

Linear-scaling local natural orbital CCSD(T) approach for open-shell systems: algorithm, benchmarks, and large-scale applications

P. Bernát Szabó,[†] József Csóka,^{†,‡,¶} Mihály Kállay,^{†,‡,¶} and Péter R. Nagy^{*,†,‡,¶}

[†]*Department of Physical Chemistry and Materials Science, Faculty of Chemical Technology and Biotechnology, Budapest University of Technology and Economics, Műegyetem rkp. 3., H-1111 Budapest, Hungary*

[‡]*HUN-REN-BME Quantum Chemistry Research Group, Műegyetem rkp. 3., H-1111 Budapest, Hungary*

[¶]*MTA-BME Lendület Quantum Chemistry Research Group, Műegyetem rkp. 3., H-1111 Budapest, Hungary*

E-mail: nagy.peter@vbk.bme.hu

Abstract

The extension of the highly-optimized local natural orbital (LNO) coupled-cluster (CC) with single-, double-, and perturbative triple excitations [LNO-CCSD(T)] method is presented for high-spin open-shell molecules based on restricted open-shell references. The techniques enabling the outstanding efficiency of the closed-shell LNO-CCSD(T) variant are adopted, including the iteration- and redundancy-free MP2 and (T) formulations, as well as the integral-direct, memory- and disk use economic, and OpenMP-parallel algorithms. For large molecules, the efficiency of our open-shell LNO-CCSD(T) method approaches that of its closed-shell parent method due to the application of restricted orbital sets for demanding integral transformations and a novel approximation

for higher-order long-range spin-polarization effects. The accuracy of open-shell LNO-CCSD(T) is extensively tested for radicals and reactions thereof, ionization processes, as well as spin-state splittings and transition-metal compounds. At the size range, where the canonical CCSD(T) reference is accessible (up to 20–30 atoms) the average open-shell LNO-CCSD(T) correlation energies are found to be 99.9–99.95% accurate, which translates into average absolute deviations of a few tenth of a kcal/mol in the investigated energy differences already with the default settings. For more extensive molecules the local errors may grow but they can be estimated and decreased via affordable systematic convergence studies. This enables the accurate modeling of large systems with complex electronic structure, as illustrated on open-shell organic radicals and transition metal complexes of up to 179 atoms, as well as on challenging biochemical systems, including up to 601 atoms and 11,000 basis functions. While the protein models involve difficulties for local approximations, such as the spin states of a bounded iron ion or an extremely delocalized singly occupied orbital, the corresponding single-node LNO-CCSD(T) computations were feasible in a matter of days with 10s to a 100 GB of memory use. Therefore, the new LNO-CCSD(T) implementation enables highly-accurate computations for open-shell systems of unprecedented size and complexity with widely accessible hardware.

1 Introduction

The accurate modeling of open-shell species remains challenging due to their potentially complicated electronic structure. Among those, the systems of interest here exhibit a high-spin open-shell ground state wave function, which can usually be described by single-reference quantum chemical methods. Such a wide range of systems include radicals appearing as stable species, or intermediates and transition states of reactions, products of ionization or electron attachment processes, etc. often relevant in redox, atmospheric, polymer, combustion, astro-, and electrochemistry, just to name a few representative fields.¹

While density functional theory (DFT) based methods remain the workhorse of computational chemistry due to their relatively affordable cost, cases posing challenges for DFT approaches occur more often for open-shell than for closed-shell systems. Therefore both the necessity and the difficulties of utilizing the wave function based treatment of electron correlation were extensively studied also for open-shell species.²⁻⁴ While the second-order Møller–Plesset (MP2)⁵ approach remains popular due to its cost-efficiency and relative simplicity, coupled-cluster (CC) methods represent a faster converging series to the chemically accurate (within 1 kcal/mol) description of processes of single-reference species.^{6,7} In particular, the “gold standard” CC model with single, double (CCSD), and perturbative triple excitations [CCSD(T)],⁸ offers a good compromise between computational-cost and robust accuracy.

Still, the steep $\mathcal{O}(\mathcal{N}^4)$ - and $\mathcal{O}(\mathcal{N}^7)$ -scaling data storage and operation count complexity of CCSD(T) with system size \mathcal{N} limits its applicability range to molecules of up to 20–25 atoms. This is an even more severe problem for open-shell species, where unrestricted CC formalisms require the solution of about three times as many equations as their restricted counterparts. Moreover, the higher technical complexity of the unrestricted CC methods also slows down the adaptation of new approaches proposed frequently only for the spin-restricted case. For example, density-fitting (DF, or resolution-of-identity) approaches can help to deal with the $\mathcal{O}(\mathcal{N}^4)$ -scaling storage complexity,⁹⁻¹³ while efficient parallelization can reduce the wall time (but not the scaling) of CCSD(T).¹¹⁻¹⁷ However, even the combination of these techniques with additional advancements in our recent integral-direct, parallel DF-CCSD(T) implementation pushes the limits of conventional CCSD(T) computations only up to 30 atoms.¹³

This motivates the introduction of reduced-scaling approximations, such as the robust frozen natural orbital (NO) approach,^{18,19} which can extend the applicability range of NO-CCSD(T) somewhat further.^{20,21} Alternatively, one can also accelerate the basis set convergence via explicitly correlated (F12) CC methods²²⁻²⁴ leading to more compact atomic

orbital (AO) basis set requirements. However, even our recent combination of the NO approach and an optimized F12 implementation²⁵ allowed us to approach the complete basis set (CBS) limit for closed-shell species of up to 50 atoms with an even smaller open-shell limit of about 30–35 atoms.²⁶

At this size range, it starts to be beneficial to take advantage of the rapid decay of electron correlation with the distance via local correlation approaches.^{27–29} Current methods still build on fundamental techniques pioneered by Pulay and Saebø,^{30,31} such as the approximation of the energy contribution of distant localized molecular orbital (LMO) pairs (pair approximation), and the restriction of the correlating orbital spaces to a spatially compact lists surrounding the strongly interacting LMOs (domain approximation). One group of methods then approximates or neglects the distant pair interactions leading to a number of decoupled subsystem MP2 or CC equations to be solved, but here, it remains challenging to deal with the significant overlap of the subsystems.^{27,28,32,33} Advanced representatives of this strategy also developed up to the CC level include the cluster-in-molecule (CIM) approach of Li, Li, Piecuch, Guo, and their co-workers,^{34–36} the divide-expand-consolidate (DEC) approach of Jørgensen, Kristensen, and their co-workers,^{37,38} the divide-and-conquer (DC) method of Li and Li³⁹ and Kobayashi and Nakai,^{40,41} the incremental method proposed by Stoll⁴² and further employed in the local correlation context by Friedrich and Dolg.^{43–45}

One can also take advantage of the wave function sparsity not only in real space but also via NO-based data compression approaches.^{46–53} Among these, the LMO pair specific pair natural orbitals (PNOs) are employed in the domain based local PNO (DLPNO) method of Neese, Valeev and co-workers,^{47,48,54–56} but PNO-based approaches were also developed up to the CCSD(T) level by Werner and Ma,^{49,50,57} as well as by Hättig and Tew.⁵¹ Compared to those, our local natural orbital (LNO) family of methods employs an LMO-specific NO set compressing both the occupied and virtual orbital spaces.^{13,52,53,58–62}

Due to the considerable challenges associated with open-shell systems and unrestricted CC formalisms, fewer local CCSD(T) methods are available for systems other than those

of a closed-shell singlet electronic structure. The incremental scheme was extended to both unrestricted Hartree–Fock (UHF)⁶³ as well as restricted open-shell HF (ROHF)⁶⁴ references. Additionally, the high-spin open-shell variants of the PNO-L methods by Werner and Ma,^{50,57} as well as the DLPNO method by Neese, Valeev, Hansen, Saitow, Guo, Kumar, and co-workers^{47,48,55,56} were also introduced recently, while here, we present the high-spin open-shell extension of our LNO-CCSD(T) approach.

To that end, here, we combine the attractive properties of two lines of recent developments within our restricted LNO-CCSD(T)^{52,53,62} as well as our open-shell local MP2 (LMP2)⁶⁵ schemes. The outstanding efficiency of these approaches originates from the Laplace-transform based, redundancy-free evaluation of the amplitudes both at the LMP2 and the (T) level.^{62,65} Moreover, the CCSD contribution is also obtained in the compact LNO space, which was further accelerated via our highly-optimized CCSD implementation designed also for the unconventional ratio of the occupied and virtual orbital dimensions occurring in the LNO basis.¹³ The resulting LMP2 and LNO-CCSD(T) algorithms are fully ab initio, i.e., free from empirical or distance based cutoff parameters, manual fragment definitions, bond cutting and capping, etc. usually associated with local correlation approaches. In fact, the LNO approximations are defined completely automatically and adopt to the complexity of the wave function of the systems. Exploiting this property, we designed an LNO approximation hierarchy of threshold combinations (e.g., Normal, Tight, ...), which form a systematically convergent series usually also suitable for extrapolation toward conventional CCSD(T) and for providing a conservative LNO error estimate.⁵³ Further unique features of our LMP2 and LNO-CCSD(T) methods include the exceptionally small memory and disk use, checkpointing, utilization of point group symmetry (even non-Abelian) and treatment of near-linear dependent AO basis sets.^{52,53,66} These capabilities were all required in our largest local CCSD(T) computation so far performed for extended supramolecular complexes⁶⁷ as well as for a protein of 1023 atoms with almost 45,000 AOs in a quadruple- ζ basis set.⁵³ Even larger systems can be targeted via various embedding approaches combining LNO-CC

based chemically active regions embedded into lower-level LNO-CC, LMP2, DFT and/or molecular mechanics (MM) environments.^{68,69}

An important goal in the generalization of the above spin-restricted local correlation methods to open-shell systems is to retain as much as possible the computational efficiency of the closed-shell scheme. Therefore, in our open-shell LMP2 method⁶⁵ as well as here for the LNO-CCSD(T) case, we employ a restricted open-shell (RO) reference determinant, RO LMO set, and RO intermediate basis sets used for the costly integral transformation steps. Interestingly, this strategy is implemented in three completely different manners in the two PNO-based and our approach. The DLPNO method employs an n -electron valence state perturbation theory Ansatz,⁴⁷ the PNO-L methods utilize a spin-adapted MP2 formulation,⁵⁷ while we employ a spin-restricted integral transformation combined with a simpler ROHF-based but unrestricted MP2 Ansatz.^{70,71} Building on that, here we construct a restricted LNO basis. However, due to the properties of the perturbative triples corrections, in the end, both the PNO and LNO methods have to use unrestricted formulae for the CCSD(T) part.

To enable the cost of the unrestricted CCSD(T) calculations of the LNO scheme to approach that of the closed-shell method, at least in the asymptotic limit, we utilize an additional approximation that we developed for our open-shell LMP2 approach.⁶⁵ Briefly, we can exploit that the long-range spin-polarization effects of localized singly occupied MOs (SOMOs) can be taken into account at the mean-field and approximated MP2 level.⁶⁵ Then, the efficient closed-shell formulae and algorithms can be utilized for the LMP2 and LNO-CCSD(T) correlation energy contributions of the LMOs that are not interacting strongly with any SOMOs. Interestingly, here, we also find that for large systems of about 200 atoms or more, up to 50–90% of the LMO correlation energy contributions can be safely evaluated with this approach at practically the cost of the closed-shell counterpart.

The capabilities of the resulting open-shell LNO-CCSD(T) code are illustrated on three-dimensional transition metal complexes of up to 179 atoms, as well as on protein models of

565 and 601 atoms. The protein models involve about twice as many (cca. 11,000) AOs as the largest open-shell local CCSD(T) computations in the literature so far.^{48,56} In addition to the more complicated electronic structure of the studied metal-complexes, the largest protein model also exhibits a highly delocalized SOMO posing a challenge to any local correlation approach. Nevertheless, these LNO-CCSD(T) computations were still feasible within wall times of about 2–4 days using a single CPU with 20 physical cores and mostly a few tens to at most 100 GBs of memory and comparable disk use.

It is even more important to retain the accuracy of restricted LNO-CCSD(T) for the more challenging open-shell applications. The closed-shell LNO-CCSD(T) method was extensively benchmarked by us^{52,53,67,72–75} as well as independently^{76–83} revealing highly competitive accuracy and efficiency compared to other local CCSD(T) approaches, e.g., for organic thermochemistry,^{53,78} non-covalent complexes,^{67,74,81} conformational and isomerization energies,^{82,83} ionic interactions^{73,80,84,85} as well as for organometallic-^{77,79} and extended π -systems^{67,76} exhibiting even moderate non-dynamic correlation. Here, we extend these benchmarks to radical stabilization energies, ionization potentials, and spin-state energies of small- to medium-sized systems and up to triple- ζ basis sets. On the average, at the range where we can compare against the canonical CCSD(T) reference (that is, ca. 20–30 atoms), we find the open-shell LNO-CCSD(T) correlation energies 99.9–99.95% accurate, which translates into a few tenth of a kcal/mol average energy difference deviations for the investigated systems. This is in accord with the accuracy of the closed-shell method, where, however, one should point out that the local errors somewhat grow with increasing system size and wave function complexity.^{52,53,67,72} In practice, the LNO error can be estimated and systematically converged close to the local approximation free limit at an affordable cost as previously demonstrated in various applications.^{52,53,67,72–75}

The discussion of the corresponding details is organized as follows. Sections 2 and 3 collect the theoretical and algorithmic details of the new LNO-CCSD(T) approach focusing on the technicalities emerging specifically for the open-shell case. The computational details

and the benchmark molecules are introduced in Section 4. The accuracy of the individual and combined local approximations is assessed in Sections 5 and 6. Finally, large-scale applications for systems of 175–601 atoms and the corresponding computational requirements are discussed in Section 7.

