ARTICLE pubs.

pubs.aip.org/aip/jcp

Assessing the domain-based local pair natural orbital (DLPNO) approximation for non-covalent interactions in sizable supramolecular complexes

Cite as: J. Chem. Phys. 161, 054114 (2024); doi: 10.1063/5.0206533 Submitted: 2 March 2024 • Accepted: 18 July 2024 • Published Online: 6 August 2024

Montgomery Gray 🔟 and John M. Herbert a 🔟

AFFILIATIONS

Department of Chemistry & Biochemistry, The Ohio State University, Columbus, Ohio 43210, USA

^{a)}Author to whom correspondence should be addressed: herbert@chemistry.ohio-state.edu

ABSTRACT

The titular domain-based local pair natural orbital (DLPNO) approximation is the most widely used method for extending correlated wave function models to large molecular systems, yet its fidelity for intermolecular interaction energies in large supramolecular complexes has not been thoroughly vetted. Non-covalent interactions are sensitive to tails of the electron density and involve nonlocal dispersion that is discarded or approximated if the screening of pair natural orbitals (PNOs) is too aggressive. Meanwhile, the accuracy of the DLPNO approximation is known to deteriorate as molecular size increases. Here, we test the DLPNO approximation at the level of second-order Møller–Plesset perturbation theory (MP2) and coupled-cluster theory with singles, doubles, and perturbative triples [CCSD(T)] for a variety of large supramolecular complexes. DLPNO-MP2 interaction energies are within 3% of canonical values for small dimers with ≤ 10 heavy atoms, but for larger systems, the DLPNO approximation is often quite poor unless the results are extrapolated to the canonical limit where the threshold for discarding PNOs is taken to zero. Counterpoise correction proves to be essential in reducing errors with respect to canonical results. For a sequence of nanoscale graphene dimers up to ($C_{96}H_{24}$)₂, extrapolated DLPNO-MP2 interaction energies agree with canonical values to within 1%, independent of system size, provided that the basis set does not contain diffuse functions; these cause the DLPNO approximation to behave erratically, such that results cannot be extrapolated in a meaningful way. DLPNO-CCSD(T) calculations are typically performed using looser PNO thresholds as compared to DLPNO-MP2, but this significantly impacts accuracy for large supramolecular complexes. Standard DLPNO-CCSD(T) settings afford errors of 2–6 kcal/mol for dimers involving coronene ($C_{24}H_{12}$) and circumcoronene ($C_{54}H_{18}$), even at the DLPNO-CCSD(T₁) level.

Published under an exclusive license by AIP Publishing. https://doi.org/10.1063/5.0206533

I. INTRODUCTION

The coupled-cluster method using singles, doubles, and perturbative triples [CCSD(T)] is the gold standard of single-reference quantum chemistry, for both thermochemistry^{1,2} and non-covalent interactions,^{3–5} yet this benchmark-level accuracy comes at nearintractable cost for systems with more than 10–20 non-hydrogen atoms. Efforts to develop localized-orbital approximations to canonical coupled-cluster theory are experiencing a renaissance,^{6–27} and among several competing approaches the domain-based local pair natural orbital (DLPNO) approximation^{7,8} has emerged as the most widely used variant,^{28–30} owing to its implementation in the ORCA software.³¹

The fidelity of the DLPNO approximation has been exhaustively evaluated for small systems, including for non-covalent interactions.^{32–42} However, there have been only a few systematic tests for non-covalent interactions in large systems.^{43–45} Neglect or approximation of distant-pair interactions is known to degrade the accuracy in such cases.⁴⁴ As a result, the accuracy of the DLPNO approximation decreases with system size,^{40,43} at least for thermochemical calculations. For benchmark-quality accuracy, results must be extrapolated to the canonical limit in which no amplitudes are neglected.^{38,46}

Even so, recent work has demonstrated discrepancies exceeding 10 kcal/mol between localized-orbital approximations to canonical CCSD(T), as compared to fixed-node diffusion Monte Carlo calculations.^{47–50} (See Ref. 49 for an up-to-date summary of stateof-the-art benchmarks.) Some of the relevant CCSD(T) benchmarks were obtained with the DLPNO approach^{48,49} while others used an alternative, localized natural orbital (LNO) implementation of





CCSD(T),²³⁻²⁶ which is available in the MRCC program.⁵¹ All of these CCSD(T) benchmarks have been extrapolated to the complete basis set (CBS) limit, yet sizable discrepancies remain in multiple cases.

Establishing the veracity of the benchmarks is vitally important for the development of low-cost methods including density functional theory (DFT). Notably, various dispersion-inclusive exchange-correlation functionals and dispersion-corrected DFT approaches afford results of wildly varying quality for supramolecular complexes with $\gtrsim 100$ atoms,⁵⁰ despite the fact that the same methods provide rather accurate interaction energies for small van der Waals dimers.^{50,52} First-principles dispersion models⁵³⁻⁵⁶ also need to be tested against high-quality *ab initio* data in large systems.

The purpose of this work is to examine the accuracy of the DLPNO approximation for large van der Waals complexes. The most thorough previous effort along these lines is a recent study of double-hybrid DFT for transition metal compounds,⁵⁷ where the localized-orbital technique was applied to the second-order Møller–Plesset (MP2) component of the double-hybrid energy expression. As compared to what is typical in thermochemical DLPNO-CCSD(T) calculations, tighter pair natural orbital (PNO) thresholds were found to be required in order that the DLPNO-MP2 interaction energy remain faithful to the canonical MP2 result. Not addressed in that study is the performance of the DLPNO approximation in the presence of diffuse basis functions, which are the bane of linear-scaling approximations but are sometimes useful for obtaining converged intermolecular interaction energies.^{58–60}

Systematic tests of the DLPNO approximation in large van der Waals complexes are missing from the literature and the present work aims to fill this gap by examining the accuracy of the DLPNO approximation as a function of molecular size for nanoscale complexes as large as $(C_{96}H_{24})_2$, along with standard large-molecule benchmark datasets, including L7⁶¹ and S12L.^{62,63} Comprehensive testing for the S66 dataset⁶⁴ of smaller dimers is used to set a baseline. Because we need canonical results for comparison, we perform many of these calculations at the MP2 level rather than the CCSD(T) level, in basis sets ranging up to aug-cc-pVQZ, although CCSD(T) results are also presented for S66 and L7. Our results establish the accuracy that can be expected when the PNO thresholds are extrapolated to the canonical limit, and we examine the impact of diffuse basis functions on that extrapolation.

II. COMPUTATIONAL PROCEDURES

All interaction energies,

$$\Delta E = E_{\rm AB} - E_{\rm A} - E_{\rm B},\tag{1}$$

are counterpoise-corrected except for some results in Sec. III E, where we examine the impact of this correction on the DLPNO errors. Even in large basis sets, counterpoise correction can have a significant impact on ΔE ,^{60,65–67} especially for large molecules.^{50,60}

The resolution-of-identity (RI) approximation is applied to all MP2 calculations but is not employed at the self-consistent field (SCF) level. CCSD(T) calculations use either the noniterative (semicanonical) triples correction,⁶⁸ which is sometimes denoted "T₀,"^{8,12,19} or else the iterative "T₁" algorithm¹² that is based on triples natural orbitals.⁶⁹ All calculations were performed using ORCA,³¹ v. 4.2.1, except for the large dimer complexes using the Karlsruhe basis sets, for which ORCA v. 5.0.3 was used.

A. Molecular datasets

Large supramolecular complexes examined in this work are presented in Fig. 1. We use standard datasets including S66⁶⁴ (not shown in Fig. 1), S12L,^{62,63} and L7,⁶¹ and we augment the latter with a "buckyball-in-a-ring" (BBR) complex, C₆₀@[6]-cycloparaphenyleneacetylene.^{47,70} Several of these large complexes exhibit differences exceeding 1 kcal/mol between diffusion Monte Carlo and CCSD(T)/CBS benchmarks, even after accounting for the uncertainties in either benchmark. These include BBR, C2C2PD, C3A, C3GC, 2a, 2b, 4a, 5a, and 6a.^{47–49}

In addition to these standard benchmarks, we also consider dimers of polybenzenoids from the circumcoronene sequence, including coronene dimer $[(C_{24}H_{12})_2]$, circumcoronene dimer $[(C_{54}H_{18})_2]$, and circumcircumcoronene dimer $[(C_{96}H_{24})_2]$. These are models of graphene nano-flakes and we place them in a cofacial arrangement ("AA stacking"),^{71–73} at 3.8 Å face-to-face separation. (This is the same separation as in the cofacial or "sandwich" isomer of the benzene dimer⁷⁴ but slightly larger than the 3.354 Å stacking distance in graphitic carbon,⁷⁵ the latter of which corresponds to offset or "AB" stacking.⁷¹) These graphene models are considered alongside the corresponding models of graphane,⁷⁶ namely, perhydrocircumcoronenes that are also shown in Fig. 1. Some previous benchmarks exist for these model complexes.^{77–79}

B. Basis sets

A variety of basis sets are examined, including (aug-)cc-pVXZ (X = D, T, and Q)^{80,81} and jun-cc-pVDZ,⁸² along with def2-SVP, def2-SVPD, def2-TZVP, and def2-TZVPD.^{83,84} Auxiliary basis sets for the RI approximation were matched to these target basis sets.⁸⁵ Although it is well-established that MP2 overestimates dispersion interactions,^{54,86–89} the primary purpose of this work is to test the accuracy of the DLPNO approximation, not the accuracy of MP2. Moreover, a CBS extrapolation of the MP2 energy often serves as the foundation upon which a smaller-basis CCSD(T) correction is added,^{43,65,90,91} in the spirit of focal-point analysis.⁹²

C. Numerical thresholds

The integral screening threshold was set to 10^{-16} a.u. for all calculations, as tight integral thresholds are often necessary for SCF convergence in large systems, especially where diffuse basis functions are involved.⁹³ The SCF convergence criterion was set to $10^{-8} E_{\rm h}$.

An important aspect of this work is the convergence of the DLPNO approximation with respect to various PNO thresholds, and several PNO-related settings have been tested. ORCA's documentation⁹⁴ suggests different hierarchical thresholds for MP2 vs CCSD(T), although both hierarchies are labeled "loose/normal/tight" so that each term means something different for MP2 than it means for CCSD(T).^{94,95} Therefore, we will distinguish between "loose-MP2" settings and "loose-CCSD(T)" settings, with similar nomenclature for the normal and tight settings. PNO thresholds used in this work are provided in Table I, specified in terms of the ORCA parameters T_{CutDO} and T_{CutPNO} .⁹⁴ For MP2, 06 August 2024 12:18:13



FIG. 1. Large supramolecular complexes investigated in this work, including the L7 complexes, ⁶¹ S12 complexes, ^{62,63} a buckyball-in-a-ring (BBR) complex, ⁴⁷ graphene dimers from the circumcoronene family, and the corresponding graphane (perhydrocircumcoronene) dimers.

these reflect the suggested values in Ref. 94 but for CCSD(T) we have set $T_{\text{CutDO}} = 10^{-2}$ in all cases, which is the "normal" value for MP2 calculations. The most important of these thresholds is T_{CutPNO} . As such, we take T_{CutDO} to be pinned to T_{CutDO} , so that if we report, for example, that $T_{\text{CutPNO}} = 10^{-9}$ (the "tight-MP2" setting), then it is implied that $T_{\text{CutDO}} = 5 \times 10^{-3}$. Note that the "tight-CCSD(T)"

TABLE I. PNO thresholds used in the present work.^a

Name	$T_{\rm CutDO}$	$T_{\rm CutPNO}$
Tight-MP2 Normal-MP2 Loose-MP2	5.0×10^{-3} 1.0×10^{-2} 2.0×10^{-2}	$1.0 imes 10^{-9}$ $1.0 imes 10^{-8}$ $1.0 imes 10^{-7}$
Tight-CCSD(T) Normal-CCSD(T) Loose-CCSD(T)	$\begin{array}{c} 1.0 \times 10^{-2} \\ 1.0 \times 10^{-2} \\ 1.0 \times 10^{-2} \end{array}$	1.0×10^{-7} 3.3×10^{-7} 1.0×10^{-6}

^aFor all calculations, the domain-size parameter is $T_{\text{cutMKN}} = 10^{-3}$, the cutoff for pair interactions is $T_{\text{CutPairs}} = 10^{-4}$, and the cutoffs for the molecular orbital (MO) and projected atomic orbital (PAO) coefficient matrices are $T_{\text{CutC}(MO)} = 10^{-3} = T_{\text{CutC}(PAO)}$.

setting is comparable to the "loose-MP2" setting, reflecting the recommendations in Ref. 94.

