Comprehensive Analysis of the Neglect of Diatomic Differential Overlap Approximation

Tamara Husch and Markus Reiher*

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

June 15, 2018

Abstract

Many modern semiempirical molecular orbital models are built on the neglect of diatomic differential overlap (NDDO) approximation. An in-depth understanding of this approximation is therefore indispensable to rationalize the success of these semiempirical molecular orbital models and to develop further improvements on them. The NDDO approximation provides a recipe to approximate electron-electron repulsion integrals (ERIs) in a symmetrically orthogonalized basis based on a far smaller number of ERIs in a locally orthogonalized basis. We first analyze the NDDO approximation by comparing ERIs in both bases for a selection of molecules and for a selection of basis sets. We find that the errors in Hartree–Fock and second-order Møller–Plesset perturbation theory energies grow roughly linearly with the number of basis functions. We then examine different approaches to correct for the errors caused by the NDDO approximation and propose a strategy to directly correct for them in the two-electron matrices that enter the Fock operator.

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

The neglect of diatomic differential overlap (NDDO) approximation is the foundation for many modern semiempirical molecular orbital (SEMO) models such as the PM\(x\) \((x = 3, 6, 7)\)\(^{2-4}\) and OM\(x\) \((x = 1, 2, 3)\)\(^{5-7}\) models. These SEMO models are chosen when, on the one hand, accurate \textit{ab initio} electronic structure models are computationally unfeasible, but when, on the other hand, a calculation with an electronic structure model is favored over a classical force field to exploit the first-principles nature of the fundamental electrostatic interactions. Examples include the simulation of very large systems such as proteins\(^{8-15}\), virtual high-throughput screening schemes for materials and drug discovery\(^{16-23}\) or real-time interactive quantum chemical calculations\(^{24-30}\).

Originally, the NDDO approximation was conceptualized as a means to reduce the number of electron-electron repulsion integrals (ERIs) which need to be explicitly calculated in the course of a Hartree–Fock (HF) calculation.\(^1\) The NDDO approximation specifies how ERIs in a symmetrically orthogonalized basis may be approximated based on a far smaller number of ERIs in a locally orthogonalized basis. However, the NDDO approximation has not found acceptance in \textit{ab initio} calculations due to the significant errors which it introduced in ERIs in the symmetrically orthogonalized basis.\(^{31-49}\) Instead, the NDDO approximation has become popular in SEMO models where it is combined with various other approximations made to the one-electron matrix and to the nucleus-nucleus repulsion energy which benefit from mutual error compensation. Obviously, the dramatically reduced computational timings of NDDO-SEMO models come at a price: they are known to be notoriously unreliable.\(^{51,52}\) For further improvements to NDDO-SEMO models, an in-depth analysis of the NDDO approximation is mandatory. Despite decades of work on NDDO-SEMO models, a fully satisfactory analysis has not been provided yet. We intend to close this gap here and study the foundations of the NDDO approximation from a state-of-the-art perspective.

First, we determine how the NDDO approximation affects ERIs evaluated in a symmetrically orthogonalized basis. Previous analyses of the NDDO approximation...
approximation were limited to few tens of molecules which consisted of few
atoms (usually less than four heavy atoms).\textsuperscript{32–49} In this work, we consider
a diverse selection of molecules that are also much larger. As the errors in
the ERI\textsc{s} will propagate to all quantities calculated in an \textsc{nddo} framework,
we study how the \textsc{nddo} approximation affects the (absolute and relative)
\textsc{hf} and \textsc{second-order m}\textsc{øller–plesset perturbation (mp2)} theory energies. In
this context, we examine different basis set choices.

In general, the \textsc{nddo} approximation is only valid for a locally orthogo-
\textsc{nal} basis set,\textsuperscript{44–46,48} which appears to restrict contemporary \textsc{nddo-semo}
models to a minimal basis set. A minimal basis set, however, is generally
unsuitable for the description of atoms in molecules because it does not
yield reliable relative energies, force constants, electric dipole moments, static
dipole polarizabilities, and other properties.\textsuperscript{53–58} It was suggested to gen-
\textsc{eralize} \textsc{nddo-semo} models to larger, e.g., double-zeta split-valence basis sets
to obtain more accurate results.\textsuperscript{60} Two studies examined\textsuperscript{60,61} the effects of
the application of a double-zeta split-valence basis set in conjunction with
the \textsc{nddo} approximation, but came to the conclusion that, contrary to what
one would expect, the results did not improve compared to the results ob-
tained with a single-zeta basis set. In this work, we dissect in detail the
origins of this counterintuitive observation and elaborate on a solution.

It does not come as a surprise that we find — in agreement with previous
results\textsuperscript{32–49} — the \textsc{nddo} approximation to cause severe errors. We therefore
examine how one can correct for these errors. We briefly review the error
compensation strategy which contemporary \textsc{nddo-semo} models apply and
then propose a way to directly correct for errors in the two-electron matrices.
We show that our approach allows for rapid extended-basis set calculations
invoking the \textsc{nddo} approximation with error control.
2 Neglect of Diatomic Differential Overlap Approximation

2.1 Basic Notation

In a basis-set representation, each spatial molecular orbital \( \psi_i = \psi_i(r) \) is approximated as a linear combination of pre-defined basis functions. Following the well-established approach for finite systems in molecular physics, we choose the basis functions to be Gaussian-type atom-centered functions \( \chi_I^\mu = \chi_I^\mu(r) \). Our notation indicates that the \( \mu \)-th basis function of type \( \chi \) is centered on atom \( I \). Additionally, we require the basis functions \( \chi \) to be locally orthogonal which means that the overlap \( \chi S_{\mu\nu} \) between the basis functions \( \chi_I^\mu \) and \( \chi_J^\nu \) must fulfill,

\[
\chi S_{\mu\nu} = \begin{cases} 
\langle \chi_I^\mu | \chi_J^\nu \rangle & I \neq J, \ \forall \mu, \nu \\
\delta_{\mu\nu} & I = J, \ \mu \neq \nu 
\end{cases}.
\]

A molecular orbital \( \psi_i \) is then given as the sum of the \( M \) basis functions \( \chi_I^\mu \) weighted with expansion coefficients \( \chi C = \{ \chi C_\mu \} \),

\[
\psi_i(r) = \sum_{\mu=1}^M \chi C_\mu \chi_I^\mu(r).
\]

Throughout this work, we denote the bases by a left superscript, i.e., \( \chi^C \). In the \( \chi \)-basis, the Roothaan–Hall equation in the spin-restricted formulation then reads,\(^34\)

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

where \( \chi F \) is the Fock matrix, which depends on \( \chi^C \), \( \chi S \) is the overlap matrix, and \( \epsilon \) the diagonal matrix of orbital energies. As \( \epsilon \) is invariant under unitary matrix transformations with which we may transform one basis into another one, it does not carry a left superscript. The evaluation of the Fock-matrix elements in the \( \chi \)-basis,\(^34\)

\[
\chi F_{\mu\nu}(\chi^C) = \langle \chi_I^\mu | h | \chi_J^\nu \rangle + \sum_{\lambda\sigma} \chi P_{\lambda\sigma}(\chi^C) \left[ \langle \chi_I^\mu \chi_J^\nu | \chi^K_L \chi^L_K \rangle - \frac{1}{2} \langle \chi_I^\mu \chi_J^\nu | \chi^K_L \chi^L_K \rangle \right],
\]

