Article

# Extrapolation to the Limit of a Complete Pair Natural Orbital Space in Local Coupled-Cluster Calculations

Ahmet Altun, Frank Neese, and Giovanni Bistoni*

Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

**ABSTRACT:** The domain-based local pair natural orbital (PNO) coupled-cluster DLPNO-CCSD(T) method allows one to perform single point energy calculations for systems with hundreds of atoms while retaining essentially the accuracy of its canonical counterpart, with errors that are typically smaller than 1 kcal/mol for relative energies. Crucial to the accuracy and efficiency of the method is a proper definition of the virtual space in which the coupled-cluster equations are solved, which is spanned by a highly compact set of pair natural orbitals (PNOs) that are specific for each electron pair. The dimension of the PNO space is controlled by the $T_{CutPNO}$ threshold: only PNOs with an occupation number greater than $T_{CutPNO}$ are included in the correlation space of a given electron pair, whilst the remaining PNOs are discarded. To keep the error of the method small, a conservative $T_{CutPNO}$ value is used in standard DLPNO-CCSD(T) calculations. This often leads to unnecessarily large PNO spaces, which limits the efficiency of the method. Herein, we introduce a new computational strategy to approach the complete PNO space limit (for a given basis set) that consists in extrapolating the results obtained with different $T_{CutPNO}$ values. The method is validated on the GMTKN55 set using canonical CCSD(T) data as the reference. Our results demonstrate that a simple two-point extrapolation scheme can be used to significantly increase the efficiency and accuracy of DLPNO-CCSD(T) calculations, thus extending the range of applicability of the technique.
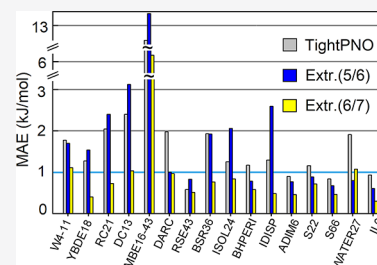
## 1. INTRODUCTION

The coupled-cluster method with singles, doubles, and perturbatively included triples excitations, i.e., CCSD(T),[1] can be used to compute relative energies with errors compared to experimental reference data that often fall within the experimental uncertainty.[2] Therefore, it is generally regarded as the "gold standard" of quantum chemistry. Unfortunately, its computational cost scales as the seventh power of the system size, and hence CCSD(T) calculations are only affordable for very small benchmark systems. To overcome this limitation, linear scaling local CCSD(T) variants that exploit the rapid decay of electron correlation with the interelectronic distance have been developed.[3−5]

In particular, the domain-based local pair natural orbital CCSD(T) method [DLPNO-CCSD(T)][5−13] has allowed for the calculation of single point energies for entire proteins.[14] To achieve such an efficiency, the coupled-cluster equations are solved in an extremely compact virtual space, which is tailored for each electron pair and spanned by a small set of pair natural orbitals (PNOs).[6] Only the PNOs with an occupation number greater than a threshold, denoted as $T_{CutPNO}$, are included in the correlation space of a given electron pair.

When TightPNO settings[5,13] are used (in this case, $T_{CutPNO}$ is set to $10^{-7}$ by default), DLPNO-CCSD(T) typically retains around 99.9% of the canonical CCSD(T) correlation energy. Thus, as shown on many benchmark data sets,[13,15−18] it provides essentially the same accuracy and reliability of the parent canonical method. On the 1505 reactions of the

GMTKN55 superset,[19] DLPNO-CCSD(T)/TightPNO results showed mean absolute errors (MAEs) compared to canonical CCSD(T) below 0.2 kcal/mol on 39 subsets, between 0.2 and 0.4 kcal/mol on 14 subsets and larger than 0.4 kcal/mol for two subsets (RC21 and MB16−43).[18] Even more accurate results could in principle be achieved by including more PNOs in the virtual space, i.e., by tightening the $T_{CutPNO}$ threshold to $10^{-8}$ or $10^{-9}$. Unfortunately, the computational cost of such calculations often approaches that of canonical CCSD(T).

Herein, we suggest an alternative computational strategy to reduce the PNO truncation error. In particular, DLPNO-CCSD(T) results obtained with different $T_{CutPNO}$ thresholds are extrapolated to the complete PNO space limit using a two-point extrapolation scheme. Two different approaches are thoroughly tested, namely $T_{CutPNO} = 10^{-5}/T_{CutPNO} = 10^{-6}$ (denoted hereafter as 5/6 extrapolation) and $T_{CutPNO} = 10^{-6}/T_{CutPNO} = 10^{-7}$ (denoted hereafter as 6/7 extrapolation). Our results demonstrate that the 5/6 extrapolation approaches the TightPNO accuracy at a fraction of its computational cost. Moreover, the 6/7 extrapolation significantly improves the

accuracy of TightPNO calculations, roughly halving the MAE with respect to canonical CCSD(T) for all the tested benchmark sets.

The paper is organized as follows. The PNO extrapolation scheme is described in Section 2. The accuracy of the method is tested on the GMTKN55 superset in Section 3.1. The basis set dependence of the extrapolated results and the efficiency of the extrapolation scheme are assessed in Section 3.2. The accuracy and efficiency of the proposed approach is further discussed in Section 3.3 on a particularly challenging application for standard computational methods, i.e., the quantification of dispersion-dominated interactions. Section 4 is devoted to the concluding remarks of this study.

## 2. BENCHMARK SETS, THEORETICAL ASPECTS, AND COMPUTATIONAL DETAILS

**2.1. Benchmark Sets.** The PNO extrapolation scheme is tested on GMTKN55 subsets for which DLPNO-CCSD(T)/TightPNO results showed an MAE of 0.2 kcal/mol or more with respect to canonical CCSD(T), i.e., on 16 subsets.[18] The subsets can be grouped in five different families:

(i) W4−11, YBDE18, RC21, and DC13: reaction energies for small systems;

(ii) MBE16−43, DARC, RSE43, BSR36, and ISOL24: reaction energies for large systems; it includes isomerization processes;

(iii) BHPERI: reaction barrier heights;

(iv) IDISP: intramolecular interactions;

(v) ADIM6, S22, S66, WATER27, and IL6: intermolecular noncovalent interactions.

Overall, the present study involves 883 molecules and 499 reactions.