2 Theoretical background

Throughout the presented derivations, restricted open-shell (RO) reference determinants are assumed, consisting of singly and doubly occupied molecular orbitals (SOMOs and DOMOs, respectively). Since the LNO method makes use of multiple orbital types, the notation of these is summarized in Table 1. The conventional and the LNO correlation energy expressions also employ unrestricted, semi-canonical (also known as pseudo-canonical) MOs. The lower (upper) case indices label orbitals with spin up (down) occupation, while $i, j, k, \dots, I, J, K, \dots$ and $a, b, c, \dots, A, B, C, \dots$ indices are used for the occupied and virtual subsets, respectively. Local approximations are introduced in the basis of localized molecular orbitals (LMOs) obtained from a restricted open-shell reference, which will be denoted in general by indices $\mathcal{I}, \mathcal{J}, \mathcal{K}, \dots$, while these LMOs will be labeled as $i', j', k', \dots (I', J', K', \dots)$, respectively, when occupied by spin up (spin down) electrons.

Table 1: Summary of index notations for orbital sets employed in Sections 2 and 3.

$i, j, k, \dots (I, J, K, \dots)$	spin up (spin down) (semi-)canonical occupied orbitals
$a, b, c, \dots (A, B, C, \dots)$	spin up (spin down) (semi-)canonical virtual orbitals
$i', j', k', \dots (I', J', K', \dots)$	spin up (spin down) localized restricted occupied orbitals
$\mathcal{I}, \mathcal{J}, \mathcal{K}, \dots$	localized restricted occupied orbitals (spatial)
$\hat{i}, \dots, \hat{a}, \dots$	restricted orbitals in the extended domain
$\tilde{i}, \dots, \tilde{a}, \dots (\tilde{I}, \dots, \tilde{A}, \dots)$	spin up (spin down) (semi-)canonical orbitals in primary/extended domains
$\bar{i}, \dots, \bar{a}, \dots$	restricted orbitals in the local interacting subspace
$\underline{i}, \dots, \underline{a}, \dots (\underline{I}, \dots, \underline{A}, \dots)$	spin up (spin down) (semi-)canonical orbitals in the local interacting subspace
μ, ν, λ, \dots	atomic orbitals
X, Y, \dots	auxiliary functions for the DF approximation

2.1 Open-Shell LNO-CCSD(T) Ansatz

Following the relevant approaches introduced for the previous members of the LNO family of methods,^{52,53,58,59,61,62,65,66} especially the open-shell LMP2⁶⁵ and the closed-shell LNO-CCSD(T),^{13,52,53,62} here, we introduce the open-shell LNO-CCSD(T) Ansatz built on restricted open-shell references. First, conventional unrestricted CCSD and (T) energy expressions^{8,86} are written in terms of semicanonical orbitals suitable for transformations to the (restricted open-shell) LMO basis due to their invariance to unitary orbital rotations.

To introduce the orbital-specific correlation energy contributions of the LNO approach, first the open-shell CCSD correlation energy expression is rewritten as a sum of contributions from the occupied spin up and spin down orbitals, δE_i^{CCSD} and δE_I^{CCSD} :

$$\begin{aligned}
E^{\text{CCSD}} &= \sum_i \delta E_i^{\text{CCSD}} + \sum_I \delta E_I^{\text{CCSD}} \\
&= \sum_i \left(\sum_a t_i^a f_i^a + \frac{1}{4} \sum_{abj} \tau_{ij}^{ab} \langle ab || ij \rangle + \frac{1}{2} \sum_{aBJ} \tilde{\tau}_{iJ}^{aB} \langle aB || iJ \rangle \right) \\
&\quad + \sum_I \left(\sum_A t_I^A f_I^A + \frac{1}{4} \sum_{ABJ} \tau_{IJ}^{AB} \langle AB || IJ \rangle + \frac{1}{2} \sum_{Abj} \tilde{\tau}_{Ij}^{Ab} \langle Ab || Ij \rangle \right), \tag{1}
\end{aligned}$$

where $\tau_{ij}^{ab} = t_{ij}^{ab} + t_i^a t_j^b - t_i^b t_j^a$, $\tilde{\tau}_{iJ}^{aB} = t_{iJ}^{aB} + t_i^a t_J^B$, and t denotes the CCSD singles and doubles cluster amplitudes. Additionally, $\langle ab || ij \rangle$ stands for antisymmetrized electron repulsion integrals (ERIs), constructed as $\langle ab || ij \rangle = \langle ab | ij \rangle - \langle ab | ji \rangle$ using the Dirac notation.

Analogously, the energy formula for the open-shell perturbative triples correction of CCSD(T) can be written as

$$\begin{aligned}
E^{(T)} &= \sum_i \delta E_i^{(T)} + \sum_I \delta E_I^{(T)} \\
&= \sum_i \left(\frac{1}{36} \sum_{abcjk} t_{ijk}^{abc} W_{ijk}^{abc} + \frac{1}{8} \sum_{aBCJK} t_{iJK}^{aBC} W_{iJK}^{aBC} + \frac{1}{8} \sum_{abCjK} t_{ijK}^{abC} W_{ijK}^{abC} \right) \\
&\quad + \sum_I \left(\frac{1}{36} \sum_{ABCJK} t_{IJK}^{ABC} W_{IJK}^{ABC} + \frac{1}{8} \sum_{Abcjk} t_{Ijk}^{Abc} W_{Ijk}^{Abc} + \frac{1}{8} \sum_{ABcJk} t_{IJk}^{ABc} W_{IJk}^{ABc} \right), \tag{2}
\end{aligned}$$

where t_{ijk}^{abc} denotes the triple excitation amplitude of the (T) method, and W_{ijk}^{abc} can be written as $t_{ijk}^{abc}D_{ijk}^{abc}$, that is, the triple excitation amplitude multiplied by the canonical orbital energy differences collected into the denominator D_{ijk}^{abc} .^{7,8} Combining eq 1 and eq 2, the full CCSD(T) correlation energy can also be written as a sum of orbital contributions:

$$\begin{aligned} E^{\text{CCSD(T)}} &= \sum_i \delta E_i^{\text{CCSD(T)}} + \sum_I \delta E_I^{\text{CCSD(T)}} \\ &= \sum_i \left(\delta E_i^{\text{CCSD}} + \delta E_i^{(\text{T})} \right) + \sum_I \left(\delta E_I^{\text{CCSD}} + \delta E_I^{(\text{T})} \right). \end{aligned} \quad (3)$$

To introduce the local CCSD(T) Ansatz, let us transform the separated occupied indices of eq 3 to the restricted open-shell LMO basis:

$$E^{\text{CCSD(T)}} = \sum_{i'} \delta E_{i'}^{\text{CCSD(T)}} + \sum_{I'} \delta E_{I'}^{\text{CCSD(T)}} = \sum_{\mathcal{I}} \delta E_{\mathcal{I}}^{\text{CCSD(T)}}, \quad (4)$$

where in the last term, \mathcal{I} denotes a spatial orbital occupied by either one or two electrons in the RO LMO basis, while i' (I') refers to orbitals with the same spatial component as LMO \mathcal{I} , but occupied by at most one spin up (spin down) electron. The explicit expressions used for the LNO method will be introduced in Sect. 3.7.

The scaling of CCSD(T) can be made asymptotically linear with respect to the system size if only a domain of asymptotically constant number of orbitals is required to evaluate the correlation energy contribution of a given orbital \mathcal{I} , that is, $\delta E_{\mathcal{I}}^{\text{CCSD(T)}}$. To construct the domain around each LMO, which is then called the central LMO of its own domain, first, those LMOs are collected that strongly interact with the central LMO. The selection of strongly interacting LMO pairs is based on multipole approximated MP2 pair energies [see the $\delta E_{\mathcal{I}\mathcal{J}}^{\text{MP2}}(\mathcal{P}_{\mathcal{I}\mathcal{J}})$ term of eq 6] evaluated in LMO pair specific domains (PD, $\mathcal{P}_{\mathcal{I}\mathcal{J}}$). LMO pairs with very small $\delta E_{\mathcal{I}\mathcal{J}}^{\text{MP2}}(\mathcal{P}_{\mathcal{I}\mathcal{J}})$ pair correlation energy estimates are considered distant and, unlike to the strong pairs, they do not enter to the higher-level computations. Next, the virtual space of the domain of LMO \mathcal{I} is constructed from local projected atomic

orbitals (PAOs) surrounding the central LMO and its strongly interacting LMO pairs. The domain obtained in this way is referred to as the extended domain (ED, $\mathcal{E}_{\mathcal{I}}$) of LMO \mathcal{I} , which is sufficiently compact to efficiently perform MP2 computations exploiting our open-shell LMP2 implementation.⁶⁵ The resulting local MP2 correlation energy contribution of the ED [$\delta E_{\mathcal{I}}^{\text{MP2}}(\mathcal{E}_{\mathcal{I}})$] is utilized as part of the correction employed to decrease the effect of the remaining approximations in the ED at the CCSD(T) level (see eq 6). To further compress the orbital spaces of the ED, LNOs are constructed using the density built from the first-order Møller–Plesset (MP1) amplitudes of the ED, yielding the local interacting subspace (LIS, $\mathcal{L}_{\mathcal{I}}$) of LMO \mathcal{I} . Consequently, the LNO-CCSD(T) correlation energy reads as

$$E^{\text{LNO-CCSD(T)}} = \sum_{\mathcal{I}} \left[\delta E_{\mathcal{I}}^{\text{CCSD(T)}}(\mathcal{L}_{\mathcal{I}}) + \Delta E_{\mathcal{I}}^{\text{MP2}} \right], \quad (5)$$

where the energy correction $\Delta E_{\mathcal{I}}^{\text{MP2}}$ is calculated at the MP2 level of theory as

$$\Delta E_{\mathcal{I}}^{\text{MP2}} = \delta E_{\mathcal{I}}^{\text{MP2}}(\mathcal{E}_{\mathcal{I}}) - \delta E_{\mathcal{I}}^{\text{MP2}}(\mathcal{L}_{\mathcal{I}}) + \frac{1}{2} \sum_{\mathcal{J}}^{\text{distant}} \delta E_{\mathcal{I}\mathcal{J}}^{\text{MP2}}(\mathcal{P}_{\mathcal{I}\mathcal{J}}). \quad (6)$$

Therefore, correlation energy terms are included in the final LNO-CCSD(T) expression for all orbital pairs of the entire molecule. The largest and most important component, which corresponds to the correlation of the strong pairs, is included at the complete CCSD(T) level in the LISs. The contribution of the frozen LNOs and the correlation energy contribution of distant LMO pairs are included at the MP2 level in the extended and pair domains, respectively.

3 Algorithm

The algorithm of the restricted open-shell LNO-CCSD(T) method is summarized in Figure 1 and described in this section step by step in detail.

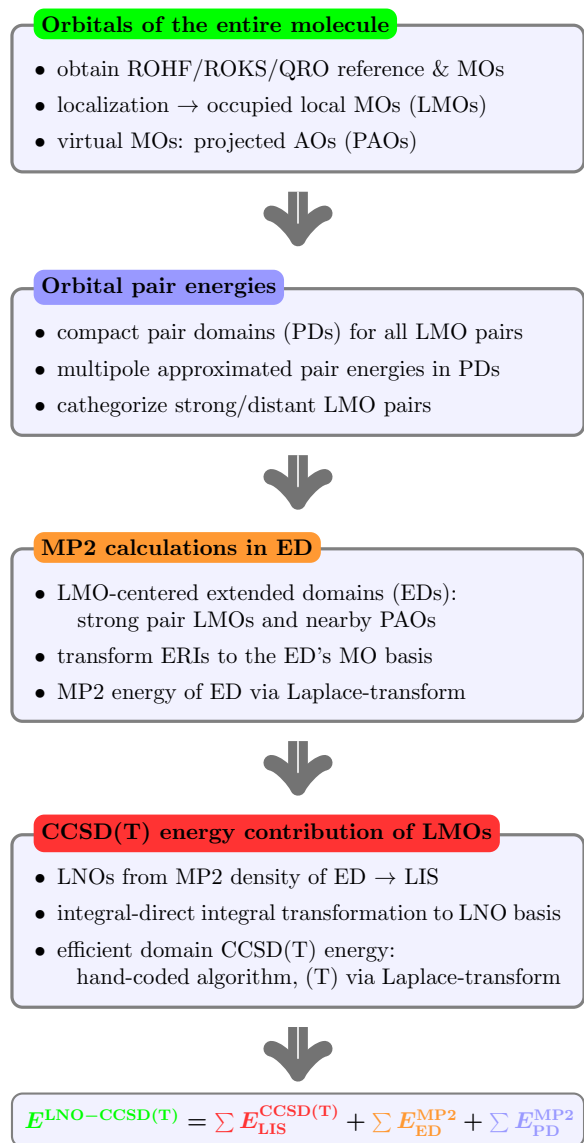


Figure 1: Major algorithmic steps of the presented restricted open-shell LNO-CCSD(T) method.

3.1 Self-Consistent Field and Orbital Localization

The considerations regarding the reference selection are analogous to those of our local RO MP2 scheme.⁶⁵ The restricted open-shell Hartree-Fock (HF) or Kohn-Sham (KS) reference determinant can be obtained via a restricted open-shell self-consistent field (SCF) calculation (ROHF/ROKS), or via unrestricted (UHF/UKS) computations followed by constructing quasi-restricted orbitals (QROs).^{65,87}

The computational cost of (RO/U)HF can become high above the few hundred atom range even with density-fitting (DF) approaches. To reduce the scaling of conventional DF-SCF calculations, for example, the local DF approach can be utilized, which restricts the lists of auxiliary functions and, for very large systems, also that of AOs at the exchange matrix evaluation to include functions that are spatially close to an LMO.^{66,88-90} In the present study, local DF is employed only for the ROHF and UHF computations in the 500- to 600-atom range.