(4)
requires the evaluation of one-electron integrals \( \langle \chi^I_{\mu} | h | \chi^J_{\nu} \rangle \), elements of the density matrix \( \chi^P(\chi C) \), and ERIs in the \( \chi \)-basis (\( \chi \)ERIs). In a spin-restricted framework, the elements of \( \chi^P(\chi C) \) are given by

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

where \( n \) is the number of electrons and real expansion coefficients, and the \( \chi \)ERIs are calculated according to

\[
\langle \chi^I_{\mu} \chi^J_{\nu} | \chi^K_{\lambda} \chi^L_{\sigma} \rangle = \int \int \chi^*_{\mu}(r_1) \chi^J_{\nu}(r_1) \frac{1}{|r_1 - r_2|} \chi^*_{\lambda}(r_2) \chi^L_{\sigma}(r_2) d^3r_1 d^3r_2.
\]

For the following discussion, it is convenient to divide the Fock matrix into a one-electron matrix \( \chi^H \) and a two-electron matrix \( \chi^G(\chi C) \) which, in HF theory, consists of a Coulomb matrix \( \chi^J(\chi C) \) and an exchange matrix \( \chi^K(\chi C) \),

\[
\chi^F(\chi C) = \chi^H + \chi^G(\chi C) = \chi^H + \chi^J(\chi C) + \chi^K(\chi C).
\]

After reaching self-consistency, the total electronic HF energy of the system is calculated from \( \chi^P(\chi C) \), \( \chi^F(\chi C) \), and the nucleus-nucleus repulsion energy \( V \),

\[
E_{el}^{HF} = \frac{1}{2} \sum_{\mu\nu} \chi^P_{\mu\nu}(\chi C) (2 \chi^H_{\mu\nu} + \chi^G_{\mu\nu}(\chi C)) + V.
\]

We need to introduce a second basis, the symmetrically orthogonalized basis \( \phi \), to discuss the NDDO approximation. The symmetrically orthogonalized basis functions \( \phi = \{ \phi_{\mu} \} \) and the locally orthogonal basis functions \( \chi = \{ \chi^I_{\mu} \} \) are related through

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

Consequently, we can calculate the ERIs in the \( \phi \)-basis (\( \phi \)ERIs) by a transformation involving the \( \chi \)ERIs,

\[
\langle \phi_{\mu} \phi_{\nu} | \phi_{\lambda} \phi_{\sigma} \rangle = \sum_{\mu'\nu'\lambda'\sigma'} (\chi_S^{-\frac{1}{2}})_{\mu'\nu'} (\chi_S^{-\frac{1}{2}})_{\nu'\lambda'} \langle \chi^I_{\mu'} \chi^J_{\nu'} | \chi^K_{\lambda'} \chi^L_{\sigma'} \rangle (\chi_S^{-\frac{1}{2}})_{\lambda'\sigma'}.
\]

This is formally a four-index transformation which scales as \( O(M^5) \).
2.2 Definition of the Approximation

The NDDO approximation provides a recipe for how to estimate \( \phi \) ERIs based on a small number of \( \chi \) ERIs,

\[
\langle \phi_\mu \phi_\nu | \phi_\lambda \phi_\sigma \rangle \approx \delta_{IJ} \delta_{KL} \langle \chi_I^{J} | \chi^K_{\lambda} \chi^L_{\sigma} \rangle .
\]

(11)

As a consequence, the formal scaling of the \( \phi \) ERI evaluation step is reduced from \( O(M^5) \) to \( O(M^2) \). It is not immediately obvious why Eq. (11) should hold true, especially in view of Eq. (9), but numerical data supports it. We examine the errors that Eq. (11) introduces in \( \phi \) ERIs in detail.

Throughout this work, we denote the error which arises from the application of Eq. (11) instead of Eq. (10) by \( \mathcal{E} \). The superscript to \( \mathcal{E} \) indicates which quantity is affected; additional specifications are then given as subscripts. For example, the error introduced by the NDDO approximation for the \( \phi \) ERI \( \langle \phi_\mu \phi_\nu | \phi_\lambda \phi_\sigma \rangle \) is denoted as \( \mathcal{E}^{\phi \text{ERI}}_{\mu \nu \lambda \sigma} \).

\[
\mathcal{E}^{\phi \text{ERI}}_{\mu \nu \lambda \sigma} = \langle \phi_\mu \phi_\nu | \phi_\lambda \phi_\sigma \rangle - \delta_{IJ} \delta_{KL} \langle \chi_I^{J} | \chi^K_{\lambda} \chi^L_{\sigma} \rangle .
\]

(12)

Obviously, \( M^4 \) different errors \( \mathcal{E}^{\phi \text{ERI}}_{\mu \nu \lambda \sigma} \) need to be accounted for.

First, we consider the effect of erroneous \( \phi \) ERIs on the HF energy when the self-consistent solution \( \phi \mathbf{P} \) obtained from an exact HF calculation is applied. The errors in the \( \phi \) ERIs affect the Coulomb matrix elements,

\[
\phi J_{\mu \nu} (\phi \mathbf{C}) \approx \chi J^{\text{NDDO}}_{\mu \nu} (\phi \mathbf{C}) = \sum_{\lambda \sigma} \phi P_{\lambda \sigma} (\phi \mathbf{C}) \delta_{IJ} \delta_{KL} \langle \chi_I^{J} | \chi^K_{\lambda} \chi^L_{\sigma} \rangle ,
\]

(13)

and the exchange matrix elements,

\[
\phi K_{\mu \nu} (\phi \mathbf{C}) \approx \chi K^{\text{NDDO}}_{\mu \nu} (\phi \mathbf{C}) = \sum_{\lambda \sigma} \phi P_{\lambda \sigma} (\phi \mathbf{C}) \delta_{IL} \delta_{JK} \langle \chi_I^{J} | \chi^K_{\lambda} \chi^L_{\sigma} \rangle .
\]

(14)

Interestingly, the matrix element \( \phi J_{\mu \nu} \) will always be exactly zero if \( \chi_I^{J} \) and \( \chi_J^{L} \) are centered on different atoms \( (I \neq J) \) irrespective of the number of atoms on which \( \chi^K_{\lambda} \) and \( \chi^L_{\sigma} \) are centered. By contrast, \( \phi K_{\mu \nu} \) will not be strictly zero in this case (see also Figure 1). By affecting the Coulomb and exchange
Figure 1: Graphical representation of $\phi_J(\phi_C)$, $\phi_K(\phi_C)$, and $\phi_G(\phi_C)$ (left to right in upper panel) and of $\chi_J^{\text{NDDO}}(\phi_C)$, $\chi_K^{\text{NDDO}}(\phi_C)$, and $\chi_G^{\text{NDDO}}(\phi_C)$ (left to right in lower panel) of the caffeine molecule (ECP-3G basis set). The matrices were evaluated with a density matrix determined from a fully converged HF calculation yielding $\phi_C$. They are colored according to their values ranging from $-0.05$ a.u. (red) to zero (white) to $0.05$ a.u. (blue).