**2.2. PNO Extrapolation.** To investigate the dependence of the DLPNO-CCSD(T) correlation energy on the dimension of the PNO space, we initially considered the benzene dimer, in which the two monomers are arranged in a parallel shifted configuration (the geometry was taken from reaction 24 of the S66 set). For this system, the dependence of the DLPNO-CCSD(T)/TightPNO/aug-cc-pVDZ-DK correlation energy on the $T_{\text{CutPNO}}$ threshold $10^{-X}$ ($X = 5, 6, 7, 8,$ and $9$) is shown in Figure 1.
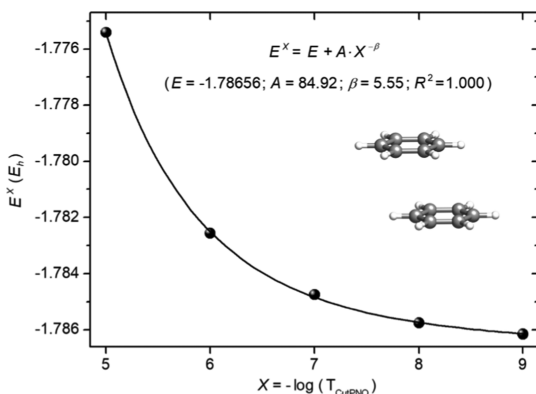


**Figure 1.** Dependence of the DLPNO-CCSD(T)/TightPNO/aug-cc-pVDZ-DK correlation energy $E^X$ on the exponent $X$ of $T_{\text{CutPNO}} = 10^{-X}$ for the benzene dimer. $E$ corresponds to the correlation energy at the complete PNO space limit.

Consistent with earlier PNO studies,[20,21] the correlation energy converges smoothly by tightening the $T_{\text{CutPNO}}$ threshold. The best fit is obtained with the following functional form

$$E^X = E + A \cdot X^{-\beta} \tag{1}$$

where $E^X$ is the correlation energy obtained with $T_{\text{CutPNO}} = 10^{-X}$ and $E$ is the target energy (for a given basis set) at the complete PNO space limit; $A$ and $\beta$ are constants ($A = 84.92$; $\beta = 5.55$). The smooth convergence of the correlation energy suggests that a simple two-point extrapolation procedure could be used to extrapolate the results obtained for small $X$ values to the complete PNO space limit.

By defining $Y = X + 1$ and $E^X$ and $E^Y$ as the correlation energies obtained with the corresponding $T_{\text{CutPNO}} = 10^{-X}$ and $T_{\text{CutPNO}} = 10^{-Y}$ values, one obtains the following expression for the correlation energy at the complete PNO space limit

$$E = \frac{Y^\beta \cdot E^Y - X^\beta \cdot E^X}{Y^\beta - X^\beta} \tag{2}$$

Using the following substitution

$$F = \frac{Y^\beta}{Y^\beta - X^\beta} \tag{3}$$

Equation 2 can be written in the compact form

$$E = E^X + F \cdot (E^Y - E^X) \tag{4}$$

Equation 4 is the basis of the PNO extrapolation scheme proposed in this work. Hence, all that is required to perform the extrapolation is the calculation of $E^X$ and $E^Y$, together with the knowledge of the single parameter $F$. Importantly, $E^X$ and $E^Y$ must be obtained from DLPNO-CCSD(T) calculations that use exactly the same computational settings (e.g., the same basis set and DLPNO settings) with the only exception being the $T_{\text{CutPNO}}$ threshold.

For the benchmark sets investigated in this work, the optimal $F$ value, i.e., the one that minimizes the error with respect to canonical CCSD(T), deviates within the $1.5 \pm 0.2$ range. It is important to note that the dependency of the MAE on the $F$ value is very small within this range. Hence, for the sake of simplicity, the recommended $F$ value is 1.5 for both 5/6 and 6/7 extrapolations.

Note that eq 4 does not make any strong assumption on the functional form that better describes the convergence of the correlation energy with respect to the $T_{\text{CutPNO}}$ threshold. For example, an exponential functional form of the type

$$E^X = E + A \cdot e^{-\alpha \sqrt{X}} \tag{5}$$

would provide the following expression for the two-point extrapolated energy

$$E = \frac{e^{\alpha \sqrt{Y}} \cdot E^Y - e^{\alpha \sqrt{X}} \cdot E^X}{e^{\alpha \sqrt{Y}} - e^{\alpha \sqrt{X}}} \tag{6}$$

However, this is equivalent to eq 4 for an $F$ value of

$$F = \frac{e^{\alpha \sqrt{Y}}}{e^{\alpha \sqrt{Y}} - e^{\alpha \sqrt{X}}} \tag{7}$$

It is worth mentioning here that eq 4 (as well as eqs 2 and 6) is analogous to that commonly used in two-point extrapolation schemes to the complete basis set (CBS) limit.[22−25] In these schemes, $E$ is the CBS extrapolated energy and $E^X$ and $E^Y$ are the energies obtained with two basis sets of consecutive $X$ and

$Y$ cardinality. In the extrapolation of HF energies,[22] it is usually assumed that $E^X - E$ is proportional to $e^{-\alpha(X)\gamma}$, where $\gamma$ is typically 1 or 0.5, and $\alpha$ is a constant obtained from fits to accurate benchmark energies. Correlation energies are usually extrapolated by considering that $E^X - E$ is proportional to $X^{-\beta}$ (with $\beta \approx 3$).[23] Importantly, when the $\beta$ value (or equivalently $F$) is obtained from a root mean square (RMS) fit to benchmark data, the extrapolated energies were shown to improve notably.[24,25]

**2.3. Efficiency Considerations.** In the present work, $E^X$ and $E^Y$ were obtained from separate DLPNO-CCSD(T)/TightPNO calculations using $T_{\mathrm{CutPNO}} = 10^{-X}$ and $T_{\mathrm{CutPNO}} = 10^{-Y}$, respectively ($Y = X + 1$). As $T_{\mathrm{CutPNO}} = 10^{-X}$ settings are significantly more efficient than $T_{\mathrm{CutPNO}} = 10^{-Y}$ settings, the overall $X/Y$ extrapolation procedure is only slightly more expensive than the corresponding $T_{\mathrm{CutPNO}} = 10^{-Y}$ calculation. For example, the computational cost of $T_{\mathrm{CutPNO}} = 10^{-6}$ calculations is typically 30–40% that of $T_{\mathrm{CutPNO}} = 10^{-7}$ calculations, as shown in Sections 3.2 and 3.3. As it will be demonstrated numerically below for a large data set, $X/Y$ extrapolation calculations are typically 2 times faster than the corresponding $T_{\mathrm{CutPNO}} = 10^{-(Y+1)}$ calculations. This indicates that the PNO extrapolation scheme provides a cost-effective alternative to tightening the $T_{\mathrm{CutPNO}}$ threshold.