The occupied orbitals of the reference determinant are localized with the algorithm proposed by Boys,⁹¹ but Pipek–Mezey,⁹² intrinsic bond orbital,⁹³ and generalized Boys methods with higher orbital variance power⁹⁴ are also implemented. The localization is carried out in a spin restricted manner, and the doubly occupied and open-shell MOs are not mixed resulting in doubly and singly occupied LMOs. The clear advantage of the restricted formalism is that the correlation energy contribution of the spin up and down electrons of a doubly occupied LMO can be obtained in a single domain corresponding to a RO LMO. The alternative way of computing $\delta E_{i'}^{\text{CCSD(T)}}$ and $\delta E_{i'}^{\text{CCSD(T)}}$ of eq 4 in separate, spin case dependent domains would be approximately twice as costly. The potential drawback of keeping the LMO space restricted is that in systems with only one or a few SOMOs (or when the mixing of the SOMOs is prohibited, e.g., due to symmetry), it may not be possible to (sufficiently) localize the SOMO(s), leading to singly occupied orbitals in the LMO basis that are (potentially) still delocalized (see below an example in Fig. 7).

3.2 PAO construction and pair energy calculation

The PAO construction and pair energy calculation algorithms follow closely those of the open-shell LMP2 scheme⁶⁵ with minor extensions needed for the LNO-CC methods discussed here in more detail. To identify the strongly interacting LMO pairs, their approximate MP2 pair correlation energies are computed in LMO pair specific domains (PDs). The PDs contain the corresponding occupied LMO pair and PAOs as virtual functions that are centered on atoms surrounding the two LMOs.

To assemble the virtual space, first, the PAOs of the entire molecule are constructed by projecting out the restricted LMOs (both DOMOs and SOMOs) from all AOs of the molecule ($|\mu\rangle$):

$$|a_\mu\rangle = \left(1 - \sum_{\mathcal{I}}^{\text{DO}\cup\text{SO}} |\mathcal{I}\rangle\langle\mathcal{I}|\right) |\mu\rangle. \quad (7)$$

The AO $|\mu\rangle$ and the corresponding atom can be considered as the center of the resulting PAO $|a_\mu\rangle$. The PAOs of eq 7 are restricted and span the virtual subspace of the spin up electrons. To span the virtual subspace of the spin down electrons, we use the union of all SOMOs and PAOs.

The pair domain of a given LMO pair is constructed as the union of the primary domains of the two LMOs. To define the primary domains, a set of atoms is assigned to each LMO and PAO using a modified^{52,66} Boughton–Pulay (BP) algorithm.⁹⁵ The BP atom list of an LMO or PAO is compiled such that the projection of the LMO/PAO onto the AOs of its BP atom list has an overlap value of at least T with the original, unprojected LMO/PAO. Thus, $1 - T$ gives an upper bound for the truncation error of the projection onto the BP atom lists. For the primary domain construction, BP atom lists are assembled using $T_{\text{PD}_0} = 0.999$ for LMOs and $T_{\text{PD}_V} = 0.98$ for PAOs. The occupied subspace of the primary domain contains a single LMO, while its virtual subspace includes those PAOs that are centered on any of the atoms in the LMO’s BP atom list. Additionally, if a BP list of a SOMO constructed with completeness criterion T_{PD_V} overlaps with the BP set of the LMO, the SOMO is also

included in the spin down virtual subspace of the primary domain. Then, the atom list of the PD contains the union of the BP atom lists of all orbitals in the primary domain. Finally, the orbitals of the primary domain are projected onto the AOs of the primary domain, the projected orbitals are orthogonalized, and the spin up and spin down orbitals are separately canonicalized.

The multipole approximated MP2 pair energy of an LMO pair is then evaluated in the primary domains' bases of the LMO pair:

$$\begin{aligned} \delta E_{\mathcal{IJ}}^{\text{MP2}}(\mathcal{P}_{\mathcal{IJ}}) &= \delta E_{i'j'}^{\text{MP2}}(\mathcal{P}_{\mathcal{IJ}}) + \delta E_{I'J'}^{\text{MP2}}(\mathcal{P}_{\mathcal{IJ}}) + \delta E_{i'j'}^{\text{MP2}}(\mathcal{P}_{\mathcal{IJ}}) + \delta E_{I'J'}^{\text{MP2}}(\mathcal{P}_{\mathcal{IJ}}) = \\ &= - \sum_{\tilde{a}_{\mathcal{I}} \tilde{b}_{\mathcal{J}}} \frac{[(\tilde{a}_{\mathcal{I}} i' | \tilde{b}_{\mathcal{J}} j')^{[4]}]^2}{\varepsilon_{\tilde{a}_{\mathcal{I}}} + \varepsilon_{\tilde{b}_{\mathcal{J}}} - F_{i'i'} - F_{j'j'}} - \sum_{\tilde{A}_{\mathcal{I}} \tilde{B}_{\mathcal{J}}} \frac{[(\tilde{A}_{\mathcal{I}} I' | \tilde{B}_{\mathcal{J}} J')^{[4]}]^2}{\varepsilon_{\tilde{A}_{\mathcal{I}}} + \varepsilon_{\tilde{B}_{\mathcal{J}}} - F_{I'I'} - F_{J'J'}} \quad (8) \\ &= - \sum_{\tilde{a}_{\mathcal{I}} \tilde{B}_{\mathcal{J}}} \frac{[(\tilde{a}_{\mathcal{I}} i' | \tilde{B}_{\mathcal{J}} J')^{[4]}]^2}{\varepsilon_{\tilde{a}_{\mathcal{I}}} + \varepsilon_{\tilde{B}_{\mathcal{J}}} - F_{i'i'} - F_{J'J'}} - \sum_{\tilde{A}_{\mathcal{I}} \tilde{b}_{\mathcal{J}}} \frac{[(\tilde{A}_{\mathcal{I}} I' | \tilde{b}_{\mathcal{J}} j')^{[4]}]^2}{\varepsilon_{\tilde{A}_{\mathcal{I}}} + \varepsilon_{\tilde{b}_{\mathcal{J}}} - F_{I'I'} - F_{j'j'}}. \end{aligned}$$

Here, the pseudocanonical orbital energy of the virtual orbitals is denoted by ε , while $F_{i'i'}$ ($F_{I'I'}$) indicates a diagonal element of the spin up (spin down) Fock matrix. The LMO subscript of the virtual orbitals indices indicates that the corresponding summations include virtual orbitals only from the primary domain of the given LMO. The ERIs, denoted here in Mulliken notation by $(\tilde{a}_{\mathcal{I}} i' | \tilde{b}_{\mathcal{J}} j')^{[4]}$ are calculated using a multipole expansion up to fourth order, including dipole–dipole, dipole–quadrupole, quadrupole–quadrupole, and dipole–octupole moment terms as discussed in Ref. 66.

Utilizing the $\delta E_{\mathcal{IJ}}^{\text{MP2}}(\mathcal{P}_{\mathcal{IJ}})$ pair energies, the \mathcal{IJ} LMO pair can be classified as a strongly interacting pair if $\delta E_{\mathcal{IJ}}^{\text{MP2}}(\mathcal{P}_{\mathcal{IJ}}) > f_w \varepsilon_w$, where ε_w is the strong pair energy threshold, and f_w is a factor of 1, $\frac{1}{2}$ or $\frac{1}{4}$ for DO–DO, DO–SO and SO–SO LMO pairs, respectively. The f_w scaling factor is introduced so that pairs between LMOs of different occupations are handled on an equal footing since SO LMOs have half as many pair energy terms in eq 8 as DO LMOs. A more detailed discussion and numerical benchmarks regarding the benefits of using the f_w factor are provided in our RO LMP2 study.⁶⁵ Note that the approximated

pair energy of those pairs which are not classified as strong is added to the MP2 correlation energy corrections as the last term of eq 6.

The pair energies of eq 8 have been extensively tested on a large number of (mainly organic) molecules,^{52,53,65,66} with consistently satisfactory performance. However, the open-shell LNO-CCSD(T) approach is expected to be applied more frequently also for transition metal complexes (see, e.g., Sections 5 and 7), exhibiting electronic structures potentially more complex than that of a typical organic molecule. In a few complicated cases, we have found that the multipole approximated MP2 pair energies can underestimate the approximation-free MP2 pair energies leading to LMO pairs classified as distant instead of strong on the border of the two categories. To remedy this issue, an additional mechanism is introduced here to extend the strong pair list. In the present approach, we investigate more closely the orbital pairs characterized originally as distant that are on the border of the distant and strong categories. More precisely, the orbital pairs with $\frac{f_w \varepsilon_w}{g_w} < \delta E_{\mathcal{I}\mathcal{J}}^{\text{MP2}}(\mathcal{P}_{\mathcal{I}\mathcal{J}}) < f_w \varepsilon_w$ are considered with parameter g_w defining the range of pair energies where potentially important LMO pairs may still appear. For these pairs, an additional measure is computed:

$$M_{\mathcal{I}\mathcal{J}} = \sqrt{\sum_A^{\text{atoms}} \left(\sum_{\mu\nu}^{\text{on } A} C_{\mu\mathcal{I}}^A S_{\mu\nu}^A C_{\nu\mathcal{I}}^A \cdot \sum_{\mu\nu}^{\text{on } A} C_{\mu\mathcal{J}}^A S_{\mu\nu}^A C_{\nu\mathcal{J}}^A \right)^2}, \quad (9)$$

where, $C_{\mu\mathcal{I}}^A$ is an orbital coefficient of truncated LMO \mathcal{I} for AOs on atom A , and $S_{\mu\nu}^A$ is the overlap matrix of these AOs. The $\sum_{\mu\nu}^{\text{on } A} C_{\mu\mathcal{I}}^A S_{\mu\nu}^A C_{\nu\mathcal{I}}^A$ part is thus the Mulliken charge of (the truncated) LMO \mathcal{I} on atom A . Therefore, measure $M_{\mathcal{I}\mathcal{J}}$ sums the products of LMO \mathcal{I} and LMO \mathcal{J} Mulliken charges for all atoms, hence it can be interpreted as a discretized measure of the overlap of the two LMOs. Orbital pairs for which $M_{\mathcal{I}\mathcal{J}}$ is large are more likely to be localized on the same or nearby atoms, and consequently may strongly interact. The strong pair list is therefore extended with those pairs which exhibit a considerable $M_{\mathcal{I}\mathcal{J}}$ measure. In practice, it is preferable to include the pairs which are indicated to be more strongly interacting compared to their pair correlation energy. The selection of the additional

strong pairs is thus controlled by the h_w parameter as $M_{\mathcal{IJ}}/\delta E_{\mathcal{IJ}}^{\text{MP2}} > h_w$ relative to the pair correlation energies. In practice, we found that $g_w = 5$ and $h_w = 50 E_h^{-1}$ are reasonable choices also for particularly challenging cases, while the strong pair list extension introduce (practically) no changes to our previous strong pair list definition for organic molecules.

3.3 Local MP2 energy in the extended domains

The MP2 correlation energy contribution of each LMO is evaluated in its ED analogously to our RO LMP2 method,⁶⁵ so we focus on the steps required in the EDs for LNO-CC computations. The occupied subspace of the ED consists of its central LMO and the strong LMO pairs of the central LMO. The atoms of the ED are defined to be the union of the extensive BP atom lists of all LMOs in the ED. These large BP atom lists are constructed with a completeness criterion of $T_{\text{EDo}} = 0.9999$ by default. Then, each LMO is projected onto the AOs of its respective BP atom list, ensuring at most $1 - T_{\text{EDo}}$ (that is, here below 0.01%) projection error for accurate computations with these projected LMOs in the ED. The projected LMOs are reorthogonalized by a specific combination of the Gram–Schmidt and Löwdin symmetric (GSL) orthogonalization algorithms.^{96,97} In the Gram–Schmidt step, the central LMO and all SOMOs are projected out from the DOMOs of the ED to ensure that they are not changed until the LNO construction step (see Section 3.4). Subsequently, the projected DOMOs are Löwdin orthogonalized and all occupied orbitals are semi-canonicalized. For the latter, the spin up and spin down Fock matrix blocks of the ED are diagonalized separately in the spin up and spin down occupied orbital bases. Next, the virtual subspace of the ED is constructed from the restricted PAOs centered on the atoms of the PAO center domain (PCD) of the ED. The PCD of an ED is defined as the union of the compact BP atom lists of the LMOs in the ED, constructed with $T_o = 0.985$. The selected PAOs are subsequently projected onto the AO basis of the ED. To span the spin down virtual subspace of the ED, the SO LMOs of the ED are appended to the spin down PAO list of the ED. Finally, the occupied subspace is projected out from the virtual space of the ED, the virtual orbitals are

Löwdin canonical orthogonalized among themselves, and are semi-canonicalized in a spin unrestricted manner.