matrices, the NDDO approximation will introduce an error compared to $E_{el}^{\text{HF}}$ which we denote by $\mathcal{E}^{\text{HF}} = E_{el}^{\text{HF}}$, $\mathcal{E}^{\phi_P}$, and $\mathcal{G}$. Hence, another error arises from the NDDO approximation by introducing errors in other quantities during the self-consistent-field (SCF) cycles, i.e., in the coefficient matrix and in the matrix of orbital energies. We denote this error by $\mathcal{G}$ and again indicate by a superscript which quantity is affected by the error (and give additional specifications as subscripts). By contrast, $\mathcal{E}$ denotes the error which is obtained when applying $\chi_P^{\text{NDDO}}$.
difference of the two self-consistent solutions produces $\mathcal{G}^{\text{HF}} = \mathcal{G}^{E_{\text{el}}}$,

\[
\mathcal{G}^{\text{HF}} = E_{\text{el}}^{\text{HF}}(\phi P) - E_{\text{el}}^{\text{HF-NDDO}}(\chi P^{\text{NDDO}}),
\]

\[
= \sum_{\mu} (\phi P_{\nu} - \chi P_{\nu}^{\text{NDDO}}) \left( \phi_{\mu} | h | \phi_{\nu} \right) + \frac{1}{2} \sum_{\mu
u} \left( \phi P_{\nu} \phi P_{\lambda} \left[ \langle \phi_{\mu} \phi_{\nu} | \phi_{\lambda} \phi_{\lambda} \rangle - \frac{1}{2} \langle \phi_{\mu} \phi_{\nu} | \phi_{\lambda} \phi_{\lambda} \rangle \right] 
\]

\[
- \chi P_{\nu}^{\text{NDDO}} \chi P_{\lambda}^{\text{NDDO}} \left[ \delta_{IJ} \delta_{KL} \langle \chi_{\mu}^{I} \chi_{\lambda}^{J} | \chi_{\lambda}^{K} \chi_{\mu}^{L} \rangle - \frac{1}{2} \delta_{IL} \delta_{JK} \langle \chi_{\mu}^{I} \chi_{\lambda}^{L} | \chi_{\lambda}^{J} \chi_{\mu}^{K} \rangle \right].
\]

(16)

The electronic HF energy does, by definition, not contain effects from electron correlation.\textsuperscript{55} Various electronic structure methods are available for calculating correlation energies.\textsuperscript{55} The prevalent single- and multi-reference method require the calculation of ERIs in the molecular orbital basis $\psi$ ($\psi$ERIs). These $\psi$ERIs are obtained through a 4-index transformation of the $\phi$ERIs (or the $\chi$ERIs),

\[
\langle \psi_{i} \psi_{j} | \psi_{k} \psi_{l} \rangle = \sum_{\mu \nu \lambda \sigma} \phi C_{\mu i} \phi C_{\nu j} \langle \phi_{\mu} \phi_{\nu} \phi_{\lambda} \phi_{\lambda} \rangle \phi C_{\lambda k} \phi C_{\sigma l}.
\]

(17)

This 4-index transformation is similar to the 4-index transformation with which the $\phi$ERIs are determined from the $\chi$ERIs (Eq. (10)). When applying Eq. (11), we may approximate the $\psi$ERIs as,

\[
\langle \psi_{i} \psi_{j} | \psi_{k} \psi_{l} \rangle \approx \sum_{\mu \nu \lambda \sigma} \phi C_{\mu i} \phi C_{\nu j} \delta_{IJ} \delta_{KL} \langle \chi_{\mu}^{I} \chi_{\lambda}^{J} | \chi_{\lambda}^{K} \chi_{\mu}^{L} \rangle \phi C_{\lambda k} \phi C_{\sigma l}.
\]

(18)

The formal scaling of the $\psi$ERI evaluation step is therefore reduced from $\mathcal{O}(M^5)$ to $\mathcal{O}(M^2)$ scaling which comes at the price of an error in the $M^4$ $\psi$ERIs, $\mathcal{E}^{\psi \text{ERI}}_{ijkl}$,

\[
\mathcal{E}^{\psi \text{ERI}}_{ijkl} = \sum_{\mu \nu \lambda \sigma} \phi C_{\mu i} \phi C_{\nu j} \mathcal{E}^{\phi \text{ERI}}_{\mu \nu \lambda \sigma} \phi C_{\lambda k} \phi C_{\sigma l}.
\]

(19)

If we determine the coefficient matrix in a self-consistent field procedure, we will introduce an additional error $\mathcal{G}^{\psi \text{ERI}}_{ijkl}$ from applying a different coefficient
matrix,

\[ G_{ijkl}^{\psi\text{ERI}} = \sum_{\mu,\nu,\lambda,\sigma} (\phi C_{\mu i}^{\psi} C_{\nu j}^{\psi} \langle \phi_{\mu} \phi_{\nu} | \phi_{\lambda} \phi_{\sigma} \rangle \phi C_{\lambda k}^{\psi} C_{\sigma l}^{\psi}) \]

\[ -\chi C_{\mu i}^{\text{NDDO}} C_{\nu j}^{\text{NDDO}} \delta_{IJ} \delta_{KL} \langle \chi^I_{\mu} \chi^J_{\nu} | \chi^K_{\lambda} \chi^L_{\sigma} \rangle \chi C_{\lambda k}^{\text{NDDO}} C_{\sigma l}^{\text{NDDO}}. \]

In this work, we demonstrate how \( E_{ijkl}^{\psi\text{ERI}} \) and \( G_{ijkl}^{\psi\text{ERI}} \) affect the MP2 correlation energies by quantifying \( E_{ijkl}^{\psi\text{MP2}} \) and \( G_{ijkl}^{\psi\text{MP2}} \), respectively.

### 2.3 Extension to Conventional Basis Sets

The \( \chi \)-basis must fulfill the condition that it is locally orthogonal for the NDDO approximation to be justifiable. This condition is, however, generally not fulfilled by ordinary basis sets. We denote ordinary Gaussian-type basis functions which are, in general, not locally orthogonal by \( \tau^I_{\mu} \) (\( \mu \)-th basis function of type \( \tau \) centered on atom \( I \)). The NDDO approximation is not straightforwardly applicable for an arbitrary \( \tau \)-basis,

\[ \langle \phi_{\mu} \phi_{\nu} | \phi_{\lambda} \phi_{\sigma} \rangle \not\approx \delta_{IJ} \delta_{KL} \langle \tau^I_{\mu} \tau^J_{\nu} | \tau^K_{\lambda} \tau^L_{\sigma} \rangle, \]

because we observe \( E_{ijkl}^{\psi\text{ERI}} > 1.0 \) atomic units (a.u.) for a large number of \( \phi \)-ERIs at the example of the water molecule (see Figure 2). To cure this problem, we propose to transform \( \{ \tau^I_{\mu} \} \) to a locally orthogonal basis \( \{ \chi^I_{\mu} \} \) by the application of the transformation matrix \( T \),

\[ T_{\mu \nu} = \delta_{IJ} (S^{-1})_{\mu \nu}, \]

so that

\[ \chi^I_{\nu} = \sum_{\mu=1}^{M} T_{\mu \nu} \tau^I_{\mu}. \]

Figure 2 illustrates that \( E_{ijkl}^{\psi\text{ERI}} \ll 1.0 \) a.u. after local orthogonalization. We demonstrate in Section 4 that this statement holds true in general.
Figure 2: \( \langle \phi_\mu \phi_\nu | \phi_\lambda \phi_\sigma \rangle \), \( \langle \tau_\mu^{I} \tau_\nu^{J} | \tau_\lambda^{K} \tau_\sigma^{L} \rangle \) and \( \langle \chi_\mu^{I} \chi_\nu^{J} | \chi_\lambda^{K} \chi_\sigma^{L} \rangle \) for the water molecule in an ECP-3G\(^{5,63}\) basis set (left) and in a def2-QZVP\(^{64}\) basis set (middle and right). The gray dashed lines indicate \( \langle \phi_\mu \phi_\nu | \phi_\lambda \phi_\sigma \rangle = 0.0 \), \( \langle \phi_\mu \phi_\nu | \phi_\lambda \phi_\sigma \rangle = \langle \chi_\mu^{I} \chi_\nu^{J} | \chi_\lambda^{K} \chi_\sigma^{L} \rangle \) (left and right), and \( \langle \phi_\mu \phi_\nu | \phi_\lambda \phi_\sigma \rangle = \langle \tau_\mu^{I} \tau_\nu^{J} | \tau_\lambda^{K} \tau_\sigma^{L} \rangle \) (middle). Red circles encode ERIs for which \( I = J \) and \( K = L \) in the \( \chi \)-basis and in the \( \tau \)-basis, and blue crosses encode ERIs for which \( I \neq J \) or \( K \neq L \) in the \( \chi \)-basis and in the \( \tau \)-basis.