It is worth mentioning here that a different computational strategy was also tested as an attempt to further increase the efficiency of the extrapolation approach proposed here. As the $T_{\mathrm{CutPNO}} = 10^{-X}$ virtual space is a subset of the $T_{\mathrm{CutPNO}} = 10^{-Y}$ virtual space if $X < Y$, an approximate estimate for the $T_{\mathrm{CutPNO}} = 10^{-X}$ energy can be obtained from the $T_{\mathrm{CutPNO}} = 10^{-Y}$ amplitudes. This could potentially save computer time by avoiding redundant operations. From a technical point of view, this was achieved by exploiting the fact that the correlation energy can always be written as a sum of "pair correlation energies" $\varepsilon_{ij}$, i.e., contributions from pairs of localized occupied orbitals $ij$. Within the PNO approximation, the strong-pair[11] component of $\varepsilon_{ij}$ reads

$$\varepsilon_{ij} = \sum_{a_{ij}b_{ij}} (ia_{ij}|jb_{ij})\tilde{\tau}^{ij}_{a_{ij}b_{ij}} \tag{8}$$

where $a_{ij}$ and $b_{ij}$ are the PNOs belonging to the $ij$ pair, $\tilde{\tau}^{ij}_{ab}$ are the contravariant amplitudes

$$\tilde{\tau}^{ij}_{a_{ij}b_{ij}} = \frac{4\tau^{ij}_{a_{ij}b_{ij}} - 2\tau^{ji}_{a_{ij}b_{ij}}}{1 + \delta_{ij}} \tag{9}$$

and $\tau^{ij}_{a_{ij}b_{ij}} = t^i_{a_{ij}} t^j_{b_{ij}} + t^{ij}_{a_{ij}b_{ij}}$ are the cluster amplitudes.

For each $ij$ pair, upon entering the PNO generation for the $10^{-Y}$ calculation, the PNOs for a threshold of $10^{-X}$ were also automatically generated. This provides us with two sets of perturbative weak-pair and PNO incompleteness corrections, namely $\Delta E_{\mathrm{wp}} (10^{-Y})$ and $\Delta E_{\mathrm{wp}} (10^{-X})$. The overlap matrix between the PNOs in the $T_{\mathrm{CutPNO}} = 10^{-Y}$ and $T_{\mathrm{CutPNO}} = 10^{-X}$ spaces $S^{ij}$ was also computed and stored.

Then, the solution of the DLPNO-CCSD equations for $T_{\mathrm{CutPNO}} = 10^{-Y}$ provided us with the corresponding $\tau^{ij}_{a_{ij}b_{ij}} (10^{-Y})$ amplitudes and hence with pair correlation energies (eq 8). The overall DLPNO-CCSD correlation energy associated with $T_{\mathrm{CutPNO}} = 10^{-Y}$ was computed by adding $\Delta E_{\mathrm{wp}} (10^{-Y})$ to the correlation energy from the strong pairs. In the next step, $S^{ij}$ was used to project the $\tau^{ij}_{a_{ij}b_{ij}} (10^{-Y})$ amplitudes from the $T_{\mathrm{CutPNO}} = 10^{-Y}$ to the $T_{\mathrm{CutPNO}} = 10^{-X}$ space

$$\boldsymbol{\tau}^{ij}(10^{-X}) = \boldsymbol{S}^{ij+} \boldsymbol{\tau}^{ij}(10^{-Y})\boldsymbol{S}^{ij} \tag{10}$$

This projection is necessary since the PNOs for a given pair are recanonicalized before entering the cluster amplitude iterations. The projected amplitudes were transformed back to the original $T_{\mathrm{CutPNO}} = 10^{-Y}$ space and used to estimate the strong-pair component of the correlation energy associated with $T_{\mathrm{CutPNO}} = 10^{-X}$. $\Delta E_{\mathrm{wp}} (10^{-X})$ was added to this energy to compute the overall DLPNO-CCSD correlation energy in the $T_{\mathrm{CutPNO}} = 10^{-X}$ space. Note that in this approach we generate exactly the same PNOs as in a separate $10^{-X}$ calculation. However, the resulting amplitudes are not identical to the "genuine" $10^{-X}$ amplitudes, since the relaxation of the amplitudes after the projection is neglected. The triples correction contribution can be computed as usual[14] starting from converged doubles amplitudes for $T_{\mathrm{CutPNO}} = 10^{-Y}$ and from projected doubles amplitudes for $T_{\mathrm{CutPNO}} = 10^{-X}$.

This procedure was implemented in ORCA and tested on the S22 and WATER27 sets. Although this scheme allows us to skip the DLPNO-CCSD part for $T_{\mathrm{CutPNO}} = 10^{-X}$ calculations, which accounts for ~25% of the overall computational time, it also introduces a non-negligible error in the $T_{\mathrm{CutPNO}} = 10^{-X}$ amplitudes that deteriorates the quality of the extrapolated results (see AutoExtr-S22 and AutoExtr-WATER27 sheets of the Supporting Information). Given the need to reconverge the amplitudes and the fact that the triples amplitudes must be iterated too with the reconverged doubles amplitudes, the computational savings that can be realized over two separate calculations are very modest. This is particularly true, since the tighter threshold calculation needs to be performed in either scheme and dominates the overall computation time. Hence, for all intent and purposes, the extrapolation procedure discussed in Section 2.2 is recommended. Its accuracy and efficiency are discussed in the following section.

**2.4. Computational Details.** All calculations were performed with a development version of the ORCA program package based on version 4.2.[26−28] CCSD(T) and DLPNO-CCSD(T) correlation energies were calculated with the default frozen core settings in ORCA.[29] For open-shell molecules, the reference energy was obtained at the quasi-restricted orbital (QRO) level.[30] The perturbative triples contributions were calculated using the recently published iterative $(T_1)$ algorithm for both closed-shell[31] and open-shell[32] systems. Since here we use the improved $(T_1)$ algorithm for the open-shell system,[32] the present open-shell DLPNO-CCSD(T) results are not exactly the same as the previously reported ones.[18] The variations are, however, very small. Integral evaluations in DLPNO-based correlation energy calculations need an auxiliary basis set. This was generated using the automated auxiliary basis set construction module of ORCA (the so-called "autoaux") with the maximum possible angular momentum.[33] Canonical CCSD(T) calculations were performed without any resolution of identity (RI) approximation. In all cases, DLPNO-CCSD(T) calculations were carried out using TightPNO settings (e.g., the $T_{\mathrm{CutPairs}}$ threshold was set to $10^{-5}$) in conjunction with different $T_{\mathrm{CutPNO}}$ thresholds.

