_J} \exp \left(-L_{Z_J} \left(\tilde{R}_{IJ} - M_{Z_J}\right)^2\right).
\end{align*}
\]

(191)

Note that MOPAC2016 applies this equation not only for N–H and O–H interactions, but also for C–H interactions, which, however, is not the intended use according to the original publication in Ref. [56]. The scaling function \( f''_{IJ}^{\text{PM6}} \) for Si–O interactions is given by

\[
\begin{align*}
  f''_{IJ}^{\text{PM6}} &= 1 + 2x_{Z_I Z_J} \exp \left(-\alpha_{Z_I Z_J} \left(\tilde{R}_{IJ} + 0.0003 \tilde{R}_{IJ}^5\right)\right) \\
  &\quad - 0.0007 \exp \left(-A^2 \left(\tilde{R}_{IJ} - 2.9\right)^2\right) \\
  &\quad + K_{Z_I} \exp \left(-L_{Z_I} \left(\tilde{R}_{IJ} - M_{Z_I}\right)^2\right) \\
  &\quad + K_{Z_J} \exp \left(-L_{Z_J} \left(\tilde{R}_{IJ} - M_{Z_J}\right)^2\right).
\end{align*}
\]

(192)

As there exists no theoretical foundation for the introduction of the modified expressions to calculate the core-core repulsion energy, we do not know why the application of these modified scaling factors yield a better agreement with experimental data.

### 10.8 Constraints on Parameters

During the parameter optimization, certain constraints have to be imposed on the parameter values to keep the parametric expressions sensible from a physical point of view. Several parameters, for instance, determine the sign of an exponential functions which depends on the internuclear distance \( \tilde{R}_{IJ} \). The sign of the argument of the exponential function must be negative so that it does not become infinite for large \( \tilde{R}_{IJ} \) which means that \( \zeta^{Z_I}_{(\mu)} > 0 \), \( \zeta^{Z_I}_{(\nu)} > 0 \), \( \alpha_{Z_I} > 0 \), \( \alpha_{Z_I Z_J} > 0 \), \( a^{Z_I}_{(\mu)} + a^{Z_J}_{(\nu)} > 0 \), and \( L_{\alpha} Z_I > 0 \). Furthermore, the scaling factors for the core-core repulsion must not become negative (which would correspond to an attractive interaction between two cores), i.e., \( x^{Z_I Z_J} \) and \( K_{Z_I} \) have to be constrained such that the scaling factors are positive. These constraints are fulfilled by all NDDO-SEMO models. To our understanding,
it is not simply possible to constrain the values of the other parameters in a meaningful way.

Interestingly, the parameters appear to vary regularly with the atomic number for lighter elements. Dewar and Thiel noted that the MNDO parameters change in a remarkably regular manner with the atomic number of hydrogen, carbon, nitrogen, and oxygen (see also Figure 20). This regular behavior is so pronounced that it was, for example, possible to estimate the parameters for fluorine to good accuracy based on the ones for hydrogen, carbon, nitrogen, and oxygen. Dewar and co-workers remarked that this 'suggest[s] that the MNDO method as a whole is suitably self-consistent'.

For heavier elements, the regularity is lost (e.g., for the second transition-metal block). Additionally, one can also see that the parameters also vary significantly between the different NDDO-SEMO models.

Figure 20: Variation of a) $\beta_{Z I}^s$ and b) $\zeta_{Z I}^p$ in a.u. with the atomic number $Z_I$ in a range of $Z_I = 1–57$: MNDO values (black crosses) PM6 values (light blue triangles), PM7 values (dark blue circles), and AM1 values (red squares). We highlight the transition metal blocks ($21 \leq Z_I \leq 30$ and $39 \leq Z_I \leq 48$) by a yellow background and indicate $Z_I$ of rare gases by gray dashed lines. The orange area marks the parameter values for which $\beta_{Z I}^s > 0$. 
10.9 Parametrization of the MNDO Model

For this work, we re-optimized the $\beta_{ij(l)}$ parameters in the MNDO model for carbon and hydrogen. The applied reference data set $D^m$ contains 12 heats of formation at 298 K ($\Delta H_{298K}^f$) of hydrocarbon compounds which are also present in the original reference data set of MNDO\(^{50}\) (see Table 8).

Table 8: Reference data set $D^m$ consisting of $\Delta H_{298K}^f$ (and the standard deviation of $\Delta H_{298K}^f$) for twelve compounds in kJ mol\(^{-1}\).

| Compound        | $\Delta H_{298K}^f$ | Ref. |
|-----------------|----------------------|------|
| dihydrogen      | 0.0 ± 0.0            | —    |
| methane         | −76.3 ± 0.3          | 290  |
| ethane          | −84.0 ± 0.4          | 290  |
| ethene          | 52.4 ± 0.5           | 290  |
| ethyne          | 227.4 ± 0.8          | 290  |
| cyclopropane    | 53.5 ± 0.6           | 291  |
| cyclobutane     | 3.0 ± 0.1            | 292  |
| benzene         | 49.0 ± 0.9           | 293  |
| neopentane      | −167.9 ± 0.6         | 294  |
| n-butane        | −125.6 ± 0.7         | 295  |
| adamantane      | −192.5 ± 0.4         | 296  |
| 1,3-butadyne     | 455.8 ± 2.0          | 297  |

We list the values for the parameters $\beta_1^s$, $\beta_6^s$, and $\beta_6^p$ which are applied within MNDO\(^{50}\) and which we obtained in our parametrizations in Table 9. We supply the parameter files which can be read in through the keyword External in MOPAC2016 to reproduce our results with standard software.

Table 9: Values for the parameters $\beta_1^s$, $\beta_6^s$, and $\beta_6^p$ which are applied within MNDO\(^{50}\) and which we obtained in our parametrizations in a.u.

| Parametrization | $\beta_1^s$ | $\beta_6^s$ | $\beta_6^p$ |
|-----------------|-------------|-------------|-------------|
| MNDO            | −0.26       | −0.70       | −0.29       |
| $p_{D^m}^m$     | −0.26       | −0.65       | −0.32       |
| $\bar{p}_{D^m}^m$ | −0.26     | −0.68       | −0.30       |

81
The results for $\Delta H_{298K}^f$ obtained with the MNDO values, the $p_{Dm}^m$ values, and the $p_{Dm}^{m\bar{m}}$ values for $\beta_1^s$, $\beta_6^s$, and $\beta_p^0$ are given in Table 10 and illustrated in Figure 21.

Table 10: $\Delta H_{298K}^f$ obtained with the original MNDO values, the $p_{Dm}^m$ values, and the bootstrapped minimum, mean, and maximum $p_{Dm}^{m\bar{m}}$ values for $\beta_1^s$, $\beta_6^s$, and $\beta_p^0$ in kJ mol$^{-1}$.

| Compound      | MNDO | $p_{Dm}^m$ | $p_{Dm}^{m\bar{m}}$ |
|---------------|------|------------|---------------------|
|               | Min. | Mean       | Max.               |
| dihydrogen    | 11.2 | 9.5        | -19.9              |
| methane       | -48.3| -54.1      | -83.2              |
| ethane        | -79.5| -95.6      | -119.6             |
| ethene        | 65.6 | 62.5       | 29.8               |
| ethyne        | 245.7| 247.1      | 213.3              |
| cyclopropane  | 55.9 | 36.9       | 18.6               |
| cyclobutane   | -13.0| -50.3      | -76.0              |
| benzene       | 91.7 | 64.7       | 20.6               |
| neopentane    | -86.5| -131.3     | -166.2             |
| n-butane      | -116.3| -152.5     | -182.6             |
| adamantane    | -76.2| -181.0     | -266.6             |
| 1,3-butadyne  | 432.7| 435.6      | 369.1              |
Figure 21: Comparison of the reference $\Delta H_{f}^{298K}$ with the predicted values for $\Delta H_{f}^{298K}$ in kJ mol$^{-1}$. $\Delta H_{f}^{298K}$ was predicted with the MNDO values for $p^{m}$ (red crosses), the $p^{m}_{Dm}$ values (blue triangles), and the $\overline{p}^{m}_{Dm}$ values (black circles). We provide the 95% confidence interval for the bootstrapped $\Delta H_{f}^{298K}$ (gray bars). The uncertainties for the experimental data are too small to be visible in this figure. It is denoted for each datapoint in green font to which entry in $D^{m}$ it belongs (cf. Figure 16 in the main text).

References

[1] Hückel, E. Quantentheoretische Beiträge zum Benzolproblem, Z. Physik 1931, 70, 204–286.

[2] Parr, R. G. A Method for Estimating Electronic Repulsion Integrals Over LCAO MO’S in Complex Unsaturated Molecules, J. Chem. Phys. 1952, 20, 1499–1499.

[3] Pople, J. A. Electron Interaction in Unsaturated Hydrocarbons, Trans. Faraday Soc. 1953, 49, 1375–1385.

[4] Pariser, R.; Parr, R. G. A Semi-Empirical Theory of the Electronic Spectra and Electronic Structure of Complex Unsaturated Molecules. I., J. Chem. Phys. 1953, 21, 466–471.