### 3 Analysis of the NDDO Approximation for Molecules in a \( \chi \)-Basis

For the first part of our analysis of the NDDO approximations, we applied the ECP-3G basis set\(^{5,63}\). We selected the ECP-3G basis set because its basis functions form a \( \chi \)-basis and because it is applied (in slightly modified forms) in the OM1, OM2, and OM3 models.\(^{5,7}\) The ECP-3G basis set specifies one \( s \)-type basis function for hydrogen and one \( s \)-type and three \( p \)-type basis functions for carbon, nitrogen, oxygen, and fluorine. We consider four different sets of molecular structures for this part of the analysis: (A) We first analyze the NDDO approximation on the simplest possible model system which is a dihydrogen molecule \( \text{H}_1 - \text{H}_2 \) with an interatomic distance \( R_{12} \). This is the simplest possible model system because NDDO is exact for the isolated atoms (where \( \phi = \chi \)) and for systems with only one electron (e.g., \( \text{H}_2^+ \)). (B) We assembled a series of linear alkane chains \( \text{C}_x \text{H}_{2x+2}, \ x = 1, 2, ..., 15 \) to study trends with an increasing molecular size. (C) We randomly selected a
subset of 5000 molecules of the QM9 data set which allows us to examine the NDDO approximation for a variety of equilibrium structures of molecules composed of H, C, N, O, and F. (D) We chose three reaction trajectories, a Diels–Alder reaction between butadiene and ethylene yielding cyclohexene (reaction A), the decomposition of azobisisobutyronitrile (reaction E), and the elimination of CO$_2$ from the benzoyl radical (reaction F) which we published in previous work.

3.1 Effect on Electron-Electron Repulsion Integrals

For the simplest possible model system, H$_2$, $2^4 = 16$ ERIs arise for a given $R_{12}$. Due to symmetry relations, only four of these 16 values are different so that it suffices to discuss $\langle \phi_1 \phi_1 | \phi_1 \phi_1 \rangle$, $\langle \phi_1 \phi_2 | \phi_1 \phi_2 \rangle$, $\langle \phi_1 \phi_1 | \phi_1 \phi_2 \rangle$, and $\langle \phi_1 \phi_2 | \phi_1 \phi_2 \rangle$ (see Figure 3). The NDDO approximation (Eq. (11)) assumes that $\langle \chi^H_1 \chi^H_1 | \chi^H_1 \chi^H_1 \rangle$ is as large as $\langle \phi_1 \phi_1 | \phi_1 \phi_1 \rangle$ and that $\langle \chi^H_1 \chi^H_1 | \chi^H_2 \chi^H_2 \rangle$ is as large as $\langle \phi_1 \phi_2 | \phi_1 \phi_2 \rangle$. Figure 3 shows that this is a poor approximation for $R_{12} < 2.0$ Å. Only if the overlap between $\chi^H_1$ and $\chi^H_2$ is small at large $R_{12}$, i.e., where the $\chi$-basis becomes a $\phi$-basis, $|\phi_{1111}|$ and $|\phi_{1122}|$ will be small.

Figure 3: Dependence of ERIs on $R_{12}$ in an H$_2$ molecule described by an ECP-3G basis. Left: $\langle \chi^H_1 \chi^H_1 | \chi^H_1 \chi^H_1 \rangle$ (red line), $\langle \chi^H_1 \chi^H_1 | \chi^H_2 \chi^H_2 \rangle$ (blue line), $\langle \chi^H_1 \chi^H_2 | \chi^H_1 \chi^H_2 \rangle$ (green line), and $\langle \chi^H_1 \chi^H_1 | \chi^H_1 \chi^H_2 \rangle$ (orange line). Middle: $\langle \phi_1 \phi_1 | \phi_1 \phi_1 \rangle$ (red line), $\langle \phi_1 \phi_1 | \phi_2 \phi_2 \rangle$ (blue line), $\langle \phi_1 \phi_2 | \phi_1 \phi_2 \rangle$ (green line), and $\langle \phi_1 \phi_1 | \phi_1 \phi_2 \rangle$ (orange line). Right: $\phi_{1111}$ (red line), $\phi_{1122}$ (blue line), $\phi_{1212}$ (green line), $\phi_{1112}$ (orange line), and $\phi_{1211}$ (Eq. (24); black line).
The \( \phi \)ERIs \( \langle \phi_1 \phi_2 | \phi_1 \phi_2 \rangle \) and \( \langle \phi_1 | \phi_1 | \phi_1 \phi_2 \rangle \) are assumed to be zero in the NDDO approximation. For \( R_{12} < 1.6 \text{ Å} \), \( |\mathcal{E}^{\phi \text{ERI}}_{1212}| \) and \( |\mathcal{E}^{\phi \text{ERI}}_{1112}| \) are smaller than \( |\mathcal{E}^{\phi \text{ERI}}_{1111}| \) and \( |\mathcal{E}^{\phi \text{ERI}}_{1212}| \). For \( R_{12} \) larger than 1.6 Å, \( \mathcal{E}^{\phi \text{ERI}}_{1112} \) is the largest individual error in an \( \phi \)ERI. This gives rise to a large cumulative error (CE) \( \mathcal{E}^{\phi \text{ERI}}_{\text{CE}} \),

\[
\mathcal{E}^{\phi \text{ERI}}_{\text{CE}} = M \sum_{\mu \nu \lambda \sigma} \mathcal{E}^{\phi \text{ERI}}_{\mu \nu \lambda \sigma},
\]

for \( 1.6 \text{ Å} < R_{12} < 3.0 \text{ Å} \). For all \( R_{12} \), we find a significant error in at least one of the \( \phi \)ERIs.

We also encounter nonnegligible \( \mathcal{E}^{\phi \text{ERI}}_{\mu \nu \lambda \sigma} \) for all structures in the sets of molecules B, C, and D. The largest absolute error of an \( \phi \)ERI in a given molecule is between 0.10 a.u. and 0.23 a.u. It is already obvious at this point that only an efficient error cancellation may yield useful observables based on these erroneous \( \phi \)ERIs (or on density matrices obtained with them in SCF procedures).

