Perturbative triples correction for local pair natural orbital based explicitly correlated CCSD(F12*) using Laplace transformation techniques

Gunnar Schmitz and Christof Hättig

Citation: The Journal of Chemical Physics **145**, 234107 (2016); doi: 10.1063/1.4972001
View online: http://dx.doi.org/10.1063/1.4972001
View Table of Contents: http://aip.scitation.org/toc/jcp/145/23
Published by the American Institute of Physics

Articles you may be interested in

Convergence of coupled cluster perturbation theory
The Journal of Chemical Physics **145**, 224104224104 (2016); 10.1063/1.4971294

Incremental full configuration interaction
The Journal of Chemical Physics **146**, 104102104102 (2017); 10.1063/1.4977727

Perspective: Explicitly correlated electronic structure theory for complex systems
The Journal of Chemical Physics **146**, 080901080901 (2017); 10.1063/1.4976974

The Laplace transformed divide-expand-consolidate resolution of the identity second-order Møller-Plesset perturbation (DEC-LT-RIMP2) theory method
The Journal of Chemical Physics **146**, 044103044103 (2017); 10.1063/1.4973710

Automatic active space selection for the similarity transformed equations of motion coupled cluster method
The Journal of Chemical Physics **146**, 074103074103 (2017); 10.1063/1.4976130

Performance of a nonempirical density functional on molecules and hydrogen-bonded complexes
The Journal of Chemical Physics **145**, 234306234306 (2016); 10.1063/1.4971853



Perturbative triples correction for local pair natural orbital based explicitly correlated CCSD(F12*) using Laplace transformation techniques

Gunnar Schmitz[a] and Christof Hättig[b]
Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, D-44801 Bochum, Germany

(Received 5 August 2016; accepted 28 November 2016; published online 19 December 2016)

We present an implementation of pair natural orbital coupled cluster singles and doubles with perturbative triples, PNO-CCSD(T), which avoids the quasi-canonical triples approximation (T0) where couplings due to off-diagonal Fock matrix elements are neglected. A numerical Laplace transformation of the canonical expression for the perturbative (T) triples correction is used to avoid an I/O and storage bottleneck for the triples amplitudes. Results for a test set of reaction energies show that only very few Laplace grid points are needed to obtain converged energy differences and that PNO-CCSD(T) is a more robust approximation than PNO-CCSD(T0) with a reduced mean absolute deviation from canonical CCSD(T) results. We combine the PNO-based (T) triples correction with the explicitly correlated PNO-CCSD(F12*) method and investigate the use of specialized F12-PNOs in the conventional triples correction. We find that no significant additional errors are introduced and that PNO-CCSD(F12*)(T) can be applied in a black box manner. *Published by AIP Publishing.* [http://dx.doi.org/10.1063/1.4972001]

I. INTRODUCTION

For a long time, it has been a pie in the sky for quantum chemistry to provide theoretical methods with accurate energy predictions for large molecules at affordable computational costs. Reaching this ambitious aim is hampered by the steep cost scaling of correlated wavefunction methods and the slow basis set convergence of the correlation energy. For example, coupled cluster singles and doubles with inclusion of perturbative triples,[1] CCSD(T), "the gold standard of quantum chemistry," has in canonical implementations operation counts that scale with $\mathcal{O}(N^7)$, where N denotes the system size. This means that the computational costs increase by a factor of 128 if the system size is doubled. On top of that, large basis sets of quadruple- or quintuple-ζ quality are required to fully exploit the accuracy of CCSD(T). This results in a huge computational burden and limits accurate canonical CCSD(T) calculations to systems with only a few atoms. Luckily, over the last decades solutions for the steep scaling[2–18] as well as for the basis set convergence[19–27] problems emerged. In the current work we present one approach to combine both strategies to make accurate CCSD(T) calculations on large molecules feasible with moderate computational costs.

The slow basis set convergence originates from a poor description of the electronic cusp provided by slater determinants of orbitals. Often extrapolation schemes as the one proposed in Ref. 28 are employed to alleviate this problem, but they do not improve the underlying wave function and sometimes fail for energy differences even for reactions of simple molecules if the basis sets are not large enough. A typical example is the hypothetical reaction of $CH_4 + 4\,H_2O_2$ to $CO_2 + 6H_2O$. The CCSD(T) results for the electronic contribution to the reaction energy in the cc-pVTZ and cc-pVQZ basis sets are, respectively, −1198 and −1188 kJ/mol. Extrapolation with $E(X) = E_{lim} - AX^{-3}$ from these two results leads to −1181 kJ/mol as an estimate for the basis set limit. This estimate deviates from the correct basis set limit (−1217 ± 0.5 kJ/mol, obtained from cc-pV5Z/cc-pV6Z extrapolation and F12 calculations) even more than the raw basis set results in the cc-pVTZ and cc-pVQZ basis sets. A more robust solution is to incorporate terms with an explicit dependence on the interelectronic distance into the wave function ansatz to enable a better description of the electron cusp as it is done in the F12 methods.[29] The latter approach enhances the basis set convergence such that F12 calculations achieve with a basis set cardinal number X typically the same accuracy as a conventional calculation with a cardinal number $X + 2$. For the above reaction CCSD(F12*)(T*) gives already in the aug-cc-pVTZ basis with −1218.5 kJ/mol a result which is converged within ±1.5 kJ/mol (in the aug-cc-VQZ and aug-cc-pV5Z basis sets results are, respectively, −1218.0 and −1217.5 kJ/mol).

The steep scaling of the computational costs of correlated wavefunction calculations is not only clearly unfavorable but also unphysical and arises from the delocalized nature of the canonical molecular orbitals (CMOs). But one has not to rely on CMOs. It is possible to transform the orbitals to a localized basis, where contributions of distant electron pairs become negligibly small. This paves the way for approximations that reduce the scaling. Among the different approaches, the pair natural orbital (PNO) formalism emerged as a very promising candidate to develop low scaling methods with a balanced error control.[9,12,30–32] In PNO methods each pair of localized occupied orbitals gets assigned a set of virtual orbitals which is pair-specific and obtained from the eigenvectors of a

[a]Author to whom correspondence should be addressed. Electronic mail: gunnar.schmitz@rub.de
[b]christof.haettig@rub.de

MP2-like pair contribution to a one-electron density matrix. Only PNOs with an eigenvalue above a user-defined threshold are kept. The number of PNOs per pair becomes quickly independent of the system size, so that the computational costs for contractions involving PNOs become asymptotically linear scaling. The main advantage of the PNO formalism is that the PNO construction from a density matrix adapts automatically to the electronic structure of the underlying system and that the accuracy can be controlled solely by the PNO truncation threshold.[12,31] In similar approaches, like the projected atomic orbital (PAO) formalism,[6,33] this is not the case. Here local domains are determined using a basis set population analysis and a distance criterion without exploiting much information on the electronic structure. After revival of the PNO approach by Neese and co-workers,[12,32] a variety of PNO-based correlation methods were presented. Many of them combine PNOs with other local representations like PAOs or orbital specific virtuals (OSVs).[13,14,32,34–38] For approximate triples methods, however, an additional problem arises for local methods because in the local basis, the Fock matrix loses its diagonal structure. The equations for the triples amplitudes have no longer a simple analytic solution and have to be solved iteratively. Unfortunately, this is not feasible for the molecular systems that are typically studied with PNO or other local correlation methods. One route to solve this issue is to remove the off-diagonal elements by diagonalizing the Fock matrix in the selected local virtual basis and to neglect the off-diagonal elements for the occupied space. This is an approximation which is often referred to as quasi-canonical (T0) correction.[39] In most cases it works quite well for energy differences. But sometimes the errors are unbalanced. Therefore, we suggest an alternative approach where the (T) correction is evaluated through a numerical Laplace-transformation of the canonical expression which does not neglect off-diagonal Fock matrix elements.

The subsequent part of the paper is organized as follows: In Sec. II the theory behind pair- and triples-specific natural orbitals, PNOs and TNOs, and the triples corrections are presented. We will introduce an alternative PAO selection scheme and the explicitly correlated PNO-CCSD(F12*) model. In Section III we describe the technical details of the benchmark calculations. This is followed in Section IV by a comparison of the accuracies of PNO-CCSD(T) and PNO-CCSD(T0) with and without intermediate PAO basis and an assessment of the use of F12-PNOs in the triples correction for PNO-CCSD[F12] and PNO-CCSD(F12*). In Sec. VI we draw our conclusion and comment on future work.

II. THE PNO-CCSD(F12*) MODEL

A. Index and naming conventions

In order to make the paper as self contained as possible, we summarize the index conventions for orbital spaces in Table I. In general we denote PAOs, PNOs, and TNOs with overbars. OSVs or pre-PNOs are indicated with a tilde. If it is not obvious from the context to which localized occupied molecular orbital (LMO) i, pair ij, or triple ijk the PAOs, PNOs, or TNOs belong, we additionally make use of subscripts. Table I also lists the naming convention for the specialized PNOs used in combination with F12 theory which will be explained in more detail in Section II D.

TABLE I. Summary on index convention.

$p, q, r, \ldots$	HF orbitals
$i, j, k, \ldots$	Active occupied orbitals
$a, b, c, \ldots$	Active virtual orbitals
$K, L, M, \ldots$	Inactive and active occupied orbitals
$P, Q, R, \ldots$	Auxiliary functions for the DF approximation
$a', b', c', \ldots$	Orbitals in the complementary space (CABS)
$a'', b'', c'', \ldots$	Union of virtual orbitals and CABS orbitals
$p'', q'', r'', \ldots$	Union of HF orbitals and CABS orbitals
$\mu, \nu, \kappa, \ldots$	Atomic orbitals of the MO basis
$\mu', \nu', \kappa', \ldots$	Atomic orbitals of the CABS basis
$\mu'', \nu'', \kappa'', \ldots$	Atomic orbitals of the MO and CABS basis
$\bar{\mu}, \bar{\nu}, \bar{\kappa}, \ldots$	PAO basis
$\bar{\mu}_i, \bar{\nu}_i, \bar{\kappa}_i, \ldots$	PAO basis specific for a LMO i
$\bar{a}_{ij}, \bar{b}_{ij}, \bar{c}_{ij}, \ldots$	PNOs/F12-PNOs for the virtual space
$\bar{a}_{ijk}, \bar{b}_{ijk}, \bar{c}_{ijk}, \ldots$	TNOs for the virtual space
$\tilde{a}_i, \tilde{b}_i, \tilde{c}_i, \ldots$	OSVs for the virtual space
$\tilde{a}_{ij}, \tilde{b}_{ij}, \tilde{c}_{ij}, \ldots$	Pre-PNOs for the virtual space
$\bar{\bar{K}}_{ij}, \bar{\bar{L}}_{ij}, \bar{\bar{M}}_{ij}, \ldots$	OPNOs-1: Specialized PNOs for the occupied space.
$\bar{K}_{ij}, \bar{L}_{ij}, \bar{M}_{ij}, \ldots$	OPNOs-2: Specialized PNOs for the occupied space.
$\bar{a}''_{ij}, \bar{b}''_{ij}, \bar{c}''_{ij}, \ldots$	CA-PNOs-1: Specialized PNOs for the union of virtual and CABS space.
$\bar{a}'_{ij}, \bar{b}'_{ij}, \bar{c}'_{ij}, \ldots$	CA-PNOs-2: Specialized PNOs for the union of virtual and CABS space.