Unless stated otherwise, both canonical and DLPNO-CCSD(T) calculations were performed with the relativistic second-order Douglas–Kroll–Hess Hamiltonian (DKH2)[34,35] in conjunction with the appropriate aug-cc-pVDZ-DK basis set,[36−38] consistent with the previous benchmark study on the GMTKN55 superset.[18] Although correlation energies obtained with double-$\zeta$ quality basis sets might suffer from the severe

basis set incompleteness error, the comparison of DLPNO-CCSD(T) and CCSD(T) results obtained with the same basis set makes it possible to assess the error inherent to the DLPNO approximation.[18]

For the W4−11,[39] YBDE18,[40] and S66[41] sets, DLPNO-CCSD(T) and CCSD(T) results were also compared for larger basis sets, as discussed in Section 3.2. For the W4−11 set, the aug-cc-pVTZ-DK and aug-cc-pVQZ-DK basis sets[36−38] were used with the same settings described above. For the YBDE18 and S66 sets, canonical CCSD(T) calculations are not possible with quadruple-$\zeta$ quality basis sets. Therefore, previously published[40,42] explicitly correlated CCSD(T) energies were used as reference data in these cases. DLPNO-CCSD(T) interaction energies were computed using aug-cc-pVTZ and aug-cc-pVQZ basis sets. In the correlation part, the auxiliary basis sets constructed with the autoaux module (see above) were used for the YBDE18 set, while the matching aug-cc-pVTZ/C and aug-cc-pVQZ/C auxiliary basis sets were used for the S66 set. Consistent with the reference data,[42] the interaction energies of the S66 set were corrected for the BSSE. CBS(3/4) extrapolation of the correlation energies was carried out as described previously.[43] Hence, eq 2 was used with $\beta = 3.05$ (or, equivalently, eq 4 with $F = 1.712$).

## 3. RESULTS AND DISCUSSION

In this section, the accuracy and efficiency of the PNO extrapolation (eq 4 with $F = 1.5$) is discussed on the GMTKN55 superset. As shown in the "SUMMARY" sheet of the Supporting Information, PNO extrapolation improves the accuracy of both the CCSD and the (T) components of the correlation energy. For the sake of simplicity, only the accuracy of the extrapolated DLPNO-CCSD(T) energies is discussed in the following.

**3.1. Accuracy of the PNO Extrapolation.** Figure 2 shows the MAEs for the computed DLPNO-CCSD(T)/TightPNO/
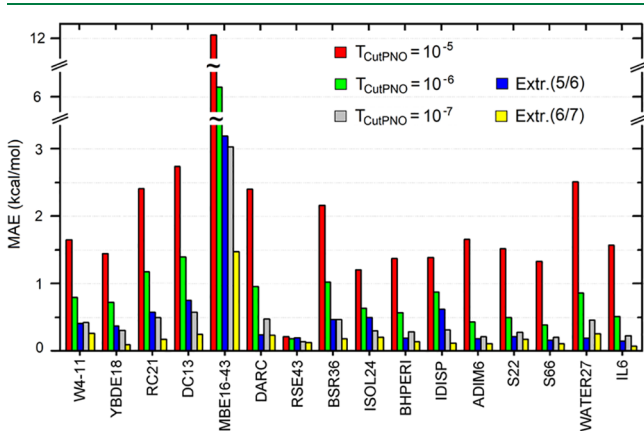


**Figure 2.** MAEs associated with DLPNO-CCSD(T)/TightPNO/aug-cc-pVDZ-DK reaction energies for the GMTKN55 subsets with $T_{CutPNO} = 10^{-X}$ ($X = 5$, 6, and 7) and with the 5/6 and 6/7 extrapolations relative to the canonical CCSD(T)/aug-cc-pVDZ-DK results.

aug-cc-pVDZ-DK reaction energies using different $T_{CutPNO}$ thresholds and PNO extrapolation schemes. In particular, $T_{CutPNO} = 10^{-X}$ ($X = 5$, 6, and 7) as well as the 5/6 and 6/7 extrapolations were tested. Canonical CCSD(T)/aug-cc-pVDZ-DK results were used as the reference, as described in Section 2.2.

As expected, $T_{CutPNO} = 10^{-5}$ and $10^{-6}$ calculations (red and green bars in Figure 2, respectively) are extremely efficient but yield large errors. Note that $T_{CutPNO} = 10^{-6}$ is the default for the so-called "LoosePNO" settings, which are only recommended for preliminary DLPNO-CCSD(T) calculations, whilst $T_{CutPNO} = 10^{-5}$ is never recommended for standard DLPNO-CCSD(T) calculations. The MAEs obtained with $T_{CutPNO} = 10^{-5}$ and $10^{-6}$ are generally between 1.5 and 3 kcal/mol, with the error reaching up to 12 kcal/mol ($T_{CutPNO} = 10^{-5}$) and 6 kcal/mol ($T_{CutPNO} = 10^{-6}$) for the MBE16−43 set of artificial molecules.

In contrast, the 5/6 extrapolation (blue bars in Figure 2) provides MAEs within chemical accuracy, i.e., less than 1 kcal/mol, for all benchmark sets. The only exception is the challenging MBE16−43 set, for which an MAE of about 3 kcal/mol was observed. These figures are similar to that obtained for the more expensive $T_{CutPNO} = 10^{-7}$ calculations (gray bars in Figure 2). As mentioned above, $T_{CutPNO} = 10^{-7}$ is the default for TightPNO calculations and it is often recommended for the study of noncovalent interactions.[44] For noncovalent interactions (the last five subsets in Figure 2), the 5/6 extrapolation provides results that are more accurate than those from $T_{CutPNO} = 10^{-7}$ calculations when the aug-cpVDZ-DK basis set is used. However, in general, both approaches provide similar accuracy. Therefore, the 5/6 extrapolation should be considered as a cost-effective—but still reasonably accurate—alternative to TightPNO calculations on large molecular systems.

Excluding the challenging MB16−43 subset, standard TightPNO calculations give MAEs less than 0.6 kcal/mol, whilst 5/6 extrapolated reaction energies show MAEs less than 0.8 kcal/mol. For these subsets, the 6/7 extrapolation (yellow bars in Figure 2) is significantly more accurate, providing a near sub-kJ/mol accuracy for all subsets (with MAEs of less than 0.27 kcal/mol). For example, on the S66 set for noncovalent interactions, the 6/7 extrapolation (MAE = 0.11 kcal/mol) reduces the TightPNO error (MAE = 0.20 kcal/mol) by a factor of two. Interestingly, despite the significant increase in the computational cost when tightening $T_{CutPNO}$ to $10^{-8}$ and $10^{-9}$, the associated 7/8 (MAE = 0.08 kcal/mol) or 8/9 (MAE = 0.06 kcal/mol) extrapolations bring only a slight improvement to the 6/7 extrapolated results (see the S66 sheet of the Supporting Information). Hence, the 6/7 extrapolation already provides almost converged results with respect to the $T_{CutPNO}$ threshold, and the residual error is likely to be associated with the other approximations used in DLPNO-CCSD(T) calculations.