Figure 4 shows that the cumulative absolute error (CAE) in the \( \phi \)ERIs,

\[
\mathcal{E}^{\phi \text{ERI}}_{\text{CAE}} = \sum_{\mu \nu \lambda \sigma} |\mathcal{E}^{\phi \text{ERI}}_{\mu \nu \lambda \sigma}|,
\]

grows roughly linearly with the number of basis functions for reaction B (Figure 4). We observe a linear growth not only in the overall cumulative absolute error, but also in individual contributions to it when we break down the corresponding \( \chi \)ERIs in one-, two-, three-, and four-center ERIs. In agreement with previous studies,\[^{40,44-48}\] we find that assuming \( \langle \phi_\mu | \phi_\nu | \phi_\lambda | \phi_\sigma \rangle \) to be as large as \( \langle \chi_\mu | \chi_\nu | \chi_\lambda | \chi_\sigma \rangle \) for \( I \neq J \neq K \) and \( \langle \phi_\mu | \phi_\nu | \phi_\lambda | \phi_\sigma \rangle \) to be as large as \( \langle \chi_\mu | \chi_\nu | \chi_\lambda | \chi_\sigma \rangle \) for \( I \neq J \) introduces significant errors (purple and orange circles in Figure 4, respectively). These contributions account for 60–65% of the overall cumulative absolute error for B. Figure 4 shows that \( \mathcal{E}^{\phi \text{ERI}}_{\text{CAE}} \) also grows approximately linearly with the number of basis functions for C and that the spread of the individual \( \mathcal{E}^{\phi \text{ERI}}_{\text{CAE}} \) is large. The change of \( \mathcal{E}^{\phi \text{ERI}}_{\text{CAE}} \) with reaction progress for D in Figure 4 shows that the cumulative absolute error in the \( \phi \)ERIs crucially depends on the arrangement of the atomic nuclei, i.e. on the underlying nuclear framework that generates the external potential.
Figure 4: Dependence of the cumulative absolute error $\mathcal{E}_{\text{CAE}}^{\phi\text{ERI}}$ (Eq. (25)) on the number of basis functions for the molecule sets B (left) and C (middle). The cumulative absolute error $\mathcal{E}_{\text{CAE}}^{\phi\text{ERI}}$ (black squares) is broken down into the corresponding $\chi\text{ERIs}$ for the molecule set B: $\langle \chi_I^I | \chi_I^J | \chi_I^\lambda | \chi_I^\sigma \rangle$ (red circles), $\langle \chi_I^I | \chi_I^J | \chi_K^L | \chi_J^\lambda | \chi_K^\sigma \rangle$, $I \neq J$ (blue circles), $\langle \chi_I^I | \chi_J^J | \chi_I^\lambda | \chi_J^\sigma \rangle$, $I \neq J$ (green circles), $\langle \chi_I^I | \chi_J^J | \chi_K^K | \chi_J^\lambda | \chi_K^\sigma \rangle$, $I \neq J \neq K$ (purple circles), $\langle \chi_I^I | \chi_J^J | \chi_K^L | \chi_J^\lambda | \chi_K^\sigma \rangle$, $I \neq J \neq K$ (cyan circles), and $\langle \chi_I^I | \chi_J^J | \chi_K^L | \chi_J^\lambda | \chi_K^\sigma \rangle$, $I \neq J \neq K \neq L$ (gray circles). Dependence of $\mathcal{E}_{\text{CAE}}^{\phi\text{ERI}}$ on the reaction progress (right) for reaction A (34 basis functions; red circles), reaction E (60 basis functions; blue circles), and reaction F (41 basis functions; green circles). All calculations were carried out with the ECP-3G basis.

3.2 Error Propagation: the Hartree–Fock Energy

For $\text{H}_2$, $\mathcal{E}^{\text{HF}}$ depends linearly on the cumulative error of the $\phi\text{ERI}s$ because $\phi\text{P}$ can be determined analytically:

$$
\mathcal{E}^{\text{HF}} = \frac{1}{4} \mathcal{E}_{\text{CE}}^{\phi\text{ERI}} = \frac{1}{2} \mathcal{E}_{1111}^{\phi\text{ERI}} + \frac{1}{2} \mathcal{E}_{1122}^{\phi\text{ERI}} + \mathcal{E}_{1212}^{\phi\text{ERI}} + \frac{2}{3} \mathcal{E}_{1112}^{\phi\text{ERI}}.
$$

Previous results show that the NDDO approximation introduces an error of 0.002 a.u. for $\text{H}_2$ with $R_{12} = 0.84 \text{ Å}$ which we can reproduce. In our detailed analysis however, we also see that this value of $R_{12}$ falls into the region where $\mathcal{E}_{\text{CE}}^{\phi\text{ERI}}$ is small due to a fortunate error cancellation of $\mathcal{E}_{1111}^{\phi\text{ERI}} + \frac{1}{2} \mathcal{E}_{1122}^{\phi\text{ERI}}$, $\mathcal{E}_{1212}^{\phi\text{ERI}}$, and $\frac{2}{3} \mathcal{E}_{1112}^{\phi\text{ERI}}$ (see Figure 3). For larger or smaller $R_{12}$, we encounter larger $\mathcal{E}^{\text{HF}}$ because the errors in the $\phi\text{ERI}s$ do not cancel as efficiently.

For the calculation of $E^{\text{HF}}$, the $\phi\text{ERI}s$ are contracted with elements of the
density matrix. The elements of the density matrix are therefore a central ingredient for an efficient error cancellation. We see how differently the errors in the $\phi$ERIs add up by comparing $\mathcal{E}^{HF}$ and $\mathcal{G}^{HF}$ for the molecule set B in Figure 5. Interestingly, $|\mathcal{E}^{HF}|$ and $\mathcal{G}^{HF}$ increase in an almost perfectly linear fashion with the number of basis functions. For a given number of basis functions, $\mathcal{G}^{HF}$ is significantly larger than $|\mathcal{E}^{HF}|$. Not surprisingly, we observe large $\mathcal{E}^{HF}$ and $\mathcal{G}^{HF}$ for the vast majority of molecules in B and C. In line with previous studies, $^{33,40,44–49}$ we see that the NDDO approximation is a rather crude approximation. We discuss strategies to overcome this situation in Section 5.

Figure 5: Dependence of $\mathcal{E}^{HF}$ (blue squares) and $\mathcal{G}^{HF}$ (red circles) on the number of basis functions for molecules in B (left). Dependence of $\mathcal{G}^{HF}$ on the number of basis functions for molecules in C (middle). The squares are colored according to the cumulative absolute error $E_{\text{CAE}}^{\phi\text{ERI}}$ in the $\phi$ERIs. Dependence of $\mathcal{E}^{HF}$ on the reaction progress (right) for reaction A (red squares), reaction E (blue squares), and reaction F (green squares). All calculations were carried out with an ECP-3G basis set.

3.3 Error Propagation: the MP2 Energy

The error in the $\phi$ERIs also propagates to the $\psi$ERIs (Eq. (18)). We show the effect of the NDDO approximation on selected $\psi$ERIs, $\langle \phi_1 \psi_1 | \phi_1 \psi_1 \rangle$, $\langle \psi_1 \phi_1 | \psi_1 \phi_1 \rangle$, $\langle \phi_2 \phi_2 | \phi_2 \phi_2 \rangle$, and $\langle \phi_1 \phi_2 | \phi_1 \phi_2 \rangle$, in H$_2$ in Figure 6. We choose to
study these $\psi$ERIs because they are applied in the calculation of the MP2 correlation energy ($\langle \psi_1 \psi_1 | \psi_2 \psi_2 \rangle$) and of the full-configuration interaction (FCI) correlation energy\(^{54}\) (all of these $\psi$ERIs). All of these $\psi$ERIs are significantly affected by the NDDO approximation. Hence, it comes as no surprise that the MP2 and FCI correlation energies are deteriorated by the NDDO approximation. The general shape of $E^{\text{MP2}}$ follows that of $E^{\psi\text{ERI}}_{1122}$ and has a minimum at $R_{12} = 0.86$ Å where $E^{\text{MP2}} = -0.009$ a.u. As a consequence, instead of $E^{\text{MP2}} = -0.016$ a.u. a $E^{\text{MP2}} = -0.007$ a.u. is obtained when the NDDO approximation is invoked. The MP2 correlation energy is therefore significantly underestimated. Also the FCI correlation energy is also significantly underestimated for $R_{12} < 0.72$ Å. An unfortunate addition of the errors in $\langle \psi_2 \psi_2 | \psi_2 \psi_2 \rangle$ and $\langle \psi_1 \psi_1 | \psi_1 \psi_1 \rangle$, however, leads to an overestimation of $E^{\text{FCI}}$ for $R_{12} \geq 0.72$ Å. The errors in the $\psi$ERIs and, hence, in $E^{\text{FCI}}$, only vanish in the limit of very large $R_{12}$ ($R_{12} > 25.0$ Å).