B. PNO construction

1. Hybrid OSV-PNO approach

In our code we use a hybrid OSV-PNO approach with an iterative generation of the OSVs. Since it has already been described elsewhere, we give here only a brief summary and refer to Refs. 36 and 40 for further details. Orbital specific virtuals and pair natural orbitals provide both tensor factorizations of the doubles space. While OSVs are specific for a given occupied localized molecular orbital (LMO), PNOs are specific for pairs of LMOs. Both are generated by diagonalization of approximate MP2 one-particle densities. Only such eigenvectors are kept which have eigenvalues above a user-defined truncation threshold. In a PNO construction without intermediate OSV or PAO basis, the amplitudes t_{ab}^{ij} have to be generated in the full virtual basis, which results in $\mathcal{O}(N_{pair}N_{vir}^3)$ scaling costs for the PNO construction. This steep scaling is avoided in hybrid schemes where the amplitudes are built in an already truncated doubles space. In the hybrid OSV-PNO scheme first OSVs are generated, which are then used to span the PNOs. This reduces the cost-scaling for the PNO construction to $\mathcal{O}(N_{pair}N_{aux}n_{OSV}^2)+\mathcal{O}(N_{pair}N_{aux}N_{basis}n_{OSV})$, where N_{aux} is the size of the auxiliary basis for the density fitting (DF) approximation. If combined with local density fitting and the iterative procedure for the OSV generation described in Ref. 36, the whole procedure scales only quadratically with the system size. In the iterative OSV generation we pick up an idea of Almlöf and Häser[41] and replace the MP2-like amplitude equation by an equivalent Laplace transformed expression, which is evaluated numerically on a grid with n_L, points t_z, and weights

ω_z. This allows us to factorize the expression for the doubles amplitudes t_{ac}^{ii} and to derive a simple expression for a matrix-vector product σ_a^k with a trial vector b_c^k,

$$\sigma_a^k = \sum_c t_{ac}^{ii} b_c^k = -\sum_z^{n_L} \omega_z \sum_\nu C_{\nu a}^z \sum_Q (Q|\nu i^z) \times \sum_P \left[{}^{(ii)}V^{-1}\right]_{QP} \sum_\mu (P|\mu i^z) \sum_c C_{\mu c}^z b_c^k, \quad (1)$$

as required for the implementation of an iterative block Davidson algorithm to solve only for the eigenvectors with eigenvalues above the OSV truncation threshold. In the last equation we introduced transformed MOs $\phi_{i^z} = \chi_\mu C_{\mu i}^z$ and modified occupied and virtual MO coefficients defined as

$$C_{\mu i}^z = \sum_j C_{\mu j} e^{\varepsilon_j t_z} T_{ij}, \quad (2)$$

$$C_{\mu c}^z = C_{\mu c} e^{-\varepsilon_c t_z}, \quad (3)$$

with the canonical MO coefficients $C_{\mu p}$ and orbital energies ε_p and the LMO coefficients T_{ij} in the CMO basis. ${}^{(ii)}V_{QP} = (Q|P)$ is the matrix with the coulomb integrals in the LMO-specific DF basis.

2. PAOs as intermediate basis

For the triples correction (*vide infra*), we analyzed how a hybrid PAO-OSV-PNO approach can be used to reduce the computational costs for the AO to PNO (or TNO) integral transformation. This idea is not new and was already pursued by Neese and co-workers[37,42] as well as by Werner and co-workers,[43] but in our approach the local PAO domains are not selected solely based on information on the spatial structure and the basis set but also on the electronic structure. This is achieved by selecting the PAO domains on the fly during the OSV generation in a way which ensures that the PAO space is large enough to solve the OSV eigenvalue problem with a pre-defined accuracy. In a first step, the OSV eigenvalue problem

$$\sum_b t_{ab}^{ii} d_{b\tilde{a}} = \tau^{\tilde{a}} d_{a\tilde{a}}, \quad (4)$$

is expressed in PAOs $\bar{\nu}$ as

$$\sum_{\bar{\nu}} t_{\bar{\mu}\bar{\nu}}^{ii} d_{\bar{\nu}\tilde{a}} = \tau^{\tilde{a}} \sum_{\bar{\nu}} S_{\bar{\mu}\bar{\nu}}^{\mathrm{PAO}} d_{\bar{\nu}\tilde{a}}, \quad (5)$$

where $d_{\bar{\nu}\tilde{a}}$ and $t_{\bar{\mu}\bar{\nu}}^{ii}$ are, respectively, the eigenvectors and the amplitudes in the PNO basis,

$$d_{\bar{\nu}\tilde{a}} = \sum_c C_{\nu c} d_{c\tilde{a}}, \quad (6)$$

$$t_{\bar{\mu}\bar{\nu}}^{ii} = \sum_a S_{\bar{\mu}a} \sum_b t_{ab}^{ii} S_{\bar{\nu}b}. \quad (7)$$

$S_{\bar{\mu}a}$ is the overlap matrix between PAOs $\bar{\mu}$ and virtuals MOs a. The residual of the eigenvalue equation can be expressed as

$$R_{\bar{\mu}}^{\tilde{a}} = \sum_{\bar{\nu}} \left(t_{\bar{\mu}\bar{\nu}}^{ii} - \tau^{\tilde{a}} S_{\bar{\mu}\bar{\nu}}^{\mathrm{PAO}}\right) d_{\bar{\nu}\tilde{a}} = \sum_k \left(\sigma_{\bar{\mu}}^k d_{k\tilde{a}}^{red} - s_{\bar{\mu}}^k d_{k\tilde{a}}^{red} \tau^{\tilde{a}}\right), \quad (8)$$

where $d_{b\tilde{a}}^{red}$ are the eigenvectors in the reduced space, and $\sigma_{\bar{\mu}}^k$ and $s_{\bar{\mu}}^k$, respectively, the sigma and contravariant basis vectors,

$$\sigma_{\bar{\mu}}^k = \sum_{\bar{\nu}} t_{\bar{\mu}\bar{\nu}}^{ii} b_{\bar{\nu}}^k, \quad (9)$$

$$s_{\bar{\mu}}^k = \sum_{\bar{\nu}} S_{\bar{\mu}\bar{\nu}} b_{\bar{\nu}}^k, \quad (10)$$

for the current set of basis vectors $b_{\bar{\nu}}^k$.

The PAO selection proceeds as followed: For each LMO, i in a first step a Löwdin population analysis is carried out. If a PAO has a Löwdin charge >0.01, all PAOs at the same atom are assigned to the local PAO domain. This initial set is orthogonalized as

$$\sum_{\bar{\mu}\bar{\nu}} S_{\bar{\mu}\bar{\nu}} b_{\bar{\mu}}^k b_{\bar{\nu}}^l = \delta_{kl} \quad (11)$$

and the $b_{\bar{\mu}}^k$ are taken as initial basis vectors. Then additional PAOs are added in an iterative procedure until the trace of the density $\mathbf{D}^{ii} = t^{ii} \cdot t^{ii}$ is converged. The contribution of the not (yet) selected PAOs to the trace of $\mathbf{D}^{ii}$ can be decomposed as $\Delta(\mathrm{Tr}\mathbf{D}^{ii}) = 2\Delta_1 + \Delta_2$, where Δ_1 is linear and Δ_2 is quadratic in the not (yet) selected PAOs. Δ_2 is negligible compared to Δ_1 as soon as a reasonable start space has been chosen and the first-order contribution can be split into contributions of individual PAOs as $\Delta_1 = \sum_{\bar{\mu}} \delta_{\bar{\mu}}$ with

$$\delta_{\bar{\mu}} = \sum_{\tilde{a}} R_{\bar{\mu}}^{\tilde{a}} \Big(\sum_{k\bar{\nu}} D_{\bar{\mu}\bar{\nu}}^{vir} \sigma_{\bar{\nu}}^k d_{k\tilde{a}}^{red} \Big). \quad (12)$$

In the last equation we used the short-hand notation $D_{\bar{\mu}\bar{\nu}}^{vir} = \sum_a C_{\bar{\mu}a} C_{\bar{\nu}a}$. Since $R_{\bar{\mu}}^{\tilde{a}}$ is orthogonal to the already selected PAO space, this also holds for $\delta_{\bar{\mu}}$. PAOs with $\delta_{\bar{\mu}}$ larger than a user defined threshold T_{PAO} are orthogonalized onto the selected PAOs and added to the reduced space for the next iteration. This procedure is repeated until no additional PAOs are selected. Afterwards the OSV eigenvalue problem is solved in the selected PAO domain and the OSVs, and later the PNOs are expanded in these PAOs. The process of the PAO selection is sketched in the pseudocode in Fig. 1.

C. F12-PNOs

For F12 methods, it is convenient to construct specialized F12-PNOs which provide a more compact representation of the virtual space by optimizing them to span primarily the part of the virtual pair space which is not captured by the geminals. This is achieved by calculating the PNOs as eigenvectors of a pseudo-density build from difference amplitudes,[44]

$$\Delta_{ab}^{ij} = t_{ab}^{ij} - r_{ab}^{\widetilde{ij}}, \quad (13)$$

where $r_{ab}^{\widetilde{ij}}$ is the amplitude for the double excitation $ij \rightarrow ab$, which would be described by the geminal χ_{ij} if they are not orthogonalized to the orbital space. If Δ_{ab}^{ij} is small, the excitation is well approximated by the geminal and only few PNOs are required to span the remaining difference.

Algorithm 1 Iterative PAO selection during OSV generation

Generate initial set of PAOs for LMO i:

Loop over all $\bar{\nu}$ with a Löwdin charge > 0.01

- Add all $\bar{\mu}$ on the same atom and orthogonalize set to obtain covariant $\mathbf{d}^c$ and contravariants$\mathbf{s}_{\tilde{c}} = \mathbf{S}^{\mathrm{PAO}}\mathbf{d}_c$ basis vector in the reduced space.

End of Loop

- Reorthonormalize basis to enhance numerical stability.

Add additional PAOs in an iterative process:

Loop PAO selection

- Calculate for current basis $\mathbf{d}_c$ the sigma vectors $\sigma_c = \mathbf{t}^{ii,\mathrm{PAO}}\mathbf{d}_c$
- Calculate contravariant sigma vectors $\bar{\sigma}_c = \mathbf{D}^{vir,ao}\sigma_c$
- Calculate the amplitudes in the reduced space $T^{red}_{ab} = \bar{\sigma}_a\mathbf{d}_b$
- Calculate for each PAO the weights δ_μ and add PAOs with $\delta_\mu \geq T_{\mathrm{PAO}}$
- If no further PAOs are added exit loop

End of Loop

- Diagonalize $\mathbf{T}^{red}$ to obtain OSVs
- Select OSVs with $n^i_{\tilde{a}} \geq T_{\mathrm{OSV}}$
- Project solution of the reduced space to the full virtual space.

FIG. 1. Scheme for the PAO selection.

D. X-PNOs

Besides the F12-PNOs for the virtual space, additional auxiliary PNOs (X-PNOs) are used for a compact pair-specific representation of the strong orthogonality operator in the two approximations ${}^A\hat{Q}$ and ${}^B\hat{Q}$. For ${}^B\hat{Q}_{12}$, we use the ansatz from Ref. 44,

$$ {}^B\hat{Q}_{12} = \overline{V}_1\overline{V}'_2 + \overline{V}'_1\overline{V}_2 + \overline{V}'_1\overline{V}'_2, \tag{14}$$

while for ${}^A\hat{Q}_{12}$ we use the ansatz termed ${}^{A_3}\hat{Q}_{12}$,

$$ {}^{A_3}\hat{Q}_{12} = 1 - \overline{\overline{O}}_1\overline{\overline{O}}_2 - \overline{V}_1\overline{V}_2 - \overline{O}_1\overline{V}''_2 - \overline{V}''_1\overline{O}_2\,. \tag{15}$$

The symbols V, V', V'', and O indicate, respectively, projectors onto virtual orbitals (V), the CABS (CA) basis (V'), as well as the union of the virtuals and the CABS basis (V''), and the occupied space (O). The overbars indicate that the projector acts on PNO spaces. Since for the occupied space two kinds of X-PNOs are used, additionally a double bar is introduced to distinguish between both choices for PNOS, which will be discussed in the following. As PNOs and F12-PNOs, also the X-PNOs are constructed by diagonalizing (pseudo) densities, which are defined as

$$ \text{(OPNOs-1)} \quad \overline{\overline{K}}_{ij}: \quad D^{ij}_{KL} = 2\sum_M \left(s^{\widetilde{ij}}_{KM} r^{\widetilde{ij}}_{LM} + s^{\widetilde{ji}}_{KM} r^{\widetilde{ji}}_{LM} \right), \tag{16}$$

$$ \text{(OPNOs-2)} \quad \overline{K}_{ij}: \quad D^{ij}_{KL} = 2\sum_{c''} \left(s^{\widetilde{ij}}_{Kc''} r^{\widetilde{ij}}_{Lc''} + s^{\widetilde{ji}}_{Kc''} r^{\widetilde{ji}}_{Lc''} \right), \tag{17}$$

$$ \text{(CA-PNOs-1)} \quad \bar{a}''_{ij}: D^{ij}_{a''b''} = 2\sum_M \left(s^{\widetilde{ij}}_{a''M} r^{\widetilde{ij}}_{b''M} + s^{\widetilde{ji}}_{a''M} r^{\widetilde{ji}}_{b''M} \right), \tag{18}$$

and

$$ \text{(CA-PNOs-2)} \quad \bar{a}'_{ij}: D^{ij}_{a'_{ij}b'_{ij}} = 2\sum_{\tilde{c}_{ij}} \left(s^{\widetilde{ij}}_{a'_{ij}\tilde{c}_{ij}} r^{\widetilde{ij}}_{b'_{ij}\tilde{c}_{ij}} + s^{\widetilde{ji}}_{a'_{ij}\tilde{c}_{ij}} r^{\widetilde{ji}}_{b'_{ij}\tilde{c}_{ij}} \right), \tag{19}$$

where $s^{\widetilde{ij}}_{pq}$ is a short-hand notation for $2r^{\widetilde{ij}}_{pq} - r^{\widetilde{ji}}_{pq}$ with $r^{\widetilde{ij}}_{pq} = \langle pq|f_{12}|\widetilde{ij}\rangle$ and where the tilde on ij indicates the contraction with the geminal coefficients from the fixed amplitude ansatz,[26,45]

$$ |\widetilde{ij}\rangle = \frac{3}{8}\phi_i(1)\phi_j(2) + \frac{1}{8}\phi_j(1)\phi_i(2). \tag{20}$$

The OPNOs 1 and 2 are both PNOs, which span the occupied orbital space, but the $\overline{\overline{K}}_{ij}$ orbitals occur in the projector together with the $\bar{a}''_{ij}$ orbitals, which span the union of complementary space (CA) and virtual orbitals. The CA-PNOs-2 $\bar{a}'_{ij}$ span the union of CA and those virtual orbitals which are not included in the F12-PNOs. In this way the most important virtuals screened out in the construction of the F12-PNOs are included in an extended CA space. The construction of the X-PNOs is again combined with the hybrid OSV-PNO approach as described in Ref. 40.