In our view, it is very bad practice to substitute plain-English names for numerical thresholds,⁹⁵ because it encourages users to remain oblivious to numerical settings that may be important.⁹³ An occasional criticism of commercial software is that algorithm details may be proprietary, but in our view that criticism should be leveled against scientific writing (and reviewing), not against software *per se.*⁹⁶ To wit, the ORCA user manual suggests that some aspects of the internal PNO truncation scheme are intentionally undocumented.⁹⁷

In the present work, we explore MP2 calculations using all six sets of thresholds that are listed in Table I. This means that some of the MP2 calculations are performed using thresholds that are less conservative than the recommendations.⁹⁴ In this way, we can compare CCSD(T) and MP2 results side-by-side, using a consistent set of thresholds that is recommended for CCSD(T).

D. Extrapolations

Extrapolation of DLPNO energies to the canonical limit (where no amplitudes are neglected) is performed in two ways, using either a "loose/normal" (L/N) extrapolation or else a "normal/tight" (N/T) scheme. Both extrapolations make use of the formula

$$E = E_x + F(E_y - E_x), \qquad (2)$$

where F = 1.5 is an empirical parameter.³⁸ The quantities E_x and E_y are the energies obtained using the looser and the tighter of the two thresholds, respectively. CBS extrapolations for the correlation energy (E_{corr}) use the standard formula

$$E_{\rm corr}(\infty) = \frac{X^3 E_{\rm corr}(X) - Y^3 E_{\rm corr}(Y)}{X^3 - Y^3},$$
 (3)

where X and Y are the leading angular momentum quantum numbers (for carbon) in the two basis sets.⁹⁸

III. RESULTS AND DISCUSSION

In what follows, we benchmark the accuracy of the DLPNO approximation, primarily at the RI-MP2 level. We do not benchmark the accuracy of RI-MP2 itself. Thus, the DLPNO errors that are reported below are defined as differences with respect to the canonical result, using the same basis set. The same comments apply to CCSD(T) calculations.

A. Basis-set analysis using S66

In this section, we use the S66 dataset⁶⁴ to examine how the choice of basis impacts the accuracy of the DLPNO approximation. Error statistics for counterpoise-corrected S66 interactions energies at the RI-MP2 level are summarized in Table II, where "error" is defined with respect to the canonical RI-MP2 result. A significant amount of additional S66 data can be found in Sec. S1 of the supplementary material, including individual DLPNO errors for each S66 dimer; see Figs. S1 and S2 for Dunning basis sets

and Figs. S7 and S8 for Karlsruhe basis sets. We examine both the loose/normal/tight thresholds as typically defined for MP2 calculations,⁹⁴ and also the more aggressive thresholds that are intended for CCSD(T) calculations. (See Table I for the numerical thresholds.) This allows us to gain some insight into how these results may bear on CCSD(T) calculations.

For the cc-pVXZ sequence of basis sets, mean absolute errors (MAEs) for all PNO thresholds are consistent across the cardinality of the basis set: double- ζ , triple- ζ , or quadruple- ζ . (Individual errors can be found in Fig. S1, whereas Table II presents a statistical summary.) The most aggressive or least conservative threshold is the "loose CCSD(T)" value ($T_{\text{CutPNO}} = 10^{-6}$), for which the MAEs are 0.13, 0.15, and 0.16 kcal/mol for cc-pVDZ, cc-pVTZ, and cc-pVQZ, respectively, and the maximum DLPNO error is <0.4 kcal/mol. Histograms of the errors can be found in Fig. S5, leading to a standard deviation of <0.1 kcal/mol that is also consistent across cardinality. Similar trends are observed for Karlsruhe basis sets (Fig. S7) and will not be discussed in detail.

Similar consistency across basis sets is observed as T_{CutPNO} is tightened. By the time that $T_{\text{CutPNO}} = 10^{-9}$, MAEs have been reduced to 0.01–0.02 ± 0.01 kcal/mol, with a maximum error <0.05 kcal/mol. As a way of visualizing this convergence, Fig. 2 plots the DLPNO errors for each of these 66 dimers, along with the MAE, as a function of the PNO thresholds. (Figure S22 provides the corresponding convergence plot for DLPNO error as a percentage of ΔE , and analogous plots for Karlsruhe basis sets appear in Figs. S23 and S24.) Note that the "tight-CCSD(T)" thresholds are similar to the "loose-MP2" thresholds, with $T_{\text{CutPNO}} = 10^{-7}$ in both cases, and this similarity is reflected in the data.

Considering augmented (diffuse) basis sets, we find that the DLPNO approximation behaves less systematically in these cases, and there are outliers when aug-cc-pVDZ is used (see Figs. S2 and S6). Specifically, five of the S66 dimers exhibit DLPNO errors ≥1kcal/mol at the RI-MP2/aug-cc-pVDZ level, which persist as the

TABLE II. DLPNO error statistics for counterpoise-corrected S66 interaction energies as a function of the PNO thresholds, for calculations at the RI-MP2 level in various basis sets.^a

	DLPNO errors (kcal/mol)					
	CCS	SD(T) thresho	lds ^b	MP2 thresholds ^c		3 ^c
Basis set	Loose	Normal	Tight	Loose	Normal	Tight
cc-pVDZ cc-pVTZ cc-pVQZ	$\begin{array}{c} 0.13 \pm 0.08 \\ 0.15 \pm 0.09 \\ 0.16 \pm 0.09 \end{array}$	$\begin{array}{c} 0.08 \pm 0.05 \\ 0.09 \pm 0.06 \\ 0.11 \pm 0.06 \end{array}$	$\begin{array}{c} 0.05 \pm 0.03 \\ 0.06 \pm 0.04 \\ 0.07 \pm 0.04 \end{array}$	$\begin{array}{c} 0.05 \pm 0.04 \\ 0.07 \pm 0.05 \\ 0.07 \pm 0.04 \end{array}$	$\begin{array}{c} 0.02 \pm 0.02 \\ 0.03 \pm 0.02 \\ 0.03 \pm 0.02 \end{array}$	$\begin{array}{c} 0.01 \pm 0.01 \\ 0.01 \pm 0.01 \\ 0.02 \pm 0.01 \end{array}$
aug-cc-pVDZ aug-cc-pVTZ aug-cc-pVQZ	$\begin{array}{c} 0.23 \pm 0.41 \\ 0.15 \pm 0.08 \\ 0.16 \pm 0.09 \end{array}$	$\begin{array}{c} 0.18 \pm 0.41 \\ 0.09 \pm 0.05 \\ 0.11 \pm 0.06 \end{array}$	$\begin{array}{c} 0.15 \pm 0.04 \\ 0.06 \pm 0.04 \\ 0.07 \pm 0.04 \end{array}$	$\begin{array}{c} 0.15 \pm 0.40 \\ 0.05 \pm 0.40 \\ 0.06 \pm 0.04 \end{array}$	$\begin{array}{c} 0.12 \pm 0.40 \\ 0.02 \pm 0.20 \\ 0.03 \pm 0.02 \end{array}$	$\begin{array}{c} 0.12 \pm 0.40 \\ 0.01 \pm 0.01 \\ 0.02 \pm 0.01 \end{array}$
def2-SVP def2-TZVP	$\begin{array}{c} 0.13 \pm 0.08 \\ 0.15 \pm 0.09 \end{array}$	$\begin{array}{c} 0.08 \pm 0.05 \\ 0.09 \pm 0.06 \end{array}$	$\begin{array}{c} 0.05 \pm 0.04 \\ 0.06 \pm 0.04 \end{array}$	$\begin{array}{c} 0.05 \pm 0.04 \\ 0.06 \pm 0.04 \end{array}$	$\begin{array}{c} 0.02 \pm 0.02 \\ 0.03 \pm 0.03 \end{array}$	$\begin{array}{c} 0.01 \pm 0.01 \\ 0.02 \pm 0.02 \end{array}$
def2-SVPD def2-TZVPD	$\begin{array}{c} 0.11 \pm 0.07 \\ 0.15 \pm 0.09 \end{array}$	$\begin{array}{c} 0.06 \pm 0.05 \\ 0.10 \pm 0.06 \end{array}$	$\begin{array}{c} 0.03 \pm 0.04 \\ 0.06 \pm 0.04 \end{array}$	$\begin{array}{c} 0.08 \pm 0.05 \\ 0.04 \pm 0.04 \end{array}$	$\begin{array}{c} 0.02 \pm 0.02 \\ 0.02 \pm 0.02 \end{array}$	$\begin{array}{c} 0.01 \pm 0.02 \\ 0.01 \pm 0.01 \end{array}$

^aEach entry represents the mean absolute error with respect to canonical RI-MP2 interaction energies, and the uncertainty represents one standard deviation.

^bThresholds for CCSD(T) calculations are used (see Table I), even though the calculations reported here are MP2.

^cUsing PNO thresholds for MP2 (Table I).



FIG. 2. Absolute DLPNO errors for counterpoise-corrected S66 interaction energies computed at the RI-MP2/cc-pVXZ level for (a) X = D, (b) X = T, and (c) X = Q. Data are shown for six different sets of PNO thresholds (see Table I), typifying values used for either CCSD(T) or MP2 calculations, as indicated. Note that all interaction energies are RI-MP2, even when thresholds typical of CCSD(T) are used. The gray traces represent individual S66 dimers and the orange trace represents the average.

PNO thresholds are tightened but disappear in aug-cc-pVTZ and aug-cc-pVQZ calculations. Similar (albeit smaller) outliers arise using def2-SVPD but are absent using def2-TZVPD. Thus, there seems to be a peculiar instability problem with augmented double- ζ basis sets, which is not observed in basis sets of higher cardinality. If diffuse functions are required (say, for calculations on a molecular anion), then double- ζ basis sets ought to be avoided. This may pose an issue for a computational procedure that consists of adding a double- ζ CCSD(T) correction to the MP2/CBS result, which is a common approach. The effect of diffuse functions will be examined more closely in Sec. III I.

Setting aside aug-cc-pVDZ and def2-SVPD, the accuracy of the DLPNO approximation in the (aug-)cc-pVXZ and Karlsruhe basis set families is very good, even for the most aggressive ["loose CCSD(T)"] threshold of $T_{\text{CutPNO}} = 10^{-6}$. To consider whether this accuracy benefits from error cancellation, Figs. S13–S18 present DLPNO errors for the individual single-point energy calculations E_{AB} , E_{A} , and E_{B} [Eq. (1)], across all of the S66 dimers. For the loosest thresholds, we observe individual errors up to 5 kcal/mol in some cases, even while the DLPNO errors in ΔE are ~100× smaller. This is an error cancellation of sorts, though not an accidental or fortuitous one. Rather, it is akin to the error introduced by making the RI approximation, which introduces a small but size-extensive error per atom, yet one that is roughly constant across the potential energy surface.⁹⁹ As such, energy differences remain faithful to the canonical calculation with almost negligible error. This is what we observe for the DLPNO errors, at least in these small dimers.

B. Detailed examination of cc-pVDZ results for S66

The data in Table II and Fig. 2 establish that DLPNO errors obtained using cc-pVDZ are representative of the errors in larger basis sets. We can thus use RI-MP2/cc-pVDZ calculations for S66 to establish a baseline expectation before proceeding to larger systems, and DLPNO errors at the RI-MP2/cc-pVDZ level are plotted in Fig. 3 for all of the S66 dimers. The data are partitioned into three standard subsets of S66: hydrogen-bonded dimers (for which electrostatics dominates dispersion), dispersion-dominated complexes, and dimers where dispersion and electrostatics are comparable in

pubs.aip.org/aip/jcp



FIG. 3. DLPNO errors for the S66 dimers at the RI-MP2/cc-pVDZ level using (a) loose-MP2, (b) normal-MP2, and (c) tight-MP2 thresholds. Note the change in energy scale between panels.

magnitude. Even for the loose-MP2 thresholds, all of the errors are $<\!0.16$ kcal/mol.