![Figure 6: Dependence of $E^{\psi\text{ERI}}_{ijkl}$ (left), $E^{\text{MP2}}$ (black line; right), and $E^{\text{FCI}}$ (red line; right) on $R_{12}$ in the H$_2$ molecule described in an ECP-3G basis. We consider $E^{\psi\text{ERI}}_{1111}$ (red line, left), $E^{\psi\text{ERI}}_{1122}$ (blue line), $E^{\psi\text{ERI}}_{2222}$ (green line), and $E^{\psi\text{ERI}}_{1212}$ (orange line).](image-url)

The NDDO approximation also deteriorates the MP2 energies for the other molecules which we investigated. Interestingly, the NDDO approximation always produced too small MP2 correlation energies. The amount by which the MP2 energy is underestimated depends roughly linearly on the number of basis functions so that the MP2 correlation energy is underestimated by $-0.15$ a.u. for a molecule with 25 basis functions and by $-0.30$ a.u. for a molecule with 60 basis functions.
a.u. for a molecule with 50 basis functions. The significance of these results becomes apparent in the context of previous studies which reported far too small correlation energies determined for NDDO-SEMO reference wave functions. Our results can be taken as an indication that the low correlation energies arise as a direct consequence of the NDDO approximation and not of the introduction of other parametrized expressions when assembling NDDO-SEMO models.

Figure 7: Left: Dependence of $\mathcal{E}^{\text{MP2}}$ (blue squares) and $\mathcal{G}^{\text{MP2}}$ (red circles) on the number of basis functions for molecules in set B. Middle: Dependence of $\mathcal{E}^{\text{MP2}}$ on the number of basis functions for molecules in set C. The squares are colored according to the cumulative absolute error $E^{\phi_{\text{ERI}} \text{CAE}}$ in the $\phi$ERIs. Right: Dependence of $\mathcal{E}^{\text{MP2}}$ on the reaction progress for reaction A (red circles), reaction E (blue circles), and reaction F (green circles). All calculations were carried out with the ECP-3G basis set.

4 Analysis of the NDDO Approximation for Molecules in the $\tau$-Basis

When applying ordinary $\tau$-basis sets, large errors in the $\phi$ERIs arise. For $\text{H}_2$ and $R_{12} = 0.74$ Å described in a def2-QZVP basis set for example, the largest absolute error in an $\phi$ERI amounts to 1.12 a.u. The application of these erroneous $\phi$ERIs even leads to a positive $E^{\text{HF}} = 3.48$ a.u. By contrast,
the largest absolute error in an $\phi$ERI is reduced to 0.22 a.u. when we locally orthogonalize the basis (Eq. (23)). After the local orthogonalization of the basis set, we obtain $G^{\text{HF}} = 0.01$ a.u. A prior local orthogonalization led to significantly smaller largest absolute errors in the $\phi$ERIs and cumulative errors in the $\phi$ERIs for all molecules (B, C, and D) by up to two orders of magnitude.

We found a fundamental limitation of the NDDO approximation in the course of our analysis of different $\tau$-basis sets (see Table 1): Electronic energies calculated with the NDDO approximation do not converge with respect to the basis set size. The HF limit for H$_2$ for $R_{12} = 0.74$ Å was determined to be $E_{\text{el}}^{\text{HF}} = -1.133629$ a.u. No convergence for $E_{\text{el}}^{\text{HF}}$ of H$_2$ with increasing basis-set size is observed when invoking the NDDO approximation (see Table 1). We also found this behavior for 50 randomly selected molecules out of molecule set C.

5 Improving on the NDDO Approximation

The NDDO approximation introduces significant errors in the $\phi$ERIs, observed here for the simplest possible neutral molecule, H$_2$, and for a diverse selection of medium-sized organic compounds. Obviously, these errors are too large and too unsystematic for the NDDO approximation to be useful in purely ab initio electronic structure models.

5.1 Modifications

In ab initio electronic structure calculations, it is customary to apply screening techniques to determine which $\chi$ERIs are negligibly small. Errors are controlled by thresholds with respect to which the $\chi$ERIs are neglected. By contrast, the NDDO approximation cannot be applied as a screening tool for $\chi$ERIs because it can only predict whether the corresponding ERIs in the $\phi$-basis are negligibly small or not. Each $\phi$ERI, however, encodes information on all $\chi$ERIs (Eq. (10)). It is, however, computationally difficult to improve on the NDDO approximation by an explicit transformation of the
Table 1: Basis set incompleteness error (BSIE) with reference to the HF limit $E_{\text{el}}^{\text{HF}} = -1.133629$ a.u.\textsuperscript{20} for $H_2$ for $R_{12} = 0.74$ Å in a.u. calculated without (3rd column) and with (4th column) the NDDO approximation in different basis sets which activate a different number of basis functions $M$. Additionally, we specify $G^{\text{HF}}$ (Eq. (16)) for each basis set.

| Basis Set   | $M$ | BSIE(HF)  | BSIE(HF-NDDO) | $G^{\text{HF}}$ |
|-------------|-----|-----------|---------------|----------------|
| STO-3G      | 2   | −0.017478 | −0.015052     | 0.002425       |
| 3-21G       | 4   | −0.010834 | +0.023863     | 0.034697       |
| 6-311G      | 6   | −0.005824 | +0.030428     | 0.036252       |
| def2-SVP    | 10  | −0.004707 | −0.008348     | −0.003641      |
| 6-311G*     | 12  | −0.001285 | +0.002197     | 0.003482       |
| def2-TZVP   | 12  | −0.001224 | +0.003359     | 0.004583       |
| 6-311+G*    | 14  | −0.001270 | +0.002789     | 0.004059       |
| def2-SVPD   | 16  | −0.004693 | −0.009206     | −0.004513      |
| def2-TZVPP  | 28  | −0.000804 | −0.008063     | −0.007259      |
| def2-TZVPPD | 34  | −0.000745 | −0.006796     | −0.006051      |
| def2-QZVP   | 60  | −0.000223 | −0.008566     | −0.008343      |

ERIs from the $\chi$- to the $\phi$-basis because this is a 4-index transformation that scales as $O(M^5)$. Some attempts\textsuperscript{49,71} were made to improve on the approximation of the $\phi$ERIs for which the corresponding $\chi$ERIs are one-center ERIs ($\langle \chi^I_\mu | \chi^I_\nu | \chi^J_\lambda | \chi^J_\sigma \rangle$) and two-center ERIs ($\langle \chi^I_\mu | \chi^I_\nu | \chi^J_\lambda | \chi^J_\sigma \rangle$ for $I \neq J$). However, these suggestions did not find widespread use. Moreover, our results suggest that a correction of these $\phi$ERIs does not suffice to obtain a reliable HF energy $E_{\text{el}}^{\text{HF}}$. A large portion of the error originates from the $\phi$ERIs which are assumed to be zero, but are not exactly zero. If we had to estimate these $\phi$ERIs, we would again need close to $O(M^4)$ operations. These estimates would have to be accurate due to the plethora of small $\phi$ERIs which would again compromise the computational efficiency of the NDDO approximation.
5.2 Capitalizing on Error Cancellation