Finally, it is worth mentioning that the relatively large error in the reaction energies of the MBE16−43 set does not change significantly if all the electron pairs are included in the coupled-cluster treatment ($T_{CutPairs} = 0$) (see the MBE16−43 sheet of the Supporting Information). Hence, the main source of error in this case is the PNO truncation error, as demonstrated by the excellent performances of the extrapolation schemes.

**3.2. Basis Set Dependence and Efficiency.** In this section, the basis set dependence and the efficiency of the PNO extrapolation scheme have been investigated on the W4−11, YBDE18, and S66 subsets.

For the W4−11 set (Figure 3a), reaction energies computed using the 5/6 extrapolation are more accurate than those obtained using $T_{CutPNO} = 10^{-6}$ settings for all basis sets. Analogously, the 6/7 extrapolation results are generally more accurate than $T_{CutPNO} = 10^{-7}$ results. Importantly, the accuracy

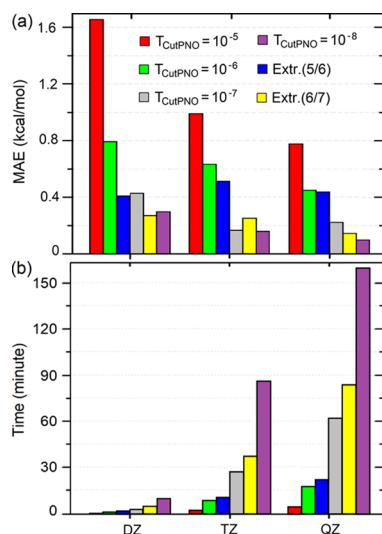**Figure 3.** (a) MAE of the DLPNO-CCSD(T)/TightPNO/aug-cc-pV$n$Z-DK ($n$ = D, T, and Q) reaction energies for the W4−11 set with $T_{CutPNO} = 10^{-X}$ ($X$ = 5, 6, 7, and 8) and with the 5/6 and 6/7 extrapolations relative to the corresponding canonical CCSD(T)/aug-cc-pV$n$Z-DK energies. (b) Total computational time of the correlation interaction energy over all molecules in the set by using 4 cores from a single cluster node equipped with 4 Intel Xeon CPUs. The timing associated with the $X/Y$ extrapolation is the sum of that of $T_{CutPNO} = 10^{-X}$ and $T_{CutPNO} = 10^{-Y}$ ($Y = X + 1$) calculations.



**Figure 4.** (a) MAE of the DLPNO-CCSD(T)/TightPNO/CBS correlation dissociation energies for the YBDE18 set with $T_{CutPNO} = 10^{-X}$ ($X$ = 5, 6, 7, and 8) and with the 5/6 and 6/7 extrapolations relative to the explicitly correlated CCSD(T)-F12a/VTZ-F12 reference energies.[40] (b) Average computational time of the correlation interaction energy per molecule in the set by using 4 cores from a single cluster node equipped with 4 Intel Xeon CPUs. The timing associated with the $X/Y$ extrapolation is the sum of that of $T_{CutPNO} = 10^{-X}$ and $T_{CutPNO} = 10^{-Y}$ ($Y = X + 1$) calculations.
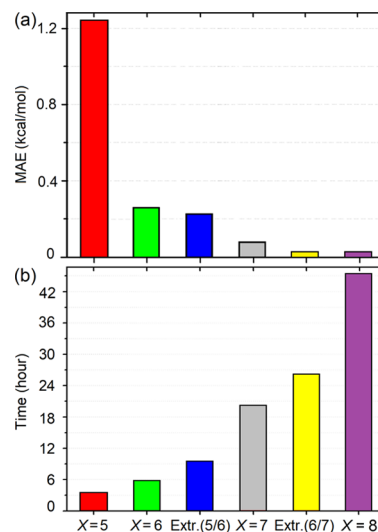


**Figure 5.** (a) MAE of the BSSE-corrected DLPNO-CCSD(T)/TightPNO/CBS correlation interaction energies for the S66 set with $T_{CutPNO} = 10^{-X}$ ($X$ = 5, 6, 7, and 8) and with the 5/6 and 6/7 extrapolations relative to the correlation part of the composite explicitly correlated BSSE-corrected MP2/CCSD(T) reference energies with the SILVER[42] settings. (b) The computational time of the correlation interaction energy per reaction by using 16 cores from a single cluster node equipped with 4 Intel Xeon CPUs. The timing associated with the $X/Y$ extrapolation is the sum of that of $T_{CutPNO} = 10^{-X}$ and $T_{CutPNO} = 10^{-Y}$ ($Y = X + 1$) calculations.

of the extrapolated results is largely independent of the basis set size. In fact, the MAEs associated with the 5/6 and 6/7 extrapolated results are approximately constant for all basis sets. In contrast, the MAE obtained with $T_{CutPNO} = 10^{-7}$ oscillates by increasing the basis set size, reaching a minimum with the triple-$\zeta$ basis set and increasing again with the quadruple-$\zeta$ basis set.

For the YBDE18 set (Figure 4a), the 5/6 extrapolation provides an accuracy between that of $T_{CutPNO} = 10^{-6}$ and of $T_{CutPNO} = 10^{-7}$. The 6/7 extrapolation provides results that are very close to those obtained using $T_{CutPNO} = 10^{-8}$. However, since the results obtained with $T_{CutPNO} = 10^{-7}$ and $10^{-8}$ are also very close to each other, the 6/7 extrapolation reduces the MAE with respect to $T_{CutPNO} = 10^{-7}$ by only 0.01 kcal/mol.

For the S66 set (Figure 5a), the 5/6 extrapolation (MAE = 0.23 kcal/mol) slightly improves $T_{CutPNO} = 10^{-6}$ results (MAE: 0.26 kcal/mol), whilst the 6/7 extrapolation (MAE = 0.03 kcal/mol) is significantly more accurate than $T_{CutPNO} = 10^{-7}$ results (MAE = 0.09 kcal/mol), providing results of essentially $T_{CutPNO} = 10^{-8}$ accuracy (MAE = 0.03 kcal/mol).

The timing associated with the various DLPNO-CCSD(T) calculations for the W4−11, YBDE18, and S66 subsets is shown in Figures 3b, 4b, and 5b, respectively. In average, the computational cost of the 5/6 extrapolation is only 46% higher than that of $T_{CutPNO} = 10^{-6}$ calculations. Compared to $T_{CutPNO} = 10^{-7}$ calculations, the 5/6 extrapolation is faster by a factor of 2.2. Analogously, the computational cost of the 6/7 extrapolation is only 38% higher than that of $T_{CutPNO} = 10^{-7}$ calculations. Again, the 6/7 extrapolation calculations are 2.0 times faster than those of $T_{CutPNO} = 10^{-8}$. In summary, $X/Y$ extrapolation calculations are about twice more efficient than the corresponding $10^{-(Y+1)}$ calculations.