This observation is notable because the *loose* PNO thresholds for MP2 are comparable to the *tight* thresholds for CCSD(T), meaning that DLPNO-CCSD(T) calculations typically involve more aggressive truncation of the PNOs as compared to DLPNO-MP2 calculations. We have performed both sets of calculations for the S66 dimers, and Fig. 4 juxtaposes the DLPNO errors in either case as a



FIG. 4. Absolute DLPNO errors for (a) CCSD(T₀)/cc-pVDZ and (b) RI-MP2/ccpVDZ interaction energies, for the S66 complexes. In both cases, the PNO thresholds are the CCSD(T) ones from Table I.

function of T_{CutPNO} for the CCSD(T) set of thresholds. (The corresponding convergence plots as a percentage of ΔE can be found in Fig. S25.)

For a given pair of thresholds T_{CutDO} and T_{CutPNO} , the DLPNO errors are larger at the CCSD(T₀) level than they are at the RI-MP2 level. The most likely explanation is the additional approximations that go into DLPNO-CCSD(T), namely, a perturbative treatment of the weak pairs,⁷ whereas DLPNO-MP2 calculations simply involve truncation of the correlated orbital space. Whatever the reason, DLPNO-CCSD(T₀) calculations exhibit outliers up to 1.3 kcal/mol when using loose thresholds, but all of these errors are reduced below 0.8 kcal/mol for the normal CCSD(T) thresholds and below 0.3 kcal/mol for tight CCSD(T) thresholds. Meanwhile, even for the loose thresholds the average error is <0.3 kcal/mol. However, these data do suggest that DLPNO errors are smaller at the RI-MP2 level, for a given set of PNO thresholds, meaning that DLPNO errors reported at the RI-MP2 level are likely lower bounds to errors that might be expected at the DLPNO-CCSD(T) level.

C. Canonical extrapolations for S66

We next examine extrapolation of the DLPNO results to the canonical limit, $T_{\text{CutPNO}} \rightarrow 0$, using the L/N and N/T procedures described in Sec. II D. Table III summarizes DLPNO errors in various basis sets, for extrapolations that use either the CCSD(T) or the MP2 values of T_{CutDO} and T_{CutPNO} . (Regardless of thresholds, all of the calculations summarized in Table III are RI-MP2 calculations.) Individual data for the entire S66 set can be found in Sec. S1 of the supplementary material.

With the exception of the ill-behaved aug-cc-pVDZ, errors in the extrapolated interaction energies are consistent across basis sets, differing by no more than 0.02 kcal/mol across basis-set cardinalities. At least for these small dimers, this implies that double- ζ basis sets provide a representative assessment of the accuracy of the DLPNO approximation. That said, the outliers noted in Sec. III A for aug-cc-pVDZ lead to extrapolated results that are less accurate in that case. In that particular basis set, there is no guarantee that

		DLPNO errors (kcal/mol)					
	CCSD(T)	thresholds ^b	MP2 thresholds ^c				
Basis set	L/N	N/T	L/N	N/T			
cc-pVDZ	0.05 ± 0.03	0.04 ± 0.03	0.01 ± 0.01	0.01 ± 0.01			
cc-pVTZ	0.07 ± 0.05	0.05 ± 0.03	0.01 ± 0.02	0.01 ± 0.01			
cc-pVQZ	0.08 ± 0.05	0.06 ± 0.03	0.02 ± 0.02	0.01 ± 0.01			
aug-cc-pVDZ	0.16 ± 0.40	0.14 ± 0.40	0.12 ± 0.40	0.12 ± 0.40			
aug-cc-pVTZ	0.06 ± 0.04	0.04 ± 0.03	0.02 ± 0.03	0.01 ± 0.01			
aug-cc-pVQZ	0.08 ± 0.05	0.05 ± 0.03	0.02 ± 0.01	0.01 ± 0.01			
def2-SVP	0.05 ± 0.04	0.04 ± 0.03	0.01 ± 0.02	0.01 ± 0.01			
def2-TZVP	0.07 ± 0.05	0.04 ± 0.03	0.02 ± 0.02	0.01 ± 0.02			
def2-SVPD	0.04 ± 0.04	0.02 ± 0.03	0.03 ± 0.05	0.02 ± 0.03			
def2-TZVPD	0.07 ± 0.05	0.05 ± 0.03	0.02 ± 0.03	0.01 ± 0.01			

TABLE III. Summary of DLPNO error statistics for S66 interaction energies computed at the RI-MP2 level, following extrapolation to the canonical limit.^a

^aEach entry represents the MAE with respect to canonical RI-MP2 interaction energies, with an uncertainty equal to one standard deviation.

^bThresholds for CCSD(T) calculations are used (see Table I), even though the calculations reported here are MP2.

^cUsing PNO thresholds for MP2 (Table I).

an extrapolation using tighter PNO thresholds will be more accurate than results obtained using looser thresholds.

Extrapolated DLPNO[N/T]-MP2 errors for the individual S66 dimers are shown in Fig. 5; these should be compared to the errors prior to extrapolation (Fig. 3). For the extrapolated results, all of the errors are <0.04 kcal/mol. This is comparable to what is obtainable using the normal-MP2 threshold ($T_{CutPNO} = 10^{-8}$) without extrapolation. For these small systems, one has reached a point of diminishing returns at this value of T_{CutPNO} .



FIG. 5. (a) Absolute and (b) percent differences between the DLPNO-MP2[N/T]/ccpVDZ interaction energies and canonical RI-MP2/cc-pVDZ values, for the S66 dimers. Colors indicate subsets of S66, as in Fig. 3. As a percentage of ΔE , the DLPNO errors are largest for the dispersion-bound subset of S66 and smallest for the hydrogenbonded subset; see Fig. 5(b). In part, this reflects smaller interaction energies for the dispersion-dominated dimers, although it is also consistent with the idea that dispersion is sensitive to tails of the density and to the "weak" (or distant) pairs.²⁶ Even for the dispersion-dominated complexes, however, DLPNO errors amount to less than 4% of ΔE .

D. Basis-set comparisons for larger complexes

Tests for the S66 dimers suggest that DLPNO errors are quite small and consistent across basis sets up to at least cc-pVQZ, provided that diffuse functions are avoided. The primary purpose of this work is to examine the accuracy of the DLPNO approximation as a function of molecular size, so we next investigate whether large basis sets are necessary to gauge the fidelity of the DLPNO approximation in larger systems.

Figure 6 compares absolute DLPNO errors for RI-MP2/ccpVXZ calculations (up to X = Q) using the L7 dataset,⁶¹ to which we add the **BBR** complex from Ref. 47. Interaction energies computed with the DLPNO approximation are systematically less attractive than canonical values but the error decreases as the PNO thresholds are tightened, in all basis sets. MAEs behave consistently across the cc-pVXZ sequence, although cc-pVQZ results are omitted for C3GC and BBR due to hardware constraints. The latter two complexes afford the largest DLPNO errors and this is why the MAEs for RI-MP2/cc-pVQZ [Fig. 6(c)] are about 2× smaller than those for either cc-pVTZ.

Table IV presents a statistical summary of DLPNO errors for this dataset. MAEs, with respect to canonical RI-MP2/cc-pVXZ

pubs.aip.org/aip/jcp



FIG. 6. DLPNO errors for counterpoise-corrected L7 + BBR interaction energies, evaluated at the RI-MP2/cc-pVXZ level with (a) X = D, (b) X = T, and (c) X = Q. In panel (c), the results for C3GC and BBR are omitted due to hardware constraints.

calculations, exceed 1 kcal/mol when the thresholds are set to looseor normal-CCSD(T) values. For cc-pVTZ, MAEs exceed 1 kcal/mol for all three sets of CCSD(T) thresholds as well as for the loose MP2 thresholds. MAEs appear to be smaller for RI-MP2/cc-pVQZ calculations but this is almost certainly an artifact of excluding the BBR and C3GC complexes, which afford the largest DLPNO errors in smaller basis sets. For example, the DLPNO error for BBR at the RI-MP2/cc-pVDZ level ranges from 8.4 kcal/mol [loose-CCSD(T)] to 0.6 kcal/mol (tight-MP2), and at the RI-MP2/cc-pVTZ level it ranges from 9.5 kcal/mol [loose-CCSD(T)] to 0.8 kcal/mol (tight-MP2). Overall, however, the trends illustrated in Fig. 6 are similar to those observed for the S66 dataset; cf. Fig. 2. In particular, we observe smooth convergence to zero error as $T_{CutPNO} \rightarrow 0$, across the whole cc-pVXZ sequence.

The analog of Fig. 6, with DLPNO errors expressed as a percentage of ΔE , can be found in Fig. S32. The largest percentage errors are associated with the guanine trimer (GGG) and can be as large as 35% or as small as 1%, depending on the thresholds and basis set. GGG has the smallest interaction energy in the L7 dataset (at ≈ 2 kcal/mol), and if this complex is excluded then the remaining percentage errors are remarkably consistent across thresholds and basis sets (see Fig. S32). For the normal-MP2 and tight-MP2 thresholds, the mean absolute percentage error is <2%.

A similar analysis for the L7 + **BBR** dataset using augmented basis sets (RI-MP2/aug-cc-pVXZ with X = D or T) is shown in Fig. 7. Addition of diffuse functions significantly reduces the DLPNO errors, eliminating the outliers that are evident in Figs. 6(a) and 6(b), and for $T_{\text{CutPNO}} \ge 3.3 \times 10^{-7}$ all errors are ≤ 1.5 kcal/mol. That said, these errors do oscillate as a function of T_{CutPNO} when the aug-cc-pVDZ basis set is used, as was seen with S66. This undesirable behavior is substantially mitigated, though not entirely eliminated, using aug-cc-pVTZ. A more comprehensive examination of the impact of diffuse functions is postponed until Sec. III I.

Results for the def2-SVP and def2-TZVP basis sets show similar trends as T_{CutPNO} is tightened (see Fig. S27), with double- ζ errors that are comparable to triple- ζ errors. For the Karlsruhe basis sets, these errors are slightly larger than those obtained using Dunning basis sets and approach or exceed 10% of ΔE in two cases (GGG and BBR), but only when $T_{\text{CutPNO}} \ge 10^{-7}$. For BBR, the DLPNO error is

TABLE IV. Error statistics for interaction energies in the L7 + BBR dataset computed at the RI-MP2/cc-pVXZ level.

		DLPNO errors (kcal/mol)			
	X =	X = D		X = T	
PNO thresholds	MAE	Max	MAE	Max	MAE
Loose-CCSD(T)	2.3	8.4	2.6	9.5	1.4
Normal-CCSD(T)	1.5	5.6	1.7	6.3	0.9
Tight-CCSD(T)	1.0	3.6	1.1	4.2	0.6
Loose-MP2	1.8	6.3	1.8	6.0	0.8
Normal-MP2	0.5	1.8	0.6	2.1	0.3
Tight-MP2	0.2	0.6	0.2	0.8	0.1

^aExcluding BBR and C3GC. Maximum error is omitted because BBR affords the largest error in smaller basis sets.



FIG. 7. DLPNO errors for counterpoise-corrected L7 + **BBR** interaction energies computed at the RI-MP2/aug-cc-pVXZ level for (a) X = D and (b) X = T.

ARTICLE



FIG. 8. Average number of PNOs per occupied MO pair for the S66 and the L7 + **BBR** dataset, computed at the RI-MP2/cc-pVXZ level for (a) X = D or (b) X = T. Colored boxes encapsulate 50% of the data (two quartiles), with the median indicated by a horizontal line. Whiskers represent $1.5 \times$ the interquartile range, which is a 99% confidence interval for a Gaussian distribution. Circles indicate data points that lie beyond the whiskers.