[5] Pariser, R.; Parr, R. G. A Semi-Empirical Theory of the Electronic Spectra and Electronic Structure of Complex Unsaturated Molecules. II, J. Chem. Phys. 1953, 21, 767–776.
[6] Parr, R. G. Three Remarks on Molecular Orbital Theory of Complex Molecules, *J. Chem. Phys.* **1960**, *33*, 1184–1199.

[7] Ohno, K. Some Remarks on the Pariser-Parr-Pople Method, *Theor. Chim. Acta* **1964**, *2*, 219–227.

[8] Pople, J. A.; Santry, D. P.; Segal, G. A. Approximate Self-Consistent Molecular Orbital Theory. I. Invariant Procedures, *J. Chem. Phys.* **1965**, *43*, 129–135.

[9] Dewar, M. J. S. *The Molecular Orbital Theory of Organic Chemistry*; McGraw-Hill Book Co.: New York, 1969.

[10] Pople, J. A.; Beveridge, D. L. *Approximate Molecular Orbital Theory*; McGraw-Hill Book Co.: New York, 1970.

[11] Dykstra, C.; Frenking, G.; Kim, K.; Scuseria, G. *Theory and Applications of Computational Chemistry: The First Forty Years*; Elsevier: Amsterdam, 2005.

[12] Kohn, W.; Sham, L. J. Self-Consistent Equations Including Exchange and Correlation Effects, *Phys. Rev.* **1965**, *140*, 1133–1138.

[13] Čížek, J. On the Correlation Problem in Atomic and Molecular Systems. Calculation of Wavefunction Components in Ursell-Type Expansion Using Quantum-Field Theoretical Methods, *J. Chem. Phys.* **1966**, *45*, 4256–4266.

[14] Akimov, A. V.; Prezhdo, O. V. Large-Scale Computations in Chemistry: A Bird’s Eye View of a Vibrant Field, *Chem. Rev.* **2015**, *115*, 5797–5890.

[15] Senn, H. M.; Thiel, W. QM/MM Methods for Biological Systems, *Top. Curr. Chem.* **2006**, *268*, 173–290.

[16] Alexandrova, A. N.; Röthlisberger, D.; Baker, D.; Jorgensen, W. L. Catalytic Mechanism and Performance of Computationally Designed Enzymes for Kemp Elimination, *J. Am. Chem. Soc.* **2008**, *130*, 15907–15915.

[17] Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems, *Angew. Chem. Int. Ed.* **2009**, *48*, 1198–1229.
[18] Alexandrova, A. N.; Jorgensen, W. L. Origin of the Activity Drop with the E50D Variant of Catalytic Antibody 34E4 for Kemp Elimination, *J. Phys. Chem. B* 2009, 113, 497–504.

[19] Stewart, J. J. P. Application of the PM6 Method to Modeling Proteins, *J. Mol. Model.* 2009, 15, 765–805.

[20] Acevedo, O.; Jorgensen, W. L. Advances in Quantum and Molecular Mechanical (QM/MM) Simulations for Organic and Enzymatic Reactions, *Acc. Chem. Res.* 2010, 43, 142–151.

[21] Doron, D.; Major, D. T.; Kohen, A.; Thiel, W.; Wu, X. Hybrid Quantum and Classical Simulations of the Dihydrofolate Reductase Catalyzed Hydride Transfer Reaction on an Accurate Semi-Empirical Potential Energy Surface, *J. Chem. Theory Comput.* 2011, 7, 3420–3437.

[22] Polyak, I.; Reetz, M. T.; Thiel, W. Quantum Mechanical/Molecular Mechanical Study on the Mechanism of the Enzymatic Baeyer–Villiger Reaction, *J. Am. Chem. Soc.* 2012, 134, 2732–2741.

[23] B. Gerber, R.; Shemesh, D.; E. Varner, M.; Kalinowski, J.; Hirschberg, B. Ab Initio and Semi-Empirical Molecular Dynamics Simulations of Chemical Reactions in Isolated Molecules and in Clusters, *Phys. Chem. Chem. Phys.* 2014, 16, 9760–9775.

[24] Weber, V.; Laino, T.; Pozdneev, A.; Fedulova, I.; Curioni, A. Semiempirical Molecular Dynamics (SEMD) I: Midpoint-Based Parallel Sparse Matrix–Matrix Multiplication Algorithm for Matrices with Decay, *J. Chem. Theory Comput.* 2015, 11, 3145–3152.

[25] Husch, T.; Yilmazer, N. D.; Balducci, A.; Korth, M. Large-Scale Virtual High-Throughput Screening for the Identification of New Battery Electrolyte Solvents: Computing Infrastructure and Collective Properties, *Phys. Chem. Chem. Phys.* 2015, 17, 3394–3401.

[26] Husch, T.; Korth, M. Charting the Known Chemical Space for Non-Aqueous Lithium–air Battery Electrolyte Solvents, *Phys. Chem. Chem. Phys.* 2015, 17, 22596–22603.

[27] Lepšík, M.; Řezáč, J.; Kolář, M.; Pecina, A.; Hobza, P.; Fanfrlík, J. The Semiempirical Quantum Mechanical Scoring Function for In Silico Drug Design, *ChemPlusChem* 2013, 78, 921–931.
[28] Brahmkshatriya, S. P.; Dobes, P.; Fanfrlík, J.; Rezac, J.; Paruch, K.; Bronowska, A.; Lepsík, M.; Hobza, P. Quantum Mechanical Scoring: Structural and Energetic Insights into Cyclin-Dependent Kinase 2 Inhibition by Pyrazolo[1,5-a]Pyrimidines, *Curr. Comput. Aided Drug Des.* **2013**, *9*, 118–129.

[29] Yilmazer, N. D.; Korth, M. Enhanced Semiempirical QM Methods for Biomolecular Interactions, *Comput. Struct. Biotechnol. J.* **2015**, *13*, 169–175.

[30] Vorlová, B.; Nachtigallová, D.; Jirásková-Vaničková, J.; Ajani, H.; Jansa, P.; Řezáč, J.; Fanfrlík, J.; Otyepka, M.; Hobza, P.; Konvalinka, J.; Lepšík, M. Malonate-Based Inhibitors of Mammalian Serine Racemase: Kinetic Characterization and Structure-Based Computational Study, *Eur. J. Med. Chem.* **2015**, *89*, 189–197.

[31] Yilmazer, N. D.; Korth, M. Prospects of Applying Enhanced Semiempirical QM Methods for Virtual Drug Design, *Curr. Med. Chem.* **2016**, *23*, 2101–2111.

[32] Sulimov, A. V.; Kutov, D. C.; Katkova, E. V.; Ilin, I. S.; Sulimov, V. B. New generation of docking programs: Supercomputer validation of force fields and quantum-chemical methods for docking, *J. Mol. Graph. Model.* **2017**, *78*, 139–147.

[33] Marti, K. H.; Reiher, M. Haptic quantum chemistry, *J. Comput. Chem.* **2009**, *30*, 2010–2020.

[34] Haag, M. P.; Marti, K. H.; Reiher, M. Generation of Potential Energy Surfaces in High Dimensions and Their Haptic Exploration, *ChemPhysChem* **2011**, *12*, 3204–3213.

[35] Haag, M. P.; Reiher, M. Real-Time Quantum Chemistry, *Int. J. Quantum Chem.* **2013**, *113*, 8–20.

[36] Haag, M. P.; Vaucher, A. C.; Bosson, M.; Redon, S.; Reiher, M. Interactive Chemical Reactivity Exploration, *ChemPhysChem* **2014**, *15*, 3301–3319.

[37] Haag, M. P.; Reiher, M. Studying chemical reactivity in a virtual environment, *Faraday Discuss.* **2014**, *169*, 89–118.

[38] Vaucher, A. C.; Haag, M. P.; Reiher, M. Real-Time Feedback from Iterative Electronic Structure Calculations, *J. Comput. Chem.* **2016**, *37*, 805–812.
Mühlbach, A. H.; Vaucher, A. C.; Reiher, M. Accelerating Wave Function Convergence in Interactive Quantum Chemical Reactivity Studies, *J. Chem. Theory Comput.* 2016, 12, 1228–1235.

Heuer, M. A.; Vaucher, A. C.; Haag, M. P.; Reiher, M. Integrated Reaction Path Processing from Sampled Structure Sequences, *J. Chem. Theory Comput.* 2018, 14, 2052–2062.

Thiel, W. Semiempirical Quantum-Chemical Methods, *WIREs Comput. Mol. Sci.* 2014, 4, 145–157.

Elstner, M.; Frauenheim Th.; Kaxiras E.; Seifert G.; Suhai S., A Self-Consistent Charge Density-Functional Based Tight-Binding Scheme for Large Biomolecules, *Physica Status Solidi B* 2000, 217, 357–376.

Gaus, M.; Cui, Q.; Elstner, M. DFTB3: Extension of the Self-Consistent-Charge Density-Functional Tight-Binding Method (SCC-DFTB), *J. Chem. Theory Comput.* 2011, 7, 931–948.

Seifert, G.; Joswig, J.-O. Density-Functional Tight Binding—an Approximate Density-Functional Theory Method, *WIREs Comput. Mol. Sci.* 2012, 2, 456–465.

Elstner, M.; Seifert, G. Density Functional Tight Binding, *Phil. Trans. R. Soc. A* 2014, 372, 20120483.