E. Correction for neglected PNOs and pairs

Besides the truncation of the virtual space also LMO pairs are discarded. Initially a set of pairs for which PNOs are constructed is selected based on spin opposite scaled[15] (SOS) MP2 pair energies evaluated in the OSV basis,[34]

$$ e^{\text{SOS-MP2}}_{ij} \approx c_{OS}\left(1+\delta_{ij}\right)\sum_{\tilde{a}_i\tilde{b}_j} \frac{2K^{ij}_{\tilde{a}_i\tilde{b}_j}K^{ij}_{\tilde{a}_i\tilde{b}_j}}{F_{ii} - F_{\tilde{a}_i\tilde{a}_i} + F_{jj} - F_{\tilde{b}_j\tilde{b}_j}}. \tag{21}$$

$K^{ij}_{\tilde{a}_i\tilde{b}_j}$ are the exchange integrals in the (semi-canonical) OSV basis $(\tilde{a}_i i|\tilde{b}_j j)$ and F_{ii} and $F_{\tilde{a}_i\tilde{a}_i}$ the diagonal elements of the Fock matrix. A further truncation of the pair lists is then done based on the semi-canonical MP2 energy (SC-MP2) expression during the PNO construction. The effect of the pair truncation can be controlled via a pair truncation threshold T_{pair}. To reduce the error due to this truncation, the final energy is corrected by adding the estimates for the pair energies of the discarded pairs. Furthermore, we add for the non-F12 methods a correction for discarded PNOs. This correction is calculated as the difference between the SC-MP2 pair energies in the PNO and pre-PNO basis, which is the initial basis in the PNO construction obtained by merging the OSVs of two LMOs. For the F12 methods the latter correction is not accessible without substantial additional costs, since it would require the evaluation of several F12 intermediates in the pre-PNO basis. However, for F12 methods the geminal contributions automatically correct partially for the truncation of the virtual space.

F. Explicitly correlated contributions for PNO-CCSD(F12*)

Already in Ref. 40 we presented an implementation of PNO-CCSD[F12]. Starting from the PNO-CCSD[F12] Lagrangian $L_{[\text{F12}]}$ as well as the singles $\Omega^i_{a,[\text{F12}]}$ and doubles $\Omega^{ij}_{\bar{a}\bar{b},[\text{F12}]}$ residuals only a few terms are missing to arrive at PNO-CCSD(F12*). These are

$$L_{(\text{F12}^*)} = L_{[\text{F12}]} + \frac{1}{2}\sum_{abij}\left(2\mathcal{V}^{ij}_{ab} - \mathcal{V}^{ji}_{ab}\right) t^i_a t^j_b, \tag{22}$$

$$\Omega^i_{a,(\text{F12}^*)} = \Omega^i_{a,[\text{F12}]} - \sum_k \mathcal{V}^i_k t^k_a, \tag{23}$$

and

$$\begin{aligned}\Omega^{ij}_{\bar{a}\bar{b},(\text{F12}^*)} = \Omega^{ij}_{\bar{a}\bar{b},[\text{F12}]} &+ \sum_{kl}\left(2\mathcal{V}^{ij}_{kl} - \mathcal{V}^{ji}_{kl}\right)\left(\frac{1}{2}t^{kl}_{ab} + t^k_a t^l_b\right) \\ &- \hat{P}^{ij}_{ab}\sum_k\left(2\mathcal{V}^{ij}_{ak} - \mathcal{V}^{ji}_{ak}\right) t^k_b - \hat{P}^{ij}_{ab}\sum_k \mathcal{V}^j_k t^{ik}_{ab}.\end{aligned} \tag{24}$$

Only additional $\mathcal{V}$ intermediates appear. They can be evaluated with relatively low costs from intermediates which are already constructed for PNO-CCSD[F12]. For example $\mathcal{V}^j_k$ can be obtained from $\mathcal{V}^j_\mu$ (where μ is an AO index), which is used in PNO-CCSD[F12] as an intermediate in the calculation of $\mathcal{V}^j_a$.

The additional contribution to the singles residual in Eq. (23) is computed straightforwardly without truncation of the singles amplitudes. The structure of the last term in Eq. (24) is analogous to a non-F12 contribution which involves a dressed Fock matrix[46,47] and is in PNO-CCSD(F12*) incorporated by adding $-\mathcal{V}^j_k$ as a F12 correction to the dressed Fock matrix.

In contrast to the terms discussed so far, the second term in Eq. (24) is more involved because it depends on the indices of three LMOs, which complicates the screening. It is known from canonical implementations[48] that this contribution is small. Therefore we neglect it for the current PNO implementation. It is a F12 correction to a conventional term which is also neglected in the LPNO-CCSD implementation described by Neese and co-workers.[47]

Due to the four occupied indices, the first term also causes complications for a PNO-based implementation, but it cannot be neglected as it is the most important correction beyond CCSD[F12]. Its structure has similarities with the conventional disjoint pair interaction term to which it is a basis set incompleteness correction. As for its conventional counterpart, the pair information is used to restrict the set $\{kl, ij\}$ to allow an efficient implementation. Furthermore, the exact evaluation of the integrals $g^{\bar{p}_{ij}\bar{q}_{ij}}_{kl}$ with $kl \neq ij$ is omitted by projection of the integrals $g^{\bar{p}_{kl}\bar{q}_{kl}}_{kl}$ onto the F12-PNO and X-PNO basis sets of the pair kl. In a first step, the intermediate $\bar{\mathcal{V}}^{ij}_{kl}$ is computed as

$$\begin{aligned}\bar{\mathcal{V}}^{ij}_{kl} = &\left\langle kl|f_{12}g_{12}|\widetilde{ij}\right\rangle \\ &- (1+\hat{P}^{ij}_{kl})\sum_{\bar{L}_{ij}\bar{c}''_{ij}} r^{\tilde{ij}}_{\bar{L}_{ij}\bar{c}''_{ij}} \sum_{\bar{L}_{kl}\bar{c}''_{kl}} S^{ij,kl}_{\bar{L}_{ij}\bar{L}_{kl}} S^{ij,kl}_{\bar{c}''_{ij}\bar{c}''_{kl}} g^{kl}_{\bar{L}_{kl}\bar{c}''_{kl}} \\ &- \sum_{\bar{\bar{K}}_{ij}\bar{\bar{L}}_{ij}} r^{\tilde{ij}}_{\bar{\bar{K}}_{ij}\bar{\bar{L}}_{ij}} \sum_{\bar{\bar{K}}_{kl}\bar{\bar{L}}_{kl}} S^{ij,kl}_{\bar{\bar{K}}_{ij}\bar{\bar{K}}_{kl}} S^{ij,kl}_{\bar{\bar{L}}_{ij}\bar{\bar{L}}_{kl}} g^{kl}_{\bar{\bar{K}}_{kl}\bar{\bar{L}}_{kl}} \\ &- \sum_{\bar{a}_{ij}\bar{b}_{ij}} r^{\tilde{ij}}_{\bar{a}_{ij}\bar{b}_{ij}} \sum_{\bar{a}_{kl}\bar{b}_{kl}} S^{ij,kl}_{\bar{a}_{ij}\bar{a}_{kl}} S^{ij,kl}_{\bar{b}_{ij}\bar{b}_{kl}} g^{kl}_{\bar{a}_{kl}\bar{b}_{kl}}.\end{aligned} \tag{25}$$

The similarity of $\bar{\mathcal{V}}^{ij}_{kl}$ with its conventional counterpart is exploited by using the same routines to evaluate it — just with different orbital spaces for the contraction — and by adding it to its conventional counterpart so that the contraction with the amplitudes is included without additional costs. The coulomb integrals required for the computation of $\bar{\mathcal{V}}^{ij}_{kl}$ are already available from the initial PNO-MP2-F12 step that precedes the PNO-CCSD[F12] and PNO-CCSD(F12*) calculations. The contribution from the identity in the projector is computed using local density fitting and the half transformed integrals $(Q|f_{12}g_{12}|\mu i)$, which are also available from the PNO-MP2-F12 part. To reduce the operation count by a factor of 2, we exploit the permutational symmetry of $\mathcal{V}^{ij}_{kl}$ and the integrals using a trick that has been used earlier in CCSD implementations.[46,49] The integral matrices are split into their symmetric (+) and antisymmetric (−) parts,

$$\begin{aligned} r^{ij,(+)}_{\bar{p}\bar{q}} &= r^{ij}_{\bar{p}\bar{q}} + r^{ji}_{\bar{p}\bar{q}} \\ \text{and} \qquad r^{ij,(-)}_{\bar{p}\bar{q}} &= r^{ij}_{\bar{p}\bar{q}} - r^{ji}_{\bar{p}\bar{q}} \end{aligned} \tag{26}$$

and similar for $g^{kl}_{\bar{p}\bar{q}}$. The contribution to $\bar{\mathcal{V}}^{ij}_{kl} = 2\mathcal{V}^{ij}_{kl} - \mathcal{V}^{ij}_{lk}$ from the projection onto the product space $\bar{p}\bar{q}$ can then be calculated as

$$\begin{aligned}\bar{\mathcal{V}}^{ij}_{kl} = \bar{\mathcal{V}}^{ij}_{kl} &+ \frac{1}{2}\sum_{\bar{p}\geq\bar{q}} \frac{1}{1+\delta_{\bar{p}\bar{q}}} r^{ij,(+)}_{\bar{p}\bar{q}} g^{\bar{p}\bar{q},(+)}_{kl} \\ &+ \frac{3}{2}\sum_{\bar{p}\geq\bar{q}} \frac{1}{1+\delta_{\bar{p}\bar{q}}} r^{ij,(-)}_{\bar{p}\bar{q}} g^{\bar{p}\bar{q},(-)}_{kl}.\end{aligned} \tag{27}$$

The remaining contribution to the Lagrangian can again be obtained from quantities already evaluated for PNO-CCSD[F12]. During the computation of the Lagrangian, $\mathcal{V}^{ij}_{\bar{a}\bar{b}}$ is read in and contracted with projected singles amplitudes,

$$L_{(\mathrm{F12}^*)} = L_{[\mathrm{F12}]} + \frac{1}{2}\sum_{\bar{a}\bar{b}ij}\left(2\mathcal{V}^{ij}_{\bar{a}\bar{b}} - \mathcal{V}^{ji}_{\bar{a}\bar{b}}\right)\sum_a d^{ij}_{a\bar{a}} t^i_a \sum_b d^{ij}_{b\bar{b}} t^j_b. \tag{28}$$