6–7 kcal/mol for RI-MP2 calculations using def2-SVP or def2-TZVP with loose-MP2 thresholds ($T_{\text{CutPNO}} = 10^{-7}$), but those errors are reduced below 2 kcal/mol for normal thresholds ($T_{\text{CutPNO}} = 10^{-8}$) and below 1 kcal/mol for tight-MP2 thresholds ($T_{\text{CutPNO}} = 10^{-9}$).

Figure 8 compares how the number of PNOs changes as a function of threshold and basis set, comparing S66 to L7 + BBR. The DLPNO approximation eliminates more pairs for the larger systems (L7 and BBR) than it does for the smaller complexes (S66) because the number of weakly interacting PNOs grows with system size. This remains true even as the number of PNOs grows larger in a larger basis set. That the average number of PNOs per pair is smaller for larger complexes than it is for the S66 dimers, while DLPNO errors can still be brought to sub-kcal/mol levels with appropriately tight thresholds, suggests that basis-dependent trends observed for the S66 dimers carry over to much larger complexes. As such, it appears that tests using the cc-pVDZ basis set are sufficient to examine the behavior of the DLPNO error in larger complexes, for which the use of more complete basis sets is challenging if tight thresholds are used. Instead, we can use cc-pVDZ to examine the DLPNO error as a function of system size and as a function of T_{CutPNO} , with the ultimate aim being to determine whether extrapolation can be used to avoid the need for the most conservative thresholds.

E. Effect of counterpoise correction

The interaction energies presented in this work are all counterpoise-corrected. In this section, we examine the effect of

the DLPNO approximation on that correction. Figure 9 compares DLPNO errors for the L7 + BBR dataset, with or without counterpoise correction, evaluated at the RI-MP2/cc-pVXZ level with X = D or T. Errors are much smaller when ΔE is counterpoise-corrected, indicating significant cancellation. For the counterpoise-corrected results, both basis sets behave similarly and the mean DLPNO error is about 2 kcal/mol for loose-CCSD(T) thresholds, although individual errors range up to 8.4 kcal/mol.

DLPNO errors are much larger for uncorrected interaction energies, however. This behavior has no precedent in smaller complexes such as the S66 dimers, where counterpoise correction modifies the error statistics by ≤ 0.2 kcal/mol, even in double- ζ basis sets.²⁶ For L7 + BBR, the DLPNO errors range up to 20 kcal/mol for cc-pVDZ and up to 16 kcal/mol for cc-pVTZ, in the absence of counterpoise correction, suggesting that this correction should be considered mandatory for non-covalent interaction energies evaluated within the DLPNO approximation. Agreement between counterpoise-corrected and uncorrected values of ΔE has sometimes been used as a criterion to indicate convergence to the CBS limit,^{60,100} but that is not a viable test for DLPNO calculations in large systems due to sizable errors in the uncorrected results. The "half-counterpoise" procedure^{26,65,67,101} (averaging corrected and uncorrected values of ΔE) is similarly ill-advised under these circumstances.

The fidelity of the DLPNO approximation thus appears to be sensitive to basis set only in the absence of counterpoise correction. Overall, the similarity of counterpoise-corrected RI-MP2/cc-pVDZ and RI-MP2/cc-pVTZ results affirms the conclusion in Sec. III D that a smaller basis set can safely be used to evaluate DLPNO errors in larger systems. In what follows, we will use the cc-pVDZ basis set to examine DLPNO errors as a function of system size, extending those studies to larger systems than would otherwise be feasible with available computing resources.

F. DLPNO approximation in larger systems

DLPNO errors are quite small for the S66 complexes but at the same time $|\Delta E|$ ranges only from 1 to 19 kcal/mol for these small dimers,⁶⁴ whereas van der Waals complexes with $\gtrsim 100$ atoms may have interaction energies that exceed 100 kcal/mol.^{50,62,102} In this section we examine the L7 + BBR and S12L datasets in detail, using calculations at the RI-MP2 level. In particular, we examine the extent to which the DLPNO error can be extrapolated to zero as a function of $T_{\rm CutPNO}$.

DLPNO errors for the L7 + BBR complexes are plotted in Fig. 10, computed at the RI-MP2/cc-pVDZ level. These errors are much larger than the corresponding errors for S66, except (debatably) when tight-MP2 thresholds are employed. Mean and maximum DLPNO errors can be found in Table IV as a function of $T_{\rm CutPNO}$, and these statistics indicate that thresholds looser than normal-MP2 afford MAEs larger than 1 kcal/mol. Notably, this includes the tight-CCSD(T) thresholds, where the MAE is close to 1 kcal/mol but the maximum error (for the BBR complex) is about 4 kcal/mol. In addition, percentage errors for the GGG complex (with $|\Delta E| \approx 2$ kcal/mol) and for BBR are not reduced below 10% unless normal- or tight-MP2 thresholds are used [Fig. 10(b)]. The results with tight-MP2 thresholds are nearly



FIG. 9. DLPNO errors at the RI-MP2/cc-pVXZ level, examining the effect of counterpoise correction using the L7 + BBR dataset: (a) counterpoise-corrected results for X = D, (b) counterpoise-corrected results for X = T, (c) uncorrected results for X = D, and (d) uncorrected results for X = T.

indistinguishable from canonical RI-MP2 interaction energies. It is worth remembering, however, that "loose-MP2" thresholds are roughly equivalent to "tight-CCSD(T)" thresholds.

Extrapolations to $T_{CutPNO} \rightarrow 0$ using Eq. (2) are shown in Fig. 11. Here, we compare L/N and N/T extrapolations using both the MP2 and the CCSD(T) thresholds, although all calculations are



FIG. 10. (a) Absolute and (b) percentage DLPNO errors for the L7 + BBR interaction energies computed at the RI-MP2/cc-pVDZ level. All interaction energies are counterpoise-corrected.

performed at the RI-MP2/cc-pVDZ level. Using MP2 thresholds, both extrapolation schemes reduce the DLPNO errors below 0.5 kcal/mol in every single case, even for the sizable **BBR** complex where the extrapolated error is about 0.5 kcal/mol for DLPNO[$(L/N)_{MP2}$] and is 0.1 kcal/mol for DLPNO[$(N/T)_{MP2}$]. These errors are below 1% except for the weakly bound guanine trimer.



FIG. 11. (a) Absolute and (b) percentage differences between extrapolated values and canonical RI-MP2/cc-pVDZ interaction energies for the L7 + BBR dataset. L/N and N/T extrapolations [Eq. (2)] were performed using either CCSD(T) or MP2 thresholds (as indicated by the subscripts), applied to RI-MP2 interaction energies in either case.

Extrapolated errors remain larger when data obtained with CCSD(T) thresholds are used, however. DLPNO[(L/N)_{CCSD(T)}] extrapolation leaves residual errors larger than 1 kcal/mol in several cases, while even DLPNO[(N/T)_{CCSD(T)}] extrapolation does not reduce the error for BBR below 2 kcal/mol, although percentage errors are <4% in that case. The DLPNO[(N/T)_{CCSD(T)}] approach thus offers a cheaper alternative to DLPNO[(L/N)_{MP2}] extrapolation with minimal effects on accuracy, at least in percentage terms. This is relevant for potential DLPNO-CCSD(T) calculations, where we would expect DLPNO[(N/T)_{CCSD(T)}] extrapolation to provide errors at or below the 1 kcal/mol level, except for very large complexes such as BBR. DLPNO-CCSD(T) for larger systems is discussed in Sec. III G.

The S12L complexes present a similar story, as shown by DLPNO errors in Fig. 12 that were again computed at the RI-MP2/cc-pVDZ level. (Similar errors are obtained at the RI-MP2/def2-SVP level; see Fig. S28.) Errors are large for thresholds less conservative than loose-MP2, suggesting in particular that the CCSD(T) thresholds may not be appropriate for benchmark calculations in systems of this size. Normal- and tight-MP2 thresholds afford MAEs of 0.7 and 0.2 kcal/mol, respectively, with maximum DLPNO errors of 2.6 kcal/mol (normal-MP2) and 1.0 kcal/mol (tight-MP2). The largest errors arise in the dispersion-bound systems **4a** and **4b**, which are complexes of C₆₀ and C₇₀ with the corannulene-based "buckycatcher" molecule, C₆₀H₂₈.¹⁰³ Even for these challenging cases, however, setting $T_{\text{CutPNO}} \leq 10^{-8}$ affords an interaction energy that is very close to the canonical RI-MP2 value.

Extrapolated results for S12L are presented in Fig. 13. For L/N and N/T extrapolation using MP2 thresholds, the errors are uniformly <0.6 kcal/mol even for the challenging **4a** and **4b** cases. Thus, benchmark-quality results are obtainable for dispersion-bound systems of this size, but only when tight thresholds are combined with extrapolation. Extrapolation based on the less conservative CCSD(T) thresholds is somewhat less accurate, especially for the difficult **4a** and **4b** complexes where DLPNO[(N/T)_{CCSD(T)}] extrapolation leaves residual errors of about 4 kcal/mol. For each of the other S12L complexes, the future DLPNO[(N/T)_{CCSD(T)}] error is



FIG. 12. (a) Absolute and (b) percentage DLPNO errors for S12L interaction energies computed at the RI-MP2/cc-pVDZ level.



FIG. 13. (a) Absolute and (b) percentage differences between extrapolated values and canonical RI-MP2/cc-pVDZ interaction energies for the S12L dataset. L/N and N/T extrapolations were performed using either CCSD(T) or MP2 thresholds (as indicated by the subscripts), applied to RI-MP2 interaction energies in either case.

<1 kcal/mol. Looking ahead to future DLPNO-CCSD(T) calculations, it would seem that benchmark-quality results are only obtainable using tight-CCSD(T) thresholds. Even then, large dispersionbound complexes may be problematic.

It is worth noting that tight PNO thresholds add significantly to the cost of DLPNO calculations. For RI-MP2 calculations on our hardware, the $(C_{96}H_{24})_2$ system that is explored in Sec. III H represents an approximate crossover point for loose-MP2 thresholds. That calculation requires 614 min for RI-MP2/cc-pVDZ with the DLPNO approximation, running on 40 processors with 4.5 Gb of memory per processor, which should be compared to 617 min for the canonical calculation on 48 processors with 5.0 Gb per processor. As the PNO thresholds are tightened, that 614 min calculation on 40 processors for $T_{\text{CutPNO}} = 10^{-7}$ becomes 4825 min for $T_{\text{CutPNO}} = 10^{-8}$ and 6531 min for $T_{\text{CutPNO}} = 10^{-9}$. It should be noted that we have set the integral threshold as conservatively as possible (10^{-16} a.u.) for the purpose of benchmarking, and more aggressive integral screening may shift the crossover point in favor of the DLPNO approach. Furthermore, CCSD(T) calculations will reach the crossover point in smaller system sizes. A systematic timing study is beyond the scope of the present work.

G. DLPNO-CCSD(T) calculations

As a final point of comparison between small-molecule results (exemplified by the S66 dataset) and those for larger complexes, we consider DLPNO-CCSD(T) calculations for the L7 dataset. For systems of this size, the memory requirement for canonical CCSD(T) exceeds the 3 Tb available on our hardware,¹⁰⁴ so we will use DLPNO-CCSD(T) calculations with tight-CCSD(T) thresholds ($T_{\text{CutPNO}} = 10^{-7}$) as reference values. We will present DLPNO-CCSD(T₀) results in detail, but we find that nearly identical interaction energies are obtained using the iterative T₁ approximation for the triples, with a maximum difference of 0.17 kcal/mol for the GGG complex. (See Table S1 for a side-by-side comparison of T₀ and T₁ results.) Other studies have found similarly small differences

between T_0 and T_1 for intermolecular interactions.¹⁰⁵ It is reported (though not quantified) that the T_0 approximation fails for cumulenes,¹² which might be attributed to extended conjugation except that we observe no such failure for C2C2PD.