Grimme, S.; Bannwarth, C.; Shushkov, P. 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.

Bursch, M.; Hansen, A.; Grimme, S. Fast and Reasonable Geometry Optimization of Lanthanoid Complexes with an Extended Tight Binding Quantum Chemical Method, *Inorg. Chem.* 2017, 56, 12485–12491.

Krishnapriyan, A.; Yang, P.; Niklasson, A. M. N.; Cawkwell, M. J. Numerical Optimization of Density Functional Tight Binding Models: Application to Molecules Containing Carbon, Hydrogen, Nitrogen, and Oxygen, *J. Chem. Theory Comput.* 2017, 13, 6191–6200.

Dewar, M. J. S.; Thiel, W. A Semiempirical Model for the Two-Center Repulsion Integrals in the NDDO Approximation, *Theor. Chim. Acta* 1976, 46, 89–104.
[50] Dewar, M. J. S.; Thiel, W. Ground States of Molecules. 38. The MNDO Method. Approximations and Parameters, J. Am. Chem. Soc. 1977, 99, 4899–4907.

[51] Dewar, M. J. S.; Zoebisch, E. G.; Healy, E. F.; Stewart, J. J. P. Development and Use of Quantum Mechanical Molecular Models. 76. AM1: A New General Purpose Quantum Mechanical Molecular Model, J. Am. Chem. Soc. 1985, 107, 3902–3909.

[52] Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods I. Method, J. Comput. Chem. 1989, 10, 209–220.

[53] Thiel, W.; Voityuk, A. A. Extension of the MNDO Formalism To d-Orbitals: Integral Approximations and Preliminary Numerical Results, Theor. Chim. Acta 1991, 81, 391–404.

[54] Thiel, W.; Voityuk, A. A. Erratum: Extension of the MNDO Formalism To d-Orbitals: Integral Approximations and Preliminary Numerical Results, Theor. Chim. Acta 1996, 93, 315.

[55] Rocha, G. B.; Freire, R. O.; Simas, A. M.; Stewart, J. J. P. RM1: A Reparameterization of AM1 for H, C, N, O, P, S, F, Cl, Br, and I, J. Comput. Chem. 2006, 27, 1101–1111.

[56] Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods V: Modification of NDDO Approximations and Application to 70 Elements, J. Mol. Model. 2007, 13, 1173–1213.

[57] Stewart, J. J. P. Optimization of Parameters for Semiempirical Methods VI: More Modifications to the NDDO Approximations and Re-Optimization of Parameters, J. Mol. Model. 2012, 19, 1–32.

[58] Kolb, M.; Thiel, W. Beyond the MNDO Model: Methodical Considerations and Numerical Results, J. Comput. Chem. 1993, 14, 775–789.

[59] Weber, W.; Thiel, W. Orthogonalization Corrections for Semiempirical Methods, Theor. Chem. Acc. 2000, 103, 495–506.

[60] Dral, P. O.; Wu, X.; Spörkel, L.; Koslowski, A.; Weber, W.; Steiger, R.; Scholten, M.; Thiel, W. Semiempirical Quantum-Chemical Orthogonalization-Corrected Methods: Theory, Implementation, and Parameters, J. Chem. Theory Comput. 2016, 12, 1082–1096.
[61] Sattelmeyer, K. W.; Tubert-Brohman, I.; Jorgensen, W. L. NO-MNDO: Reintroduction of the Overlap Matrix into MNDO, *J. Chem. Theory Comput.* **2006**, *2*, 413–419.

[62] Fiedler, L.; Gao, J.; Truhlar, D. G. Polarized Molecular Orbital Model Chemistry. 1. Ab Initio Foundations, *J. Chem. Theory Comput.* **2011**, *7*, 852–856.

[63] Zhang, P.; Fiedler, L.; Leverentz, H. R.; Truhlar, D. G.; Gao, J. Polarized Molecular Orbital Model Chemistry. 2. The PMO Method, *J. Chem. Theory Comput.* **2011**, *7*, 857–867.

[64] Zhang, P.; Fiedler, L.; Leverentz, H. R.; Truhlar, D. G.; Gao, J. Erratum: “Polarized Molecular Orbital Chemistry. 2. The PMO Method”, *J. Chem. Theory Comput.* **2012**, *8*, 2983.

[65] Isegawa, M.; Fiedler, L.; Leverentz, H. R.; Wang, Y.; Nachimuthu, S.; Gao, J.; Truhlar, D. G. Polarized Molecular Orbital Model Chemistry 3. The PMO Method Extended to Organic Chemistry, *J. Chem. Theory Comput.* **2013**, *9*, 33–45.

[66] Fiedler, L.; Leverentz, H. R.; Nachimuthu, S.; Friedrich, J.; Truhlar, D. G. Nitrogen and Sulfur Compounds in Atmospheric Aerosols: A New Parametrization of Polarized Molecular Orbital Model Chemistry and Its Validation against Converged CCSD(T) Calculations for Large Clusters, *J. Chem. Theory Comput.* **2014**, *10*, 3129–3139.

[67] Dral, P. O.; von Lilienfeld, O. A.; Thiel, W. Machine Learning of Parameters for Accurate Semiempirical Quantum Chemical Calculations, *J. Chem. Theory Comput.* **2015**, *11*, 2120–2125.

[68] Thomas, H. B.; Hennemann, M.; Kibies, P.; Hoffgaard, F.; Güssregen, S.; Hessler, G.; Kast, S. M.; Clark, T. The hpCADD NDDO Hamiltonian: Parametrization, *J. Chem. Inf. Model.* **2017**, *57*, 1907–1922.

[69] Citation analysis of Refs. [50–53, 55–60] https://scholar.google.com, (Accessed: 25. April 2018).

[70] Dewar, M. J. S. Development and status of MINDO/3 and MNDO, *J. Mol. Struct.* **1983**, *100*, 41–50.

[71] Dewar, M. J. S. The Semiempirical Approach to Chemistry, *Int. J. Quantum Chem.* **1992**, *44*, 427–447.
[72] Thiel, W. Semiempirical Methods: Current Status and Perspectives, *Tetrahedron* 1988, 44, 7393–7408.

[73] Clark, T. Semiempirical Molecular Orbital Theory: Facts, Myths and Legends. In *Recent Experimental and Computational Advances in Molecular Spectroscopy*, Vol. 406; Fausto, R., Ed.; Springer: Dordrecht, 1993.

[74] Thiel, W. Perspectives on Semiempirical Molecular Orbital Theory, *Adv. Chem. Phys.* 1996, 93, 703–757.

[75] Thiel, W. Thermochemistry from Semiempirical Molecular Orbital Theory. In *Computational Thermochemistry: Prediction and Estimation of Molecular Thermodynamics*, Vol. 677; Irikura, K. K.; Frurip, D. J., Eds.; American Chemical Society: Washington, DC, 1998.

[76] Clark, T. Quo Vadis Semiempirical MO-Theory?, *J. Mol. Struct. Theochem* 2000, 530, 1–10.

[77] Bredow, T.; Jug, K. Theory and Range of Modern Semiempirical Molecular Orbital Methods, *Theor. Chem. Acc.* 2004, 113, 1–14.

[78] Stewart, J. P. P. Semiempirical Molecular Orbital Methods. In *Reviews in Computational Chemistry*; Lipkowitz, K. B.; Boyd, D. B., Eds.; John Wiley & Sons: New York, 2007.

[79] Clark, T.; Stewart, J. J. P. MNDO-Like Semiempirical Molecular Orbital Theory and Its Application to Large Systems. In *Computational Methods for Large Systems: Electronic Structure Approaches for Biotechnology and Nanotechnology*; Reimers, J. R., Ed.; John Wiley & Sons: New York, 2011.

[80] Jug, K. *Zweihundert Jahre Entwicklung der Theoretischen Chemie im deutschsprachigen Raum*; Springer: Berlin, 2015, pp. 135–164.

[81] Lewars, E. G. *Introduction to the Theory and Applications of Molecular and Quantum Mechanics*; Springer: Cham, 2016, pp. 391–444.

[82] Bredow, T.; Jug, K. Semiempirical Molecular Orbital Methods. In *Handbook of Solid State Chemistry*; Dronskowski, R.; Kikkawa, S.; Stein, A., Eds.; John Wiley & Sons: New York, 2017.

[83] Condon, E. U.; Shortley, G. H. *The Theory of Atomic Spectra*; Cambridge University Press: Cambridge, 1959, pp. 174–177.
[84] Pelikán, P.; Turi Nagi, L. Expression of monocentric repulsion integrals on the basis of s,p,d atomic orbitals, *Chem. Pap.* 1974, 28, 594–598.

[85] Kumar, A.; Mishra, P. C. Evaluation of one-centre electron interaction integrals over Slater type atomic orbitals, *Pramana* 1987, 29, 385–390.

[86] Glaeske, H.-J.; Reinhold, J.; Volkmer, P. *Quantenchemie - Ein Lehrgang. Band 5: Ausgewählte mathematische Methoden der Chemie*; VEB Deutscher Verlag der Wissenschaften: Leipzig, 1987, pp. 607–609.