III. PERTURBATIVE CONNECTED TRIPLES

A. Triples selection and construction of TNOs

Crucial for an efficient triples correction in a local framework is the prescreening of triples entering the computational demanding steps. For that purpose it is important to take a look at the equation for the triples amplitudes,

$$\varepsilon^{abc}_{ijk} t^{ijk}_{abc} = \hat{P}^{ijk}_{abc}\Big(\sum_d^{N_v} g^{id}_{ab} t^{jk}_{dc} - \sum_l^{N_o} g^{ij}_{al} t^{lk}_{bc}\Big), \tag{29}$$

where ε^{abc}_{ijk} are the orbital energy differences, t^{jk}_{dc} the converged CCSD doubles amplitudes, and $\hat{P}^{ijk}_{abc}$ is an operator, which creates a sum over all six permutations of the index pairs ai, bj, and kc. The restriction in the pair lists for the doubles amplitudes directly transfers to the triples. Furthermore it is evident that the sum of occupied orbitals l is restricted to LMOs, which form non-vanishing pairs with the LMOs i, j, and k. If additionally the virtual space is spanned by TNOs, the ranges of the summations in the local basis, $N_{\bar{v}}$ and $N_{\bar{o}}$, will for large systems become independent of the system size. At this point the computational costs increase only linearly with the number of selected triples. Unfortunately there is, to the best of our knowledge, no way to compute cheap estimates for the integrals in the amplitude equation. Therefore we base our screening on the CCSD pair energies. The products of the pair energies should be a measure for the significance for a given triple. The direct use of the arithmetic mean value of the pair energy products gave no satisfactory results. We also tested other averages. Fig. 2 shows for some of them the correlation with the correlation with the exact PNO triples energy corrections. After intensive numerical tests, we came up with a simple estimate for the importance of a LMO triple based on a modified logarithmic mean of the product of pair energies e_i. Its performance for the Gly4 examples is shown in Fig. 2(d). In contrast to the other averages, it shows a better correlation and the data points are distributed quite narrowly on the line down to almost 10^{-7} a.u. Nevertheless it is solely an empiric estimate and therefore we decided to couple T_{trip} conservatively to T_{TNO} so that no decline in accuracy can be expected. Improvements to these estimates are subject of ongoing research in our laboratories since they pave the way for substantial computational savings if the initial triples list can be further truncated. If all three products differ at least by 10^{-12}, the proposed estimate is defined as

$$E_{\mathrm{guess}} = 3.5\left[\frac{(e_1-e_2)\cdot(e_2-e_3)\cdot(e_3-e_1)}{2\cdot\left[(e_2-e_3)\log e_1 + (e_3-e_1)\log e_2 + (e_1-e_2)\log e_3\right]}\right]^{0.4}. \tag{30}$$

If only two products are not equal in this sense, it is given by

$$E_{\mathrm{guess}} = 3.5\left[\frac{e_1-e_2}{\log e_1 - \log e_2}\right]^{0.9}. \tag{31}$$

In our program the estimate is compared to a user defined threshold T_{trip}. After several test calculation, we chose it as $T_{\mathrm{trip}} = 10\cdot T_{\mathrm{TNO}}$ for routine applications.

The triples correction is in the spirit of the PNO framework computed in a triple-specific triple natural orbital (TNO) basis, which was introduced by Riplinger and Neese.[37] With some modifications, their idea can be combined with the hybrid OSV-PNO approach. For the triple specific TNO construction, we use as an intermediate step pre-TNOs, which are constructed by merging the OSVs of the LMOs i,j, and k. This basis for triples ijk is than used to build the averaged density matrix

$$D^{ijk}_{\tilde{a}_{ijk}\tilde{b}_{ijk}} = \frac{1}{3}\left(D^{ij}_{\tilde{a}_{ijk}\tilde{b}_{ijk}} + D^{ik}_{\tilde{a}_{ijk}\tilde{b}_{ijk}} + D^{jk}_{\tilde{a}_{ijk}\tilde{b}_{ijk}}\right), \tag{32}$$

calculated from the converged CCSD doubles amplitudes projected to the pre-TNO basis. This averaged density is afterwards diagonalized to get initial TNOs. As for the PNOs only TNOs with occupation numbers above the truncation threshold T_{TNO} are retained. In the next step, the Fock matrix is projected onto this preliminary TNO basis and diagonalized to obtain quasi-canonical TNOs.

B. Non-iterative (T0) triples correction

In the CCSD(T) model the perturbative triples correction is defined in the canonical MO basis as

$$E_{(\mathrm{T})} = E_{[\mathrm{T}]} + E_{5\mathrm{S}}, \tag{33}$$

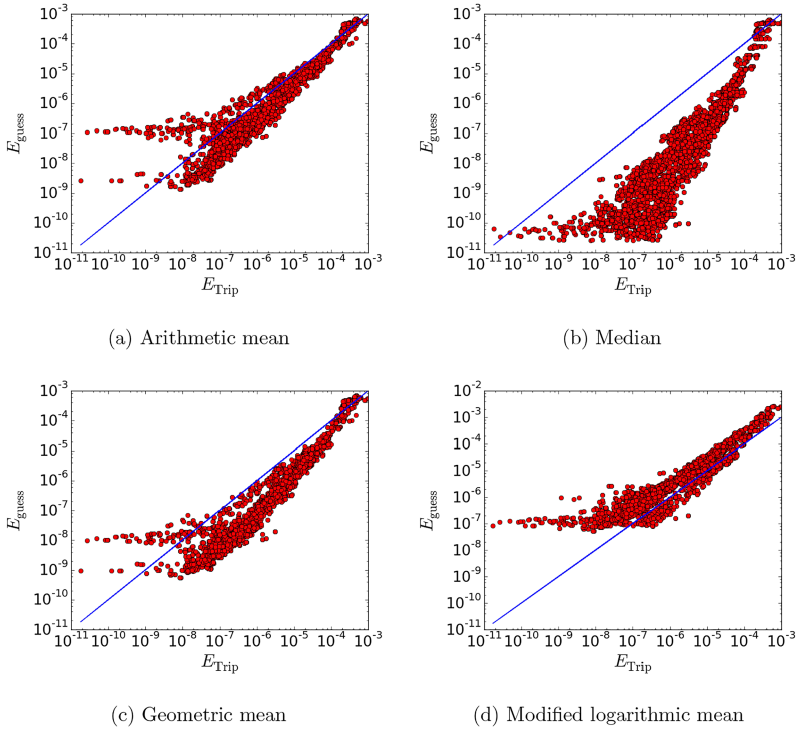
$$E_{[\mathrm{T}]} = \sum_{i\geq j\geq k}\sum_{a,b,c} \varepsilon^{abc}_{ijk} t^{ijk}_{abc}\left(4t^{ijk}_{abc} + t^{jki}_{abc} + t^{kij}_{abc} - 2t^{jik}_{abc} - 2t^{ikj}_{abc} - 2t^{kij}_{abc}\right), \tag{34}$$

and

$$\begin{aligned} E_{5\mathrm{S}} = -\sum_{i\geq j\geq k}\sum_{a,b,c} &\left(t^{ijk}_{abc} - t^{ijk}_{bac}\right)\left(g^{jk}_{bc} t^i_a - g^{ik}_{bc} t^j_a\right) \\ &+ \left(t^{ijk}_{abc} - t^{ijk}_{acb}\right)\left(g^{ki}_{ca} t^j_b - g^{ji}_{ca} t^k_b\right) \\ &+ \left(t^{ijk}_{abc} - t^{ijk}_{cba}\right)\left(g^{ij}_{ab} t^k_c - g^{kj}_{ab} t^i_c\right). \end{aligned} \tag{35}$$

For the construction of the triples amplitudes $t^{ijk}_{\bar{a}\bar{b}\bar{c}}$, we follow the approach proposed by Noga and Valiron.[50] A central point of their algorithm is the arrangement of the triples amplitude equations into a super-matrix form

$$\varepsilon^{abc}_{ijk} t^{ijk}_{abc} = \hat{P}^{ijk}_{abc}\sum_r^N K^{ir}_{ab} L^{jk}_{rc} = \hat{P}^{ijk}_{abc}\Lambda^{ijk}_{abc}. \tag{36}$$



(a) Arithmetic mean (b) Median

(c) Geometric mean (d) Modified logarithmic mean

FIG. 2. Tested averages (in a.u.) of pair energy products for Gly_4 in the def2-TZVP basis.

This leads to a better performance of matrix-matrix multiplications (dgemm) and allows furthermore for a more efficient arrangement of the computational steps with the auxiliary super-matrices $\mathbf{\Lambda}$ constructed from the $\mathbf{K}$ and $\mathbf{L}$ matrices. We do not repeat here the arguments and the derivation given by Noga and Valiron and refer instead to their original work.[50] Projecting Eq. (29) onto the quasi-canonical TNO basis leads to

$$\varepsilon_{ijk}^{\bar{a}_{ijk}\bar{b}_{ijk}\bar{c}_{ijk}} t_{\bar{a}_{ijk}\bar{b}_{ijk}\bar{c}_{ijk}}^{ijk} = \hat{P}_{\bar{a}_{ijk}\bar{b}_{ijk}\bar{c}_{ijk}}^{ijk} \Bigg(\sum_{\bar{d}_{ijk}}^{N_{\bar{v}}} g_{\bar{a}_{ijk}\bar{b}_{ijk}}^{i\bar{d}_{ijk}} \times \sum_{\bar{d}_{jk}\bar{c}_{jk}}^{N_{\bar{v}}} S_{\bar{d}_{ijk}\bar{d}_{jk}}^{ijk,jk} S_{\bar{c}_{ijk}\bar{c}_{jk}}^{ijk,jk} t_{\bar{d}_{jk}\bar{c}_{jk}}^{jk} - \sum_{l}^{N_o} g_{\bar{a}_{ijk}l}^{ij} \sum_{\bar{b}_{lk}\bar{c}_{lk}}^{N_{\bar{v}}} S_{\bar{b}_{ijk}\bar{b}_{lk}}^{ijk,jk} S_{\bar{c}_{ijk}\bar{c}_{lk}}^{ijk,jk} t_{\bar{b}_{lk}\bar{c}_{lk}}^{lk} \Bigg). \quad (37)$$

In the last equation, we introduced the overlap matrices between TNOs and PNOs,

$$S_{\bar{a}_{ijk}\bar{a}_{jk}}^{ijk,jk} = \sum_{c} d_{c\bar{a}_{ijk}}^{ijk} d_{c\bar{a}_{jk}}^{jk}. \quad (38)$$

Eq. (37) can be cast into the form of Eq. (36) by defining $\mathbf{K}$ and $\mathbf{L}$ in the TNO basis as

$$K_{\bar{a}_{ijk}\bar{b}_{ijk}}^{ir} = \begin{cases} -t_{\bar{a}_{ijk}\bar{b}_{ijk}}^{ir} & r \in [1, N_{\bar{o}}] \\ g_{\bar{a}_{ijk}\bar{b}_{ijk}}^{i(r-N_o)_{ijk}} & r \in [N_{\bar{o}}+1, \bar{N}] \end{cases}, \quad (39)$$

$$L_{r\bar{c}_{ijk}}^{jk} = \begin{cases} g_{r\bar{c}_{ijk}}^{jk} & r \in [1, N_{\bar{o}}] \\ t_{(r-N_o)_{ijk}\bar{c}_{ijk}}^{jk} & r \in [N_{\bar{o}}+1, \bar{N}] \end{cases}, \quad (40)$$

with the doubles amplitudes projected onto the TNO space as

$$t_{\bar{a}_{ijk}\bar{b}_{ijk}}^{il} \approx \sum_{\bar{a}_{il}\bar{b}_{il}}^{N_{\bar{v}}} S_{\bar{a}_{ijk}\bar{a}_{il}}^{ijk,il} S_{\bar{b}_{ijk}\bar{b}_{il}}^{ijk,il} t_{\bar{a}_{il}\bar{b}_{il}}^{il}. \quad (41)$$

With these adjustments, the algorithm from Ref. 50 can be applied in the TNO basis, but Eq. (37) still assumes canonical occupied orbitals. In the LMO basis, the Fock matrix is no longer diagonal and the equations would have to be solved iteratively which would require the storage of at least a few sets of triples amplitudes. At this point often the semi-canonical (T0) approximation is applied where the off-diagonal Fock matrix elements in the occupied block are neglected.[39,51] This avoids an expensive iterative solution, but it is an additional approximation. Reaction and interaction energies of CCSD(T) and CCSD(T0) are often very close, but there are also cases where they deviate significantly. Therefore it is desirable to have the possibility to check if the (T0) correction might be too incorrect. Schütz and co-workers proposed the (T1) correction for such cases, which can be seen as a modified first iteration in an iterative solution of the amplitude equations, and also demonstrated that it is in principle possible to solve the amplitudes in a full iterative way, although this is extremely costly.[39]