The ERIs of the ED are transformed to the MO bases of the ED using the DF approximation^{98,99} and highly-optimized AO integral¹⁰⁰ and local integral-transformation algorithms.^{65,66} As demonstrated previously,⁶⁶ the set of auxiliary basis functions of the DF approximation can also be restricted to include those residing on atoms of the PCD. The three-center DF integrals $(\tilde{\mu}\tilde{\nu}|X)$ are therefore computed only for a significantly restricted AO list $(\tilde{\mu})$ of the ED and for auxiliary functions (X) in the PCD. The transformation to the MO bases of the ED is started with an intermediate transformation step to the restricted occupied LMO basis of the ED. Since this first step is the most demanding, the overall cost of the open-shell DF integral transformation is kept comparable to that of our closed-shell LMP2 algorithm.⁶⁵

The four-center antisymmetrized ERIs are then only assembled for the single central LMO, e.g., a single i' index value, retaining the formally at most fourth-power-scaling operational complexity in the ED in terms of the dimensions of the ED bases:

$$\langle \tilde{a}\tilde{b}||i'\tilde{j} \rangle = (\tilde{a}i'|\tilde{b}\tilde{j}) - (\tilde{a}\tilde{j}|\tilde{b}i') = K_{\tilde{a}i',\tilde{b}\tilde{j}} - K_{\tilde{a}\tilde{j},\tilde{b}i'}, \quad (10)$$

where the \mathbf{K} tensors are computed from the DF integrals according to

$$\mathbf{K} = \mathbf{IV}^{-1}\mathbf{I}^T = \mathbf{JI}^T. \quad (11)$$

In the above equation, \mathbf{I} contains three-center DF integrals transformed to the ED MO basis $[(\tilde{a}\tilde{j}|X) = I_{\tilde{a}\tilde{j},X}]$, while $J_{\tilde{a}i',Y} = \sum_X I_{\tilde{a}i',X}V_{XY}^{-1}$, and $V_{XY} = (X|Y)$ is the two-center auxiliary integral, whose inverse is computed using Cholesky-decomposition: $\mathbf{V}^{-1} = (\mathbf{LL}^T)^{-1} = (\mathbf{L}^{-1})^T \mathbf{L}^{-1}$.

Similar to the four-center ERIs, the assembly of the MP1 amplitudes is only required for the fixed i' or I' values, therefore, it is advantageous to compute them using either

Cholesky-decomposition or Laplace-transform. This way, the amplitudes can be written in closed form also in the non-canonical basis used here to avoid the redundant computation for other occupied index values.^{65,66} Accordingly, the MP1 amplitudes of the ED are computed as

$$t_{i'\tilde{J}}^{\tilde{a}\tilde{B}[1]} = \sum_{\omega} \sum_X \bar{I}_{\tilde{a}i',X}^{\omega} \bar{J}_{\tilde{B}\tilde{J},X}^{\omega}, \quad (12)$$

where the index ω runs over the Cholesky-vectors or the integration quadrature of the Laplace-transform. The bar over integrals I and J denotes that these have been multiplied with the factor corresponding to the factorized energy denominator. For example, $\bar{J}_{\tilde{B}\tilde{J},X}^{\omega} = J_{\tilde{B}\tilde{J},X} c_{\tilde{B}\tilde{J}}^{\omega}$, where $c_{\tilde{B}\tilde{J}}^{\omega}$ is either a Cholesky-vector element or if Laplace-transform is used,

$$c_{\tilde{B}\tilde{J}}^{\omega} = \sqrt{w_{\omega}} \exp(-(\varepsilon_{\tilde{B}} - \varepsilon_{\tilde{J}}) t_{\omega}) \quad (13)$$

with t_{ω} and w_{ω} as a quadrature point and the corresponding weight, respectively. It is important to note that, for example, the Laplace-quadrature is also determined in a spin independent manner⁶⁵ enabling the assembly also of the mixed spin MP1 amplitudes, such as in eq 12. Building the four-center integrals and MP1 amplitudes for all required spin cases leads to the ED MP2 energy contribution of the central LMO as

$$\begin{aligned} \delta E_{\mathcal{I}}^{\text{MP2}}(\mathcal{E}_{\mathcal{I}}) &= \delta E_{i'}^{\text{MP2}}(\mathcal{E}_{\mathcal{I}}) + \delta E_{I'}^{\text{MP2}}(\mathcal{E}_{\mathcal{I}}) = \\ &\sum_{\tilde{a}} t_{i'\tilde{J}}^{\tilde{a}[1]} f_{i'\tilde{J}}^{\tilde{a}} + \frac{1}{2} \sum_{\tilde{a} < \tilde{b}, \tilde{J}} t_{i'\tilde{J}}^{\tilde{a}\tilde{b}[1]} \langle \tilde{a}\tilde{b} || i'\tilde{J} \rangle + \frac{1}{2} \sum_{\tilde{a}\tilde{B}\tilde{J}} t_{i'\tilde{J}}^{\tilde{a}\tilde{B}[1]} \langle \tilde{a}\tilde{B} || i'\tilde{J} \rangle + \\ &\sum_{\tilde{A}} t_{I'\tilde{J}}^{\tilde{A}[1]} f_{I'\tilde{J}}^{\tilde{A}} + \frac{1}{2} \sum_{\tilde{A} < \tilde{B}, \tilde{J}} t_{I'\tilde{J}}^{\tilde{A}\tilde{B}[1]} \langle \tilde{A}\tilde{B} || I'\tilde{J} \rangle + \frac{1}{2} \sum_{\tilde{A}\tilde{b}, \tilde{J}} t_{I'\tilde{J}}^{\tilde{A}\tilde{b}[1]} \langle \tilde{A}\tilde{b} || I'\tilde{J} \rangle. \end{aligned} \quad (14)$$

Here, it is important to note how the contributions from single excitations to the ED MP2 energy originate from two sources. First, as we employ the restricted open-shell determinant, the unrestricted Fock matrices are not self-consistent and thus have off-diagonal elements even in the unrestricted canonical MO basis of the entire molecule. Second, as a consequence of truncating the occupied and virtual MOs of the ED, a small portion of the exact virtual

space is mixed into the occupied ED MOs and vice versa. The latter contribution to the singles MP1 amplitudes and hence to the MP2 energy of the ED is discarded in our closed-shell LMP2 variant, which, however, would undesirably discard the first contribution as well in the open-shell case. Since the first contribution originates from the off-diagonal block of the MO Fock matrix, while the second one appears mostly in the diagonal part of the MO Fock, the two parts can be separated.⁶⁵ For that purpose, we also need to separate the off-diagonal Fock matrix blocks in the AO basis as

$$\mathbf{F}^{\text{OD}} = \mathbf{F} - \mathbf{C}\boldsymbol{\epsilon}\mathbf{C}^{\text{T}}, \quad (15)$$

where \mathbf{F}^{OD} and \mathbf{F} are the off-diagonal and the full Fock matrices in the AO basis, respectively, while \mathbf{C} is the matrix of the unrestricted MO coefficients, and $\boldsymbol{\epsilon}$ is a diagonal matrix with the corresponding orbital energies. Then, we use \mathbf{F}^{OD} to evaluate off-diagonal Fock matrix elements in the MP2 energy of the ED in eq 14. This \mathbf{F}^{OD} matrix vanishes if the reference orbitals of the molecules are exact eigenfunctions of the full Fock matrix, such as in the case of our closed-shell methods where self-consistent HF reference is employed.

Finally, we also note that the complete local MP2 energy can be constructed as a side product of an LNO-CC computation as

$$E^{\text{LMP2}} = \sum_{\mathcal{I}} \delta E_{\mathcal{I}}^{\text{MP2}}(\mathcal{E}_{\mathcal{I}}) + \sum_{\substack{\text{distant} \\ \mathcal{I} < \mathcal{J}}} \delta E_{\mathcal{I}\mathcal{J}}^{\text{MP2}}(\mathcal{P}_{\mathcal{I}\mathcal{J}}). \quad (16)$$

This is beneficial as this local MP2 energy can be used, for example, for various composite energy expressions, such as local MP2 level basis set corrections.

3.4 Local natural orbitals

Due to the extensive cost of CCSD(T), the orbital spaces of the EDs should be further decreased after the MP2 part of the ED computation is completed. To that end, local natural

orbitals (LNOs) are constructed as the eigenvectors of the second-order density matrix blocks of the ED built from the MP1 amplitudes of eq 12. The so obtained LNO list is truncated by retaining the most important occupied and virtual LNOs for the domain correlation energy contribution with occupation numbers above a threshold.

For this purpose, first, the occupied-occupied density matrix block contribution of central LMO \mathcal{I} is obtained in the semi-canonical ED basis for both spin cases as

$$\begin{aligned}
 D_{\tilde{j}\tilde{k}}^{\mathcal{I}} &= \sum_{\tilde{a}<\tilde{b}} t_{i'\tilde{j}}^{\tilde{a}\tilde{b}[1]} t_{i'\tilde{k}}^{\tilde{a}\tilde{b}[1]} + \sum_{\tilde{A}\tilde{b}} t_{i'\tilde{j}}^{\tilde{A}\tilde{b}[1]} t_{i'\tilde{k}}^{\tilde{A}\tilde{b}[1]}, \\
 D_{\tilde{J}\tilde{K}}^{\mathcal{I}} &= \sum_{\tilde{A}<\tilde{B}} t_{i'\tilde{J}}^{\tilde{A}\tilde{B}[1]} t_{i'\tilde{K}}^{\tilde{A}\tilde{B}[1]} + \sum_{\tilde{a}\tilde{B}} t_{i'\tilde{J}}^{\tilde{a}\tilde{B}[1]} t_{i'\tilde{K}}^{\tilde{a}\tilde{B}[1]}.
 \end{aligned} \tag{17}$$

In order to reduce the computational cost of the integral transformation step (see Section 3.5), it is beneficial to first construct an intermediate spin-restricted orbital set also for the LNOs. To obtain spin-restricted LNOs, we transform the spin up and spin down density matrix blocks of eq 17 to the restricted occupied basis of the ED and add together the resulting terms:

$$D_{i\tilde{m}}^{\mathcal{I}} = \sum_{\tilde{j}\tilde{k}} C_{i\tilde{j}}^{\uparrow} D_{\tilde{j}\tilde{k}}^{\mathcal{I}} C_{\tilde{m}\tilde{k}} + \sum_{\tilde{J}\tilde{K}} C_{i\tilde{J}}^{\downarrow} D_{\tilde{J}\tilde{K}}^{\mathcal{I}} C_{\tilde{m}\tilde{K}}. \tag{18}$$

Here, $C_{i\tilde{j}}^{\uparrow}$ ($C_{i\tilde{j}}^{\downarrow}$) contains the spin up (spin down) semi-canonical to restricted LMO transformation coefficients of the ED.

Note that the restricted occupied basis contains both the central LMO and the SOMOs without mixing them with the rest of the truncated LMOs of the ED. This is the intentional result of using the GSL orthogonalization as described in Section 3.3 since the central LMO has to be kept exactly in the LIS, and we also want to keep the DO and SO subspaces separated after the truncation of the LNO basis. Therefore, before diagonalizing the restricted density matrix, its $D_{i\tilde{m}}^{\mathcal{I}}$ and $D_{i\tilde{I}}^{\mathcal{I}}$ elements are replaced with zeros to avoid mixing the central LMO with the rest of the occupied ED orbitals upon diagonalization. If the ED contains any SOMOs, the SOMO-DOMO blocks of the density matrix are also overwritten with zeros

in order to obtain a single set of restricted occupied LNOs by compressing only the DO subspace of the ED. After diagonalizing this modified density matrix, the (restricted) occupied basis of the LIS ($\{\bar{i}\}$) consists of the central LMO, all SOMOs, and those doubly occupied LNOs that exhibit occupation numbers higher than the occupied LNO threshold, ε_o . Then, the retained orbitals are semi-canonicalized using the spin up and spin down Fock matrices, yielding the final, unrestricted semi-canonical occupied LNO basis ($\{\underline{i}\}, \{\underline{I}\}$) in which the CCSD(T) contribution of the LIS is evaluated.

Notice that in domains where the central LMO is singly occupied, all MP1 amplitudes with an occupied index of I' are zero because there is no spin down electron on such orbitals. Thus, for SO central LMOs, two of the four terms of eq 17 also vanish, which approximately halves the number of non-zero density matrix terms compared to the case of DO central LMOs. To treat the LNO construction with both SO and DO central LMOs on the same footing, we scale the density matrix corresponding to SO central LMOs by a factor of two. With this setting, we have found that the same ε_o occupied LNO threshold provides balanced accuracy for both domain types.

Having the occupied LNO space at hand, we recommended to take advantage of the fact that the virtual LNOs only need to describe the correlation energy contribution of the retained occupied LNOs.⁵² To achieve this, the occupied indices of the MP1 amplitudes in the ED are transformed to the retained LNO basis before the virtual density matrix construction. This decreases the number of terms contributing to the virtual-virtual density matrix elements and thus leads to a smaller number of virtual LNO occupation numbers being above the corresponding threshold, ε_v .⁵²

It is worth noting a difference in the definition of the density matrix contribution of the central LMO to its occupied (eq 17) and virtual (eq 19 below) blocks. Namely, for the occupied-occupied block only a single occupied index can be selected to be the central LMO as the other two indices (\tilde{j} and \tilde{k} of eq 17) are the indices of the density matrix block. In contrast to that, virtual indices (\tilde{a} and \tilde{b} of eq 19 below) label the elements of the

virtual-virtual density matrix block, thus there is some freedom of choice regarding which occupied index is set to be the central LMO and at what point of the derivation should one introduce the restriction for the central LMO index. This leads to three different local virtual density matrix fragment variants, for which we provide derivations and further analysis in the Appendix. Out of the three choices the definition with the most balanced contribution for all spin cases was selected:

$$\begin{aligned}
 D_{\bar{a}\bar{b}}^{\mathcal{I}} &= \frac{1}{2} \left(\sum_{\bar{c}_j} t_{i'j}^{\bar{a}\bar{c}[1]} t_{i'j}^{\bar{b}\bar{c}[1]} + \sum_{\bar{c}_J} t_{i'J}^{\bar{a}\bar{c}[1]} t_{i'J}^{\bar{b}\bar{c}[1]} + \sum_{\bar{c}_j} t_{i'j}^{\bar{a}\bar{c}[1]} t_{i'j}^{\bar{b}\bar{c}[1]} \right), \\
 D_{\bar{A}\bar{B}}^{\mathcal{I}} &= \frac{1}{2} \left(\sum_{\bar{c}_J} t_{i'J}^{\bar{A}\bar{c}[1]} t_{i'J}^{\bar{B}\bar{c}[1]} + \sum_{\bar{c}_j} t_{i'j}^{\bar{A}\bar{c}[1]} t_{i'j}^{\bar{B}\bar{c}[1]} + \sum_{\bar{c}_J} t_{i'J}^{\bar{A}\bar{c}[1]} t_{i'J}^{\bar{B}\bar{c}[1]} \right).
 \end{aligned}
 \tag{19}$$

Let us note that for closed-shell systems this virtual density matrix expression does not completely match the density matrix in our closed-shell LNO-CC method^{52,53,58,59} as our previous definition reduces to one of the other two expressions derived in the Appendix. Therefore the virtual density matrix expression of the closed-shell algorithm is updated to match the open-shell form defined here in the limit of applying the open-shell code to closed-shell molecules. This introduces a very small change to the closed-shell LNO-CCSD(T) energies much below the uncertainties corresponding to the LNO approximations.