These computational experiments demonstrate that the suggested PNO extrapolation scheme is valid irrespective of
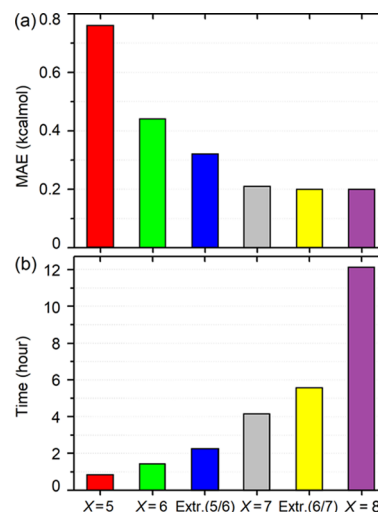
the basis set size and provides a cost-effective but still very accurate alternative to tightening the $T_{CutPNO}$ threshold.

**3.3. Dispersion-Dominated Interactions.** Herein, we discuss the accuracy and efficiency of the extrapolation scheme on three representative examples of dispersion-dominated interactions, i.e., the stacking interaction in the uracil dimer (reaction 26 of the S66 set, see Figure 6), the interaction of
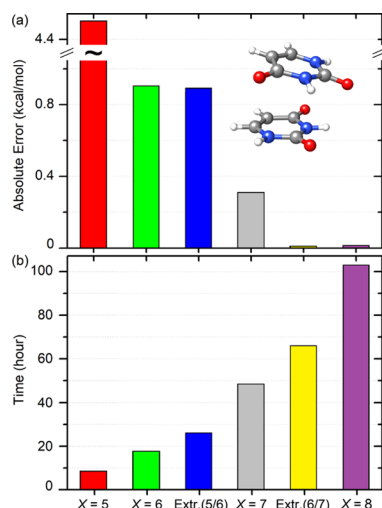
**Figure 6.** (a) Error in the DLPNO-CCSD(T)/TightPNO/CBS correlation interaction energy of the uracil dimer with $T_{CutPNO} = 10^{-X}$ ($X$ = 5, 6, 7, and 8) and with the 5/6 and 6/7 extrapolations relative to the correlation part of the composite explicitly correlated BSSE-corrected MP2/CCSD(T) reference energies with the SILVER[42] settings. (b) The computational time of the correlation part of the interaction energy with the corresponding computational settings by using 16 cores from a single cluster node equipped with 4 Intel Xeon CPUs. Computational time for the extrapolations is obtained as the sum of the time required with the involving $T_{CutPNO}$ settings and basis sets.

two pentane molecules (reaction 34 of the S66 set, see Figure 7), and the bowl to cage isomerization reaction of $C_{20}$ (reaction 2 of the DC13 set, see Figure 8).
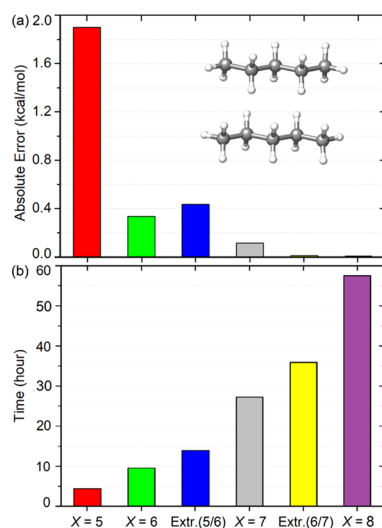


**Figure 7.** (a) Error in the DLPNO-CCSD(T)/TightPNO/CBS correlation interaction energy of the pentane dimer with $T_{CutPNO} = 10^{-X}$ ($X$ = 5, 6, 7, and 8) and with the 5/6 and 6/7 extrapolations relative to the correlation part of the composite explicitly correlated BSSE-corrected MP2/CCSD(T) reference energies with the SILVER[42] settings. (b) The computational time of the correlation part of the interaction energy with the corresponding computational settings by using 16 cores from a single cluster node equipped with 4 Intel Xeon CPUs. Computational time for the extrapolations is obtained as the sum of the time required with the involving $T_{CutPNO}$ settings and basis sets.
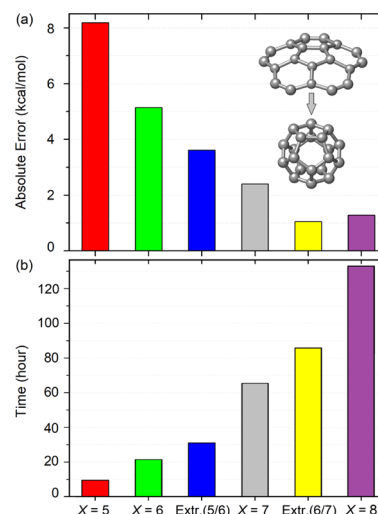


**Figure 8.** (a) Error in the DLPNO-CCSD(T)/TightPNO/aug-cc-pVDZ-DK bowl to cage isomerization energy of $C_{20}$ with $T_{CutPNO} = 10^{-X}$ ($X$ = 5, 6, 7, and 8) and with the 5/6 and 6/7 extrapolations relative to the canonical CCSD(T)/aug-cc-pVDZ-DK isomerization energy. (b) Computational time of the correlation part of the isomerization energy with the corresponding computational settings by using 4 cores from a single cluster node equipped with 4 Intel Xeon CPUs. Computational time for the extrapolations is obtained as the sum of the time required with the involving $T_{CutPNO}$ settings.

For these systems, the 5/6 extrapolation does not generally provide a significant improvement over $T_{CutPNO} = 10^{-6}$ because the $T_{CutPNO} = 10^{-5}$ results are dominated by the perturbative contribution, which largely overestimates the strength of dispersion interactions. The default TightPNO calculations also feature relatively large errors, i.e., 0.31, 0.12, and 2.41 kcal/mol for the uracil dimer interaction, pentane dimer interaction, and the $C_{20}$ isomerization reaction, respectively. In contrast, the 6/7 extrapolation reduces the error to 0.01, 0.01, and 1.05 kcal/mol, respectively. With a ∼40% smaller computational cost, the 6/7 extrapolation is even slightly more accurate than $T_{CutPNO} = 10^{-8}$. As the latter becomes unaffordable already for medium-sized systems (also due to its huge memory requirements), the 6/7 extrapolation provides a cost-effective alternative to tightening the $T_{CutPNO}$ threshold, yielding essentially canonical CCSD(T) accuracy even for such challenging reactions.

## 4. CONCLUSIONS

We proposed a computational strategy to approach the complete PNO space limit in DLPNO-CCSD(T) calculations. Results obtained using different $T_{CutPNO}$ values were extrapolated using a two-point extrapolation scheme, while keeping all the remaining parameters of the calculations fixed to the TighPNO default. Our scheme was validated on the most challenging subsets of the GMTKN55 superset, including a total of 883 molecules and 499 reactions.