C. Laplace transformed CCSD(T)

We follow here a third alternative which applies instead of the Laplace transformation technique. It has the advantage that it requires only few modifications to a code for the (T0) correction and allows us to solve the triples equations within an accuracy which can be adjusted by a single threshold without storage of the triples amplitudes. As will be shown below, an accuracy which is sufficient for most practical purposes is typically achieved with computational costs which are only by a factor of 3–4 higher than for the (T0) correction. This is still acceptable in cases where the (T0) correction is not sufficient. Applying the numerical Laplace transformation to Eq. (29) leads to

$$t^{ijk}_{abc} = \hat{P}^{ijk}_{abc} \sum_{z=1}^{n_L} \omega_z \Big(\sum_d e^{(\varepsilon_j+\varepsilon_k-\varepsilon_c)\cdot t_z} t^{jk}_{dc}(ai|bd) e^{(\varepsilon_i-\varepsilon_a-\varepsilon_b)t_z} - \sum_l e^{(\varepsilon_k-\varepsilon_b-\varepsilon_c)t_z} t^{lk}_{bc}(ai|lj) e^{(\varepsilon_i+\varepsilon_j-\varepsilon_a)t_z} \Big), \tag{42}$$

in the CMO basis. For the expression in the LMO and PNO basis, we include the exponential factors in the integrals and amplitudes and obtain grid point dependent quantities,

$$t^{ijk}_{\bar{a}\bar{b}\bar{c}} \approx \hat{P}^{ijk}_{\bar{a}\bar{b}\bar{c}} \sum_{z=1}^{n_L} \omega_z \Big(\sum_{\bar{d}} t^{j^z k^z}_{\bar{d}\bar{c}^z} (\bar{a}^z i^z | \bar{b}^z \bar{d}) - \sum_l t^{lk^z}_{\bar{b}^z \bar{c}^z} (\bar{a}^z i^z | l j^z) \Big). \tag{43}$$

To evaluate the modified integrals, we use for the LMOs again the grid point dependent MO coefficients defined in Eq. (2) and for the (modified) TNOs,

$$C^z_{\mu\bar{c}} = \sum_c C_{\mu c} e^{-\varepsilon_c t_z} d^{ijk}_{c\bar{c}}, \tag{44}$$

where $d^{ijk}_{c\bar{c}}$ are the TNO coefficients. To reduce the prefactor for the evaluation of the modified TNO coefficients in the AO basis, they are computed from grid point dependent OSV coefficients in the AO basis. The computation of the modified amplitudes is slightly more involved. The modification of the TNO index can be easily achieved by adapting the overlap of the PNO and TNO spaces,

$$S^{ij,ijk,z}_{\bar{a}_{ij},\bar{a}_{ijk}} = \sum_a d^{ij}_{a,\bar{a}_{ij}} e^{-\varepsilon_a t_z} d^{ijk}_{a,\bar{a}_{ijk}} \tag{45}$$

The transformation of the LMO indices is best carried out in the PNO basis to reduce the generation of overlap matrices to a minimum. We need amplitudes with just one and with both LMO indices transformed to the grid point dependent basis. The transformation of one LMO index can be expressed as

$$t^{ij^z}_{\bar{a}\bar{b}} = t^{ij}_{\bar{a}\bar{b}} T^z_{jj} + \sum_{k\neq j} T^z_{kj} \sum_{\bar{a}_{ik},\bar{b}_{ik}} S^{ik,ij}_{\bar{a}_{ik},\bar{a}_{ij}} S^{ik,ij}_{\bar{b}_{ik},\bar{b}_{ij}} t^{ik}_{\bar{a}_{ik},\bar{b}_{ik}} \tag{46}$$

with the transformation matrix from the LMO space to the Laplace modified LMO space,

$$T^z_{kj} = \sum_j T_{ij} e^{\varepsilon_j t_z} T_{ij}. \tag{47}$$

To obtain $\mathbf{t}^{i^z j^z}$, this step is repeated with $\mathbf{t}^{ij}$ replaced by $\mathbf{t}^{ij^z}$. In that way we avoid the calculation of overlap matrices for disjoint pairs $\mathbf{S}^{ij,kl}$ by approximating them implicitly as $\mathbf{S}^{ij,kl} \approx \mathbf{S}^{ij,ik}\mathbf{S}^{ik,kl}$. After the evaluation of the grid point dependent integrals and doubles amplitudes, the calculation of the (T) correction in the TNO basis proceeds as in the semi-canonical approximation, just that these steps have to be repeated for every grid point.

D. TNO integral generation

For the semi canonical as well as the Laplace-transformed triples correction, we need the integrals $(i\bar{a}_{ijk}|\bar{b}_{ijk}\bar{c}_{ijk})$ and $(i\bar{a}_{ijk}|kl)$. These are calculated in the DF approximation from the 3-index precursors $(Q_{ijk}|\bar{a}_{ijk}i)$, $(Q_{ijk}|\bar{b}_{ijk}\bar{c}_{ijk})$, and $(Q_{ijk}|kl)$. In $(Q_{ijk}|kl)$ the index l can be restricted to the subset with non-vanishing pairs il, kl, or jl. We use with a minor modification the local density fitting approach that is also used for PNO-MP2 and PNO-CCSD.[36] Since no appropriate triples energy estimate is available, the maximal CCSD doubles norm of the pairs ij, ik, jk is used to bias the selection of a triples specific auxiliary basis. A shell of auxiliary functions is selected if it fulfills the condition

$$\mathcal{Q}_{ijk} = \left\{ Q | \max\left(\mathcal{S}_{Qi}, \mathcal{S}_{Qj}, \mathcal{S}_{Qk}\right) \cdot f_{ijk} \geq T_{\mathrm{DF}} \right\}, \tag{48}$$

where $\mathcal{S}_{Qi}$ are the overlap criteria[36] between auxiliary basis functions and LMOs, T_{DF} the density fitting truncation threshold, and f_{ijk} the triple specific bias factor defined as

$$f_{ijk} = 2.5 \cdot 10^{-3} \cdot \sqrt{\max\left(||t_{ij}||, ||t_{ik}||, ||t_{jk}||\right)}. \tag{49}$$

The prefactor was determined by test calculations on large glycine chains and confirmed to not significantly affect the reaction energies presented in Section V C.

E. Triples corrections for the explicitly correlated variants

As in the canonical implementations, the explicitly correlated CCSD variants are combined with a triples correction which is computed without including geminal contributions. To correct partially for this approximation, we take up the (T*) correction proposed by Knizia *et al.*[52] which scales the so-computed triples correction by the ratio between the PNO-MP2-F12 and PNO-MP2 correlation energies. To avoid recomputation of the PNO-MP2 energy, it is exploited that the PNO-MP2-F12 energy can be separated into an explicitly correlated and a conventional part,

$$L_{\text{PNO-MP2-F12}} = \underbrace{\sum_{i\leq j} \left(2-\delta_{ij}\right) \sum_{\bar{a}\bar{b}} u^{ij}_{\bar{a}\bar{b}} g^{ij}_{\bar{a}\bar{b}}}_{E^{\mathrm{corr}}_{\text{PNO-MP2}}} + \underbrace{\sum_{i\leq j} \left(2-\delta_{ij}\right) \left(B^{ij}_{ij} - X^{ij}_{ij} \left(\varepsilon_i + \varepsilon_j\right) + 2V^{ij}_{ij} + u^{ij}_{\bar{a}\bar{b}} \mathcal{C}^{ij}_{\bar{a}\bar{b}} \right)}_{\Delta E^{\mathrm{corr}}_{\mathrm{F12}}}. \tag{50}$$

In contrast to the canonical case $E^{\mathrm{corr}}_{\text{PNO-MP2}}$ is in a PNO-MP2-F12 calculation smaller than in a not explicitly correlated PNO-MP2 calculation. This is due to the fact that in the F12 variant the virtual space is spanned by the more compact F12-PNOs. This leads to slightly larger scaling factors which implicitly corrects for the fact that the doubles amplitudes in the smaller F12-PNO basis are also used to compute the triples correction.

IV. COMPUTATIONAL DETAILS

All calculations were performed with a development version of the TURBOMOLE package.[53] The canonical CCSD[F12](T) and CCSD(F12*)(T) calculations were carried out with the ccsdf12[48] module, whereas the local development version of the pnoccsd program was used for the

PNO-CCSD[F12](T) and PNO-CCSD(F12*)(T) calculations. We used the correlation consistent cc-pVTZ and aug-cc-pVXZ basis set families[54–56] with X = D, T and the def2-XZVP basis sets with X = T, Q together with the optimized auxiliary basis sets (CABS, C, and JK basis sets) from Refs. 57–60. As correlation factor, we used in all calculations

$$f(r_{12}) = -\frac{1}{\gamma}\exp\left(-\gamma \cdot r_{12}\right), \tag{51}$$

where the exponential is approximated by a linear combination of six gaussians as described in Ref. 27 and the exponent γ was optimized for specific basis sets. In the current study a value of 1.40 was used for the different basis sets.

V. RESULTS AND DISCUSSION

A. The TNO truncation threshold

To test how to choose the threshold T_{TNO} as function of the PNO threshold T_{PNO}, calculations for a small test set of dimers were carried out, where we investigated the relative error in interaction energies and varied T_{PNO} and T_{TNO}. The test set includes the water-water, the water-peptide, $MeNH_2$-peptide, $MeNH_2$-water, and the AcOH-AcOH dimers from the S66 set.[61] The calculations were carried out in the aDZ and aTZ basis sets. Figure 3(a) shows the mean relative error as a function of the TNO truncation threshold for different values of T_{PNO}. The curves start to flatten out at $T_{\mathrm{TNO}} = T_{\mathrm{PNO}}$. After that point tighter TNO truncation thresholds no longer improve the accuracy. This observation is not unexpected since the doubles space that is used in the triples correction is restricted via T_{PNO}. The range of eigenvalues in the TNO eigenvalue problem is predetermined by the range of the PNO eigenvalues. Figure 3(b), which shows the mean relative error in the interaction energies as function of T_{PNO} for various values of T_{TNO}, demonstrates how T_{PNO} limits the accuracy. Based on these results, we decided to set T_{TNO} to the same value as T_{PNO} for the subsequent analysis of the accuracy and performance of PNO-CCSD(T) and PNO-CCSD(F12*)(T).

B. Accuracy and performance of the Laplace triples correction

The additional costs for the Laplace transformed PNO-CCSD(T) variant is directly proportional to the number of required Laplace grid points. Therefore it is important to study how many grid points are needed to ensure converged energy differences. For this purpose a test set of 11 reactions — listed in Table II — was chosen. This test set will also be used in further assessments.

In a first step, calculations with different convergence criteria for the numerical Laplace transformation were carried out in the def2-TZVP basis to study how reaction energies converge with the accuracy of the Laplace transformation. The grid points t_z and quadrature weights ω_z are determined by minimizing the Laplace error function

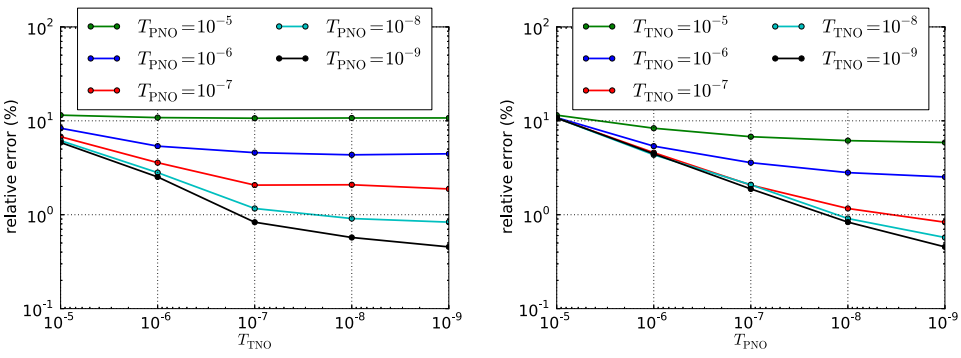
$$\mathrm{F}_{\mathrm{lap}} = \int_{x_{\min}}^{x_{\max}} dx \left(\frac{1}{x} - \sum_{z}^{n_z} \omega_z e^{-x\cdot t_z}\right)^2. \tag{52}$$

$\sqrt{\mathrm{F}_{\mathrm{lap}}}$ is the root mean square deviation for the error in the numerical integration.[17,62] $x_{\min}$ and $x_{\max}$ are set, respectively, to the smallest and largest orbital energy denominators and the number of grid points n_z is increased until the $\sqrt{\mathrm{F}_{\mathrm{lap}}}$ for the optimized t_z and ω_z decreased below the required threshold (*vide infra*).

As a reference to assess the accuracy of the numerical Laplace transformation for the (T) correction, we used calculations, where it is ensured that $\sqrt{\mathrm{F}_{\mathrm{lap}}} \leq 10^{-5}$. For the PNO calculations, a T_{PNO} and T_{TNO} threshold of 10^{-7} was chosen. Figure 4 shows the deviations relative to this reference point as a function of $\sqrt{\mathrm{F}_{\mathrm{lap}}}$. As comparison the lines for the (T0) approximation are plotted as horizontal lines. The errors systematically decrease with $\sqrt{\mathrm{F}_{\mathrm{lap}}}$. But differences in the slopes can be observed. On the first view the distribution seems to be quite broad, but this is due to the logarithmic scale. For $\sqrt{\mathrm{F}_{\mathrm{lap}}} \leq 10^{-2}$ the errors are smaller than 0.1 kJ/mol and thus smaller than the expected error due to the PNO and TNO truncations and the expected accuracy of the underlying method. At this point the number of grid points is usually in the range of 3–4.