In order to define an intermediate restricted LNO basis, the virtual density matrices of eq 19 built in the canonical ED basis are transformed to the restricted PAO basis of the ED. Then, the spin up and spin down components are added together analogously to the occupied case in eq 18. Similar to the occupied density matrix fragment, half of the terms of eq 19 vanish in domains with a SO central LMO. Therefore, in these domains the virtual density matrix fragment is also scaled by a factor of two. If SOMOs are present in the ED, the SOMO-unoccupied block of the restricted density matrix is replaced with zeros, enabling the computation of spin-restricted virtual LNOs upon the diagonalization of the density matrix modified in this way. The restricted virtual basis of the LIS ($\{\bar{a}\}$) then consists of the SOMOs

and the retained virtual LNOs with occupation numbers larger than the threshold ε_v . The resulting restricted virtual LNO basis is then semi-canonicalized separately with the spin up and spin down Fock matrices to obtain the unrestricted, semi-canonical virtual LNO basis of the LIS ($\{\underline{a}\}, \{\underline{A}\}$).

3.5 Two-external integral transformations

For the CCSD(T) computations in the LNO basis of the LIS, the integral lists $(\overline{ij}|X)$, $(\overline{ai}|X)$, and $(\overline{ab}|X)$ are required. By utilizing intermediate restricted LNO bases, this demanding step can be performed at a cost very close to the that of our closed-shell LNO-CC method. The $(\overline{ab}|X)$ two-external integral list poses a larger computational challenge due to the fact that a typical LIS contains about 4–5 times as many or more virtual LNOs as occupied ones. To overcome a potential bottleneck for large systems corresponding to naive AO (\rightarrow ED PAO) \rightarrow LNO transformation algorithms, for our closed-shell LNO-CC approaches, we implemented a significantly more efficient solution introducing approximate intermediate basis sets denoted as PAO' and LNO'.⁵² That approach is generalized here to the open-shell case focusing on the algorithmic details, while we refer to our original discussion regarding the theoretical introduction and justification of the PAO' and LNO' functions.⁵²

Briefly, an intermediate PAO' basis is introduced in each ED designed to provide a much more compact expansion of the virtual LNOs than the AO basis of the ED. For that purpose, we project out the restricted DOMOs and SOMOs of the ED from the AOs of its PCD (μ^{PCD}):

$$|\mu'\rangle = \left(1 - \sum_{\hat{j}}^{\text{ED}} |\hat{j}\rangle\langle\hat{j}| \right) |\mu^{\text{PCD}}\rangle = |\mu^{\text{PCD}}\rangle - \sum_{\hat{j}}^{\text{ED}} O_{\hat{j}\mu^{\text{PCD}}} |\hat{j}\rangle, \quad (20)$$

where $O_{\hat{j}\mu^{\text{PCD}}}$ denotes the overlap integral $\langle\hat{j}|\mu^{\text{PCD}}\rangle$. The resulting PAO' orbitals are not identical to the PAOs of the ED but the difference is small for our purposes and well controlled by the strong pair energy threshold (ε_w) and the ED occupied BP completeness criterion (T_{EDo}) as discussed in detail for the closed-shell case.⁵² Therefore, the compact expansion

in the PAO' basis instead of the AO basis of the ED can also be written for approximated virtual LNOs as

$$|\bar{a}'\rangle = \sum_{\mu'} A_{\mu'\bar{a}'} |\mu'\rangle = \sum_{\mu^{\text{PCD}}} A_{\mu^{\text{PCD}}\bar{a}'} |\mu^{\text{PCD}}\rangle - \sum_{\hat{j}}^{\text{ED}} B_{\hat{j}\bar{a}'} |\hat{j}\rangle. \quad (21)$$

Here, \mathbf{A} collects the orbital coefficients of the LNO' functions in the PAO' basis, while $\mathbf{B} = \mathbf{OA}$. Then, the corresponding two-external integrals can be written in the LNO' basis as

$$\begin{aligned} (\bar{a}'\bar{b}'|X) &= \sum_{\mu,\nu}^{\text{PCD}} A_{\mu\bar{a}'} A_{\nu\bar{b}'} (\mu\nu|X) - \sum_{\nu,\hat{j}}^{\text{PCD,ED}} B_{\hat{j}\bar{a}'} A_{\nu\bar{b}'} (\hat{j}\nu|X) \\ &\quad - \sum_{\mu,\hat{k}}^{\text{PCD,ED}} A_{\mu\bar{a}'} B_{\hat{k}\bar{b}'} (\mu\hat{k}|X) + \sum_{\hat{j},\hat{k}}^{\text{ED}} B_{\hat{j}\bar{a}'} B_{\hat{k}\bar{b}'} (\hat{j}\hat{k}|X). \end{aligned} \quad (22)$$

The main benefit is that the summations over the AOs in the above equation run over the PCD, which is usually about 2-3 times smaller than the (complete) AO basis of the ED spanning the original virtual LNOs. We showed previously that this approach can lead to about a factor of 8-9 speedup for large systems with saturated ED sizes with negligible difference compared to the use of the original LNO basis.

Compared to the two-external integrals, the $(\bar{i}\bar{j}|X)$ and $(\bar{i}\bar{b}'|X)$ integrals are obtained with relatively low computational costs by transforming, e.g., the half-transformed integrals in the ED LMO basis $[(\mathcal{I}\bar{\mu}|X)]$ to the LIS LNOs. To complete the two-external integral lists, one additionally needs to evaluate the SOMO-virtual and SOMO-SOMO blocks of $(\bar{a}'\bar{b}'|X)$, which will contribute to the two-external integrals in the spin down unrestricted basis. However, since we so far worked in a restricted basis, where the SOMOs play the dual role of occupied spin up and virtual spin down orbitals, the required blocks are equivalent to the SOMO-virtual and the SOMO-SOMO block of the spin up $(\bar{i}\bar{b}'|X)$ and $(\bar{i}\bar{j}|X)$ integrals, respectively. The complete set of restricted two-external integrals of the LIS can therefore be efficiently compiled from the three sets of integrals obtained so far. As the last step of

the two-external integral transformations, the restricted $(\bar{a}'\bar{b}'|X)$ integrals are transformed to the semi-canonical spin up and spin down bases of the LIS, which represents a negligible additional cost compared to the closed-shell algorithm. For the sake of simplifying the notation, the prime distinction of the virtual LNOs will be omitted in the remaining sections.

3.6 Natural auxiliary functions

Before utilizing the above three-center integrals to assemble the four-center ERIs of the LIS, their auxiliary function index is compressed using the natural auxiliary function (NAF) technique.^{20,101} NAFs are constructed to find the best low-rank approximation of the $J_{\underline{pq},X}$ and $J_{\underline{PQ},X}$ three-center integrals of the LIS. For efficiency considerations, instead of performing the singular value decomposition of the three-center integral tensors, we can build the following \mathbf{W} intermediates:

$$W_{XY}^\uparrow = \sum_{\underline{p} \leq \underline{q}} (\underline{pq}|X) (\underline{pq}|Y) , \quad W_{XY}^\downarrow = \sum_{\underline{P} \leq \underline{Q}} (\underline{PQ}|X) (\underline{PQ}|Y) . \quad (23)$$

It is important to note that for the assembly of four-center ERIs with mixed spin, the auxiliary basis should be the same for both the spin up and spin down DF integrals. Therefore, we diagonalize the spin averaged $\mathbf{W} = (\mathbf{W}^\uparrow + \mathbf{W}^\downarrow)/2$ matrix^{20,101} and retain its eigenvectors as NAFs with eigenvalues above ϵ_{NAF} . We also note that the NAFs obtained in this manner are equivalent to those of the closed-shell LNO-CCSD(T) algorithm in the closed-shell limit due to the factor of $\frac{1}{2}$ in the definition of \mathbf{W} .

3.7 Correlation energy calculation in the LIS

Having the DF integrals transformed to the LNO and NAF bases of the LIS, one can proceed with writing the CCSD and (T) correlation energy contributions of eqs 1 and 2 specifically for the LNO approximations in the LIS.

First, the CCSD correlation energy contribution is evaluated in the LIS as

$$\begin{aligned} \delta E_{\mathcal{I}}^{\text{CCSD}} = & \sum_{\underline{a}} t_{i'}^a f_{i'}^a + \frac{1}{4} \sum_{\underline{abj}} \tau_{i'j}^{ab} \langle \underline{ab} || i' j \rangle + \frac{1}{2} \sum_{\underline{aBJ}} \tilde{\tau}_{i'J}^{aB} \langle \underline{aB} | i' J \rangle \\ & + \sum_{\underline{A}} t_{I'}^A f_{I'}^A + \frac{1}{4} \sum_{\underline{ABJ}} \tau_{I'J}^{AB} \langle \underline{AB} || I' J \rangle + \frac{1}{2} \sum_{\underline{Abj}} \tilde{\tau}_{I'j}^{Ab} \langle \underline{Ab} | I' j \rangle. \end{aligned} \quad (24)$$

The CCSD amplitudes are obtained by solving the CCSD equations in the basis of the LIS, relying on our hand-coded, highly-optimized, and semi-integral direct unrestricted CCSD(T) implementation in MRCC.^{102,103} The computational cost of a CCSD iteration scales in total as the sixth power of the occupied (\underline{n}_o) and virtual (\underline{n}_v) LIS basis dimension [e.g., $\mathcal{O}(\underline{n}_o^2 \underline{n}_v^4)$], which may take a significant portion of the total wall time of a LNO-CCSD(T) computation. In comparison with the closed-shell alternative, the computational cost of the LIS CCSD part is roughly three times higher, stemming from the three times as many terms in the open-shell CCSD equations.

Regarding the (T) correction of the LIS, its naive implementation would scale with the seventh power of the LIS dimensions [i.e., $\mathcal{O}(\underline{n}_o^3 \underline{n}_v^4)$]. This can be reduced to sixth-power scaling in the LIS if we recognize that we only need the triples amplitudes for a fixed central LMO index to evaluate the $\delta E_{\mathcal{I}}^{(T)}$ correlation energy contribution since the perturbative (T) triples amplitudes are not coupled, at least in a semi-canonical basis. For that purpose, we utilize an orbital invariant form of the (T) expressions using the Laplace-transform in our LNO methods.⁶² In this way, the triples energy denominators in the LIS, D_{ijk}^{abc} , are factorized as

$$\frac{1}{D_{ijk}^{abc}} = \int_0^\infty e^{-D_{ijk}^{abc} s} ds \approx \sum_q^{n_q} \omega_q e^{-D_{ijk}^{abc} s_q}, \quad (25)$$

utilizing n_q pieces of quadrature points s_q , with quadrature weights ω_q . Writing the exponential of the energy denominators as products of exponentials of single orbital energies, they can be absorbed in the intermediates contributing to the triple excitation amplitudes. Then, the latter can be directly computed in a basis that contains the central LMO, enabling

an algorithm with a much more favorable scaling in the LIS [i.e., $\mathcal{O}(n_q n_o^2 n_v^4)$].

Utilizing this Laplace-transform approach, introduced in detail in Ref. 62, the (T) contribution of central LMO \mathcal{I} is given as

$$\begin{aligned} \delta E_{\mathcal{I}}^{(T)} = \frac{1}{3} \sum_q \left(\sum_{\substack{a < b < c \\ j < k}} \bar{t}_{i'jk,q}^{abc} \bar{W}_{i'jk,q}^{abc} + \sum_{\substack{a < B < C \\ J < K}} \bar{t}_{i'JK,q}^{aBC} \bar{W}_{i'JK,q}^{aBC} + \sum_{\substack{a < b < C \\ jK}} \bar{t}_{i'jK,q}^{abC} \bar{W}_{i'jK,q}^{abC} \right. \\ \left. + \sum_{\substack{A < B < C \\ J < K}} \bar{t}_{I'JK,q}^{ABC} \bar{W}_{I'JK,q}^{ABC} + \sum_{\substack{A < b < c \\ j < k}} \bar{t}_{I'jk,q}^{Abc} \bar{W}_{I'jk,q}^{Abc} + \sum_{\substack{A < B < c \\ Jk}} \bar{t}_{I'Jk,q}^{ABc} \bar{W}_{I'Jk,q}^{ABc} \right). \end{aligned} \quad (26)$$

Here, the overbar of \bar{t} and \bar{W} indicates that these quantities contain the Laplace energy denominator factors absorbed as defined in Ref. 62. The formal scaling of evaluating eq 26 is similar to that of its closed-shell analogue,⁶² with a larger prefactor of about 3 due to the increased number of spin cases. The number of quadrature points is determined system specifically and, matching our closed-shell experience,⁶² the resulting 3(-4) points provide sufficiently high accuracy for the (T) contribution.