In SEMO models, further approximations (beyond NDDO) allow for error compensation to a certain degree. All successful NDDO-SEMO models introduce various parametric expressions to assemble the one-electron matrix and to calculate the nucleus-nucleus repulsion energy. For contemporary NDDO-SEMO models, the parametric expressions were designed such that the result of the SCF optimization yields a result close to a reference energy. Overall, the results obtained with NDDO-SEMO models achieve a remarkably high accuracy with respect to these reference data. At the same time, NDDO-SEMO models are notoriously unreliable for molecules not considered in the parametrization procedure. It is virtually impossible to rationalize why errors occur due to the number and the diversity of the approximations invoked in an NDDO-SEMO model. At least some of these errors are likely to be due to the NDDO approximation. In general, we may anticipate that the improvement of the parametric expressions in NDDO-SEMO models is as complicated as the direct correction for the error introduced by the NDDO approximation as discussed in Section 5.1. Therefore, the only viable use of NDDO-SEMO models appears to be their combination with system-focused rigorous error estimation schemes as proposed in Refs. 72, 73.

5.3 Correcting the Two-Electron Matrices

In 1969, Roby and Sinanoğlu made an attempt to accelerate single-point HF calculations for a diverse set of small molecules. They suggested to scale $\chi^G_{\text{NDDO}}$ with a matrix $\Gamma$ to obtain a better estimate for $\phi G$.

\[
\phi G \approx \Gamma^\phi G_{\text{NDDO}}. \tag{27}
\]

Their attempt to define general rules to assemble $\Gamma$ turned out to be impossible.

In this Section, we reconsider and build upon the Roby–Sinanoğlu approach. We can exactly determine $\Gamma(\{\tilde{R}_i\})$ for a given structure $\{\tilde{R}_i\}$ from
a reference HF, KS-DFT, or multi-configurational SCF calculation (yielding the exact \( \phi G^n(\widetilde{R}_I^n) (\phi C) \)),

\[
\Gamma(\widetilde{R}_I^n) (\phi C) = \phi G^n(\widetilde{R}_I^n) (\phi C) \cdot \left( x G^{\text{NDDO}}(\widetilde{R}_I^n) (\phi C) \right)^{-1}.
\] 

Figure 8 now shows that \( \Gamma(\widetilde{R}_I^n) (\phi C) \) is transferable to a certain degree in a sequence of related structures. I.e., for two similar structures \( \{ \widetilde{R}_I^n \} \) and \( \{ \widetilde{R}_I^{n+1} \} \) we have

\[
\phi G^n(\widetilde{R}_I^{n+1}) (\phi C) \approx \Gamma(\widetilde{R}_I^n) (\phi C) \cdot x G^{\text{NDDO}}(\widetilde{R}_I^{n+1}) (\chi C^{\text{NDDO}}).
\] 

Eq. (29) defines a system-focused NDDO model that can be applied in connection with any Fock operator. It allows us to maintain complete error control on the resulting model because we can determine \( \Gamma(\widetilde{R}_I^{n+1}) \) for a given molecule with nuclear coordinates \( \{ \widetilde{R}_I^{n+1} \} \) in case of doubt. For reaction A (Figure 8), we see that we can update \( \Gamma(\widetilde{R}_I^n) (\phi C) \) every third step to maintain a negligible error of \( G^{\text{HF}} < 1.0 \text{ kJ mol}^{-1} \). If we tolerate a larger deviation, e.g., chemical accuracy \( G^{\text{HF}} < 4.2 \text{ kJ mol}^{-1} \), we can update \( \Gamma(\widetilde{R}_I^n) (\phi C) \) every tenth step. Obviously, this rule is specific for the example of reaction A. However, it may be easily generalized for any reaction by exploiting structural similarity measures[74,75] as demonstrated in Ref. 73.
The original Roby–Sinanoğlu approach is not the only one conceivable for the construction of correction matrices. In fact, a multiplicative correction matrix changes matrix elements through a combination of elements of the original matrix. An additive correction appears easier and more straightforward to achieve the goal of re-adjusting individual matrix elements. We may therefore define separate additive correction with the matrices \( \Gamma_J \) and \( \Gamma_K \) to \( \chi_J^{\text{NDDO}} \) and to \( \chi_K^{\text{NDDO}} \),

\[
\phi_G(\tilde{R}_i^{(n+1)}) (\phi C) \approx \Gamma_J (\tilde{R}_i^n)(\phi C) + \chi_J^{\text{NDDO}} (\tilde{R}_i^{(n+1)})(\chi C^{\text{NDDO}}),
\]

\[
+ \Gamma_K (\tilde{R}_i^n)(\phi C) + \chi_K^{\text{NDDO}} (\tilde{R}_i^{(n+1)})(\chi C^{\text{NDDO}}),
\]

or a unification of \( \Gamma_J \) and \( \Gamma_K \) as a total additive correction. For reaction A, this additive approach leads to smaller errors in general (see Figure 8) for the first part of the reaction.

Both correction approaches suffer from limitations. The nuclear coordinates and also the density matrices obviously differ in a sequence of structures. Eq. (29) will only yield sensible results when the change of both quantities remains small. We are currently exploring the possibility to apply this strategy in practice for sequences of structures as they occur during structure optimization, in Born–Oppenheimer molecular-dynamics trajectories, or in interactive reactivity studies.

6 Conclusions

The NDDO approximation is a central ingredient for many modern semiempirical molecular orbital models. We studied the effect of the NDDO approximation on the ERIs in the symmetrically orthogonalized basis for the simplest possible model system, \( \text{H}_2 \), and for a diverse set of molecules. As expected and in agreement with previous results, we find that the NDDO approximation leads to significant errors for molecules in their equilibrium structure. The errors do only vanish in the atomization limit where the overlap between different basis functions vanishes. The errors in the ERIs in the
symmetrically orthogonalized basis increase roughly linearly with the number of basis functions. Additionally, we found that the errors in the ERIs in the symmetrically orthogonalized basis depend strongly on the arrangement of the atomic nuclei. These nonnegligible errors introduced by NDDO may only be alleviated by an efficient error cancellation which is in operation in SEMO models, but not in HF theory. For HF calculations, error cancellation is unlikely to occur because of the fact that the ERIs in the symmetrically orthogonalized basis are contracted with elements of the density matrix.

We were then able to dissect how the NDDO approximation affects ERIs in the molecular orbital basis and, hence, correlation energies. We found that MP2 correlation energies are underestimated and the underestimation increases with the number of basis functions. This finding explains previous reports that correlation energies obtained with respect to an NDDO-SEMO reference wave function are far too small.

We proposed a local orthogonalization that allowed us to transgress the domain of minimal basis sets and to apply ordinary basis sets in conjunction with the NDDO approximation. While we observed a drastic reduction in the largest absolute errors in the ERIs in the symmetrically orthogonalized basis, we discovered another limitation of the NDDO approximation. Electronic energies calculated with the NDDO approximation do not converge with respect to the basis set size so that this solution to the small basis-set restriction of NDDO-SEMO models does not pay off.