C. Reaction and interaction energies

To investigate the effect of the difference between the (T0) and the Laplace-transformed (T) triples correction on the accuracy of reaction energies, we carried out calculations in the def2-TZVP and def2-QZVPP basis sets and compared them with the canonical CCSD(T) results. First we want to discuss the results in the def2-QZVPP basis, which are compiled in Table III. For analysis the mean error $\overline{\Delta}$ and the mean absolute



(a)Relative error vs. TNO truncation.

(b)Relative error vs. PNO truncation.

FIG. 3. Averaged relative error in the PNO-CCSD(T) interaction energies plotted against the PNO and TNO truncation thresholds. The errors are given with respect to the total interaction energy. The number of Laplace points was determined by requiring that the root mean square deviation in the numerical integration for the Laplace transformation $\sqrt{\mathrm{F}_{\mathrm{lap}}}$ (see Sec. V B) is below 10^{-2}.

TABLE II. Numbering of the studied reactions.

Number	Reaction
01	Neo-pentane → n-pentane
02	Ethandiole → dimethylperoxide
03	Benzene + $3H_2$ → cyclohexane
04	$3H_2 + N_2 \rightarrow 2NH_3$
05	Hexanoic acid → methylpivlate
06	Dihydrofuran-2(3H)-one → vinylacetate
07	Tetrahydropyran-2-one → acetylacetone
08	2 2,3-dimethyl-2-butene → octamethylcyclobutane
09	Pyrazole → imidazole
10	$CO + H_2O \rightarrow CO_2 + H_2$
11	Furane + NH_3 → pyrrole + H_2O

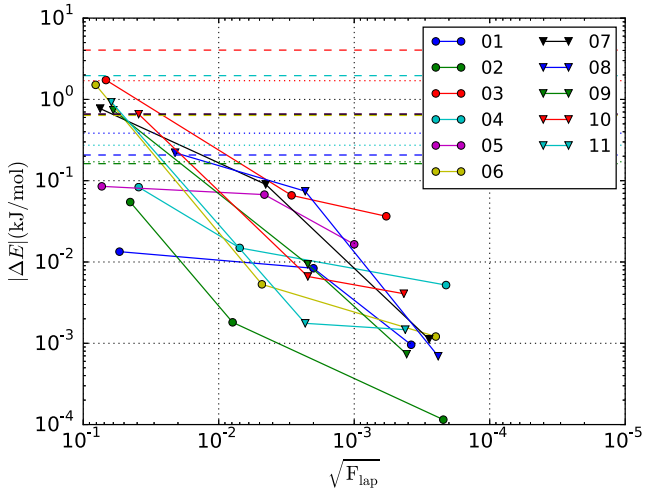


FIG. 4. Dependence of the absolute errors ΔE (kJ/mol) in reaction energies on the square root of the Laplace error function $\sqrt{F_{lap}}$. The errors of the T0 approximation are depicted as horizontal lines and share the same color as the line for the reaction. For reactions 1–7 dashed lines are used, while for reactions 8–11 dotted lines are used.

error $|\bar{\Delta}|$ are given. The absolute error $|\bar{\Delta}|$ for the whole test set is the most important measure for the accuracy. PNO-CCSD shows with 0.30 kJ/mol a quite small error; for PNO-CCSD(T) the error is with 0.80 kJ/mol substantially larger, but still in an acceptable range. One might assume that this indicates that the TNO space should be increased, but as discussed in Section V A, a decrease of T_{TNO} below T_{PNO} does not significantly improve the accuracy anymore since the TNO space is restricted by the PNO spaces of the contributing pairs. The mean absolute error of the PNO-CCSD(T0) results is about twice as large as for PNO-CCSD(T) but with 1.61 kJ/mol still small. Large deviations occur mainly if highly symmetric molecules are involved as it is the case for the reactions 03, 04, and 10. A reaction where PNO-CCSD(T0) as well as PNO-CCSD(T) have exceptional large errors is reaction 08. This is because in contrast to the educt the product, octamethylcyclobutan, is a very compact sterically overloaded molecule, so that a larger intramolecular basis set superposition error (BSSE) than in the educt occurs. The double and triple excitations which mainly cause the BSSE are the first ones which are skipped in the truncated PNO and TNO expansions. That the exceptional large deviations are related to intramolecular BSSE is corroborated by the semi-empirical geometrical counterpoise (gCP) correction developed by Kruse and Grimme.[63] The gCP correction is parameterized for different basis sets for HF and DFT. The largest basis for which gCP has been parameterized is def2-TZVP. At the HF/def2-TZVP level the gCP correction for reaction 08 is 5.5 kJ/mol. Of course it is only an empirical correction and it would be smaller in the def2-QZVPP basis. Nevertheless it shows the huge impact of intramolecular BSSE for this reaction. One could argue that the PNO-based calculations deliver for this kind of systems more accurate results, but this argumentation is dangerous since it is hard to accurately quantify the BSSE

TABLE III. Errors in PNO-CCSD, PNO-CCSD(T0), and PNO-CCSD(T) reaction energies with $T_{PNO} = T_{TNO} = 10^{-7}$ relative to canonical CCSD and CCSD(T) calculations for the def2-QZVPP basis. As comparison also the canonical reaction energies E^{reac} are given.

	PNO-CCSD	PNO-CCSD(T0)	PNO-CCSD(T)	E^{reac}_{CCSD}	$E^{reac}_{CCSD(T)}$
Reaction			kJ/mol		
01	−0.02	−0.31	−0.10	−13.33	−15.97
02	0.43	1.40	0.82	−273.15	−264.93
03	−0.31	4.03	0.33	−309.41	−291.14
04	0.23	2.16	0.20	−125.42	−118.12
05	−0.41	−1.60	−1.06	39.30	35.08
06	0.04	−0.39	0.10	59.42	57.86
07	0.01	−0.65	−0.05	28.69	27.73
08	−1.11	−4.03	−4.60	−69.45	−80.38
09	0.03	0.29	0.14	−47.81	−45.49
10	−0.42	−2.68	−1.08	−16.03	−25.68
11	0.17	0.31	0.30	−73.98	−74.42
$\bar{\Delta}$	−0.12	−0.13	−0.46		
$\lvert\bar{\Delta}\rvert$	0.29	1.62	0.80		
Δ_{max}	−1.11	−4.03	−4.60		
For $T_{PNO} = T_{TNO} = 10^{=8}$:					
08:	−0.57	−2.19	−2.71	−69.45	−80.38

TABLE IV. Deviations of PNO-CCSD, PNO-CCSD(T0), and PNO-CCSD(T) reaction energies with $T_{PNO} = T_{TNO} = 10^{-7}$ from untruncated CCSD, CCSD(T0), and CCSD(T) reference results. As comparison also the energy differences between CCSD(T) and CCSD(T0) are listed as Δ(T0) in the last column. All results are for the def2-TZVP basis.

	PNO-CCSD	PNO-CCSD(T0)	PNO-CCSD(T)	Δ(T0)
Reaction		kJ/mol		
01	0.07	−0.07	−0.04	−0.22
02	0.32	0.63	0.65	0.69
03	−0.45	−0.09	−0.10	3.58
04	0.08	0 .22	0.26	1.94
05	−0.31	−1.02	−0.97	−0.69
06	0.08	0 .17	0.15	−0.54
07	−0.13	−0.17	−0.33	−0.60
08	−1.37	−4.39	−4.41	0.28
09	0.00	0.05	0.05	0.16
10	−0.48	−1.00	−1.01	−1.77
11	0.05	0.09	0.09	0.20
$\bar{\Delta}$	−0.19	−0.51	−0.51	0.28
$\lvert\bar{\Delta}\rvert$	0.30	0.72	0.73	0.97
Δ_{max}	−1.37	−4.39	−4.41	3.58

and its relation to the PNO error. With $T_{PNO} = T_{TNO} = 10^{-8}$ the error reduces by roughly a factor of two. For this reaction the contribution of the triples correction is quite large and, although the absolute values of the errors are above the average for the test set, the relative errors in the triples correction are in the same ballpark as for the other reactions.

To investigate in more detail the relation between errors from the PNO and TNO truncation and the error due to the (T0) approximation, we repeated the calculations in def2-TZVP basis[64] for which we also computed CCSD(T0) reference results without PNO or TNO approximation. The results are listed in Table IV. If PNO-CCSD(T0) is compared with CCSD(T0) and PNO-CCSD(T) with CCSD(T), one gets essentially the same error statistics, which shows that the larger deviations between PNO-CCSD(T0) and CCSD(T) are due to error accumulation. The two errors have in the majority of cases the same sign. PNO-CCSD(T) and PNO-CCSD(T0) recover a different amount of the correlation energy (see Table V). With a threshold of $T_{PNO} = 10^{-7}$ PNO-CCSD recovers without correction for discarded pairs and PNOs on average 99.86% of the CCSD correlation energy in the def2-QZVPP basis, with a standard deviation of 0.03%. With the correction, which slightly overshoots, the PNO-CCSD correlation energy amounts on average to 100.02% of the canonical value. The correction systematically improves the results, which is also indicated by the smaller standard deviation Δ_{STD} of only 0.01%. PNO-CCSD(T0) recovers around 91.63% of the (T) contribution while PNO-CCSD(T) captures 96.77%, which is a significant improvement, but the relative errors are nevertheless larger than for the CCSD contribution. Furthermore the standard deviation of the errors is larger, which will lead to a poorer error compensation for energy differences. For the latter, however, the contributions to the errors in the total correlation energy are more important. From Table V it is evident that the error of the (T0) approximation dominates at this level the overall error and limits the accuracy, while the Laplace-transformed (T) correction reaches a similar accuracy as the uncorrected PNO-CCSD results.

So far the analysis only dealt with variants of PNO-CCSD(T) which did not use an intermediate PAO basis. Analogous calculations for the same test set in the def2-QZVPP[65] basis were carried out with the OSVs expanded in a preselected set of PAOs as described in Sec. II B 2. The results compiled in Table VI show that the errors are

TABLE V. Percentage of the CCSD correlation energy and the perturbative triples correction (T) recovered by the PNO methods for def2-QZVPP with $T_{PNO} = T_{TNO} = 10^{-7}$. E_{CCSD} is the PNO-CCSD result without and $E_{CCSD}^{corrected}$ the result with the perturbative correction for discarded PNOs and pairs (see Sec. II E). The columns (T0) and (T) list the results for the triples correction relative to the canonical (T) values as reference. The last three columns show the contributions to the relative error $\delta_Y^{PNO} = \frac{E_Y - E_{PNO\text{-}Y}}{E_{CCSD(T)}}$ in the final CCSD(T) correlation energy.

Molecule	E_{CCSD}	$E_{CCSD}^{corrected}$	(T0)	(T)	δ_{CCSD}^{PNO}	$\delta_{(T0)}^{PNO}$	$\delta_{(T)}^{PNO}$
Octamethylcyclobutane	99.80	100.01	91.68	95.25	0.19	0.29	0.14
Hexanoic acid	99.87	100.02	91.18	96.30	0.13	0.26	0.09
n-pentane	99.90	100.04	93.22	96.76	0.09	0.18	0.07
Dihydrofuran-2(3H)-one	99.85	100.01	90.41	96.31	0.15	0.30	0.10
Binylacetate	99.86	100.01	90.26	96.50	0.13	0.31	0.10
Tetrahydropyran-2-one	99.84	100.01	90.29	95.95	0.15	0.29	0.10
Acetylacetone	99.86	100.01	90.18	96.13	0.14	0.31	0.11
2,3-Dimethyl-2-butene	99.87	100.03	93.16	97.01	0.12	0.25	0.09
Pyrazole	99.84	100.01	92.05	97.76	0.15	0.40	0.12
Imidazole	99.85	100.01	92.05	97.74	0.15	0.39	0.11
Furane	99.86	100.02	92.27	97.78	0.14	0.36	0.10
Pyrrole	99.86	100.02	92.74	97.79	0.13	0.35	0.10
Ethandiole	99.91	100.03	92.10	96.97	0.08	0.20	0.06
Dimethylperoxide	99.89	100.01	91.28	96.61	0.11	0.24	0.07
$\bar{\Delta}$	99.86	100.02	91.63	96.77	0.13	0.29	0.10
Δ_{STD}	0.03	0.01	0.86	0.37	0.03	0.07	0.02

TABLE VI. Errors of the PAO-based versions of PNO-CCSD, PNO-CCSD(T0) and PNO-CCSD(T) in comparison to canonical CCSD and CCSD(T) with $T_{PNO} = T_{TNO} = 10^{-7}$ in the def2-QZVPP basis.