3.8 Comparison with the closed-shell LNO-CCSD(T) method

Here, we compare the computational requirements of the most important steps of the restricted open-shell LNO-CCSD(T) algorithm to their closed-shell counterparts. The computationally demanding part of the MP2 pair energy evaluation is the transformation and contraction of the multipole approximated integrals, of which there are four times as many in the open-shell case. However, even for the largest systems of 500–600 atoms considered here, the pair energy evaluation takes only a few percent of the total run time. The next important algorithmic step is the three-center integral transformation in the ED basis. Even though there are twice as many integrals to transform as in the closed-shell case, all transformation steps of considerable demand are kept at the same cost as in the closed-shell case by utilizing the intermediate restricted ED basis. Next, the MP1 amplitudes, the MP2 energy, and the density matrices in the ED are computed from the available ERIs of the ED. As these

steps have to be carried out in the unrestricted semi-canonical basis, their operation count requirement is roughly 2–3 times as large as that of the closed-shell analogue. As the next step, the ERIs necessary for the LIS CCSD(T) calculations are transformed to the LNO and NAF bases. Again, all transformation steps in the LIS with considerable cost are performed in the intermediate restricted virtual LNO basis to keep the operation count close to that of the closed-shell case. Finally, the correlation energy contribution of the central LMO is computed in the LIS using an unrestricted CCSD(T) formalism requiring about three times as many operations as closed-shell CCSD(T). Since the LIS CCSD(T) part usually takes about half of the total LNO-CCSD(T) wall time, and our unrestricted CCSD(T) implementation is less well parallelized than the restricted one,¹³ one can expect at least twice as high costs for the open-shell LNO-CCSD(T) compared to its closed-shell counterpart for molecules of similar wave function complexity.

3.9 Approximate long-range spin polarization

While for small to medium-sized systems, the additional expense of using unrestricted CCSD(T) formalism is affordable with the present as well as other local CCSD(T) methods,^{48,50} the same might not be the case for demanding, large-scale local CCSD(T) computations. Recently, we introduced the idea of approximating the long-range spin polarization at a lower level of theory, such as our distant pair MP2 model, which turned out to be very effective in our local MP2 implementation.⁶⁵

To introduce this approach also at the LNO-CCSD(T) level, let us consider an extended system with a limited amount of SOMOs, which are often located around a well-defined, narrow region of the molecule. We activate this long-range spin polarization approximation domain specifically when the central LMO of a given domain does not couple strongly to any SOMOs, i.e., when there are only DOMOs and no SOMO within the ED (and LIS). Thus, when this approach is activated, the small, spin-polarized correlation effects between the distant LMO-SOMO pairs are taken into account in the final LNO-CCSD(T) correlation

energy via the multipole approximated MP2 pair energies. In other words, if there are no SOMOs in the EDs of the LMOs that form distant pairs with all SOMOs, then the only spin-polarization effect to take into account in those EDs comes from the spin-dependence of the domain Fock matrices. This manifests in the slight splitting of the spin-up and spin-down LMOs and PAOs of the ED due to their semi-canonicalization with the corresponding spin-dependent Fock matrices. However, since the central LMO of such EDs forms distant pairs with all SOMOs, the AOs around the SOMOs are mostly not part of the ED. Then, one can expect that the blocks of the spin-dependent Fock matrices corresponding to AOs far away from the SOMOs are moderately spin-polarized and that the magnitude of this secondary spin-polarization effect in the ED energy contribution is small. Therefore, if we approximate this spin polarization effect by replacing the spin dependent ED Fock matrix blocks with their spin-averaged counterpart, then all orbital sets of the ED and LIS become spin independent. Then, we can utilize the restricted local MP2 and LNO-CCSD(T) algorithms to evaluate the ED and LIS correlation energy contributions for the domains far away from the SOMOs. As a consequence, in large systems where the SOMOs are well localized to a single region of the molecule, we can take advantage of the more efficient restricted formalism and thus approach the computational effectiveness of a completely closed-shell calculation.⁶⁵

3.10 Scaling of the algorithm

The presented open-shell LNO-CCSD(T) method, just as its closed-shell analogue, achieves asymptotically linear scaling with the system size for its rate determining steps. Since the size of the EDs saturates for sufficiently large systems, all computations performed in these domains exhibit asymptotically linear scaling.⁶⁵ On the other hand, computations performed on the whole system, such as the occupied orbital localization, the PAO construction, and the pair energy computation, scale as the third, third, and second power of the system size, respectively. However, the cumulative run time of these computations is negligible compared to the total run time, even for the largest systems considered here. Moreover,

especially for molecules of several hundreds of atoms, the SCF computation can potentially take a considerable portion of the run time. The implementation for the HF exchange term in MRCC is in principle also asymptotically linear scaling^{66,89} (alongside an efficient cubic-scaling DF-based Coulomb term). However, the HF exchange computations reach the linear scaling system size range for much larger molecules than our local correlation methods, which is typically above a few thousand atoms for three dimensional structures. Below that size range, about cubic-scaling can be expected, as the local DF capability of our HF exchange implementation can effectively reduce the scaling already above a few hundred atoms.

In terms of memory requirements, the open-shell LNO-CCSD(T) algorithm requires the storage of six matrices with dimensions equal to the total number of AOs of the entire molecule, while the preceding SCF procedure needs eight such matrices. These quadratic-scaling arrays are only relevant above tens of thousands of AOs. Furthermore, the remaining arrays allocated within the EDs are asymptotically constant in size, and their memory requirement is also highly optimized following the ideas exploited in our restricted LNO-CCSD(T) code.^{52,53} The shared-memory parallelization model used within a single node also contributes to the highly memory economic nature of our implementation.

4 Computational details and test systems

4.1 Technical details

The open-shell LNO-CCSD(T) method as presented here is implemented in the development version of the MRCC quantum chemical program suite^{102,103} and will be made available in a forthcoming release of the MRCC package. The default or **Normal** truncation thresholds of the local approximations and their corresponding keywords are collected in Table 2. These default values are equivalent to those utilized in the most recent closed-shell formulation of LNO-CCSD(T).^{52,53}

In all HF and reference canonical CCSD(T) calculations the DF approximation was

employed using the MRCC program. To accelerate the HF calculations for the largest systems (containing more than 500 atoms), the HF exchange contribution was evaluated in local fitting domains, as described in Section 3.1. The core electrons (including the subvalence electrons of the iron and cobalt atoms) were kept frozen in all correlated computations. The remaining occupied orbitals were localized with the algorithm of Boys¹⁰⁴ separately for the DO and SO subspace.

The calculations presented make use of the triple- ζ valence basis set including polarization functions (def2-TZVP) developed by Weigend and Ahlrichs¹⁰⁵ and Dunning’s (augmented) correlation-consistent polarized valence basis sets [(aug-)cc-pVXZ, $X = D, T$ and Q],¹⁰⁶ with the revised aug-cc-pV($X+d$)Z basis sets used for second-row atoms.¹⁰⁷ Additionally, the triple- ζ , weighted core-valence basis of Balabanov and Peterson (cc-pwCVTZ) was utilized for the iron atom of the Fe(III) complex investigated in Section 5.1, following the work of Radoń.¹⁰⁸ The appropriate auxiliary basis sets of Weigend et al. were used for all AO bases.¹⁰⁹ The extrapolations of the HF¹¹⁰ and the correlation energies¹¹¹ towards the complete basis set (CBS) limit were performed according to the standard expressions.

The relative energy deviations with respect to some reference energy (E_{ref}) are obtained as $(100\%) \cdot (E_{\text{LNO-CCSD(T)}} - E_{\text{ref}}) / E_{\text{ref}}$. To characterize the achieved accuracy on various test sets the following statistical measures were utilized: the maximum absolute error (MAX), the mean absolute error (MAE), and, to measure the consistency of errors, the standard deviation of the absolute errors (STD). The presented timing measurements were performed using 2.5 GHz Intel Xeon Gold 6148 processors with 20 physical cores and at most 192 GB of total memory in dual-socket nodes.

4.2 Benchmark sets and test systems

The accuracy of the open-shell LNO-CCSD(T) correlation energies and energy differences are benchmarked on three test sets containing small to medium sized molecules. The first test set contains the radical stabilization energies (RSEs) of 30 small organic molecules

Table 2: Default threshold values used in this study.

Symbol	Keyword ^a	Value
ε_o	lnoepso	10^{-5}
ε_v	lnoepsv	10^{-6}
ε_w [E_h]	wpairtol	10^{-5}
T_{EDo}	bpedo	0.9999
ε_{NAF} [E_h]	naf_cor	10^{-2}
T_{LT}	laptol	10^{-2}
g_w	epairscale	5
h_w [E_h^{-1}]	epairestfact	50

^a Name of the corresponding keyword in the MRCC program package.^{102,103}

(RSE30).¹¹² These structures were selected from the RSE43 compilation¹¹³ and reoptimized in ref 57. Next, a test set of adiabatic ionization potentials for 21 organic molecules (IP21) is considered as defined in ref 57 (with structures reoptimized in ref 65). Lastly, singlet–triplet energy gaps of aryl carbenes (AC) are also investigated invoking the AC12 test set as compiled in ref 114.

To investigate the convergence of the local approximations as a function of the corresponding thresholds, six larger open-shell systems containing 23–80 atoms are also selected. The structures of vitamin E succinate, trityl radical, and artemisinin are obtained from ref 47, testosterone is taken from ref 57, whereas diphenylcarbene comes from the AC12 test set of ref 114. Lastly, the 4' Fe(III) complex of ref 108 is also investigated, constructed by replacing the methyl groups of $[\text{Fe}(\text{acac}_2\text{trien})]^+$ with hydrogens (where $\text{H}_2\text{acac}_2\text{trien}$ is the Schiff base obtained from the 1:2 condensation of triethylenetetramine with acetylacetone). The structures of these medium-sized systems are depicted in Figure 2.

To demonstrate the performance of the open-shell LNO-CCSD(T) method, large-scale calculations were also carried out on four systems of 175–601 atoms (see Figure 3). Among these is the 175-atom Fe(II) complex taken from ref 115, both in its triplet and quintet spin state. Moreover, the Cob^{II}alamin (Cbl) radical containing 179 atoms, a product of the homolytic bond breaking of the coenzyme B₁₂ (5'-deoxyadenosylcobalamin, dAdoCbl), was also considered.¹¹⁶ The largest systems investigated are the models of photosystem II (PSII)

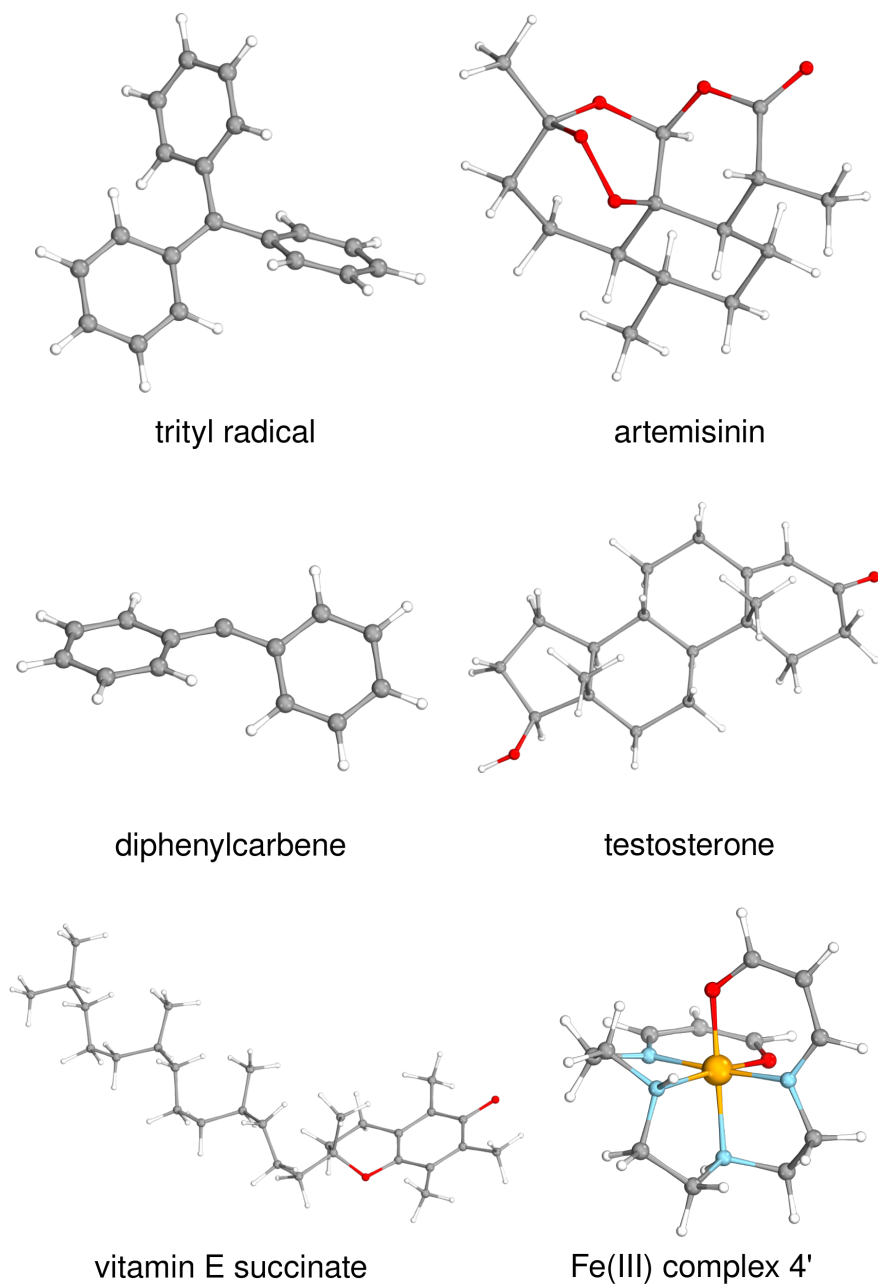


Figure 2: Structures of the six systems of intermediate size used to investigate the convergence of the local approximations. Top row: trityl radical and triplet artemisinin; middle row: diphenylcarbene in its triplet state and testosterone cation; bottom row: vitamin E succinate radical and the 4' Fe(III) complex of Radoń¹⁰⁸ considered both with doublet and sextet multiplicities.

bicarbonate⁴⁷ containing 565 atoms and of the D-amino-acid oxidase (DAAO)¹¹⁷ made up of 601 atoms. The singlet and triplet spin states of PSII bicarbonate were calculated utilizing the def2-TZVP basis set and QRO reference determinants. The particular complexities and technicalities of this reference computation are discussed in ref 65. In the case of DAAO, two steps along the oxidative half-reaction of D-alanine oxidation are taken from Kiss et al.,¹¹⁷ where the reduced flavin adenine dinucleotide (FAD) moiety of DAAO is oxidized by O_2^- to produce the oxidized form of FAD and H_2O_2 . The corresponding reactant and product structures labeled by $O1^T$ and $O3^{CSS}$ in ref 117 are also depicted in Figure 5 of ref 65.