Typically, the 5/6 extrapolation provides an accuracy between that of $T_{CutPNO} = 10^{-6}$ and $T_{CutPNO} = 10^{-7}$ calculations at a small fraction of the $T_{CutPNO} = 10^{-7}$ computational cost. It is especially recommended as a cost-effective alternative to TightPNO calculations for very large systems, for which standard TightPNO calculations are computationally too demanding.

In all cases, the 6/7 extrapolation provides results that are more accurate than those obtained from standard TightPNO calculations. Its accuracy is comparable to that obtained using $T_{CutPNO} = 10^{-8}$ but at a substantially lower computational cost. The residual error is associated with the remaining approximations made in DLPNO-CCSD(T) calculations, such as the RI and strong-pair approximations.

## ■ ASSOCIATED CONTENT

### Ⓢ Supporting Information

The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.jctc.0c00344.

> Detailed energetics and error analyses for each benchmark set (XLSX)

## ■ AUTHOR INFORMATION

### Corresponding Author

**Giovanni Bistoni** − *Max-Planck-Institut für Kohlenforschung, D-45470 Mülheim an der Ruhr, Germany;* ⓞ orcid.org/0000-0003-4849-1323; Email: giovanni.bistoni@kofo.mpg.de

### Authors

**Ahmet Altun** − *Max-Planck-Institut für Kohlenforschung, D-45470 Mülheim an der Ruhr, Germany;* ⓞ orcid.org/0000-0001-8818-9925

**Frank Neese** − *Max-Planck-Institut für Kohlenforschung, D-45470 Mülheim an der Ruhr, Germany;* ⓞ orcid.org/0000-0003-4691-0547

Complete contact information is available at:
https://pubs.acs.org/10.1021/acs.jctc.0c00344

### Notes

The authors declare no competing financial interest.

## ■ ACKNOWLEDGMENTS

## ■ REFERENCES

(1) Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M. A Fifth-Order Perturbation Comparison of Electron Correlation Theories. *Chem. Phys. Lett.* **1989**, *157*, 479−483.

(2) Ramabhadran, R. O.; Raghavachari, K. Extrapolation to the Gold-Standard in Quantum Chemistry: Computationally Efficient and Accurate CCSD(T) Energies for Large Molecules Using an Automated Thermochemical Hierarchy. *J. Chem. Theory Comput.* **2013**, *9*, 3986−3994.

(3) *Linear-Scaling Techniques in Computational Chemistry and Physics: Methods and Applications*; Zaleśny, R.; Papadopoulos, M. G.; Mezey, P. G.; Leszczynski, J., Eds.; Springer, 2011.

(4) Ma, Q.; Schwilk, M.; Köppl, C.; Werner, H.-J. Scalable Electron Correlation Methods. 4. Parallel Explicitly Correlated Local Coupled Cluster with Pair Natural Orbitals (PNO-LCCSD-F12). *J. Chem. Theory Comput.* **2017**, *13*, 4871−4896.

(5) Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. Sparse Maps—A Systematic Infrastructure for Reduced-Scaling Electronic Structure Methods. II. Linear Scaling Domain Based Pair Natural Orbital Coupled Cluster Theory. *J. Chem. Phys.* **2016**, *144*, No. 024109.

(6) Neese, F.; Hansen, A.; Liakos, D. G. Efficient and Accurate Approximations to the Local Coupled Cluster Singles Doubles Method Using a Truncated Pair Natural Orbital Basis. *J. Chem. Phys.* **2009**, *131*, No. 064103.

(7) Neese, F.; Hansen, A.; Wennmohs, F.; Grimme, S. Accurate Theoretical Chemistry with Coupled Pair Models. *Acc. Chem. Res.* **2009**, *42*, 641−648.

(8) Neese, F.; Wennmohs, F.; Hansen, A. Efficient and Accurate Local Approximations to Coupled-Electron Pair Approaches: An Attempt to Revive the Pair Natural Orbital Method. *J. Chem. Phys.* **2009**, *130*, No. 114108.

(9) Hansen, A.; Liakos, D. G.; Neese, F. Efficient and Accurate Local Single Reference Correlation Methods for High-Spin Open-Shell Molecules Using Pair Natural Orbitals. *J. Chem. Phys.* **2011**, *135*, No. 214102.

(10) Huntington, L. M. J.; Hansen, A.; Neese, F.; Nooijen, M. Accurate Thermochemistry from a Parameterized Coupled-Cluster Singles and Doubles Model and a Local Pair Natural Orbital Based Implementation for Applications to Larger Systems. *J. Chem. Phys.* **2012**, *136*, No. 064101.

(11) Riplinger, C.; Neese, F. An Efficient and near Linear Scaling Pair Natural Orbital Based Local Coupled Cluster Method. *J. Chem. Phys.* **2013**, *138*, No. 034106.

(12) Liakos, D. G.; Hansen, A.; Neese, F. Weak Molecular Interactions Studied with Parallel Implementations of the Local Pair Natural Orbital Coupled Pair and Coupled Cluster Methods. *J. Chem. Theory Comput.* **2011**, *7*, 76−87.

(13) Liakos, D. G.; Sparta, M.; Kesharwani, M. K.; Martin, J. M. L.; Neese, F. Exploring the Accuracy Limits of Local Pair Natural Orbital Coupled-Cluster Theory. *J. Chem. Theory Comput.* **2015**, *11*, 1525−1539.

(14) Riplinger, C.; Sandhoefer, B.; Hansen, A.; Neese, F. Natural Triple Excitations in Local Coupled Cluster Calculations with Pair Natural Orbitals. *J. Chem. Phys.* **2013**, *139*, No. 134101.

(15) Liakos, D. G.; Neese, F. Is It Possible To Obtain Coupled Cluster Quality Energies at near Density Functional Theory Cost? Domain-Based Local Pair Natural Orbital Coupled Cluster vs Modern Density Functional Theory. *J. Chem. Theory Comput.* **2015**, *11*, 4054−4063.

(16) Paulechka, E.; Kazakov, A. Efficient DLPNO−CCSD(T)-Based Estimation of Formation Enthalpies for C-, H-, O-, and N-Containing Closed-Shell Compounds Validated Against Critically Evaluated Experimental Data. *J. Phys. Chem. A* **2017**, *121*, 4379−4387.

(17) Koerstz, M.; Elm, J.; Mikkelsen, K. V. Benchmark Study of the Structural and Thermochemical Properties of a Dihydroazulene/Vinylheptafulvene Photoswitch. *J. Phys. Chem. A* **2017**, *121*, 3148−3154.