	PNO-CCSD	PNO-CCSD(T0)	PNO-CCSD(T)
Reaction		kJ/mol	
01	−0.07	−0.48	−0.32
02	0.43	1.45	0.88
03	−0.39	3.94	0.24
04	0.18	2.11	0.15
05	−0.49	−1.70	−1.09
06	0.06	−0.29	0.20
07	0.02	−0.51	0.28
08	−2.28	−4.50	−5.17
09	0.02	0.33	0.46
10	0.23	−2.00	−0.38
11	0.13	0.27	0.32
$\bar{\Delta}$	−0.20	−0.12	−0.40
$\lvert\bar{\Delta}\rvert$	0.39	1.60	0.86
Δ_{max}	−2.28	−4.50	−5.17

only slightly increased. The overall accuracy is not significantly affected by the restriction of the OSVs to preselected PAO domains. We conclude that the intermediate PAO basis can safely be used in production level calculations.

Finally, we studied the performance of the explicitly correlated variants for the same test set in the def2-QZVPP basis, but with reaction 08 excluded because the canonical reference calculations could not not be done within reasonable time for all CCSD(T) variants. As for the canonical methods the conventional triples correction is combined with an explicitly correlated CCSD calculation. In the canonical case it is known that the triples correction in this approach is often the limiting factor for the accuracy relative to the CCSD(T) basis set limit.[52] Due to the smaller magnitude of the doubles amplitudes in CCSD[F12] and CCSD(F12*) the triples correction is smaller as well and not exactly as accurate as in the conventional case. As discussed in Sec. III E this effect is enhanced in the PNO implementation. Therefore, we studied if the more compact F12-PNOs lead to a significant reduction of the accuracy for the connected triples correction. Table VII summarizes the results for the explicitly correlated PNO methods. As observed already in Ref. 40 the errors are somewhat larger than for the conventional PNO methods, if one compares for the latter with the result that includes the perturbative correction for discarded PNOs which is not available for the F12 methods. The mean average PNO and TNO truncation errors of the explicitly correlated PNO methods are close to those for the corresponding non-F12 PNO methods with this correction. Important is the comparison between the (T) corrections. There is no large decline in accuracy with F12-PNOs although the errors are slightly larger. Again it can be concluded that the combination with the conventional (T) correction can be done in a black box manner without further caution. As expected the errors for the scaled triples correction are slightly reduced compared to the unscaled variants since the scaling factors are larger than in the canonical case. But the effect is small.

D. Performance using an intermediate PAO basis

For the efficiency of the hybrid PAO-OSV-PNO approach, it is essential how many PAOs are on average required to obtain an accurate basis of OSVs. Table VIII lists the statistics for a series of polycyclic aromatic hydrocarbons in the cc-pVTZ and aug-cc-pVTZ basis sets. It is evident that there is a significant difference between the basis sets with and without augmentation with diffuse functions. In the cc-pVTZ basis the

TABLE VII. Errors for the explicitly correlated PNO-CCSD variants in comparison to their canonical counterparts with $T_{PNO} = T_{TNO} = 10^{-7}$ in the def2-QZVPP basis for the test set without reaction 08. The last three rows list for comparison the result for the non-F12 methods for the same set of methods—in parentheses without the correction for discarded PNOs which is not available for the F12 variants.

	PNO-CCSD							
	(F12*)	(F12*)(T0)	(F12*)(T)	[F12]	[F12](T0)	[F12](T)	(F12*)(T0*)	(F12*) (T*)
Reaction				kJ/mol				
01	−0.29	−0.83	−0.67	−0.27	−0.80	−0.64	−0.80	−0.63
02	0.76	2.30	1.79	0.51	1.99	1.47	2.23	1.68
03	−0.52	3.99	0.38	−0.77	3.64	−0.02	3.54	−0.37
04	−0.05	1.81	0.04	−0.06	1.78	−0.01	1.95	0.04
05	−1.04	−2.61	−2.12	−0.90	−2.45	−1.95	−2.49	−1.96
06	0.53	0.39	0.92	0.45	0.30	0.84	0.24	0.81
07	0.66	0.60	1.27	0.60	0.54	1.22	0.47	1.20
09	0.05	0.40	0.18	−0.03	0.31	0.08	0.30	0.06
10	−0.43	−1.86	−1.37	−0.56	−3.45	−1.96	−3.19	−1.56
11	0.19	0.40	0.26	0.21	0.43	0.29	0.49	0.32
$\bar{\Delta}$	−0.01	0.46	0.07	−0.08	0.23	−0.07	0.27	−0.04
$\lvert\bar{\Delta}\rvert$	0.41	1.38	0.82	0.40	1.43	0.77	1.43	0.78
Δ_{max}	1.04	3.99	2.12	0.90	3.64	1.96	3.54	1.96
Non-F12 methods								
$\bar{\Delta}$	−0.03 (−0.06)	0.26 (0.23)	−0.04 (−0.07)					
$\lvert\bar{\Delta}\rvert$	0.19 (0.45)	1.26 (1.36)	0.38 (0.66)					
Δ_{max}	0.43 (1.11)	4.03 (4.18)	1.08 (1.77)					

TABLE VIII. Average number of PAOs as intermediate basis for the OSVs. In parenthesis the percentage of the full AO space is given. The PAO threshold was chosen as $T_{\mathrm{PAO}} = 10^{-9}$.

	Benzene (%)	Naphthalene (%)	Anthracene (%)	Tetracene (%)	Pentacene (%)
aug-cc-pVTZ	391 (94)	585 (91)	745 (85)	796 (72)	800 (60)
cc-pVTZ	238 (90)	340 (83)	408 (73)	458 (65)	493 (58)

percentage of required PAOs declines faster. For both basis sets the asymptotic limit is not reached up to pentacene, but from the trend it is clear that the number of PAOs becomes constant at a certain system size and that a huge reduction is possible, which can be exploited in the integral transformations.

To study the effect of the PAO selection on the scaling of the computational costs, we carried out calculations for a series of glycine chains in the def2-TZVP basis with using CCSD(T), PNO-CCSD(T), and PAO-PNO-CCSD(T) and the (T0) variants of PNO methods. For PNO-CCSD(T) the number of Laplace points was 4 for Gly_1–Gly_4 and 5 for Gly_8. The wall timings are depicted in Fig. 5. Only for Gly_1 the canonical implementation is slightly faster. Already for Gly_2, the benefit of the PNO/TNO implementation is substantial. The curves for PAO-PNO-CCSD(T) and PNO-CCSD(T) are in the beginning very close, but start to diverge at around 500 basis functions. It is evident that the difference between both implementations increases with the system size, but up to 1000 basis functions it is currently not that significant. This conclusion might change if all parts of our code are adapted to an intermediate PAO basis. Currently the benefits of the PAO basis are only exploited for the computation of the connected triples correction. The comparison of the calculation with and without Laplace triples correction shows that the curves are only separated by an offset. The slopes are nearly identical.

Table IX lists the parameters for power laws of the form $10^A \times N_{bas}^B$ fitted to the data points for two different PNO and TNO thresholds. At first glance it surprises that the effective scaling for PNO-CCSD(T) decreases if the thresholds are tightened. This behaviour originates from the increased prefactor of the TNO integral constructions, which scales quadratically with the system size. For the variants with an intermediate PAO basis we observe the expected opposite behaviour which is that the exponent increases. A closer look at the power laws reveals that for canonical CCSD(T) the asymptotic scaling of $\mathcal{O}(\mathcal{N}^7)$ is not yet reached.

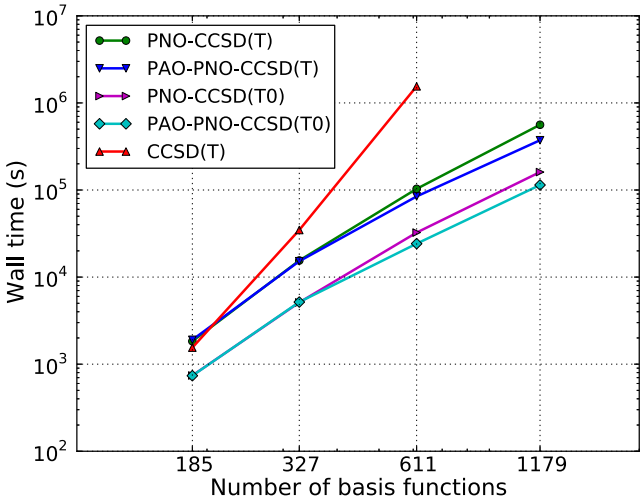


FIG. 5. Wall timings (in seconds) for glycine chains $(\mathrm{Gly})_n$ with $n = 1, 2, 4, 8$ (def2-TZVP basis) using canonical CCSD(T), PNO-CCSD(T), and PAO-PNO-CCSD(T) with a PNO- and TNO- threshold of $T_{\mathrm{TNO}} = T_{\mathrm{PNO}} = 10^{-7}$. For better comparison, the calculations were done without any parallelization.

E. Large scale example

As a large scale example we present results for a calix[4]arene-water complex (Fig. 6). Calix[n]arenes are an interesting application since they are used for molecular machines[67] and are models for ion pumps.[68] They have industrial as well as biological relevance, and theoretical methods can help to get a deeper insight into the underlying mechanisms. With PNO methods, it becomes possible to carry out accurate CCSD and CCSD(T) calculations on these large systems. For the calix[4]arene water complex, the binding energy of the water molecule is calculated as

$$E_{\mathrm{int}} = E^{\mathrm{C4A–H_2O}}_{\mathrm{C4A–H_2O}}(\mathrm{C4A–H_2O}) - E^{\mathrm{C4A}}_{\mathrm{C4A}}(\mathrm{C4A}) - E^{\mathrm{H_2O}}_{\mathrm{H_2O}}(\mathrm{H_2O}), \tag{53}$$

where the superscripts indicate the basis and the subscripts the structure. Since a large BSSE can be expected, counterpoise corrected energies were calculated as

$$E^{\mathrm{CP}}_{\mathrm{int}} = E_{\mathrm{int}} + \Delta E^{\mathrm{CP}}, \tag{54}$$

where ΔE^{CP} was obtained as

$$\begin{aligned}\Delta E^{\mathrm{CP}} = {} & E^{\mathrm{H_2O}}_{\mathrm{C4A–H_2O}}(\mathrm{H_2O}) - E^{\mathrm{C4A–H_2O}}_{\mathrm{C4A–H_2O}}(\mathrm{H_2O}) \\ & + E^{\mathrm{C4A}}_{\mathrm{C4A–H_2O}}(\mathrm{C4A–H_2O}) - E^{\mathrm{C4A–H_2O}}_{\mathrm{C4A–H_2O}}(\mathrm{C4A–H_2O}).\end{aligned} \tag{55}$$

This system was studied before by Hontama *et al.* in Ref. 66. They have done (non-F12) DF-MP2 calculations up to the aQZ basis, which allows a good comparison with our PNO-MP2-F12/aDZ results (Table X). We were able to do calculations from the PNO-MP2-F12 up to the PNO-CCSD(F12*)(T) level of theory in this basis. The PNO-MP2-F12/aDZ values agree quite well with the DF-MP2/aQZ values and show the expected enhancement in basis set convergence due to F12

TABLE IX. Fitted power laws for the wall time of the different PNO-CCSD(T) implementations and canonical CCSD(T) (def2-TZVP basis).

T_{PNO}	PAO-PNO-CCSD(T)	PNO-CCSD(T)	CCSD(T)
10^{-7}	$10^{-2.02} N_{bas}^{2.47}$	$10^{-2.82} N_{bas}^{2.79}$	$10^{-9.96} N_{bas}^{5.79}$
10^{-8}	$10^{-1.53} N_{bas}^{2.50}$	$10^{-1.98} N_{bas}^{2.67}$	$10^{-9.96} N_{bas}^{5.79}$

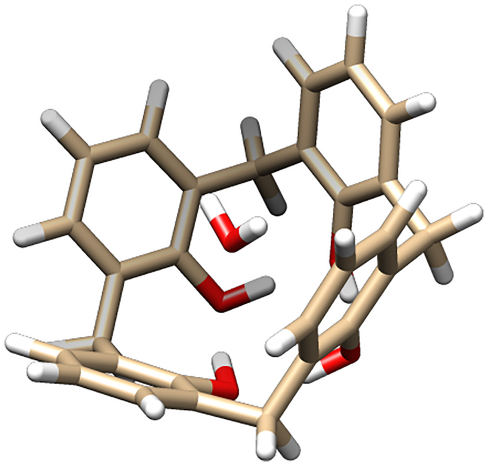


FIG. 6. The calix[4]arene water complex.

theory. For the other methods we expect a similar agreement with their canonical counterparts.