[87] Řezáč, J.; Fanfrlík, J.; Salahub, D.; Hobza, P. Semiempirical Quantum Chemical PM6 Method Augmented by Dispersion and H-Bonding Correction Terms Reliably Describes Various Types of Noncovalent Complexes, *J. Chem. Theory Comput.* 2009, 5, 1749–1760.

[88] Korth, M.; Pitonak, M.; Rezac, J.; Hobza, P. A Transferable H-Bonding Correction for Semiempirical Quantum-Chemical Methods, *J. Chem. Theory Comput.* 2010, 6, 344–352.

[89] Korth, M. Third-Generation Hydrogen-Bonding Corrections for Semiempirical QM Methods and Force Fields, *J. Chem. Theory Comput.* 2010, 6, 3808–3816.

[90] Řezáč, J.; Hobza, P. Advanced Corrections of Hydrogen Bonding and Dispersion for Semiempirical Quantum Mechanical Methods, *J. Chem. Theory Comput.* 2012, 8, 141–151.

[91] Brahmkshatriya, P. S.; Dobes, P.; Fanfrlík, J.; Rezac, J.; Paruch, K.; Bronowska, A.; Lepsík, M.; Hobza, P. Quantum Mechanical Scoring: Structural and Energetic Insights into Cyclin-Dependent Kinase 2 Inhibition by Pyrazolo[1,5-a]Pyrimidines, *Curr. Comput. Aided Drug Des.* 2013, 9, 118–129.

[92] Kromann, J. C.; Christensen, A. S.; Steinmann, C.; Korth, M.; Jensen, J. H. A Third-Generation Dispersion and Third-Generation Hydrogen Bonding Corrected PM6 Method: PM6-D3H+, *PeerJ* 2014, 2, e449.

[93] Vorlová, B.; Nachtigallová, D.; Jirásková-Vaníčková, J.; Ajani, H.; Jansa, P.; Řezáč, J.; Fanfrlík, J.; Otyepka, M.; Hobza, P.; Konvalinka, J.; Lepsík, M. Malonate-Based Inhibitors of Mammalian Serine Racemase: Kinetic Characterization and Structure-Based Computational Study, *Eur. J. Med. Chem.* 2015, 89, 189–197.
[94] Tuttle, T.; Thiel, W. OMx-D: Semiempirical Methods with Orthogonalization and Dispersion Corrections. Implementation and Biochemical Application, *Phys. Chem. Chem. Phys.* **2008**, *10*, 2159–2166.

[95] Řezáč, J.; Hobza, P. A Halogen-Bonding Correction for the Semiempirical PM6 Method, *Chem. Phys. Lett.* **2011**, *506*, 286–289.

[96] Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. PDDG/PM3 and PDDG/MNDO: Improved Semiempirical Methods, *J. Comput. Chem.* **2002**, *23*, 1601–1622.

[97] Tubert-Brohman, I.; Guimarães, C. R. W.; Repasky, M. P.; Jorgensen, W. L. Extension of the PDDG/PM3 and PDDG/MNDO Semiempirical Molecular Orbital Methods to the Halogens, *J. Comput. Chem.* **2004**, *25*, 138–150.

[98] Christensen, A. S.; Kubař, T.; Cui, Q.; Elstner, M. Semiempirical Quantum Mechanical Methods for Noncovalent Interactions for Chemical and Biochemical Applications, *Chem. Rev.* **2016**, *116*, 5301–5337.

[99] Szabo, A.; Ostlund, N. S. *Modern Quantum Chemistry: Introduction to Advanced Electronic Structure Theory*; Dover Publications: New York, 1996.

[100] Helgaker, T.; Jorgensen, P.; Olsen, J. *Molecular Electronic-Structure Theory*; John Wiley & Sons: Chichester, 2012.

[101] Löwdin, P.-O. On the Nonorthogonality Problem, *Adv. Quantum Chem.* **1970**, *5*, 185–199.

[102] Fischer-Hjalmars, I. Zero Differential Overlap in $\phi$-Electron Theories, *Adv. Quantum Chem.* **1966**, *2*, 25–46.

[103] Cook, D. B.; Hollis, P. C.; McWeeny, R. Approximate Ab Initio Calculations on Polyatomic Molecules, *Mol. Phys.* **1967**, *13*, 553–571.

[104] Roby, K. R.; Sinanoğlu, O. On the Performance and Parameter Problems of Approximate Molecular Orbital Theory, with Comparative Calculations on the Carbon Monoxide Molecule, *Int. J. Quantum Chem.* **1969**, *3*, 223–236.

[105] Roby, K. R. On the Justifiability of Neglect of Differential Overlap Molecular Orbital Methods, *Chem. Phys. Lett.* **1971**, *11*, 6–10.
[106] Roby, K. R. Fundamentals of an Orthonormal Basis Set Molecular Orbital Theory, *Chem. Phys. Lett.* **1972**, *12*, 579–582.

[107] Brown, R. D.; Burden, F. R.; Williams, G. R.; Phillips, L. F. Simplified Ab-Initio Calculations on Hydrogen-Containing Molecules, *Theor. Chim. Acta* **1971**, *21*, 205–210.

[108] Brown, R. D.; Burton, P. G. ‘Balance’ and Predictive Capability in Approximate Molecular Orbital Theory, *Chem. Phys. Lett.* **1973**, *20*, 45–49.

[109] Weinhold, F.; Carpenter, J. E. A Collection of Papers Presented at the First World Congress of Theoretical Chemists: Some Remarks on Nonorthogonal Orbitals in Quantum Chemistry, *J. Mol. Struct. Theocem* **1988**, *165*, 189–202.

[110] Neymeyr, K.; Seelig, F. F. “Neglect of Diatomic Differential Overlap” in Nonempirical Quantum Chemical Orbital Theories. I. On the Justification of the Neglect of Diatomic Differential Overlap Approximation, *Int. J. Quantum Chem.* **1995**, *53*, 515–518.

[111] Neymeyr, K.; Seelig, F. F. “Neglect of Diatomic Differential Overlap” in Nonempirical Quantum Chemical Orbital Theories. II. A Polynomial Expansion for $\Delta^{-1/2}$ in Terms of Legendre and Chebyshev Polynomials, *Int. J. Quantum Chem.* **1995**, *53*, 519–535.

[112] Neymeyr, K.; Engel, K. “Neglect of Diatomic Differential Overlap” in Nonempirical Quantum Chemical Orbital Theories. III. On the Spectrum of the Overlap Matrix for Diatomic Molecules over Locally Orthonormalized Basis Functions, *Int. J. Quantum Chem.* **1995**, *53*, 537–540.

[113] Neymeyr, K. “Neglect of Diatomic Differential Overlap” in Nonempirical Quantum Chemical Orbital Theories. IV. An Examination of the Justification of the Neglect of Diatomic Differential Overlap (NDDO) Approximation, *Int. J. Quantum Chem.* **1995**, *53*, 541–552.

[114] Neymeyr, K. “Neglect of Diatomic Differential Overlap” in Nonempirical Quantum Chemical Orbital Theories. V. A Calculus of Error Concerning the Justification of the Neglect of Diatomic Differential Overlap (NDDO) Approximation, *Int. J. Quantum Chem.* **1995**, *53*, 553–568.
[115] Koch, W. Neglect of Diatomic Differential Overlap (NDDO) in Non-Empirical Quantum Chemical Orbital Theories, Z. Naturforsch. A 1993, 48, 819–828.

[116] Husch, T.; Reiher, M. Comprehensive analysis of the neglect of diatomic differential overlap approximation, J. Chem. Theory Comput. 2018, submitted, [arXiv: 1806.05615].

[117] Stevens, W. J.; Basch, H.; Krauss, M. Compact Effective Potentials and Efficient Shared-exponent Basis Sets for the First- and Second-row Atoms, J. Chem. Phys. 1984, 81, 6026–6033.

[118] Sustmann, R.; Williams, J. E.; Dewar, M. J. S.; Allen, L. C.; von Rague Schleyer, P. Molecular Orbital Calculations on Carbonium Ions. II. Methyl, Ethyl, and Vinyl Cations. The Series C3H7+, J. Am. Chem. Soc. 1969, 91, 5350–5357.

[119] Koster, J. L.; Ruttink, P. J. A. Non-empirical approximate calculations for the ground states of H2 and H3 including complete configuration interactions, Chem. Phys. Lett. 1972, 17, 419–421.

[120] Birner, P.; Köhler, H. J.; Weiss, C. C–H Acidity Comparative CNDO/2 and NDDO Calculations on the Reactivity of Azabenzenes, Chem. Phys. Lett. 1974, 27, 347–350.

[121] Chandler, G. S.; Grader, F. E. A Re-Examination of the Justification of Neglect of Differential Overlap Approximations in Terms of a Power Series Expansion in S, Theor. Chim. Acta 1980, 54, 131–144.

[122] Duke, B. J.; Collins, M. P. S. The Ab Initio Neglect of Differential Diatomic Overlap Method, Theor. Chim. Acta 1981, 58, 233–244.

[123] Tu, Y.; Jacobsson, S. P.; Laaksonen, A. Re-examination of the NDDO approximation and introduction of a new model beyond it, Mol. Phys. 2003, 101, 3009–3015.

[124] Cao, X.; Dolg, M. Pseudopotentials and Modelpotentials, WIREs Comput. Mol. Sci. 2011, 1, 200–210.