5 Convergence of the local approximations

In this section, the effect of the local approximations of the open-shell LNO-CCSD(T) method on its accuracy is investigated individually. To this end, the thresholds associated with such approximations are systematically varied, and the resulting energy deviations from approximation-free DF-CCSD(T) or other reference methods are observed. Almost all of the employed approximations were extensively benchmarked in our related open-shell LMP2 and closed-shell LNO-CCSD(T) studies.^{52,53,60,65,66} Consequently, the presented tests focus on the parameters with the largest impact on accuracy: the strong pair energy threshold ϵ_w and the LNO occupation limits ϵ_o and ϵ_v . Additionally, approximations that were previously not employed in the context of either any open-shell local CCSD(T) implementation or our open-shell LMP2 method, such as the distant spin polarization approximation discussed in Section 3.8 and the NAFs introduced in Section 3.6, are also investigated. For the remaining truncation thresholds, affecting closed and open-shell systems similarly, such as the BP parameters of the LMO atom list definition, the previously benchmarked values are adapted, which will be tested in combination with all approximations in Section 6.

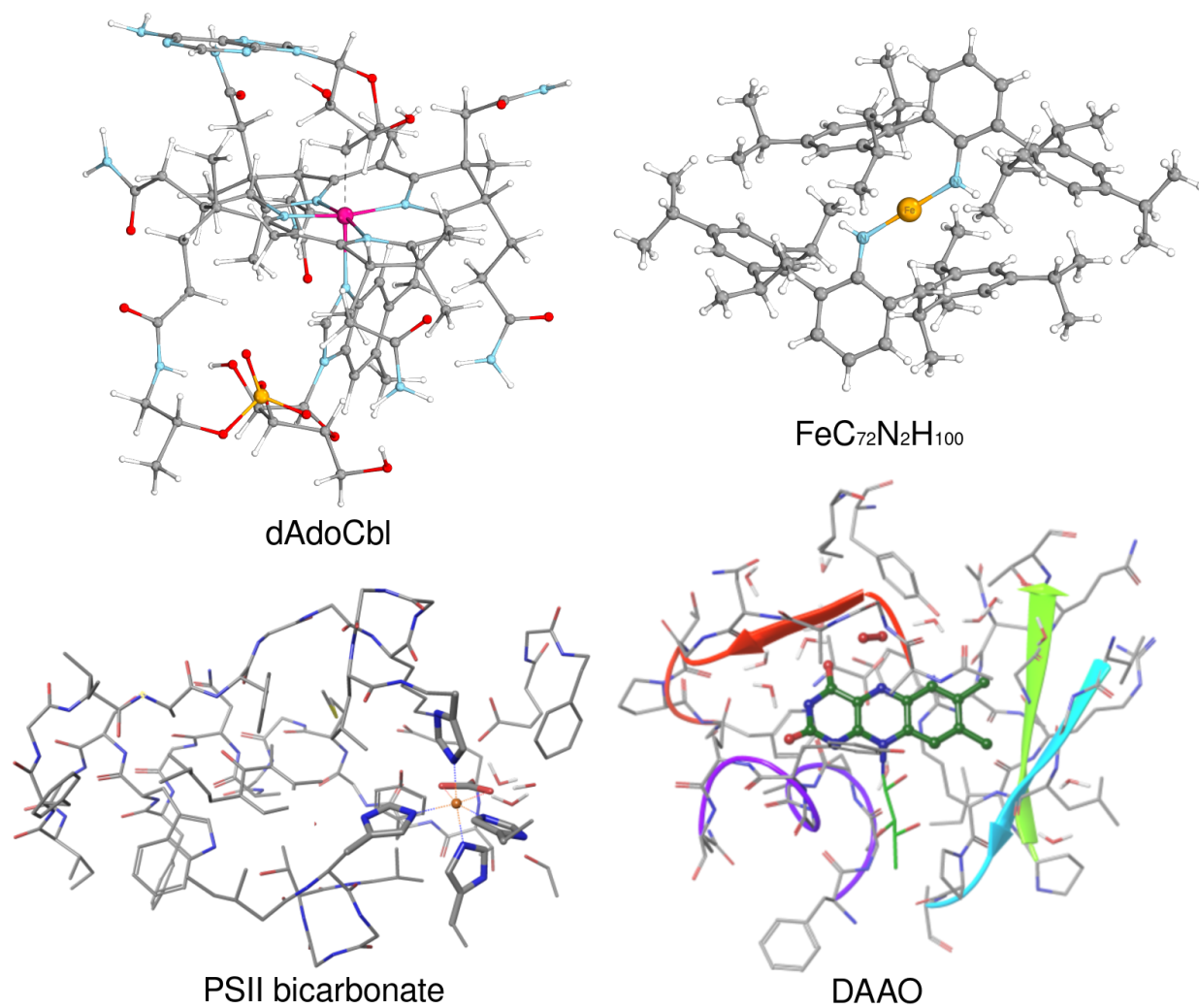


Figure 3: The structure of the four largest systems considered. Top row: the 5'-deoxyadenosylcobalamin (209 atoms, out of which the Cob^{II}alamin (Cbl) radical part has 179 atoms) and the 175-atom iron(II) complex FeC₇₂N₂H₁₀₀. Bottom row: a 565-atom model of photosystem II bicarbonate and the O1^T structure of D-amino-acid oxidase containing 601 atoms.

5.1 LNO selection

First, the errors introduced by the truncation of the LNO basis of the LISs are illustrated on three systems. The system sizes were chosen to be as large as possible so that the LNO truncation is already active in all domains (with up to 50% of the occupied and 85% of virtual LNOs discarded), while still keeping the possibility of using DF-CCSD(T) energies as reference. The one-particle basis set of cc-pVTZ is chosen for the description of the hydrogen addition of the triphenylmethyl radical, while cc-pVQZ is used for the calculation of the singlet-triplet gap of diphenylcarbene. A composite basis set composed of cc-pwCVTZ for the iron atom, cc-pVTZ for the atoms connected to the iron atom, and cc-pVDZ for all other atoms is chosen to investigate the doublet-sextet gap of the Fe(III) complex, following Radoń.¹⁰⁸ Since our previous studies found that setting the ratio of the virtual and occupied LNO selection thresholds ($\varepsilon_o/\varepsilon_v$) to 10 gives satisfactory results,^{52,53,58,59} this ratio was kept in this work.

As can be seen from the left panel of Figure 4, the correlation energy errors decrease monotonically and converge to zero when the LNO selection thresholds are tightened (the approximations other than the LNO truncation are turned off in these tests). Moreover, the 99.9% accuracy approached already with the default setting of $\varepsilon_o = 10^{-5}$ and $\varepsilon_v = 10^{-6}$ is highly satisfactory for such non-trivial systems. Considering the right panel of the same figure, the three energy differences of the associated chemical processes also exhibit good convergence. The singlet-triplet energy gap of diphenylcarbene with almost perfect error cancellation already with the loosest thresholds is probably an exception, representative only for very local properties, like spin state differences, localized mostly on a few atoms. Nonetheless, the deviations introduced by the LNO truncation for this system remain below 0.3 kcal/mol for all investigated ε_o values, and the trityl RSE and the doublet-sextet gap of the Fe(III) complex also drop below the satisfactory 0.5 kcal/mol value after $\varepsilon_o = 10^{-5}$. Thus, the default LNO threshold values of $\varepsilon_o = 10^{-5}$ and $\varepsilon_v = 10^{-6}$ are selected, which are recommended also based on our closed-shell LNO-CCSD(T) benchmarks^{52,53} as both the

correlation energy and energy difference deviations are sufficiently converged (with at most 0.14% and 0.53 kcal/mol deviations, respectively, for the examples of this section). We note in passing that this value is not directly comparable to the frozen NO^{20,21} or pair NO⁴⁸ threshold settings as only the strong pair LMOs are included in our orbital specific LNO density matrices (Sect. 3.4), while the summation is not restricted for the frozen NO and even more restricted for the orbital pair specific pair NO density matrices.

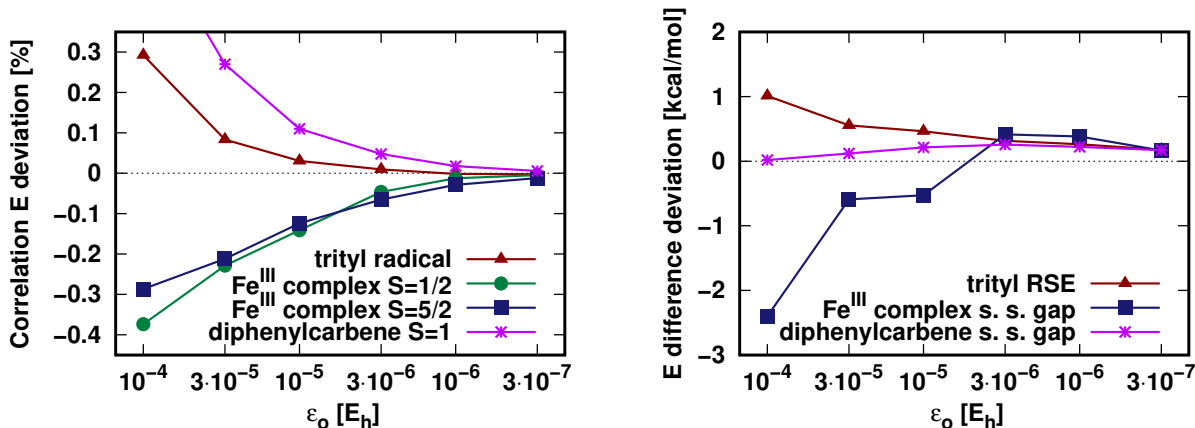


Figure 4: Deviations of the LNO-CCSD(T) energies from reference DF-CCSD(T) values as a function of the LNO selection thresholds. Left panel: relative correlation energy differences, right panel: deviations in energy differences (s. s. = spin state). The ratio of the occupied and virtual LNO selection thresholds were kept constant, $\epsilon_o/\epsilon_v = 10$. For a description of the systems and basis sets utilized, see Section 5.1.

5.2 Strong pair classification

The convergence of the LNO-CCSD(T) energies with the tightening of the pair energy threshold (ϵ_w) is also investigated. The restriction of the strong pair LMO list, controlled by this threshold, only starts to take effect for somewhat larger systems, preventing one from utilizing canonical DF-CCSD(T) calculations as reference. To overcome this difficulty, first, we recall that this approximation is the main source of error in the LMP2 correlation energy, which we benchmarked against DF-MP2 for a number of larger open-shell molecules containing 42–81 atoms. Namely, the singlet-triplet gap of artemisinin, the vertical ionization energy of testosterone, and the hydrogen addition to vitamin E succinate were investigated

using the aug-cc-pVTZ basis set.⁶⁵ There, we found that the energy differences converged already at the default $\varepsilon_w = 10^{-5} E_h$ LMP2 setting with at most 0.03% relative correlation energy and 0.05 kcal/mol energy difference deviations.

To also assess the accuracy of LNO-CCSD(T) against an affordable reference on the same systems, we set all thresholds to their default value except for the ε_w parameter, which we scan (see Figure 5). Therefore, as reference, the converged calculation with the very tight $\varepsilon_w = 10^{-6} E_h$ can be taken. With that, we find the relative correlation energy and absolute energy difference deviations to converge quickly to zero also for LNO-CCSD(T) as the pair energy threshold is tightened. Note that the absolute energy difference considered for vitamin E succinate is about four (two) times larger than that for artemisinin (testosterone), meaning that the slightly larger energy difference deviations observed for vitamin E succinate correspond to similarly high quality relative deviations for all systems. The value of $\varepsilon_w = 10^{-5} E_h$, yielding highly satisfactory LNO-CCSD(T) results of less than 0.04% correlation energy deviations and highest energy difference deviations of just above 0.1 kcal/mol, can be again safely chosen as default. Our strong pair threshold can be compared to the settings of other local approaches, and for example, matches the value employed in the TightPNO settings of the DLPNO methods.⁴⁸ This value matches our closed-shell setting as well, where the restriction of the strong pair LMO list was successfully benchmarked for even larger molecules of 92–260 atoms against DF-MP2 references.⁶⁶

5.3 Approximate long-range spin polarization

The accuracy of the distant spin polarization approximation introduced in Section 3.8 is evaluated on both local correlation energies and energy differences. Here, the reference energies are obtained from default LNO-CCSD(T) calculations without activating the distant spin polarization approximation. Since the approximation is only active in the EDs that do not contain SOMOs, reasonably large systems should be chosen to activate and test this approach. Results for three medium to large test systems of 81–179 atoms are collected in

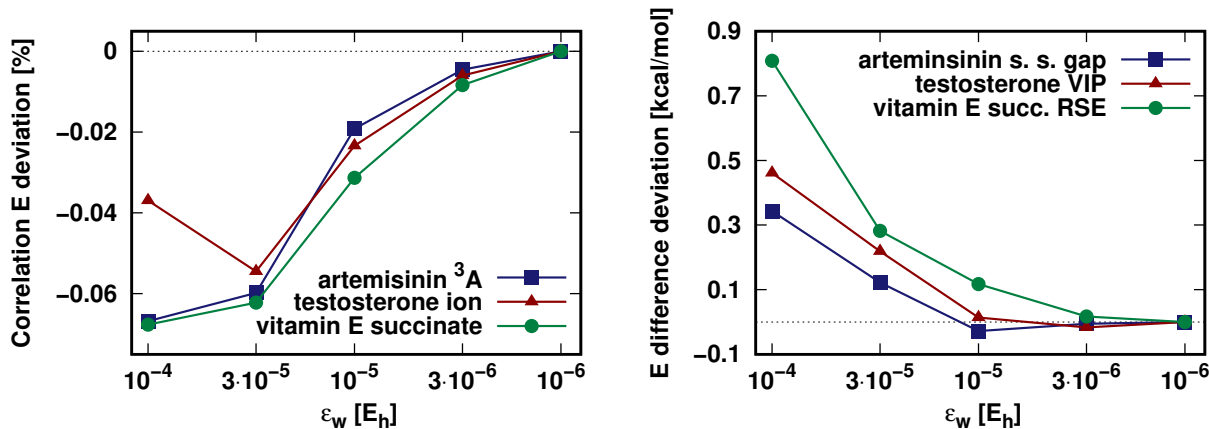


Figure 5: Relative LNO-CCSD(T) correlation energy (left) and absolute energy difference (right) deviations. A converged LNO-CCSD(T) calculation with $\varepsilon_w = 10^{-6}$ was used as reference. See Section 5.2 for a description of the utilized systems and basis sets.