From the results we can conclude that the structure of the complex is only stable because of the dispersion interaction. On the HF level we see a repulsive interaction. All other methods show an attractive interaction, but the results vary in magnitude. As often observed, MP2 overestimates the dispersion interaction. A substantial deviation from the PNO-CCSD(F12*) and PNO-CCSD(F12*)(T) results can be observed. For this system we see a close agreement of the (T) and (T0) approximations. The deviation of 0.1 kcal/mol is insignificant. On the other hand, the contribution of the counterpoise correction is still significant, especially for the CCSD variants which include a triples correction. The conventional calculation of the triples correction leads to a larger BSSE than at the CCSD level. PNO-SCS-MP2-F12 agrees reasonably well with PNO-CCSD(F12*)(T) and could be used as a cost efficient alternative if PNO-CCSD(F12*)(T) calculations are not possible. The time determining calculation on the dimer took 34 days with OpenMP parallelization on 6 CPU of an Intel Xeon X5670 CPU with 2.93 GHz and 48 GiB main memory. The PNO-CCSD(F12*)(T0) calculation could be finished in around 16 days. It has to be noted that these timings refer to an older version of our code, where some rearrangements in the integral evaluation were not included, which lead to a further increase in efficiency. But already with this preliminary implementation PNO-CCSD(F12*)(T0) calculations on this kind of systems became a standard task, which can be finished in a bit more than two weeks.

TABLE X. Counterpoise corrected $E_{\text{int}}^{\text{CP}}$ and uncorrected E_{int} binding energies for the water molecule in the calix[4]arene water complex using $T_{\text{PNO}} = 10^{-7}$. The PNO calculations were done in the aDZ basis. The DF-MP2 results are from Ref. 66.

Method	$E_{\text{int}}^{\text{CP}}$ (kcal/mol)	E_{int} (kcal/mol)
HF	4.26	2.27
PNO-MP2-F12	−7.72	−8.50
PNO-SCS-MP2-F12	−4.95	−5.70
PNO-CCSD(F12*)	−4.13	−5.03
PNO-CCSD(F12*)(T0)	−5.16	−6.56
PNO-CCSD(F12*)(T)	−5.26	−6.65
PNO-CCSD(F12*)(T*)	−5.60	−7.02
DF-MP2/aTZ	−7.97	−10.29
DF-MP2/aQZ	...	−8.94

VI. CONCLUSION AND OUTLOOK

We presented an implementation of explicitly correlated PNO-based local coupled cluster singles and doubles with inclusion of a perturbative triples correction within the hybrid OSV-PNO scheme, which does not neglect the off-diagonal elements of the Fock matrix by using Laplace transformation techniques. While the virtual pair space is spanned by PNOs and in the explicitly correlated variants by F12-PNOs, the virtual space in the triples correction is spanned by TNOs, which are obtained by merging the PNO spaces. In contrast to the often used semi-canonical (T0) approximation, our method can in the limit of no PNO and TNO truncation converge to the canonical result. For routine application, 3–4 grid points for the numerical Laplace transformation are usually sufficient to ensure converged energy differences. Compared to the (T0) approximation, the Laplace transformed (T) approach reduces the mean absolute deviation (MAD) for reaction energies substantially. The required TNO integrals can be obtained from AO or PAO integrals. For the selection of the PAO domains we use a procedure, which determines the domains on the fly during the OSV construction. The use of PAO intermediates speeds up the integral transformation step. This effect starts to become significant if more than 800 basis functions are used, while the error due to the intermediate PAO basis is negligible. The combination of the conventional triples correction in the PNO/TNO framework can be safely combined with explicitly correlated PNO-CCSD models although in that case the more compact F12-PNOs are used. The effect of smaller F12-PNO space on the accuracy of the triples accuracy is negligible. With the PNO implementation, we could perform accurate calculations on a calix[4]arene water complex from the PNO-MP2-F12/aDZ level up to PNO-CCSD(F12*)(T*)/aDZ level of theory, which yielded a binding energy of around −5.60 kcal/mol. The calculation could be finished in a bit more than a month while it is out of reach for canonical methods. A comparison of the PNO-MP2-F12 results with conventional DF-MP2 results in the aQZ basis shows a good agreement, which can also be expected to hold for the higher correlated methods.

ACKNOWLEDGMENTS

The financial support by the Deutsche Forschungsgemeinschaft through Grant No. Ha 2588/7 is gratefully acknowledged.

[1] K. Raghavachari, G. W. Trucks, J. A. Pople, and M. Head-Gordon, Chem. Phys. Lett. **157**, 479 (1989).
[2] D. S. Lambrecht, B. Doser, and C. Ochsenfeld, J. Chem. Phys. **123**, 184102 (2005).
[3] H. Stoll, Phys. Rev. B: Condens. Matter Mater. Phys. **46**, 6700 (1992).
[4] K. Kristensen, I.-M. Høyvik, B. Jansik, P. Jørgensen, T. Kjærgaard, S. Reine, and J. Jakowski, Phys. Chem. Chem. Phys. **14**, 15706 (2012).
[5] J. J. Eriksen, P. Baudin, P. Ettenhuber, K. Kristensen, T. Kjærgaard, and P. Jørgensen, J. Chem. Theory Comput. **11**, 2984 (2015).

[6]S. Saebø and P. Pulay, Chem. Phys. Lett. **113**, 13 (1985).
[7]M. Schütz, G. Hetzer, and H.-J. Werner, J. Chem. Phys. **111**, 5691 (1999).
[8]M. Schütz and H.-J. Werner, J. Chem. Phys. **114**, 661 (2001).
[9]R. Ahlrichs, H. Lischka, V. Staemmler, and W. Kutzelnigg, J. Chem. Phys. **62**, 1225 (1975).
[10]R. Ahlrichs, F. Driessler, H. Lischka, V. Staemmler, and W. Kutzelnigg, J. Chem. Phys. **62**, 1235 (1975).
[11]R. Ahlrichs, F. Keil, H. Lischka, W. Kutzelnigg, and V. Staemmler, J. Chem. Phys. **63**, 455 (1975).
[12]F. Neese, F. Wennmohs, and A. Hansen, J. Chem. Phys. **130**, 114108 (2009).
[13]B. Helmich and C. Hättig, J. Chem. Phys. **135**, 214106 (2011).
[14]J. Yang, G. K.-L. Chan, F. R. Manby, M. Schütz, and H.-J. Werner, J. Chem. Phys. **136**, 144105 (2012).
[15]Y. Jung, R. C. Lochan, A. D. Dutoi, and M. Head-Gordon, J. Chem. Phys. **121**, 9793 (2004).
[16]Y. M. Rhee and M. Head-Gordon, J. Phys. Chem. A **111**, 5314 (2007).
[17]N. O. C. Winter and C. Hättig, J. Chem. Phys. **134**, 184101 (2011).
[18]E. G. Hohenstein, R. M. Parrish, and T. J. Martínez, J. Chem. Phys. **137**, 044103 (2012).
[19]E. A. Hylleraas, Z. Phys. **48**, 469 (1928).
[20]R. Bukowski, B. Jeziorski, and K. Szalewicz, "Gaussian geminals in coupled cluster and many-body perturbation theories," in *Explicitly Correlated Wave Functions in Chemistry and Physics*, Progress in Theoretical Chemistry and Physics, edited by J. Rychlewski (Springer, Netherlands, 2003), Vol. 13, pp. 185–248.
[21]W. Kutzelnigg, Theor. Chim. Acta **68**, 445 (1985).
[22]W. Klopper and W. Kutzelnigg, J. Phys. Chem. **94**, 5625 (1990).
[23]W. Klopper and C. C. M. Samson, J. Chem. Phys. **116**, 6397 (2002).
[24]E. F. Valeev, Chem. Phys. Lett. **395**, 190 (2004).
[25]S. Ten-no, Chem. Phys. Lett. **398**, 56 (2004).
[26]S. Ten-no, J. Chem. Phys. **121**, 117 (2004).
[27]D. P. Tew and W. Klopper, J. Chem. Phys. **123**, 074101 (2005).
[28]A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, and J. Olsen, Chem. Phys. Lett. **302**, 437 (1999).
[29]C. Hättig, W. Klopper, A. Köhn, and D. P. Tew, Chem. Rev. **112**, 4 (2012).
[30]P.-O. Löwdin and H. Shull, Phys. Rev. **101**, 1730 (1956).
[31]W. Meyer, Int. J. Quantum Chem. **5**, 341 (1971).
[32]A. Hansen, D. G. Liakos, and F. Neese, J. Chem. Phys. **135**, 214102 (2011).
[33]S. Saebø and P. Pulay, Annu. Rev. Phys. Chem. **44**, 213 (1993).
[34]C. Hättig, D. P. Tew, and B. Helmich, J. Chem. Phys. **136**, 204105 (2012).
[35]C. Krause and H.-J. Werner, Phys. Chem. Chem. Phys. **14**, 7591 (2012).
[36]G. Schmitz, B. Helmich, and C. Hättig, Mol. Phys. **111**, 2463 (2013).
[37]C. Riplinger and F. Neese, J. Chem. Phys. **138**, 034106 (2013).
[38]C. Riplinger, B. Sandhoefer, A. Hansen, and F. Neese, J. Chem. Phys. **139**, 134101 (2013).
[39]M. Schütz, J. Chem. Phys. **113**, 9986 (2000).
[40]G. Schmitz, C. Hättig, and D. P. Tew, Phys. Chem. Chem. Phys. **16**, 22167 (2014).
[41]M. Häser and J. Almlöf, J. Chem. Phys. **96**, 489 (1992).
[42]C. Riplinger, P. Pinski, U. Becker, E. F. Valeev, and F. Neese, J. Chem. Phys. **144**, 024109 (2016).
[43]H.-J. Werner, G. Knizia, C. Krause, M. Schwilk, and M. Dornbach, J. Chem. Theory Comput. **11**, 484 (2015).
[44]D. P. Tew and C. Hättig, Int. J. Quantum Chem. **113**, 224 (2013).
[45]S. Ten-no, J. Chem. Phys. **126**, 014108 (2007).
[46]H. Koch, A. Sánchez de Merás, T. Helgaker, and O. Christiansen, J. Chem. Phys. **104**, 4157 (1996).
[47]F. Neese, A. Hansen, and D. G. Liakos, J. Chem. Phys. **131**, 064103 (2009).
[48]C. Hättig, D. P. Tew, and A. Köhn, J. Chem. Phys. **132**, 231102 (2010).
[49]T. J. Lee and J. E. Rice, Chem. Phys. Lett. **150**, 406 (1988).
[50]J. Noga and P. Valiron, Mol. Phys. **103**, 2123 (2005).
[51]M. Schütz and H.-J. Werner, Chem. Phys. Lett. **318**, 370 (2000).
[52]G. Knizia, T. B. Adler, and H.-J. Werner, J. Chem. Phys. **130**, 054104 (2009).
[53]TURBOMOLE version 6.6 2014, a Development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989-2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.
[54]C. W. J. Bauschlicher, S. R. Langhoff, and P. R. Taylor, J. Chem. Phys. **88**, 2540 (1988).
[55]D. E. Woon and T. H. J. Dunning, J. Chem. Phys. **98**, 1358 (1993).
[56]T. H. J. Dunning, J. Chem. Phys. **90**, 1007 (1989).
[57]F. Weigend, A. Köhn, and C. Hättig, J. Chem. Phys. **116**, 3175 (2002).
[58]F. Weigend, M. Häser, H. Patzelt, and R. Ahlrichs, Chem. Phys. Lett. **294**, 143 (1998).
[59]K. E. Yousaf and K. A. Peterson, Chem. Phys. Lett. **476**, 303 (2009).
[60]F. Weigend, J. Comput. Chem. **29**, 167 (2008).
[61]J. Rezac, K. E. Riley, and P. Hobza, J. Chem. Theory Comput. **7**, 2427 (2011).
[62]N. O. Winter and C. Hättig, Chem. Phys. **401**, 217 (2012).
[63]H. Kruse and S. Grimme, J. Chem. Phys. **136**, 154101 (2012).
[64]A. Schäfer, C. Huber, and R. Ahlrichs, J. Chem. Phys. **100**, 5829 (1994).
[65]F. Weigend, F. Furche, and R. Ahlrichs, J. Chem. Phys. **119**, 12753 (2003).
[66]N. Hontama, Y. Inokuchi, T. Ebata, C. Dedonder-Lardeux, C. Jouvet, and S. S. Xantheas, J. Phys. Chem. A **114**, 2967 (2010).
[67]V. Balzani, A. Credi, F. M. Raymo, and J. F. Stoddart, Angew. Chem. **112**, 3484 (2000).
[68]J. de Mendoza, F. Cuevas, P. Prados, E. S. Meadows, and G. W. Gokel, Angew. Chem. **110**, 1650 (1998).