[125] Dolg, M.; Cao, X. Relativistic Pseudopotentials: Their Development and Scope of Applications, Chem. Rev. 2012, 112, 403–480.

[126] Zerner, M. C. Removal of Core Orbitals in ‘Valence Orbital Only’ Calculations, Mol. Phys. 1972, 23, 963–978.
[127] Ridley, J.; Zerner, M. An Intermediate Neglect of Differential Overlap Technique for Spectroscopy: Pyrrole and the Azines, *Theor. Chim. Acta* **1973**, *32*, 111–134.

[128] Freed, K. F. Is There a Bridge between Ab Initio and Semiempirical Theories of Valence?, *Acc. Chem. Res.* **1983**, *16*, 137–144.

[129] Freed, K. F. Building A Bridge Between Ab Initio and Semiempirical Theories of Molecular Electronic Structure. In *Structure and Dynamics of Atoms and Molecules: Conceptual Trends*; Calais, J. L.; Kryachko, E. S., Eds.; Springer: Dordrecht, 1995.

[130] Kolos, W. Possible Improvements of the Interaction Energy Calculated Using Minimal Basis Sets, *Theor. Chim. Acta* **1979**, *51*, 219–240.

[131] Francl, M. M.; Pietro, W. J.; Hehre, W. J.; Binkley, J. S.; Gordon, M. S.; DeFrees, D. J.; Pople, J. A. Self-consistent Molecular Orbital Methods. XXIII. A Polarization-type Basis Set for Second-row Elements, *J. Chem. Phys.* **1982**, *77*, 3654–3665.

[132] Davidson, E. R.; Feller, D. Basis Set Selection for Molecular Calculations, *Chem. Rev.* **1986**, *86*, 681–696.

[133] Giese, T. J.; York, D. M. Improvement of Semiempirical Response Properties with Charge-Dependent Response Density, *J. Chem. Phys.* **2005**, *123*, 164108.

[134] Li, A.; Muddana, H. S.; Gilson, M. K. Quantum Mechanical Calculation of Noncovalent Interactions: A Large-Scale Evaluation of PMx, DFT, and SAPT Approaches, *J. Chem. Theory Comput.* **2014**, *10*, 1563–1575.

[135] Pople, J. A. A Priori Geometry Predictions. In *Modern Theoretical Chemistry*, Vol. 4; Schaefer III, H. F., Ed.; Springer: New York, 1977.

[136] Kulik, H. J.; Luehr, N.; Ufimtsev, I. S.; Martinez, T. J. Ab Initio Quantum Chemistry for Protein Structures, *J. Phys. Chem. B* **2012**, *116*, 12501–12509.

[137] Sure, R.; Grimme, S. Corrected Small Basis Set Hartree-Fock Method for Large Systems, *J. Comput. Chem.* **2013**, *34*, 1672–1685.

[138] Dewar, M. J. S.; Thiel, W. Ground States of Molecules. 39. MNDO Results for Molecules Containing Hydrogen, Carbon, Nitrogen, and Oxygen, *J. Am. Chem. Soc.* **1977**, *99*, 4907–4917.
[139] Stewart, J. J. P. Optimization of parameters for semiempirical methods IV: extension of MNDO, AM1, and PM3 to more main group elements, *J. Mol. Model.* **2004**, *10*, 155–164 and the references cited herein.

[140] Nanda, D. N.; Narasimhan, P. T. On Invariance Requirements in Approximate SCF MO Theory, *Int. J. Quantum Chem.* **1977**, *12*, 215–223.

[141] Klopman, G. A Semiempirical Treatment of Molecular Structures. II. Molecular Terms and Application to Diatomic Molecules, *J. Am. Chem. Soc.* **1964**, *86*, 4550–4557.

[142] Dewar, M. J. S.; Klopman, G. Ground States of \(\sigma\)-Bonded Molecules. I. Semiempirical S.C.F. Molecular Orbital Treatment of Hydrocarbons, *J. Am. Chem. Soc.* **1967**, *89*, 3089–3098.

[143] Voigt, B. On Bridging the Gap between the INDO and the NDDO Schemes, *Theor. Chim. Acta* **1973**, *31*, 289–295.

[144] Dewar, M. J. S.; Lo, D. H. Ground States of \(\sigma\)-Bonded Molecules. XVII. Fluorine Compounds, *J. Am. Chem. Soc.* **1972**, *94*, 5296–5303.

[145] Denton, J.; McCourt, M.; McIver, J. W. New Formulas for Integrals in Semiempirical Molecular Orbital Methods: Part 2. Coulomb Integrals and Their First and Second Cartesian Derivatives in the NDDO Theory, *J. Mol. Struct. Theochem* **1988**, *163*, 355–388.

[146] Dewar, M. J. S.; Jie, C.; Yu, J. SAM1; The First of a New Series of General Purpose Quantum Mechanical Molecular Models, *Tetrahedron* **1993**, *49*, 5003–5038.

[147] Holder, A. J.; Dennington, R. D.; Jie, C. Addendum to SAM1 Results Previously Published, *Tetrahedron* **1994**, *50*, 627–638.

[148] Holder, A. J.; Evleth, E. M. SAM1: General Description and Performance Evaluation for Hydrogen Bonds. In *Modeling the Hydrogen Bond*, Vol. 569; Smith, D. A., Ed.; American Chemical Society: Washington, DC, 1994.

[149] Stewart, J. J. P. Application of the PM6 Method to Modeling the Solid State, *J. Mol. Model.* **2008**, *14*, 499–535.

[150] Margraf, J. T.; Hennemann, M.; Meyer, B.; Clark, T. EMPIRE: a highly parallel semiempirical molecular orbital program: 2: periodic boundary conditions, *J. Mol. Model.* **2015**, *21*, 144.
[151] Weber, W. *Ein neues semiempirisches NDDO-Verfahren mit Orthogonalisierungskorrekturen: Entwicklung des Modells, Implementierung, Parametrisierung und Anwendungen*, Thesis, Universität Zürich, 1996.

[152] Scholten, M. *Semiemirische Verfahren mit Orthogonalisierungskorrekturen: Die OM3 Methode*, Thesis, Heinrich-Heine-Universität Düsseldorf, 2003.

[153] Brown, R. D.; Roby, K. R. Approximate Molecular Orbital Theory for Inorganic Molecules, *Theor. Chim. Acta* 1970, 16, 175–193.

[154] Spanget-Larsen, J. On Bridging the Gap between Extended Hückel and NDO Type LCAO-MO Theories, *Theor. Chim. Acta* 1980, 55, 165–172.

[155] De Bruijn, S. Analysis of the Inadequacies of Some Semi-Empirical MO Methods as Theories of Structure and Reactivity, *Int. J. Quantum Chem.* 1984, 25, 367–390.

[156] Pople, J. A.; Segal, G. A. Approximate Self-Consistent Molecular Orbital Theory. III. CNDO Results for AB2 and AB3 Systems, *J. Chem. Phys.* 1966, 44, 3289–3296.

[157] Goeppert-Mayer, M.; Sklar, A. L. Calculations of the Lower Excited Levels of Benzene, *J. Chem. Phys.* 1938, 6, 645–652.

[158] Coffey, P. Potential Energy Integrals in Semiempirical MO Methods, *Int. J. Quantum Chem.* 1974, 8, 263–266.

[159] Kollmar, C.; Böhm, M. C. An Analysis of the Zero Differential Overlap Approximation. Towards an Improved Semiempirical MO Method beyond It, *Theor. Chim. Acta* 1995, 92, 13–47.

[160] Mulliken, R. S. The theory of molecular orbitals, *J. Chim. Phys.* 1949, 46, 497.

[161] Jug, K. Operator Equations in Approximate Molecular Orbital Theories, *Theor. Chim. Acta* 1971, 23, 183–194.

[162] Coffey, P.; Jug, K. Semiempirical Molecular Orbital Calculations and Molecular Energies. A New Formula for the $\beta$ Parameter, *J. Am. Chem. Soc.* 1973, 95, 7575–7580.

[163] De Bruijn, S. Resonance Integrals in Semi-Empirical MO Theories, *Chem. Phys. Lett.* 1978, 54, 399–406.
[164] Horn, A. H. C.; Lin, Jr-H.; Clark, T. Multipole electrostatic model for MNDO-like techniques with minimal valence spd-basis sets, *Theor. Chem. Acc.* 2005, 114, 159–168.

[165] Horn, A. H. C.; Lin, Jr-H.; Clark, T. Erratum: Multipole electrostatic model for MNDO-like techniques with minimal valence spd-basis sets, *Theor. Chem. Acc.* 2007, 117, 461–465.

[166] Burstein, K. Y.; Isaev, A. N. MNDO Calculations on Hydrogen Bonds. Modified Function for Core-Core Repulsion, *Theor. Chim. Acta* 1984, 64, 397–401.

[167] Voityuk, A. A.; Rösch, N. AM1/d Parameters for Molybdenum, *J. Phys. Chem. A* 2000, 104, 4089–4094.

[168] Winget, P.; Horn, A. H. C.; Selçuki, C.; Martin, B.; Clark, T. AM1* Parameters for Phosphorus, Sulfur and Chlorine, *J. Mol. Model.* 2003, 9, 408–414.