The present calculations are intended to explore the difference between tight-CCSD(T) thresholds and loose- or normal-CCSD(T) thresholds, and to do the same at the RI-MP2 level of theory. That way, we can assess whether conclusions drawn from RI-MP2 calculations can be expected to hold for CCSD(T) calculations as well. Results in Fig. 14 make it clear these differences are much larger for CCSD(T) calculations than they are for RI-MP2 calculations, even when the PNO thresholds are the same. Even with normal-CCSD(T) thresholds, differences with respect to the reference values range up to almost 6 kcal/mol (for the C2C2PD complex) in the CCSD(T) case the DLPNO errors do not always decrease monotonically with $T_{\rm CutPNO}$. There is precedent for this behavior in the S66 data (see Fig. 2), but the energy scale of the fluctuations is much larger in these complexes.

These results affirm our previous conclusion, based on S66 data, that DLPNO errors obtained at the RI-MP2 level are likely lower bounds to those obtained at the CCSD(T) level, probably due to the approximate treatment of the weak pairs in DLPNO-CCSD(T).⁷ To obtain benchmark-quality CCSD(T) interaction energies whose fidelity approaches 1 kcal/mol with respect to canonical results, PNO thresholds more conservative than tight-CCSD(T) (i.e., $T_{\text{CutPNO}} < 10^{-7}$) may be needed in large supramolecular complexes. For reasons of cost, however, the loose/normal/tight hierarchy for T_{CutPNO} is set in the opposite sense, i.e., the cutoffs are more aggressive at the DLPNO-CCSD(T) level than they are for DLPNO-MP2. This may not be appropriate for large van der Waals complexes.

H. DLPNO errors vs system size

Prompted by the observation that the dispersion-bound subset of S66 affords the largest DLPNO errors, along with the



FIG. 14. DLPNO errors for L7 interaction energies computed at (a) the CCSD(T)/cc-pVDZ level or (a) the RI-MP2/cc-pVDZ level. In either case, the reference value is the corresponding DLPNO calculation performed with tight-CCSD(T) thresholds, whereas these calculations use loose- or normal-CCSD(T) thresholds.

documented dependence of DLPNO errors on molecular size,^{40,43} we next consider how the DLPNO approximation behaves for homologous systems of increasing size. These tests are performed at the RI-MP2/cc-pVDZ level.

We first examine linear alkanes $C_n H_{2n+2}$, as considered recently using an alternative local-orbital approximation to MP2.¹⁰⁶ That work demonstrated a constant fractional error in the correlation energy as a function of the chain length *n*, meaning that a consistent fraction of the (size-extensive) correlation energy is recovered. The same is true within the DLPNO approximation as demonstrated by results in Fig. 15, which plots the DLPNO error per carbon atom as a function of *n*. Size-extensivity implies that the canonical result is constant for large systems, and that behavior is indeed observed for each set of PNO thresholds. Loose thresholds require a somewhat longer carbon chain to reach the asymptotic result.

Extensivity also implies that the DLPNO error will increase in magnitude as the system size grows, just as the RI error increases with system size.⁹⁹ In the present examples, the absolute error increases from 3.5 kcal/mol for n = 20 to 18.0 kcal/mol for n = 100, when loose-MP2 thresholds are employed. For normal-MP2 thresholds, the absolute errors are 1.3 kcal/mol (n = 20) and 7.6 kcal/mol (n = 100), while for tight-MP2 thresholds they are 0.6 kcal/mol (n = 20) and 2.8 kcal/mol (n = 100). Absolute errors are not transferable to systems of very different size, which is worth considering in the context of large-molecule benchmarking.

Just as the RI approximation performs well for energy differences despite absolute errors that grow with system size, it is not immediately clear whether larger DLPNO errors (as a function of system size) will manifest as larger errors in ΔE . Thus, we next consider a sequence of increasingly large polybenzenoid dimers in the circumcoronene family, ranging from $(C_6H_6)_2$ up to $(C_{96}H_{24})_2$. DLPNO errors for these dimers, evaluated at the RI-MP2/cc-pVDZ level, are plotted in Fig. 16. Accumulation of error with respect to system size is especially evident for the two largest systems.

As with the linear alkanes, this error is systematically reduced (though not eliminated) as T_{CutPNO} is tightened. For the largest system considered, $(C_{96}H_{24})_2$, the DLPNO error can exceed 20 kcal/mol. For all of these systems, we find that tight-CCSD(T) thresholds afford much smaller DLPNO errors as compared to loose-MP2 thresholds, which is interesting because both approaches



FIG. 15. DLPNO errors in the correlation energy per carbon atom for linear alkanes at the RI-MP2/cc-pVDZ level.

The Journal of Chemical Physics



FIG. 16. DLPNO errors for polybenzenoid dimers of increasing size, computed at the RI-MP2/cc-pVDZ level using various PNO thresholds.

use the same value of T_{CutPNO} but a 2× more conservative value of T_{CutDO} in the tight-CCSD(T) case; see Table I. Although T_{CutPNO} is generally viewed as the main PNO threshold,⁹⁴ other thresholds may assume more importance in large systems.

To reduce the DLPNO error for $(C_{96}H_{24})_2$ below 10 kcal/mol requires thresholds of normal-MP2 quality or tighter. Even then, the error for $(C_{54}H_{18})_2$ remains on the order of several kcal/mol and it is that large for $(C_{96}H_{24})_2$ even with tight-MP2 thresholds. Absent extrapolation (which is considered next), the DLPNO approximation does not afford benchmark-quality results in systems of this size.

Extrapolated DLPNO errors for these systems are presented in Fig. 17. For $(C_{96}H_{24})_2$, the $(L/N)_{MP2}$ scheme affords an error of almost 5 kcal/mol and even larger errors are incurred using the $(L/N)_{CCSD(T)}$ and $(N/T)_{CCSD(T)}$ schemes that employ CCSD(T) thresholds. For this largest system, $(N/T)_{MP2}$ extrapolation is required to reach a benchmark-quality result, with a DLPNO error of only 0.5 kcal/mol or <1% of ΔE . The $(L/N)_{MP2}$ extrapolation scheme might be considered acceptable for $(C_{54}H_{18})_2$. Unfortunately, this consideration depends on system size and it seems



FIG. 17. (a) Absolute and (b) percentage errors in L/N and N/T extrapolations of RI-MP2/cc-pVDZ interaction energies for polybenzenoid dimers, using either MP2 or CCSD(T) thresholds. Reference values are canonical RI-MP2/cc-pVDZ interaction energies.

unwise to make a blanket recommendation for all systems. Results for $(C_{96}H_{24})_2$ suggest that $(N/T)_{MP2}$ extrapolation (i.e., the N/T procedure using MP2 thresholds) is a viable option for benchmark-quality results on nanoscale dispersion-bound complexes with more than 100 atoms. Absent extrapolation, however, the DLPNO results are clearly unacceptable as benchmarks in systems of this size.

ARTICLE

To examine whether error accumulation in these graphene nano-flakes is merely a function of molecular size or whether extended conjugation may also play a role, we next examine analogous dimers of perhydrocircumcoronenes, which are the structures whose basic building block is cyclohexane rather than benzene. A structure for the dimer of [24]graphane (also known as perhydrocoronene,¹⁰⁷ C₂₄H₃₆) was obtained from Ref. 78 and larger structures were constructed around it, maintaining the interlayer separation at 4.7 Å as determined for (C₂₄H₃₆)₂.⁷⁸ DLPNO errors for these graphane dimers are listed in Table V, where they are juxtaposed with errors for graphene dimers having the same number of carbon atoms.

Although the DLPNO errors do increase with system size for both the circumcoronene ([*n*]graphene) and perhydrocircumcoronene ([*n*]graphane) dimers, the effect is much larger for the former, up to 5 kcal/mol larger for n = 54 when loose-MP2 thresholds are employed. This may result from many-body dispersion screening effects that arise in the conjugated systems, or what Dobson calls "type B" nonadditivity,^{108,109} although this effect is ordinarily considered not to be captured at the MP2 level.¹⁰⁸ Alternatively, it might simply reflect the more delocalized molecular orbitals in graphene as compared to graphane, leading to larger errors in a localized-orbital implementation. Whatever the origin, these data suggest that systems with extensive conjugation may represent some of the most problematic cases for the DLPNO approximation, for which the use of tight PNO thresholds should be considered mandatory.

I. Effect of diffuse basis functions

Finally, we return to the issue of troublesome outliers in the aug-cc-pVDZ basis set. These can be seen for the S66 complexes in Figs. S2 and S6, with the effect that the mean DLPNO error oscillates

 TABLE V. DLPNO errors for dimers of polycyclic hydrocarbons, computed at the RI-MP2/cc-pVDZ level.

	DLPNO error (kcal/mol)			
Complex	Loose ^a	Normal ^a	Tight ^a	
(Benzene) ₂	0.08	0.04	0.01	
(Cyclohexane) ₂	0.06	0.03	0.01	
Difference ^b	-0.02(25%)	-0.01(25%)	0.00(0%)	
(Coronene) ₂	1.79	0.71	0.26	
([24]Graphane) ₂	0.81	0.51	0.15	
Difference ^b	-0.98(55%)	-0.20(28%)	-0.11(42%)	
(Circumcoronene) ₂ ([54]Graphane) ₂ Difference ^b	8.74 3.27 -5.47(63%)	3.49 1.95 -1.54(43%)	1.23 0.78 -0.45(37%)	

^aMP2 thresholds (see Table I).

^bDifference between the graphene and graphane errors (in parentheses as a percentage of the graphene value).

as T_{CutPNO} is tightened and does not converge to zero, although this behavior disappears in the aug-cc-pVTZ basis set. In this section, we analyze DLPNO errors in the presence of diffuse basis functions, using RI-MP2 calculations for the L7 and S12L datasets.

Figure 18 shows the convergence behavior as a function of T_{CutPNO} for RI-MP2/(jun-)cc-pVDZ calculations on the L7 complexes. In five of seven systems, the DLPNO error exhibits non-monotonic behavior as the PNO thresholds are tightened, which is not observed in the absence of diffuse functions. This wreaks havoc with extrapolations, which need not be more accurate than even the loose-MP2 threshold result. (See Fig. S33 for the L/N- and N/T-extrapolated results.) The same behavior is observed for the S12L complexes; see Fig. 19 for the DLPNO errors and Fig. S34 for the extrapolated results.

For both datasets, convergence with respect to T_{CutPNO} is monotonic at the RI-MP2/cc-pVDZ level but erratic at the RI-MP2/jun-cc-pVDZ level, so the anomalies are clearly attributable to the diffuse basis functions. As such, DLPNO calculations with diffuse basis functions should not be used in $T_{\text{CutPNO}} \rightarrow 0$ extrapolations. On the other hand, when loose-MP2 thresholds are used, the DLPNO error for jun-cc-pVDZ is consistently (and in some cases, significantly) smaller than the error obtained using cc-pVDZ for both the L7 and the S12L complexes. This is something of an artifact for loose thresholds, however. For normal-MP2 thresholds, the errors are typically (though not always) smaller for cc-pVDZ than for jun-cc-pVDZ, and for tight-MP2 thresholds, the error is always smaller for the non-augmented basis set.

It is worth noting that jun-cc-pVDZ is only minimally augmented, with no diffuse functions at all on hydrogen and only diffuse *s*- and *p*-functions on second-row elements.⁸² To investigate whether the anomalies just noted might arise from an unbalanced selection of diffuse functions, we performed RI-MP2/def2-SVPD calculations on the L7 + **BBR** and S12L datasets, with DLPNO errors reported in Fig. 20. (For errors as a percentage of ΔE , see Figs. S30 and S31.)

For the L7 + BBR dataset, DLPNO errors range up to 2.8 kcal/mol using normal PNO thresholds (with the largest error



FIG. 18. (a) Absolute and (b) percent DLPNO errors for RI-MP2/(jun-)cc-VDZ interaction energies for the L7 complexes.