We then continued to propose how one could still capitalize on the efficiency enabled by the NDDO approximation without significant loss of accuracy in a system-focused manner for similar structures. Specifically, we proposed a strategy to correct for the errors introduced by the NDDO approximation in the two-electron matrices which was inspired by work of Roby and Sinanoğlu. The two-electron matrix obtained within the NDDO approximation is modified with a correction matrix obtained from a reference HF, KS-DFT, or multi-configurational SCF calculation. These correction matrices are transferable to a certain degree within sequences of related structures.
Appendix: Computational Methodology

All calculations in this work were carried out with our modified version of PySCF (version 1.4). The ERIs in the $\tau$-basis were transformed to the corresponding ERIs in the $\chi$-basis or in the $\phi$-basis with the AO2MO integral transformation module of PySCF. We applied the ECP-3G, STO-3G, 3-21G, 6-31G, and def2 basis sets in calculations.

Raymond and co-workers assembled a database considered to be representative of chemical space. We randomly chose 5000 entries of the QM9 data set (set C) to study the error of the NDDO approximation across a large set of molecules. Additionally, we worked with linearly growing alkane chains (set B) with the stoichiometry $C_xH_{2x+2}$, ($x = 1, 2, ..., 15$). We include the optimized structures as Supplementary material. Finally, we selected three reactions (set D) which we had considered for interactive reactivity explorations in the framework of real-time quantum chemistry; they can be found in the supplementary material in Ref. 29.

Acknowledgements

This work was supported by the Schweizerischer Nationalfonds.

References

[1] 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.

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

[3] 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.
[4] 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.

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

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

[7] 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.

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

[9] 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.

[10] Senn, H. M.; Thiel, W. QM/MM Methods for Biomolecular Systems, *Angew. Chem. Int. Ed.* 2009, 48, 1198–1229.

[11] 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.

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

[13] 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.
[14] 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.

[15] 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.

[16] 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.

[17] 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.

[18] 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.

[19] Brahmkshatriya, S. P.; Dobes, P.; Fanfrlík, J.; Rezac, J.; Paruch, K.; Bronowska, A.; Lepšík, 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.

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

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

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

[23] 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.

[24] Bosson, M.; Richard, C.; Plet, A.; Grudinin, S.; Redon, S. Interactive quantum chemistry: A divide-and-conquer ASED-MO method, *J. Comput. Chem.* 2012, 33, 779–790.

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

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

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

[28] Vaucher, A. C.; Haag, M. P.; Reiher, M. Real-Time Feedback from Iterative Electronic Structure Calculations, *J. Comput. Chem.* 2016, 37, 805–812.

[29] 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.

[30] 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.
[31] Cook, D. B.; Hollis, P. C.; McWeeny, R. Approximate Ab Initio Calculations on Polyatomic Molecules, *Mol. Phys.* 1967, 13, 553–571.

[32] 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.

[33] 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 C$_3$H$_7^+$, *J. Am. Chem. Soc.* 1969, 91, 5350–5357.

[34] 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.

[35] Roby, K. R. On the Justifiability of Neglect of Differential Overlap Molecular Orbital Methods, *Chem. Phys. Lett.* 1971, 11, 6–10.

[36] 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.

[37] Roby, K. R. Fundamentals of an Orthonormal Basis Set Molecular Orbital Theory, *Chem. Phys. Lett.* 1972, 12, 579–582.

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

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

[40] 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.
[41] Duke, B. J.; Collins, M. P. S. The Ab Initio Neglect of Differential Diatomic Overlap Method, *Theor. Chim. Acta* 1981, 58, 233–244.

[42] 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. Theochem* 1988, 165, 189–202.

[43] Koch, W. Neglect of Diatomic Differential Overlap (NDDO) in Non-Empirical Quantum Chemical Orbital Theories, *Z. Naturforsch. A* 1993, 48, 819–828.

[44] 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.

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

[46] 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 Orthogonalized Basis Functions, *Int. J. Quantum Chem.* 1995, 53, 537–540.

[47] 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.

[48] 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.
[49] 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.

[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] 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.

[52] 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.

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

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

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

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

[57] 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.
[58] Davidson, E. R.; Feller, D. Basis Set Selection for Molecular Calculations, *Chem. Rev.* **1986**, *86*, 681–696.

[59] Dewar, M. J. S. The Semiempirical Approach to Chemistry, *Int. J. Quantum Chem.* **1992**, *44*, 427–447.

[60] Thiel, W. Semiempirical NDDO Calculations with STO-3G and 4-31G Basis Sets, *Theor. Chim. Acta* **1981**, *59*, 191–208.

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

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

[63] 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.

[64] Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Ru: Design and assessment of accuracy, *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297–3305.

[65] Ruddigkeit, L.; van Deursen, R.; Blum, L. C.; Reymond, J.-L. Enumeration of 166 billion organic small molecules in the chemical universe database GDB-17, *J. Chem. Inf. Model.* **2012**, *52*, 2864–2875.

[66] Ramakrishnan, R.; Dral, P. O.; Rupp, M.; von Lilienfeld, O. A. Quantum chemistry structures and properties of 134 kilo molecules, *Sci. Data* **2014**, *1*, 140022.

[67] Koster, J. L.; Ruttink, P. J. A. Non-empirical approximate calculations for the ground states of \( \text{H}_2 \) and \( \text{H}_3 \) including complete configuration interactions, *Chem. Phys. Lett.* **1972**, *17*, 419–421.

[68] Thiel, W. The MNDOC Method, a Correlated Version of the MNDO Model, *J. Am. Chem. Soc.* **1981**, *103*, 1413–1420.
[69] 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.

[70] Jensen, F. The basis set convergence of the Hartree–Fock energy for H₂, *J. Chem. Phys* **1999**, *110*, 6601–6605.

[71] 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.

[72] Simm, G. N.; Reiher, M. Systematic Error Estimation for Chemical Reaction Energies, *J. Chem. Theory Comput.* **2016**, *12*, 2762-2773.

[73] Simm, G.; Reiher, M. Error-Controlled Exploration of Chemical Reaction Networks with Gaussian Processes, *J. Chem. Theory Comput.* **2018**, submitted, [arXiv: 1805.09886].

[74] De, S.; Bartók, A. P.; Csányi, G.; Ceriotti, M. Comparing Molecules and Solids across Structural and Alchemical Space, *Phys. Chem. Chem. Phys.* **2016**, *18*, 13754-13769.

[75] Huang, B.; von Lilienfeld, O. A. Understanding molecular representations in machine learning: The role of uniqueness and target similarity, *J. Chem. Phys.* **2016**, *145*, 161102.

[76] 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.

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

[78] Hehre, W. J.; Stewart, R. F.; Pople, J. A. Self-Consistent Molecular-Orbital Methods. I. Use of Gaussian Expansions of Slater-Type Atomic Orbitals, *J. Chem. Phys.* **1969**, *51*, 2657–2664.
[79] Binkley, J. S.; Pople, J. A.; Hehre, W. J. Self-Consistent Molecular Orbital Methods. 21. Small Split-Valence Basis Sets for First-Row Elements, *J. Am. Chem. Soc.* 1979, **102**, 939–947.

[80] Krishnan, R.; Binkley, J. S.; Seeger, R.; Pople, J. A. Self-consistent molecular orbital methods. XX. A basis set for correlated wave functions, *J. Chem. Phys.* 1980, **72**, 650–654.

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

[82] 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.