[169] Winget, P.; Clark, T. AM1* Parameters for Aluminum, Silicon, Titanium and Zirconium, *J. Mol. Model.* 2005, 11, 439–456.

[170] Kayi, H.; Clark, T. AM1* Parameters for Copper and Zinc, *J. Mol. Model.* 2007, 13, 965–979.

[171] Kayi, H.; Clark, T. AM1* Parameters for Vanadium and Chromium, *J. Mol Model* 2009, 15, 253–269.

[172] Kayi, H.; Clark, T. AM1* Parameters for Bromine and Iodine, *J Mol Model* 2009, 15, 295–308.

[173] Kayi, H.; Clark, T. AM1* Parameters for Manganese and Iron, *J Mol Model* 2010, 16, 1109–1126.

[174] Kayi, H.; Clark, T. AM1* Parameters for Cobalt and Nickel, *J Mol Model* 2010, 16, 29–47.

[175] Kayi, H. AM1* Parameters for Gold, *J Mol Model* 2010, 16, 1029–1038.

[176] Kayi, H.; Clark, T. AM1* Parameters for Palladium and Silver, *J Mol Model* 2011, 17, 2585–2600.

[177] Hennemann, M.; Clark, T. EMPIRE: A highly parallel semiempirical molecular orbital program: 1: Self-Consistent Field Calculations *J. Mol. Model.* 2014, 20, 2331–2342.
[178] Stewart, J. P. P. MOPAC2016, version 17.231, Stewart Computational Chemistry, Colorado Springs, CO, USA, [http://OpenMOPAC.net](http://OpenMOPAC.net) (Accessed: 20. April 2018).

[179] Grimme, S.; Hansen, A.; Brandenburg, J. G.; Bannwarth, C. Dispersion-Corrected Mean-Field Electronic Structure Methods, *Chem. Rev.* **2016**, *116*, 5105–5154.

[180] Brandenburg, J. G.; Hochheim, M.; Bredow, T.; Grimme, S. Low-Cost Quantum Chemical Methods for Noncovalent Interactions, *J. Phys. Chem. Lett.* **2014**, *5*, 4275–4284.

[181] Sure, R.; Grimme, S. Comprehensive Benchmark of Association (Free) Energies of Realistic Host–Guest Complexes, *J. Chem. Theory Comput.* **2015**, *11*, 3785–3801.

[182] Kolb, M. *Ein neues semiempirisches Verfahren auf Grundlage der NDDO-Näherung: Entwicklung der Methode, Parametrisierung und Anwendungen*, Thesis, Universität Wuppertal, 1991.

[183] Oleari, L.; Sipio, L. D.; Michelis, G. D. The Evaluation of the One-Centre Integrals in the Semi-Empirical Molecular Orbital Theory, *Mol. Phys.* **1966**, *10*, 97–109.

[184] Gleghorn, J. T.; McConkey, F. W. Extended Basis NDDO Calculations on Diatomic Molecules, *Theor. Chim. Acta* **1982**, *61*, 283–293.

[185] Zhidomirov, G. M.; Zhanpeisov, N. U.; Zilberberg, I. L.; Yudanov, I. V. On Some Ways of Modifying Semiempirical Quantum Chemical Methods, *Int. J. Quantum Chem.* **1996**, *58*, 175–184.

[186] Spanget-Larsen, J. The Alternant Hydrocarbon Pairing Theorem and All-Valence Electrons Theory. An Approximate LCAO Theory for the Electronic Absorption and MCD Spectra of Conjugated Organic Compounds, Part 2, *Theor. Chem. Acc.* **1997**, *98*, 137–153.

[187] Kollmar, C. A New Concept for an Approximate MO Formalism, *Chem. Phys. Lett.* **1997**, *269*, 215–221.

[188] Nanda, D. N.; Jug, K. SINDO1. A Semiempirical SCF MO Method for Molecular Binding Energy and Geometry I. Approximations and Parametrization, *Theor. Chim. Acta* **1980**, *57*, 95–106.
[189] Filatov, M. J.; Gritsenko, O. V.; Zhidomirov, G. M. CNDO-S2—a Semiempirical SCF MO Method for Transition Metal Organometallics, *Theor. Chim. Acta* 1987, 72, 211–222.

[190] Jug, K.; Iffert, R.; Schulz, J. Development and Parametrization of SINDO1 for Second-Row Elements, *Int. J. Quantum Chem.* 1987, 32, 265–277.

[191] Gray, N. A. B.; Stone, A. J. Justifiability of the ZDO approximation in terms of a power series expansion, *Theor. Chim. Acta* 1970, 18, 389–390.

[192] Wadt, W. R.; Hay, P. J. Ab Initio Effective Core Potentials for Molecular Calculations. Potentials for Main Group Elements Na to Bi, *J. Chem. Phys.* 1985, 82, 284–298.

[193] Chandrasekhar, J.; Mehrotra, P. K.; Subramanian, S.; Manoharan, P. T. NDDO MO Calculations, *Theor. Chim. Acta* 1976, 41, 243–256.

[194] Chandrasekhar, J.; Mehrotra, P. K.; Subramanian, S.; Manoharan, P. T. NDDO MO Calculations, *Theor. Chim. Acta* 1979, 52, 303–310.

[195] Laikov, D. N. Neglect of Four- and Approximation of One-, Two-, and Three-Center Two-Electron Integrals in a Symmetrically Orthogonalized Basis, *J. Comput. Chem.* 2007, 28, 698–702.

[196] Chang, D. T.; Schenter, G. K.; Garrett, B. C. Self-Consistent Polarization Neglect of Diatomic Differential Overlap: Application to Water Clusters, *J. Chem. Phys.* 2008, 128, 164111.

[197] Laikov, D. N. Intrinsic Minimal Atomic Basis Representation of Molecular Electronic Wavefunctions, *Int. J. Quantum Chem.* 2011, 111, 2851–2867.

[198] Laikov, D. N. A New Parametrizable Model of Molecular Electronic Structure, *J. Chem. Phys.* 2011, 135, 134120.

[199] Jug, K. Mechanism of Cyclopropane-Propene Isomerization, *Theor. Chim. Acta* 1976, 42, 303–310.

[200] Jug, K.; Geudtner, G. Treatment of Hydrogen Bonding in SINDO1, *J. Comput. Chem.* 1993, 14, 639–646.
[201] Zuber, G.; Hug, W. Rarefied Basis Sets for the Calculation of Optical Tensors. 1. The Importance of Gradients on Hydrogen Atoms for the Raman Scattering Tensor, *J. Phys. Chem. A* 2004, 108, 2108–2118.

[202] Hinze, J.; Jaffé, H. H. Electronegativity. I. Orbital Electronegativity of Netural Atoms, *J. Am. Chem. Soc.* 1962, 84, 540–546.

[203] Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Challenges for Density Functional Theory, *Chem. Rev.* 2012, 112, 298–320.

[204] Perdew, J. P.; Parr, R. G.; Levy, M.; Balduz, Jr., J. L. Density-Functional Theory for Fractional Particle Number: Derivative Discontinuities of the Energy, *Phys. Rev. Lett.* 1982, 49, 1691–1694.

[205] Cohen, A. J.; Mori-Sánchez, P.; Yang, W. Insights into Current Limitations of Density Functional Theory, *Science* 2008, 321, 792–794.

[206] Mori-Sánchez, P.; Cohen, A. J. The derivative discontinuity of the exchange-correlation functional, *Phys. Chem. Chem. Phys.* 2014, 16, 14378–14387.

[207] Sichel, J. M.; Whitehead, M. A. Atomic Parameters for Semi-Empirical SCF-LCAO-MO Calculations, *Theor. Chim. Acta* 1967, 7, 32–40.

[208] Di Sipio, L.; Tondello, E.; De Michelis, G.; Oleari, L. Semi-Empirical Molecular Orbital Theory. The One-Centre Quantities for the Elements of the First and Second Transition Series, *Chem. Phys. Lett.* 1971, 11, 287–289.

[209] Margraf, J. T.; Claudino, D.; Bartlett, R. J. Determination of Consistent Semiempirical One-Centre Integrals Based on Coupled-Cluster Theory, *Mol. Phys.* 2017, 115, 538–544.

[210] Bingham, R. C.; Dewar, M. J. S.; Lo, D. H. Ground States of Molecules. XXV. MINDO/3. Improved Version of the MINDO Semiempirical SCF-MO Method, *J. Am. Chem. Soc.* 1975, 97, 1285–1293.

[211] Hicks, M. G.; Thiel, W. Reference energies in semiempirical parametrizations, *J. Comput. Chem.* 1986, 7, 213–218.

[212] Rossi, I.; Truhlar, D. G. Parameterization of NDDO Wavefunctions Using Genetic Algorithms. An Evolutionary Approach to Parameterizing Potential Energy Surfaces and Direct Dynamics Calculations for Organic Reactions, *Chem. Phys. Lett.* 1995, 233, 231–236.
[213] Brothers, E. N.; Merz, K. M. Sodium Parameters for AM1 and PM3 Optimized Using a Modified Genetic Algorithm, *J. Phys. Chem. B* 2002, 106, 2779–2785.