FIG. 19. (a) Absolute and (b) percent DLPNO errors for RI-MP2/(jun-)cc-VDZ interaction energies for the S12 complexes. In the legend, the value of T_{CutPNO} is indicated in parentheses.

for **BBR**), although errors are smaller than 0.1 kcal/mol for some of the other complexes. Percent errors (Fig. S30) are actually on par with, or even smaller than, those observed using cc-pVDZ. However, two of the S12L complexes (**4a** and **4b**) exhibit DLPNO errors larger than 3 kcal/mol using normal-MP2 thresholds. More importantly, diffuse functions in def2-SVPD preclude extrapolation because they cause errors to oscillate as a function of T_{CutPNO} , just



FIG. 20. DLPNO errors at the RI-MP2/def2-SVPD level for (a) the L7 + **BBR** dataset and (b) the S12L dataset, using MP2 values of T_{CutPNO} , as indicated in parentheses in the legend.

The Journal of Chemical Physics



FIG. 21. DLPNO errors for the L7 + **BBR** dataset, for RI-MP2 calculations using (a) def2-SVP and (b) def2-SVPD. MP2 thresholds are used (as indicated by values of T_{CuIPNO} in the legend), and two extrapolations are also shown.

as we saw with jun-cc-pVDZ. This behavior is not observed in the def2-SVP basis set [Fig. 21(a)], just as it is not observed with cc-pVDZ, and is inherent to the use of diffuse functions.

This oscillatory behavior has a dramatic and detrimental effect on extrapolation to the canonical limit, as shown for L7 + BBR in Fig. 21. There, we compare the extrapolated DLPNO errors based on either def2-SVP or def2-SVPD data. (For the corresponding percentage errors, see Fig. S29.) The extrapolations based on def2-SVPD calculations are clearly unreliable and afford much larger errors than the input data, for every single complex in the dataset. This is not observed for def2-SVP, where the (L/N)_{MP2} and (N/T)_{MP2} extrapolations are typically (though not always) more accurate than the input data. An exception is the BBR complex, where the loose-MP2 calculation is of such low quality that L/N extrapolation deteriorates the quality of the normal-MP2 calculation, and N/T extrapolation degrades the tight-MP2 calculation albeit to a much lesser degree. Unlike the case of def2-SVPD, however, in the absence of diffuse functions the extrapolation procedure itself does not generate any outliers.

Together, these data strongly suggest that diffuse functions should be avoided in DLPNO calculations, at least in cases where canonical extrapolation ($T_{\text{CutPNO}} \rightarrow 0$) is required. This is potentially problematic for DLPNO calculations involving anions (e.g., recent ionic liquid benchmarks),^{41,110} for which diffuse functions are crucial. More generally, diffuse functions may also be important for obtaining converged non-covalent interaction energies.

To put this on a quantitative footing, we performed canonical RI-MP2/(aug-)cc-pVXZ calculations for the S66 dataset and then estimated the MP2/CBS limit by combining the Hartree–Fock/

TABLE VI. Error statistics for various	MP2/CBS	extrapolations,	averaged	over	the
S66 interaction energies.			•		

Extrapolation		Error (kcal/mo	ol) ^a
Scheme	MAE	Max	std. dev.
cc-pV[D/T]Z aug-cc-pV[D/T]Z cc-pV[T/Q]Z aug-cc-pV[T/Q]Z	0.27 0.07 0.04 0.01	0.81 0.39 0.23 0.06	$0.16 \\ 0.08 \\ 0.04 \\ 0.01$

^aRelative to MP2/CBS reference values from Ref. 26.

(aug-)cc-pVQZ energy with a two-point extrapolation of the correlation energy [Eq. (3)]. For nanoscale van der Waals complexes, our preferred extrapolation scheme would be cc-pV[D/T]Z, for reasons of computational cost, since this procedure requires neither quadruple- ζ calculations nor augmented basis sets. We test this against several other variants for S66 interaction energies.

The cc-pV[T/Q]Z and aug-cc-pV[T/Q]Z extrapolation schemes prove to be negligibly different (see Fig. S19), with a mean absolute difference of just 0.04 kcal/mol, suggesting that diffuse functions are unnecessary if cc-pVQZ calculations are feasible. Comparing the cc-pV[D/T]Z and aug-cc-pV[D/T]Z extrapolation procedures (Fig. S20), we observe differences ranging up to 0.7 kcal/mol (12% of $|\Delta E|$) and averaging 0.2 kcal/mol (0.9% of $|\Delta E|$), meaning that diffuse functions matter only at the sub-kcal/mol level. Table VI provides a statistical summary of all four extrapolation procedures in comparison to well-converged MP2/CBS benchmarks.²⁶ The most affordable extrapolation scheme (cc-pV[D/T]Z) yields sub-kcal/mol accuracy for every single S66 dimer, with an average deviation of 0.3 kcal/mol from MP2/CBS reference values.

This is notable given the considerable computational expense of augmenting the basis set. For the largest S66 system (pentane dimer), the RI-MP2/aug-cc-pVDZ calculation required 1.2 h on 12 processors compared to only 0.2 h for RI-MP2/cc-pVDZ on the same hardware. For small systems, this additional effort may or may not be justifiable to capture <1 kcal/mol of interaction energy, but for larger systems the addition of diffuse basis functions may simply be impossible. Regardless, the contribution of the diffuse functions to ΔE is likely extensive with system size,⁶⁰ and it remains to be determined whether diffuse functions are important in sizable van der Waals complexes. Our own experience with protein–ligand interaction energies computed using DFT suggests that this may be the case.^{50,60} However, more effort is needed along these lines in order to establish reliable computational protocols for non-covalent interactions in sizable supramolecular complexes.⁵⁰

IV. CONCLUSIONS

At the RI-MP2 level, the DLPNO approximation reproduces canonical intermolecular interaction energies with relatively small errors, even in percentage terms, but DLPNO errors increase with molecular size. Whereas "loose-MP2" PNO thresholds⁹⁵ ($T_{CutPNO} = 10^{-7}$) are faithful to within 3% percent of the canonical result for small dimers, much tighter thresholds ($T_{CutPNO} = 10^{-9}$) are necessary in order to come within 1–2 kcal/mol of the canonical RI-MP2 result for nanoscale van der Waals complexes with $\gtrsim 100$ atoms.

pubs.aip.org/aip/jcp

Alternatively, extrapolation to the canonical limit $(T_{CutPNO} \rightarrow 0)$ affords good results when only $T_{CutPNO} = 10^{-8}$ is feasible. This is true even for large complexes involving C_{60} and C_{70} with an equally large partner. Notably, counterpoise correction significantly reduces the DLPNO error for large systems and should be considered mandatory when DLPNO approximation is used for intermolecular interaction energies.

However, for systems such as $(C_{96}H_{24})_2$, which is a nanoscale model of π -stacking in graphene, the "loose/normal" $(L/N)_{MP2}$ extrapolation procedure (requiring only $T_{CutPNO} \ge 10^{-8}$) is insufficient to obtain benchmark-quality results, affording a DLPNO error of 5 kcal/mol. For this system, a "normal/tight" $(N/T)_{MP2}$ extrapolation (requiring $T_{CutPNO} = 10^{-9}$) must be used instead, if the DLPNO error is to be reduced below 1 kcal/mol or about 0.5% of $|\Delta E|$.

The conventional understanding of loose, normal, and tight thresholds is different for CCSD(T) than it is for MP2, and thresholds suggested as "loose" for MP2 calculations are comparable to what is considered "tight" for CCSD(T).^{94,95} DLPNO-CCSD(T) calculations for the S66 and L7 datasets, using either the semi-canonical T₀ or iterative T₁ approximations,¹² suggest that DLPNO errors at the CCSD(T) level are larger than those at the RI-MP2 level, even when the same set of PNO thresholds is used for both calculations. Moreover, whereas $T_{\text{CutPNO}} = 10^{-7}$ is suggested as the "tight" threshold for DLPNO-CCSD(T) calculations,⁹⁴ that choice affords errors of 2–6 kcal/mol for large π -stacked complexes in the L7 dataset, specifically those that involve the highly conjugated coronene and circumcoronene molecules.

Finally, we have shown that canonical extrapolation $(T_{\text{CutPNO}} \rightarrow 0)$ is often ill-posed in augmented basis sets because diffuse basis functions sometimes lead to errors that behave erratically as a function of T_{CutPNO}. Operationally, diffuse functions may reduce the absolute DLPNO error while simultaneously introducing behavior that precludes extrapolation to $T_{\text{CutPNO}} \rightarrow 0$. For small complexes, such as those in the S66 dataset, addition of diffuse functions modifies interaction energies by <5% upon two-point extrapolation to the MP2/CBS limit, and even the modest cc-pV[D/T]Z extrapolation procedure affords sub-kcal/mol accuracy, without requiring diffuse functions or quadruple- ζ basis sets. However, it is not yet clear whether this result generalizes to larger supramolecular complexes. More work is needed on this front, to decide whether diffuse functions can sensibly be avoided in benchmark-quality calculations involving large van der Waals complexes.

SUPPLEMENTARY MATERIAL

See the supplementary material for coordinates for the linear alkane, graphene, and graphane test systems created for this work, along with a significant number of additional figures analyzing the data, and also the raw data (electronic energies) themselves.

ACKNOWLEDGMENTS

This work was supported by the National Institute of General Medical Sciences of the National Institutes of Health under Award No. R43GM148095-01A1. M.G. further acknowledges a Presidential Fellowship from The Ohio State University. Calculations were performed at the Ohio Supercomputer Center.¹⁰⁴

AUTHOR DECLARATIONS

Conflict of Interest

The authors have no conflicts to disclose.

Author Contributions

Montgomery Gray: Conceptualization (contributing); Data curation (lead); Formal analysis (lead); Investigation (lead); Methodology (equal); Validation (lead); Visualization (equal); Writing – original draft (lead); Writing – review & editing (equal). **John M. Herbert:** Conceptualization (lead); Formal analysis (contributing); Funding acquisition (lead); Methodology (equal); Project administration (lead); Supervision (lead); Visualization (equal); Writing – review & editing (equal).

DATA AVAILABILITY

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

REFERENCES

¹J. D. Watts, "Chemical reactions: Thermochemical calculations," in *Handbook of Computational Chemistry*, edited by J. Leszczynski (Springer, Dordrecht, 2012), Chap. 13, pp. 467–482.

² A. Karton, "Quantum mechanical thermochemical predictions 100 years after the Schrödinger equation," Annu. Rep. Comput. Chem. 18, 123–166 (2022).

³J. Řezáč and P. Hobza, "Describing noncovalent interactions beyond the common approximations: How accurate is the 'gold standard,' CCSD(T) at the complete basis set limit?," J. Chem. Theory Comput. **9**, 2151–2155 (2013).

⁴K. Patkowski, "Benchmark databases of intermolecular interaction energies: Design, construction, and significance," Annu. Rep. Comput. Chem. **13**, 3–91 (2017).

⁵M. Kodrycka and K. Patkowski, "Platinum, gold, and silver standards of intermolecular interaction energy calculations," J. Chem. Phys. **151**, 070901 (2019).

⁶F. Neese, A. Hansen, and D. G. Liakos, "Efficient and accurate approximations to the local coupled cluster singles doubles method using a truncated pair natural orbital basis," J. Chem. Phys. **131**, 064103 (2009).

⁷C. Riplinger and F. Neese, "An efficient and near linear scaling pair natural orbital based local coupled cluster method," J. Chem. Phys. **138**, 034106 (2013).

⁸C. Riplinger, B. Sandhoefer, A. Hansen, and F. Neese, "Natural triple excitations in local coupled cluster calculations with pair natural orbitals," J. Chem. Phys. 139, 134101 (2013).

⁹P. Pinski, C. Riplinger, E. F. Valeev, and F. Neese, "Sparse maps—A systematic infrastructure for reduced-scaling electronic structure methods. I. An efficient and simple linear scaling local MP2 method that uses an intermediate basis of pair natural orbitals," J. Chem. Phys. **143**, 034108 (2015).

¹⁰C. Riplinger, P. Pinski, U. Becker, E. F. Valeev, and F. Neese, "Sparse maps—A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory," J. Chem. Phys. **144**, 024109 (2016).

¹¹F. Pavošević, C. Peng, P. Pinski, C. Riplinger, F. Neese, and E. F. Valeev, "SparseMaps—A systematic infrastructure for reduced scaling electronic structure methods. V. Linear scaling explicitly correlated coupled-cluster method with pair natural orbitals," J. Chem. Phys. **146**, 174108 (2017).