(18) Liakos, D. G.; Guo, Y.; Neese, F. Comprehensive Benchmark Results for the Domain Based Local Pair Natural Orbital Coupled Cluster Method (DLPNO-CCSD(T)) for Closed- And Open-Shell Systems. *J. Phys. Chem. A* **2020**, *124*, 90−100.

(19) Goerigk, L.; Hansen, A.; Bauer, C.; Ehrlich, S.; Najibi, A.; Grimme, S. A Look at the Density Functional Theory Zoo with the Advanced GMTKN55 Database for General Main Group Thermochemistry, Kinetics and Noncovalent Interactions. *Phys. Chem. Chem. Phys.* **2017**, *19*, 32184−32215.

(20) Nyden, M. R.; Petersson, G. A. Complete Basis Set Correlation Energies. I. The Asymptotic Convergence of Pair Natural Orbital Expansions. *J. Chem. Phys.* **1981**, *75*, 1843−1862.

(21) Montgomery, J. A.; Ochterski, J. W.; Petersson, G. A. A Complete Basis Set Model Chemistry. IV. An Improved Atomic Pair Natural Orbital Method. *J. Chem. Phys.* **1994**, *101*, 5900−5909.

(22) Halkier, A.; Helgaker, T.; Jørgensen, P.; Klopper, W.; Olsen, J. Basis-Set Convergence of the Energy in Molecular Hartree−Fock Calculations. *Chem. Phys. Lett.* **1999**, *302*, 437−446.

(23) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-Set Convergence of Correlated Calculations on Water. *J. Chem. Phys.* **1997**, *106*, 9639−9646.

(24) Schwenke, D. W. The Extrapolation of One-Electron Basis Sets in Electronic Structure Calculations: How It Should Work and How It Can Be Made to Work. *J. Chem. Phys.* **2005**, *122*, No. 014107.

(25) Varandas, A. J. C. Straightening the Hierarchical Staircase for Basis Set Extrapolations: A Low-Cost Approach to High-Accuracy

Computational Chemistry. *Annu. Rev. Phys. Chem.* **2018**, *69*, 177−203.

(26) Neese, F. The ORCA Program System. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 73−78.

(27) Neese, F. Software Update: The ORCA Program System, Version 4.0. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, No. e1327.

(28) Neese, F.; Wennmohs, F.; Becker, U.; Riplinger, C. The ORCA Quantum Chemistry Program Package. *J. Chem. Phys.* **2020**, *152*, No. 224108.

(29) Bistoni, G.; Riplinger, C.; Minenkov, Y.; Cavallo, L.; Auer, A. A.; Neese, F. Treating Subvalence Correlation Effects in Domain Based Pair Natural Orbital Coupled Cluster Calculations: An Out-of-the-Box Approach. *J. Chem. Theory Comput.* **2017**, *13*, 3220−3227.

(30) Neese, F. Importance of Direct Spin−Spin Coupling and Spin-Flip Excitations for the Zero-Field Splittings of Transition Metal Complexes: A Case Study. *J. Am. Chem. Soc.* **2006**, *128*, 10213−10222.

(31) Guo, Y.; Riplinger, C.; Becker, U.; Liakos, D. G.; Minenkov, Y.; Cavallo, L.; Neese, F. An Improved Linear Scaling Perturbative Triples Correction for the Domain Based Local Pair-Natural Orbital Based Singles and Doubles Coupled Cluster Method [DLPNO-CCSD(T)]. *J. Chem. Phys.* **2018**, *148*, No. 011101.

(32) Guo, Y.; Riplinger, C.; Liakos, D. G.; Becker, U.; Saitow, M.; Neese, F. Linear Scaling Perturbative Triples Correction Approximations for Open-Shell Domain-Based Local Pair Natural Orbital Coupled Cluster Singles and Doubles Theory [DLPNO-CCSD(T0/T)]. *J. Chem. Phys.* **2020**, *152*, No. 024116.

(33) Stoychev, G. L.; Auer, A. A.; Neese, F. Automatic Generation of Auxiliary Basis Sets. *J. Chem. Theory Comput.* **2017**, *13*, 554−562.

(34) Nakajima, T.; Hirao, K. The Douglas−Kroll−Hess Approach. *Chem. Rev.* **2012**, *112*, 385−402.

(35) Reiher, M. Relativistic Douglas-Kroll-Hess Theory. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2012**, *2*, 139−149.

(36) Dunning, T. H. Gaussian Basis Sets for Use in Correlated Molecular Calculations. I. The Atoms Boron through Neon and Hydrogen. *J. Chem. Phys.* **1989**, *90*, 1007−1023.

(37) Balabanov, N. B.; Peterson, K. A. Systematically Convergent Basis Sets for Transition Metals. I. All-Electron Correlation Consistent Basis Sets for the 3d Elements Sc−Zn. *J. Chem. Phys.* **2005**, *123*, No. 064107.

(38) Peterson, K. A.; Dunning, T. H. Accurate Correlation Consistent Basis Sets for Molecular Core−Valence Correlation Effects: The Second Row Atoms Al−Ar, and the First Row Atoms B−Ne Revisited. *J. Chem. Phys.* **2002**, *117*, 10548−10560.

(39) Lein, M.; Szabó, A.; Kovács, A.; Frenking, G. Energy Decomposition Analysis of the Chemical Bond in Main Group and Transition Metal Compounds. *Faraday Discuss.* **2003**, *124*, 365−378.

(40) Zhao, Y.; Ng, H. T.; Peverati, R.; Truhlar, D. G. Benchmark Database for Ylidic Bond Dissociation Energies and Its Use for Assessments of Electronic Structure Methods. *J. Chem. Theory Comput.* **2012**, *8*, 2824−2834.

(41) Řezáč, J.; Riley, K. E.; Hobza, P. S66: A Well-Balanced Database of Benchmark Interaction Energies Relevant to Biomolecular Structures. *J. Chem. Theory Comput.* **2011**, *7*, 2427−2438.

(42) Kesharwani, M. K.; Karton, A.; Sylvetsky, N.; Martin, J. M. L. The S66 Non-Covalent Interactions Benchmark Reconsidered Using Explicitly Correlated Methods Near the Basis Set Limit. *Aust. J. Chem.* **2018**, *71*, 238−248.

(43) Altun, A.; Neese, F.; Bistoni, G. HFLD: A Nonempirical London Dispersion-Corrected Hartree−Fock Method for the Quantification and Analysis of Noncovalent Interaction Energies of Large Molecular Systems. *J. Chem. Theory Comput.* **2019**, *15*, 5894−5907.

(44) Bistoni, G. Finding Chemical Concepts in the Hilbert Space: Coupled Cluster Analyses of Noncovalent Interactions. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2020**, *10*, No. e1442.