[214] Hutter, M. C.; Reimers, J. R.; Hush, N. S. Modeling the Bacterial Photosynthetic Reaction Center. 1. Magnesium Parameters for the Semiempirical AM1 Method Developed Using a Genetic Algorithm, *J. Phys. Chem. B* 1998, 102, 8080–8090.

[215] Cox, J. D.; Wagman, D. D.; Medvedev, V. A. *CODATA Key Values for Thermodynamics*; Hemisphere Publishing Corporation: New York, 1989.

[216] McQuarrie, D. A. *Statistical Mechanics*; University Science Books: Sausalito, 2000.

[217] Repasky, M. P.; Chandrasekhar, J.; Jorgensen, W. L. Improved Semiempirical Heats of Formation through the Use of Bond and Group Equivalents, *J. Comput. Chem.* 2002, 23, 498–510.

[218] Kromann, J. C.; Welford, A.; Christensen, A. S.; Jensen, J. H. Random Versus Systematic Errors in Reaction Enthalpies Computed Using Semi-Empirical and Minimal Basis Set Methods, *ACS Omega* 2018, 3, 4372–4377.

[219] Winget, P.; Selçuki, C.; Horn, A. H. C.; Martin, B.; Clark, T. Towards a “next Generation” Neglect of Diatomic Differential Overlap Based Semiempirical Molecular Orbital Technique, *Theor. Chem. Acc.* 2003, 110, 254–266.

[220] Proppe, J.; Reiher, M. Reliable Estimation of Prediction Uncertainty for Physicochemical Property Models, *J. Chem. Theory Comput.* 2017, 13, 3297–3317.

[221] Weymuth, T.; Proppe, J.; Reiher, M. Statistical Analysis of Semi-classical Dispersion Corrections, *J. Chem. Theory Comput.* 2018, 14, 2480–2494.

[222] Efron, B. Bootstrap Methods: Another Look at the Jackknife, *Ann. Stat.* 1979, 7, 1–26.

[223] Hastie, T.; Tibshirani, R. J.; Friedman, J. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*; Springer: New York, 2016.
[224] Oreluk, J.; Liu, Z.; Hegde, A.; Li, W.; Packard, A.; Frenklach, M.; Zubarev, D. Diagnostics of Data-Driven Models: Uncertainty Quantification of PM7 Semi-Empirical Quantum Chemical Method, 2018, submitted, [arXiv: 1806.04813].

[225] Jiang, W.; DeYonker, N. J.; Wilson, A. K. Multireference Character for 3d Transition-Metal-Containing Molecules, J. Chem. Theory Comput. 2012, 8, 460–468.

[226] http://openmopac.net/PM7_accuracy/molecules.html (Accessed: 20. April 2018).

[227] Nieke, C.; Reinhold, J. NDDO Study of the Jahn—Teller Distortions of Octahedral Transition Metal Compounds, J. Mol. Struct. Theocem 1985, 124, 87–92.

[228] Nieke, C.; Reinhold, J. NDDO Study of the Coordination Structure of M2+(S2CNH2)2 Complexes (M = Ni, Cu), J. Mol. Struct. Theocem 1986, 139, 241–245.

[229] Filatov, M. J.; Zilberberg, I. L.; Zhidomirov, G. M. NDDO/MC: A New Semiempirical SCFMO Method for Transition Metal Complexes, Int. J. Quantum Chem. 1992, 44, 565–585.

[230] Bosque, R.; Maseras, F. Performance of the Semiempirical PM3 (Tm) Method in the Geometry Optimization of Transition Metal Complexes, J. Comput. Chem. 2000, 21, 562–571.

[231] Minenkov, Y.; Sharapa, D. I.; Cavallo, L. Application of Semiempirical Methods to Transition Metal Complexes: Fast Results but Hard-to-Predict Accuracy, J. Chem. Theory Comput. doi:10.1021/acs.jctc.8b00018.

[232] Husch, T.; Freitag, L.; Reiher, M. Calculation of Ligand Dissociation Energies in Large Transition-Metal Complexes, J. Chem. Theory Comput. 2018, 14, 2456–2468.

[233] Weymuth, T.; Couzijn, E. P. A.; Chen, P.; Reiher, M. New Benchmark Set of Transition-Metal Coordination Reactions for the Assessment of Density Functionals, J. Chem. Theory Comput. 2014, 10, 3092–3103.

[234] Korth, M.; Thiel, W. Benchmarking Semiempirical Methods for Thermochemistry, Kinetics, and Noncovalent Interactions: OMx Methods Are Almost As Accurate and Robust As DFT-GGA Methods for Organic Molecules, J. Chem. Theory Comput. 2011, 7, 2929–2936.
[235] Dral, P. O.; Wu, X.; Spörkel, L.; Koslowski, A.; Thiel, W. Semiempirical Quantum-Chemical Orthogonalization-Corrected Methods: Benchmarks for Ground-State Properties, *J. Chem. Theory Comput.* **2016**, *12*, 1097–1120.

[236] Lopez, X.; York, D. M. Parameterization of Semiempirical Methods to Treat Nucleophilic Attacks to Biological Phosphates: AM1/d Parameters for Phosphorus, *Theor. Chem. Acc.* **2003**, *109*, 149–159.

[237] Giese, T. J.; Sherer, E. C.; Cramer, C. J.; York, D. M. A Semiempirical Quantum Model for Hydrogen-Bonded Nucleic Acid Base Pairs, *J. Chem. Theory Comput.* **2005**, *1*, 1275–1285.

[238] Nam, K.; Cui, Q.; Gao, J.; York, D. M. Specific Reaction Parametrization of the AM1/d Hamiltonian for Phosphoryl Transfer Reactions: H, O, and P Atoms, *J. Chem. Theory Comput.* **2007**, *3*, 486–504.

[239] Tejero, I.; González-Lafont, À.; Lluch, J. M. A PM3/d Specific Reaction Parameterization for Iron Atom in the Hydrogen Abstraction Catalyzed by Soybean Lipoxygenase-1, *J. Comput. Chem.* **2007**, *28*, 997–1005.

[240] Wu, X.; Thiel, W.; Pezeshki, S.; Lin, H. Specific Reaction Path Hamiltonian for Proton Transfer in Water: Reparameterized Semiempirical Models, *J. Chem. Theory Comput.* **2013**, *9*, 2672–2686.

[241] Liang, S.; Roitberg, A. E. AM1 Specific Reaction Parameters for Reactions of Hydroxide Ion with Halomethanes in Complex Environments: Development and Testing, *J. Chem. Theory Comput.* **2013**, *9*, 4470–4480.

[242] Zhou, Y.; Pu, J. Reaction Path Force Matching: A New Strategy of Fitting Specific Reaction Parameters for Semiempirical Methods in Combined QM/MM Simulations, *J. Chem. Theory Comput.* **2014**, *10*, 3038–3054.

[243] Saito, T.; Kitagawa, Y.; Takano, Y. Reparameterization of PM6 Applied to Organic Diradical Molecules, *J. Phys. Chem. A* **2016**, *120*, 8750–8760.

[244] Fredin, L. A.; Allison, T. C. Predicting Structures of Ru-Centered Dyes: A Computational Screening Tool, *J. Phys. Chem. A* **2016**, *120*, 2135–2143.
Bartók, A. P.; Kondor, R.; Csányi, G. On representing chemical environments, *Phys. Rev. B* 2013, 87, 184115.

Kee, R. J.; Miller, J. A.; Jefferson, T. H. CHEMKIN: a general-purpose, problem-independent, transportable, FORTRAN chemical kinetics code package, Technical Report SAND–80-8003, Sandia National Labs., Livermore, CA (USA), 1980.

Glowacki, D. R.; Liang, C.-H.; Morley, C.; Pilling, M. J.; Robertson, S. H. MESMER: An Open-Source Master Equation Solver for Multi-Energy Well Reactions, *J. Phys. Chem. A* 2012, 116, 9545–9560.

Proppe, J.; Husch, T.; Simm, G. N.; Reiher, M. Uncertainty quantification for quantum chemical models of complex reaction networks, *Faraday Discuss.* 2017, 195, 497–520.

Proppe, J.; Reiher, M. Mechanism Deduction from Noisy Chemical Reaction Networks, *J. Chem. Theory Comput.* 2018, submitted, [arXiv: 1803.09346].

Maeda, S.; Ohno, K.; Morokuma, K. Systematic Exploration of the Mechanism of Chemical Reactions: The Global Reaction Route Mapping (GRRM) Strategy Using the ADDF and AFIR Methods, *Phys. Chem. Chem. Phys.* 2013, 15, 3683–3701.

Magoon, G. R.; Green, W. H. Design and Implementation of a Next-Generation Software Interface for on-the-Fly Quantum and Force Field Calculations in Automated Reaction Mechanism Generation, *Computers & Chemical Engineering* 2013, 52, 35–45.

Rappoport, D.; Galvin, C. J.; Zubarev, D. Y.; Aspuru-Guzik, A. Complex Chemical Reaction Networks from Heuristics-Aided Quantum Chemistry, *J. Chem. Theory Comput.* 2014, 10, 897–907.

Zimmerman, P. M. Navigating Molecular Space for Reaction Mechanisms: An Efficient, Automated Procedure, *Mol. Simul.* 2015, 41, 43–54.