¹²Y. Guo, C. Riplinger, U. Becker, D. G. Liakos, Y. Minenkov, L. Cavallo, and F. Neese, "Communication: 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. **148**, 011101 (2018).

¹³Y. Guo, C. Riplinger, D. G. Liakos, U. Becker, M. Saitow, and F. Neese, "Linear scaling perturbative triples correction approximations for open-shell domain-based local pair natural orbital coupled cluster singles and doubles theory [DLPNO-CCSD(T_0/T)]," J. Chem. Phys. **152**, 024116 (2020).

¹⁴W. Li, Z. Ni, and S. Li, "Cluster-in-molecule local correlation method for post-Hartree-Fock calculations of large systems," Mol. Phys. **114**, 1447–1460 (2016).

¹⁵Y. Guo, U. Becker, and F. Neese, "Comparison and combination of 'direct' and fragment based local correlation methods: Cluster in molecules and domain based local pair natural orbital perturbation and coupled cluster theories," J. Chem. Phys. **148**, 124117 (2018).

¹⁶G. Schmitz and C. Hättig, "Perturbative triples correction for local pair natural orbital based explicitly correlated CCSD(F12)^{*} using Laplace transformation techniques," J. Chem. Phys. **145**, 234107 (2016).

¹⁷G. Schmitz and C. Hättig, "Accuracy of explicitly correlated local PNO-CCSD(T)," J. Chem. Theory Comput. 13, 2623–2633 (2017).

¹⁸M. Schwilk, Q. Ma, C. Köppl, and H.-J. Werner, "Scalable electron correlation methods. 3. Efficient and accurate parallel local coupled cluster with pair natural orbitals (PNO-LCCSD)," J. Chem. Theory Comput. **13**, 3650–3675 (2017).

¹⁹Q. Ma and H.-J. Werner, "Scalable electron correlation methods. 5. Parallel perturbative triples correction for explicitly correlated local coupled cluster with pair natural orbitals," J. Chem. Theory Comput. 14, 198–215 (2018).

²⁰Q. Ma and H.-J. Werner, "Explicitly correlated local coupled-cluster methods using pair natural orbitals," WIREs Comput. Mol. Sci. 8, e1371 (2018).

²¹Q. Ma and H.-J. Werner, "Scalable electron correlation methods. 7. Local open-shell coupled-cluster methods using pair natural orbitals: PNO-RCCSD and PNO-UCCSD," J. Chem. Theory Comput. **16**, 3135–3151 (2020).

²²Q. Ma and H.-J. Werner, "Scalable electron correlation methods. 8. Explicitly correlated open-shell coupled-cluster with pair natural orbitals PNO-RCCSD(T)-F12 and PNO-UCCSD(T)-F12," J. Chem. Theory Comput. **17**, 902–926 (2021).

²³ P. R. Nagy and M. Kállay, "Optimization of the linear-scaling local natural orbital CCSD(T) method: Redundancy-free triples correction using Laplace transform," J. Chem. Phys. 146, 214106 (2017).
 ²⁴ P. R. Nagy, G. Samu, and M. Kállay, "Optimization of the linear-scaling

²⁴P. R. Nagy, G. Samu, and M. Kállay, "Optimization of the linear-scaling local natural orbital CCSD(T) method: Improved algorithm and benchmark applications," J. Chem. Theory Comput. 14, 4193–4215 (2018).

²⁵P. R. Nagy and M. Kállay, "Approaching the basis set limit of CCSD(T) energies for large molecules with local natural orbital coupled-cluster methods," J. Chem. Theory Comput. 15, 5275–5298 (2019).

²⁶ P. R. Nagy, L. Gyevi-Nagy, B. D. Lörincz, and M. Kállay, "Pursuing the basis set limit of CCSD(T) non-covalent interaction energies for medium-sized complexes: Case study on the S66 compilation," Mol. Phys. **121**, e2109526 (2022).

²⁷A. Jiang, Z. L. Glick, D. Poole, J. M. Turney, C. D. Sherrill, and H. F. Schaefer III, "Accurate and efficient open-source implementation of domain-based local pair natural orbital (DLPNO) coupled-cluster theory using a t₁-transformed Hamiltonian," chemRxiv:2024-cvs8h (2024).

²⁸M. Sparta and F. Neese, "Chemical applications carried out by local pair natural orbital based coupled-cluster methods," Chem. Soc. Rev. 43, 5032–5041 (2014).

²⁹D. Folmsbee and G. Hutchison, "Assessing conformer energies using electronic structure and machine learning methods," Int. J. Quantum Chem. **121**, e26381 (2021).

³⁰I. Stylianakis, N. Zervos, J.-H. Lii, D. A. Pantazis, and A. Kolocouris, "Conformational energies of reference organic molecules: Benchmarking of common efficient computational methods against coupled cluster theory," J. Comput.-Aided Mol. Des. **37**, 607–656 (2023).

³¹F. Neese, F. Wennmohs, U. Becker, and C. Riplinger, "The ORCA quantum chemistry program package," J. Chem. Phys. **152**, 224108 (2020).

³²D. G. Liakos, M. Sparta, M. K. Kesharwani, J. M. L. Martin, and F. Neese, "Exploring the accuracy limits of local pair natural orbital coupled-cluster theory," J. Chem. Theory Comput. 11, 1525–1539 (2015).

³³Y. Minenkov, E. Chermak, and L. Cavallo, "Accuracy of DLPNO-CCSD(T) method for noncovalent bond dissociation enthalpies from coinage metal cation complexes," J. Chem. Theory Comput. **11**, 4664–4676 (2015).

³⁴Y. Minenkov, H. Wang, Z. Wang, S. M. Sarathy, and L. Cavallo, "Heats of formation of medium-sized organic compounds from contemporary electronic structure methods," J. Chem. Theory Comput. 13, 3537–3560 (2017). ³⁵Y. Minenkov, G. Bistoni, C. Riplinger, A. A. Auer, F. Neese, and L. Cavallo, "Pair natural orbital and canonical coupled cluster reaction enthalpies involving light to heavy alkali and alkaline earth metals: The importance of sub-valence correlation," Phys. Chem. Chem. Phys. **19**, 9374–9391 (2017).

³⁶B. Maity, Y. Minenkov, and L. Cavallo, "Evaluation of experimental alkali metal ion-ligand noncovalent bond strengths with DLPNO-CCSD(T) method," J. Chem. Phys. 151, 014301 (2019).

³⁷D. G. Liakos, Y. Guo, and F. Neese, "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 **124**, 90-100 (2020).

³⁸A. Altun, F. Neese, and G. Bistoni, "Extrapolation to the limit of a complete pair natural orbital space in local coupled-cluster calculations," J. Chem. Theory Comput. 16, 6142–6149 (2020).

³⁹G. Schmitz and J. Elm, "Assessment of the DLPNO binding energies of strongly noncovalent bonded atmospheric molecular clusters," ACS Omega 5, 7601–7612 (2020).

⁴⁰I. Sandler, J. Chen, M. Taylor, S. Sharma, and J. Ho, "Accuracy of DLPNO-CCSD(T): Effect of basis set and system size," J. Phys. Chem. A **125**, 1553–1563 (2021).

⁴¹Z. L. Seeger and E. I. Izgorodina, "A DLPNO-CCSD(T) benchmarking study of intermolecular interactions of ionic liquids," J. Comput. Chem. **43**, 106–120 (2022).

⁴²A. Altun, C. Riplinger, F. Neese, and G. Bistoni, "Exploring the accuracy limits of PNO-based local coupled-cluster calculations for transition-metal complexes," J. Chem. Theory Comput. **19**, 2039–2047 (2023).

⁴³ J. Chen, T. Sun, Y. Wang, and W. Wang, "Toward a less costly but accurate calculation of the CCSD(T)/CBS noncovalent interaction energy," J. Comput. Chem. 41, 1252–1260 (2020).

⁴⁴Z. Ni, Y. Guo, F. Neese, W. Li, and S. Li, "Cluster-in-molecule local correlation method with an accurate distant pair correction for large systems," J. Chem. Theory Comput. 17, 756–766 (2021).

⁴⁵E. Brémond, H. Li, J. C. Sancho-García, and C. Adamo, "Double hybrids and noncovalent interactions: How far can we go?," J. Phys. Chem. A **126**, 2590–2599 (2022).

⁴⁶K. Sorathia and D. P. Tew, "Basis set extrapolation in pair natural orbital theories," J. Chem. Phys. **153**, 174112 (2020).

⁴⁷Y. S. Al-Hamdani, P. R. Nagy, A. Zen, D. Barton, M. Kállay, J. G. Brandenburg, and A. Tkatchenko, "Interactions between large molecules pose a puzzle for reference quantum mechanical methods," Nat. Commun. **12**, 3927 (2021).

⁴⁸F. Ballesteros, S. Dunivan, and K. U. Lao, "Coupled cluster benchmarks of large noncovalent complexes: The L7 dataset as well as DNA-ellipticine and buckycatcher-fullerene," J. Chem. Phys. **154**, 154104 (2021).

 49 C. Villot, F. Ballesteros, D. Wang, and K. U. Lao, "Coupled cluster benchmarking of large noncovalent complexes in L7 and S12L as well as the C_{60} dimer, DNA-ellipticine, and HIV-indinavir," J. Phys. Chem. A **126**, 4326–4341 (2022).

⁵⁰ M. Gray and J. M. Herbert, "Density functional theory for van der Waals complexes: Size matters," Annu. Rep. Comput. Chem. **20**, 1–61 (2024).

⁵¹ M. Kállay, P. R. Nagy, D. Mester, Z. Rolik, G. Samu, J. Csontos, J. Csóka, P. B. Szabó, L. Gyevi-Nagy, B. Hégely, I. Ladjánszki, B. Ladóczki, B. Ladóczki, L. Szegedy, M. Farkas, P. D. Mezei, and A. Ganyecz, "The MRCC program system: Accurate quantum chemistry from water to proteins," J. Chem. Phys. **152**, 074107 (2020).

⁵²C. J. Nickerson, K. R. Bryenton, A. J. A. Price, and E. R. Johnson, "Comparison of density-functional theory dispersion corrections for the DES15K database," J. Phys. Chem. A **127**, 8712–8722 (2023).

⁵³J. Hermann, R. A. DiStasio, Jr., and A. Tkatchenko, "First-principles models for van der Waals interactions in molecules and materials: Concepts, theory, and applications," Chem. Rev. 117, 4714–4758 (2017).

⁵⁴K. U. Lao and J. M. Herbert, "Accurate and efficient quantum chemistry calculations of noncovalent interactions in many-body systems: The XSAPT family of methods," J. Phys. Chem. A **119**, 235–253 (2015).

⁵⁵K. Carter-Fenk, K. U. Lao, K.-Y. Liu, and J. M. Herbert, "Accurate and efficient *ab initio* calculations for supramolecular complexes: Symmetry-adapted

J. Chem. Phys. **161**, 054114 (2024); doi: 10.1063/5.0206533 Published under an exclusive license by AIP Publishing perturbation theory with many-body dispersion," J. Phys. Chem. Lett. 10, 2706-2714 (2019).

⁵⁶K. Carter-Fenk, K. U. Lao, and J. M. Herbert, "Predicting and understanding non-covalent interactions using novel forms of symmetry-adapted perturbation theory," Acc. Chem. Res. 54, 3679-3690 (2021).

⁵⁷H. Neugebauer, P. Pinski, S. Grimme, F. Neese, and M. Bursch, "Assessment of DLPNO-MP2 approximations in double-hybrid DFT," J. Chem. Theory Comput. 19, 7695-7703 (2023).

⁵⁸J. Witte, J. B. Neaton, and M. Head-Gordon, "Push it to the limit: Characterizing the convergence of common sequences of basis sets for intermolecular interactions as described by density functional theory," J. Chem. Phys. 144, 194306 (2016).

⁵⁹M. Gray and J. M. Herbert, "Comprehensive basis-set testing of extended symmetry-adapted perturbation theory and assessment of mixed-basis combinations to reduce cost," J. Chem. Theory Comput. 18, 2308-2330 (2022).