Habershon, S. Automated Prediction of Catalytic Mechanism and Rate Law Using Graph-Based Reaction Path Sampling, *J. Chem. Theory Comput.* 2016, 12, 1786–1798.

Bergeler, M.; Simm, G. N.; Proppe, J.; Reiher, M. Heuristics-Guided Exploration of Reaction Mechanisms, *J. Chem. Theory Comput.* 2015, 11, 5712–5722.
[256] Simm, G. N.; Reiher, M. Context-Driven Exploration of Complex Chemical Reaction Networks, *J. Chem. Theory Comput.* **2017**, *13*, 6108–6119.

[257] Marx, D.; Hutter, J. *Ab Initio Molecular Dynamics: Basic Theory and Advanced Methods*; Cambridge University Press: Cambridge, 2009.

[258] Wang, L.-P.; Titov, A.; McGibbon, R.; Liu, F.; Pande, V. S.; Martínez, T. J. Discovering Chemistry with an *Ab Initio* Nanoreactor, *Nat. Chem.* **2014**, *6*, 1044–1048.

[259] Martínez, T. J. *Ab Initio Reactive Computer Aided Molecular Design*, *Acc. Chem. Res.* **2017**, *50*, 652–656.

[260] Schweig, A.; Thiel, W. MNDOC Study of Excited States, *J. Am. Chem. Soc.* **1980**, *103*, 1425–1431.

[261] Thiel, W. The MNDOC Method, a Correlated Version of the MNDO Model, *J. Am. Chem. Soc.* **1981**, *103*, 1413–1420.

[262] Clark, T.; Chandrasekhar, J. NDDO-Based CI Methods for the Prediction of Electronic Spectra and Sum-Over-States Molecular Hyperpolarization, *Isr. J. Chem.* **1993**, *33*, 435–448.

[263] Liotard, D. A.; Holder, A. An Implementation of Configuration Interaction in a General Purpose Semiempirical Context, *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 587–593.

[264] Toniolo, A.; Ben-Nun, M.; Martínez, T. J. Optimization of Conical Intersections with Floating Occupation Semiempirical Configuration Interaction Wave Functions, *J. Phys. Chem. A* **2002**, *106*, 4679–4689.

[265] Toniolo, A.; Granucci, G.; Martínez, T. J. Conical Intersections in Solution: A QM/MM Study Using Floating Occupation Semiempirical Configuration Interaction Wave Functions, *J. Phys. Chem. A* **2003**, *107*, 3822–3830.

[266] Koslowski, A.; Beck, M. E.; Thiel, W. Implementation of a General Multireference Configuration Interaction Procedure with Analytic Gradients in a Semiempirical Context Using the Graphical Unitary Group Approach, *J. Comput. Chem.* **2003**, *24*, 714–726.

[267] Toniolo, A.; Thompson, A. L.; Martínez, T. J. Excited state direct dynamics of benzene with reparameterized multi-reference semiempirical configuration interaction methods, *Chem. Phys.* **2004**, *304*, 133–145.
[268] Lei, Y.; Suo, B.; Dou, Y.; Wang, Y.; Wen, Z. New Implementations of MRCI in Semiempirical Frameworks, *J. Comput. Chem.* 2010, 31, 1752-1758.

[269] Dral, P. O.; Clark, T. Semiempirical UNO–CAS and UNO–CI: Method and Applications in Nanoelectronics, *J. Phys. Chem. A* 2011, 115, 11303–11312.

[270] Liu, J.; Thiel, W. An efficient implementation of semiempirical quantum-chemical orthogonalization-corrected methods for excited-state dynamics, *J. Chem. Phys.* 2018, 148, 154103.

[271] Thiel, W. MNDOC study of reactive intermediates and transition states, *J. Am. Chem. Soc.* 1981, 103, 1420–1425.

[272] Silva-Junior, M. R.; Thiel, W. Benchmark of Electronically Excited States for Semiempirical Methods: MNDO, AM1, PM3, OM1, OM2, OM3, INDO/S, and INDO/S2, *J. Chem. Theory Comput.* 2010, 6, 1546–1564.

[273] Tuna, D.; Lu, Y.; Koslowski, A.; Thiel, W. Semiempirical Quantum-Chemical Orthogonalization-Corrected Methods: Benchmarks of Electronically Excited States, *J. Chem. Theory Comput.* 2016, 12, 4400–4422.

[274] Grimme, S.; Waletzke, M. A combination of Kohn–Sham density functional theory and multi-reference configuration interaction methods, *J. Chem. Phys.* 1999, 111, 5645–5655.

[275] Silva-Junior, M. R.; Schreiber, M.; Sauer, S. P. A.; Thiel, W. Benchmarks for electronically excited states: Time-dependent density functional theory and density functional theory based multireference configuration interaction, *J. Chem. Phys.* 2008, 129, 104103.

[276] Simm, G. N.; Proppe, J.; Reiher, M. Error Assessment of Computational Models in Chemistry, *Chimia Int. J. Chem.* 2017, 71, 202–208.

[277] Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach, *J. Chem. Theory Comput.* 2015, 11, 2087–2096.

[278] Simm, G.; Reiher, M. Error-Controlled Exploration of Chemical Reaction Networks with Gaussian Processes, *J. Chem. Theory Comput.* 2018, submitted, [arXiv: 1805.09886].
[279] Software for Chemical Interaction and Networks (SCINE), www.scine.ethz.ch (Accessed 6. June 2018).

[280] Sun, Q.; Berkelbach, T. C.; Blunt, N. S.; Booth, G. H.; Guo, S.; Li, Z.; Liu, J.; McClain, J. D.; Sayfutyarova, E. R.; Sharma, S.; Wouters, S.; Chan, G. K.-L. PySCF: the Python-based simulations of chemistry framework, WIREs Comput. Mol. Sci 2017, 8, e1340.

[281] Sun, Q. Libcint: An efficient general integral library for Gaussian basis functions, J. Comput. Chem. 2015, 36, 1664–1671.

[282] Thiel, W. MNDO2005, Max-Planck-Institut für Kohlenforschung: Mülheim an der Ruhr, Germany, 2005.

[283] Dewar, M. J. S.; Zoebisch, E. G. Extension of AM1 to the Halogens, Journal of Molecular Structure: THEOCHEM 1988, 180, 1–21.

[284] Dewar, M. J. S.; Jie, C.; Zoebisch, E. G. AM1 Calculations for Compounds Containing Boron, Organometallics 1988, 7, 513–521.

[285] Kayi, H.; Clark, T. AM1* Parameters for Copper and Zinc, J Mol Model 2007, 13, 965–979.

[286] Stewart, J. J. P. MOPAC: A semiempirical molecular orbital program, J. Comput. Aided Mol. Des. 1990, 4, 1–105.

[287] http://openmopac.net/manual/hcore.html (Accessed: 20. April 2018).

[288] Steinmann, C.; Blædel, K. L.; Christensen, A. S.; Jensen, J. H. Interface of the Polarizable Continuum Model of Solvation with Semi-Empirical Methods in the GAMESS Program, PLoS ONE 2013, 8, e67725.

[289] Dewar, M. J. S.; Yamaguchi, Y.; Suck, S. H. MNDO Calculations of Molecular Electric Polarizabilities, Hyperpolarizabilities, and Nonlinear Optical Coefficients, Chem. Phys. Lett. 1978, 59, 541–544.

[290] Manion, J. A. Evaluated Enthalpies of Formation of the Stable Closed Shell C1 and C2 Chlorinated Hydrocarbons, J. Phys. Chem. Ref. Data 2002, 31, 123–172.

[291] Knowlton, J. W.; Rossini, F. D. Heats of Combustion and Formation of Cyclopropane, J. Res. Natl. Bur. Stand. 1949, 43, 113–115.
[292] Kaarsemaker, S.; Coops, J. Thermal Quantities of Some Cycloparaffins. Part III. Results of Measurements, *Recl. Trav. Chim. Pays-Bas* 1952, 71, 261–276.

[293] Roux, M. V.; Temprado, M.; Chickos, J. S.; Nagano, Y. Critically Evaluated Thermochemical Properties of Polycyclic Aromatic Hydrocarbons, *J. Phys. Chem. Ref. Data* 2008, 37, 1855–1996.

[294] Good, W. D. The Enthalpies of Combustion and Formation of the Isomeric Pentanes, *J. Chem. Thermodyn.* 1970, 2, 237–244.

[295] Pittam, D. A.; Pilcher, G. Measurements of Heats of Combustion by Flame Calorimetry. Part 8.—Methane, Ethane, Propane, n-Butane and 2-Methylpropane, *J. Chem. Soc. Faraday Trans.* 1972, 68, 2224–2229.

[296] Clark, T.; Knox, T. M.; McKervey, M. A.; Mackle, H.; Rooney, J. J. Thermochemistry of Bridged-Ring Substances. Enthalpies of Formation of Some Diamondoid Hydrocarbons and of Perhydroquinacene. Comparisons with Data from Empirical Force Field Calculations, *J. Am. Chem. Soc.* 1979, 101, 2404–2410.

[297] Rogers, D. W.; Matsunaga, N.; Zavitsas, A. A.; McLafferty, F. J.; Liebman, J. F. The Conjugation Stabilization of 1,3-Butadiyne Is Zero, *Org. Lett.* 2003, 5, 2373–2375.