⁶⁰M. Gray, P. E. Bowling, and J. M. Herbert, "Systematic evaluation of counterpoise correction in density functional theory," J. Chem. Theory Comput. 18, 6742-6756 (2022).

⁶¹ R. Sedlak, T. Janowski, M. Pitoňák, J. Řezáč, P. Pulay, and P. Hobza, "Accuracy of quantum chemical methods for large noncovalent complexes," J. Chem. Theory Comput. 9, 3364-3374 (2013).

⁶²S. Grimme, "Supramolecular binding thermodynamics by dispersion-corrected density functional theory," Chem. - Eur. J. 18, 9955-9964 (2012).

⁶³T. Risthaus and S. Grimme, "Benchmarking of London dispersion-accounting density functional theory methods on very large molecular complexes," J. Chem. Theory Comput. 9, 1580–1591 (2013).

⁶⁴J. Řezác, K. E. Riley, and P. Hobza, "S66: A well-balanced database of benchmark interaction energies relevant to biomolecular structures," J. Chem. Theory Comput. 7, 2427–2438 (2011); Erratum 10, 1359–1360 (2014).

⁶⁵R. M. Richard, K. U. Lao, and J. M. Herbert, "Achieving the CCSD(T) basis-set limit in sizable molecular clusters: Counterpoise corrections for the many-body expansion," J. Phys. Chem. Lett. 4, 2674-2680 (2013).

⁶⁶R. M. Richard, K. U. Lao, and J. M. Herbert, "Approaching the complete-basis limit with a truncated many-body expansion," J. Chem. Phys. 139, 224102 (2013). ⁶⁷L. A. Burns, M. S. Marshall, and C. D. Sherrill, "Comparing counterpoisecorrected, uncorrected, and averaged binding energies for benchmarking noncovalent interactions," J. Chem. Theory Comput. 10, 49-57 (2014).

⁶⁸M. Schütz and H.-J. Werner, "Local perturbative triples correction (T) with linear cost scaling," Chem. Phys. Lett. 318, 370-378 (2000).

⁶⁹M. Schütz, "Low-order scaling local electron correlation methods. III. Linear scaling local perturbative triples correction (T)," J. Chem. Phys. 113, 9986-10001 (2000).

⁷⁰T. Kawase, K. Tanaka, N. Fujiwara, H. R. Darabi, and M. Oda, "Complexation of a carbon nanoring with fullerenes," Angew. Chem., Int. Ed. Engl. 115, 1662-1666 (2003).

⁷¹ M. Birowska, K. Milowska, and J. A. Majewski, "Van der Waals density functionals for graphene layers and graphite," Acta Phys. Pol., A 120, 845-848 (2011).

⁷²A. V. Rozhkov, A. O. Sboychakov, A. L. Rakhmanov, and F. Nori, "Electronic properties of graphene-based bilayer systems," Phys. Rep. **648**, 1–104 (2016). ⁷³ M. Gray and J. M. Herbert, "Origins of offset-stacking in porous frameworks,"

J. Phys. Chem. C 127, 2675–2686 (2023).

⁷⁴K. Carter-Fenk and J. M. Herbert, "Electrostatics does not dictate the slipstacked arrangement of aromatic π - π interactions," Chem. Sci. 11, 6758–6765 (2020).

⁷⁵R. E. Franklin, "The structure of graphitic carbons," Acta Crystallogr. 4, 253–261 (1951).

⁷⁶J. O. Sofo, A. S. Chaudhari, and G. D. Barber, "Graphane: A two-dimensional hydrocarbon," Phys. Rev. B 75, 153401 (2007).

 77 A. A. Fokin, D. Gerbig, and P. R. Schreiner, " σ/σ - and π/π -interactions are equally important: Multilayered graphanes," J. Am. Chem. Soc. 133, 20036-20039 (2011).

⁷⁸T. Janowski and P. Pulay, "A benchmark comparison of σ/σ and π/π dispersion: The dimers of naphthalene and decalin, and coronene and perhydrocoronene," J. Am. Chem. Soc. 134, 17520-17525 (2012).

⁷⁹K. Carter-Fenk and J. M. Herbert, "Reinterpreting π -stacking," Phys. Chem. Chem. Phys. 22, 24870-24886 (2020).

⁸⁰T. H. Dunning, Jr., "Gaussian basis sets for use in correlated molecular calculations. I. The atoms boron through neon and hydrogen," J. Chem. Phys. 90, 1007-1023 (1989).

⁸¹D. E. Woon and T. H. Dunning, Jr., "Gaussian basis sets for use in correlated molecular calculations. IV. Calculation of static electrical response properties," J. Chem. Phys. 100, 2975-2988 (1994).

82 E. Papajak, J. Zheng, X. Xu, H. R. Leverentz, and D. G. Truhlar, "Perspectives on basis sets beautiful: Seasonal plantings of diffuse basis functions," J. Chem. Theory Comput. 7, 3027-3034 (2011).

83 F. Weigend and R. Ahlrichs, "Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy," Phys. Chem. Chem. Phys. 7, 3297-3305 (2005).

⁸⁴D. Rappoport and F. Furche, "Property-optimized Gaussian basis sets for molecular response calculations," J. Chem. Phys. 133, 134105 (2010).

85 F. Weigend, A. Köhn, and C. Hättig, "Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations," J. Chem. Phys. 116, 3175-3183 (2002).

⁸⁶P. Jurečka, J. Šponer, J. Černý, and P. Hobza, "Benchmark database of accurate (MP2 and CCSD(T) complete basis set limit) interaction energies of small model complexes, DNA base pairs, and amino acid pairs," Phys. Chem. Chem. Phys. 8, 1985-1993 (2006).

87 A. Tkatchenko, R. A. DiStasio, Jr., M. Head-Gordon, and M. Scheffler, "Dispersion-corrected Møller-Plesset second-order perturbation theory," J. Chem. Phys. 131, 094106 (2009).

⁸⁸D. Cremer, "Møller–Plesset perturbation theory: From small molecule methods to methods for thousands of atoms," WIREs Comput. Mol. Sci. 1, 509-530 (2011). 89 B. Nguyen, G. P. Chen, M. M. Agee, A. M. Burow, M. Tang, and F. Furche, "Divergence of many-body perturbation theory for noncovalent interactions of large molecules," J. Chem. Theory Comput. 16, 2258-2273 (2020).

90 M. S. Marshall, L. A. Burns, and C. D. Sherrill, "Basis set convergence of the coupled-cluster correction, $\delta_{MP2}^{CCSD(T)}$: Best practices for benchmarking noncovalent interactions and the attendant revision of the S22, NBC10, and HSG databases," J. Chem. Phys., 135, 194102 (2011).

⁹¹ J. Pogrebetsky, A. Siklitskaya, and A. Kubas, "MP2-based correction scheme to approach the limit of a complete pair natural orbitals space in DLPNO-CCSD(T) calculations," J. Chem. Theory Comput. 19, 4023-4032 (2023).

⁹²A. G. Császár and W. D. Allen, "The composite focal-point analysis (FPA) approach," in Molecular Quantum Mechanics: From Methylene to DNA and beyond, Selected Papers of Henry, edited by F. Schaefer, R. J. Bartlett, T. D. Crawford, M. Head-Gordon and C. D. Sherrill (Brandon's Printing, Atlanta, 2010), Chap. 23, pp. 261-265.

93 M. Gray, P. E. Bowling, and J. M. Herbert, "In defense of (certain) Pople-type basis sets," chemRxiv:2024-jz473 (2024).

94 ORCA user manual, v. 5.0.4, https://www.kofo.mpg.de/970316, accessed 05-07-2024.

95 We strongly dislike the idea of ascribing names like "tight" or "loose" to numerical thresholds. Such terminology unduly insulates users from numerical details of the algorithm, leading to overreliance upon (and overconfidence in) softwarespecific default values. These defaults may change without warning in new versions of the software, so explicating the actual numerical values of all thresholds is also important for reproducibility. Adding to the confusion, OCRA uses "loose," "normal," and "tight" to mean different numerical settings for DLPNO-MP2 as compared to DLPNO-CCSD(T).94 Furthermore, the loose/normal/tight nomenclature has been adopted in altogether different implementations of local CCSD(T),^{25,27} with no guarantee that the meaning is the same in each case. We use this terminology here, grudgingly, because it has become ubiquitous. However, we have carefully explicated the values of T_{CutDO}, T_{CutPNO}, T_{cutMKN}, and T_{CutPairs} that are used; see Table I.

96 A. I. Krylov, J. M. Herbert, F. Furche, M. Head-Gordon, P. J. Knowles, R. Lindh, F. R. Manby, P. Pulay, C.-K. Skylaris, and H.-J. Werner, "What is the price of open-source software?," J. Phys. Chem. Lett. 6, 2751-2754 (2015).

⁹⁷ As stated in the ORCA user manual,⁹⁴ "All DLPNO-CCSD truncations are bound to these three truncation parameters [T_{CutPNO}, T_{CutPairs}, and T_{CutMKN}]

J. Chem. Phys. 161, 054114 (2024); doi: 10.1063/5.0206533 Published under an exclusive license by AIP Publishing

and should not almost be touched (Hence they are also not documented :-)." (Emoticon in the original).

⁹⁸A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olsen, and A. K. Wilson, "Basis-set convergence in correlated calculations on Ne, N₂, and H₂O," Chem. Phys. Lett. **286**, 243–252 (1998).

⁹⁹R. Ahlrichs, S. D. Elliott, and U. Hunlar, "*Ab initio* treatment of large molecules," in *Modern Methods and Algorithms of Quantum Chemistry, NIC Series*, edited by J. Grotendorst, 2nd ed. (John von Neumann Institute for Computing Jülich, 2000), Vol. 3, pp. 7–25.

¹⁰⁰ V. Fishman, E. Semidalas, and J. M. L. Martin, "Basis set extrapolation from the vanishing counterpoise correction condition," arXiv:2405.04658 (2024).
 ¹⁰¹ B. Brauer, M. K. Kesharwani, and J. M. L. Martin, "Some observations on counterpoise corrections for explicitly correlated calculations on noncovalent interactions," J. Chem. Theory Comput. 10, 3791–3799 (2014).

102 J. Gorges, B. Bädorf, A. Hansen, and S. Grimme, "Efficient computation of the interaction energies of very large non-covalently bound complexes," Synlett 34, 1135–1146 (2023).

¹⁰³ A. Sygula, F. R. Fronczek, R. Sygula, P. W. Rabideau, and M. M. Olmstead, "A double concave hydrocarbon buckycatcher," J. Am. Chem. Soc. **129**, 3842–3843 (2007). 104 Ohio Supercomputer Center, http://osc.edu/ark:/19495/f5s1ph73 (accessed 06-07-2024).

¹⁰⁵G. J. O. Beran, C. C. Cook, and P. A. Unzueta, "Contrasting conformational behaviors of molecules XXXI and XXXII in the seventh blind test of crystal structure prediction," Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater. (in press).

¹⁰⁶Z. Wang, A. Aldossary, T. Shi, Y. Liu, X. S. Li, and M. Head-Gordon, "Local second-order Møller–Plesset theory with a single threshold using orthogonal virtual orbitals: Theory, implementation, and assessment," J. Chem. Theory Comput. **19**, 7577–7591 (2023).

¹⁰⁷J. C. Fetzer, W. R. Biggs, and M. Zander, "Notizen: The preparation and purification of perhydrocoronene for use as a spectroscopic matrix," Z. Naturforsch. A 46, 291–292 (1991).

¹⁰⁸J. F. Dobson, "Beyond pairwise additivity in London dispersion interactions," Int. J. Quantum Chem. **114**, 1157–1161 (2014).

¹⁰⁹K. U. Lao and J. M. Herbert, "A simple correction for nonadditive dispersion within extended symmetry-adapted perturbation theory (XSAPT)," J. Chem. Theory Comput. **14**, 5128–5142 (2018).

¹¹⁰K. Carter-Fenk and J. M. Herbert, "Appraisal of dispersion damping functions for the effective fragment potential method," Mol. Phys. **121**, e2055504 (2023).