Table 3. As shown in the last column of Table 3, already at this relatively moderate system size, one half to two thirds of the EDs without SOMOs can be treated with the more efficient closed-shell formalism. At the same time, all relative correlation energy errors are well below 10^{-4} %, indicating only a marginal loss of accuracy. The absolute and relative errors in the computed energy differences of about 0.1–0.4 cal/mol or up to $5.4 \cdot 10^{-4}$ % are similarly negligible. It is also worthwhile noting that here, the RSE and bond breaking processes in the first and third lines involve also large closed-shell species, where this approximation does not have an analogue eliminating the possibility of error cancellation. In conclusion, the errors introduced by the distant spin polarization approximation are substantially smaller than those brought in by the other local approximations employed. These results are in line with the performance of the analogous approximation in our open-shell LMP2 approach, where we have also performed such tests up to 500-600 atoms.⁶⁵ Notably, in that size range, 80–90% of the EDs are free of SOMOs and can be treated with the more efficient restricted algorithms.

Table 3: Accuracy of the long-range spin polarization approximation compared to reference LNO-CCSD(T) correlation energies and energy differences obtained without this approximation.

	atoms	LMOs	$E^{\text{LNO-CCSD(T)}}$ error [%]	error in energy difference		EDs without SOMOs [%]
				[cal/mol]	[%]	
vitamin E succinate	81	89	$1.2 \cdot 10^{-5}$	0.44	$1.1 \cdot 10^{-4}$	54
FeC ₇₂ N ₂ H ₁₀₀	5A 3A	205	$3.3 \cdot 10^{-5}$	0.22	$5.4 \cdot 10^{-4}$	54
		204	$3.5 \cdot 10^{-5}$			54
Cbl radical	179	250	$2.0 \cdot 10^{-6}$	0.087	$1.6 \cdot 10^{-4}$	68

5.4 Natural auxiliary functions

In this section, the effect of the NAF truncation is investigated on three systems of up to 81 atoms with various basis sets. These include the vertical ionization energy of testosterone (cc-pVTZ), the hydrogen addition process of vitamin E succinate (aug-cc-pVTZ), and the singlet-triplet gap of diphenylcarbene (cc-pVQZ). Since the NAF approximation was found very effective in the analogous closed-shell context,^{52,53} and the only difference compared to the latter is that here, both spin up and spin down integrals have to be expanded in the same NAF basis, we expect very similar accuracy.

To evaluate the performance of NAFs, LNO-CCSD(T) calculations with default settings except for the scanned ε_{NAF} parameter are compared against the reference obtained with $\varepsilon_{\text{NAF}} = 0$. Figure 6 shows the convergence of the correlation energy (left) and energy difference (right) deviations. The correlation energy deviations converge almost completely already with $\varepsilon_{\text{NAF}} = 10^{-2} E_h$, while energy difference deviations approach the reference also almost completely monotonically. Since the largest energy difference deviation is already below 0.01 kcal/mol with $\varepsilon_{\text{NAF}} = 10^{-2} E_h$, this value is chosen as the default, again matching the default setting of closed-shell LNO-CCSD(T).

6 Benchmarks for small and medium sized systems

In this section, the cumulative effect of all local truncation thresholds employed at their default values is benchmarked against canonical DF-CCSD(T) references. Statistical measures

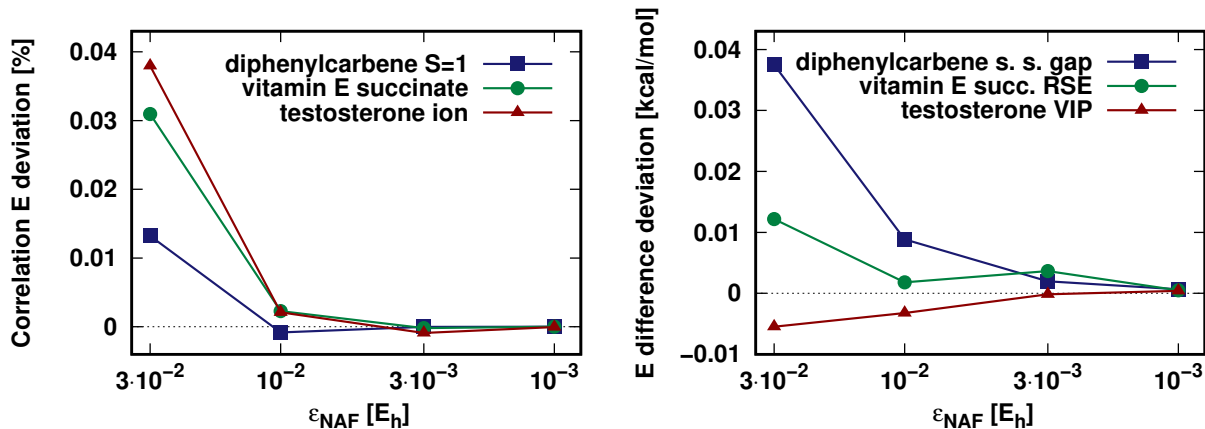


Figure 6: Errors of LNO-CCSD(T) energies and energy differences computed with various NAF truncation thresholds. The reference energies are obtained from LNO-CCSD(T) calculations with $\epsilon_{\text{NAF}} = 0$. Left panel: relative correlation energy differences, right panel: absolute deviations in energy differences.

for the correlation energy and energy difference deviations are presented for three test sets containing radical stabilization energies, ionization potentials, and spin-state energies. The reference data utilized in this section is available in the Supporting Information.

6.1 Accuracy of local correlation energies

Here, the accuracy of the open-shell LNO-CCSD(T) correlation energies is evaluated against reference DF-CCSD(T) values. The investigated systems include 30 radicals (from the RSE30 test set), 21 cations from IP21, and the 12 aryl carbenes in their triplet state of the AC12 test set.

The statistical measures of the correlation energy deviations of the three test sets are collected in the third columns of Tables 4–6. For all test sets, the MAEs of the relative correlation energy deviations with the triple- ζ quality basis sets are below 0.05%, with all errors not exceeding 0.11%. The highest triple- ζ deviation of 0.11% is observed for the oxalic acid cation, where a large (0.35) single excitation amplitude of the reference DF-CCSD(T) solution suggests some multireference character. Apart from this challenging system lying at the boundary of single reference CC’s applicability, all triple- ζ local correlation energy deviations

are below 0.1%, which is highly satisfactory already with the default LNO-CCSD(T) settings. When moving to the quadruple- ζ and triple- ζ -quadruple- ζ extrapolated [CBS(T,Q)] results of the RSE30 test set, one observes a slight increase in the MAE and MAX local correlation energy deviations. It is worth noting that the largest relative error is observed in both cases for the methyl radical. Due to its small size, the reference correlation energy values for this system are tiny, greatly increasing the relative correlation energy deviations even for the smallest of absolute errors. Compared to the high quality of MAEs for the correlation energies, it is important to point out that their STD is always as good as or, in a number of cases, 2–3 times smaller than the MAE for all three test sets. This indicates that one can expect excellent cancellation between the correlation energy errors upon computing various energy differences, as found in the next section.

Table 4: Relative correlation energy deviations and absolute errors of the LNO-CCSD(T) reaction energies for radical stabilization energies of the RSE30 test set using the default thresholds.

basis	error measure	error in $E^{\text{LNO-CCSD(T)}}$ [%]	error in RSE [kcal/mol]
aug-cc-pV(T+d)Z	MAX	0.081	0.111
	MAE	0.029	0.042
	STD	0.034	0.041
aug-cc-pV(Q+d)Z	MAX	0.195	0.136
	MAE	0.099	0.037
	STD	0.050	0.050
CBS(T,Q)	MAX	0.292	0.173
	MAE	0.173	0.050
	STD	0.063	0.069

Table 5: Relative correlation energy deviations and absolute errors of the LNO-CCSD(T) vertical ionization energies of the IP21 test set using the default thresholds.

basis	error measure	error in $E^{\text{LNO-CCSD(T)}}$	error in IP	
		[%]	[eV]	[kcal/mol]
aug-cc-pV(T+d)Z	MAX	0.109	0.027	0.62
	MAE	0.047	0.007	0.16
	STD	0.058	0.009	0.21

Table 6: Relative correlation energy deviations and absolute errors of the LNO-CCSD(T) singlet-triplet energy gaps of the AC12 test set using the default thresholds.

basis	error measure	error in $E^{\text{LNO-CCSD(T)}}$ [%]	error in S-T gap [kcal/mol]
cc-pVDZ	MAX	0.073	0.337
	MAE	0.041	0.106
	STD	0.017	0.118
cc-pVTZ	MAX	0.084	0.452
	MAE	0.043	0.244
	STD	0.021	0.159

6.2 Radical stabilization energies

The RSE30 test set contains radical stabilization reactions of the form:



where $\text{R} \cdot$, stands for the different radicals containing C, N, O, F, P, and S atoms.⁵⁷ The deviations of LNO-CCSD(T) radical stabilization energies from reference DF-CCSD(T) ones are collected in the last column of Table 4. With MAEs of at most 0.05 kcal/mol and MAX errors below 0.2 kcal/mol, even for the CBS extrapolated energy differences, the results are highly satisfactory. Here, the 0.050 % and 0.063% STDs of the quadruple- ζ and CBS(T,Q) correlation energies, respectively, are important indicators to explain that the triple- ζ , quadruple- ζ , and CBS(T,Q) RSE values are of the same high quality. One might also compare the results obtained with the aug-cc-pV(T+d)Z basis to those of the PNO-UCCSD(T)-F12⁵⁰ method, calculated for the same RSE30 test set. Naturally, the PNO-UCCSD(T)-F12 and the LNO-CCSD(T) are not directly comparable as the former contains additional explicitly correlated terms. Nonetheless, the LNO-CCSD(T) results appear to be comparable to or slightly better than the PNO-UCCSD(T)-F12 results with defaults settings providing 0.076 kcal/mol RMS, 0.064 kcal/mol MAE, and 0.192 kcal/mol MAX errors with the aug-cc-pVTZ basis set.⁵⁰ It is always important to keep in mind at such comparisons that both the PNO-UCCSD(T)- and the LNO-CC-type methods can be converged to the same local approximation free limit,

and here, we compare only the results obtained with the default settings of each method determined according to considerations that are not identical for the two methods.

6.3 Vertical ionization potentials

The statistical measures of the vertical ionization potentials of the IP21 test set are shown in Table 5. Considering the fact that the IPs represent quite large energy differences in the range of 8–14 eV (184–323 kcal/mol), the mean average error of 0.007 eV (0.16 kcal/mol) is excellent. The MAX IP deviation of 0.027 eV can be attributed to the benzoquinone cation, the reference DF-CCSD(T) solution of which exhibits a moderately large maximum singles amplitude of 0.15. Removing this outlier, the MAX IP error drops to below 0.02 eV (0.04 kcal/mol), and the standard deviation also decreases to 0.007 eV (0.16 kcal/mol). The PNO-UCCSD(T)-F12 method achieves comparable or slightly more accurate results for this test set with its default settings and the aug-cc-pVTZ basis set (0.17 kcal/mol RMS, 0.11 kcal/mol MAE, and 0.50 kcal/mol MAX).⁵⁰

6.4 Singlet–triplet energy gaps

Finally, the singlet–triplet spin state energy gaps of the aryl carbenes are benchmarked, and the results are collected in the last column of Table 6. The MAE of 0.24 kcal/mol for the cc-pVTZ basis is somewhat larger than those for the other two test sets but is still only a fraction of a kcal/mol marking chemical accuracy. The slightly increased errors are at least partly explained by the fact that the system sizes of this test set are larger than that of the other two, with the average AC12 system containing more than twice as many heavy atoms as either RSE30 or IP21. To put the results in perspective, let us consider that the basis set incompleteness error of canonical CCSD(T) for the AC12 test set found to be 2.74 and 0.98 kcal/mol, respectively, for the cc-pVDZ and cc-pVTZ bases,¹¹⁴ which are significantly larger than the MAX local errors reported here.

7 Performance and computational requirements for larger systems

The current capabilities of the open-shell LNO-CCSD(T) method are illustrated on large-scale calculations on four three-dimensional systems containing 175–601 atoms (see Table 7). Of these systems, the Cbl radical and the $\text{FeC}_{72}\text{N}_2\text{H}_{100}$ complex represent the higher end of the typical size range, e.g., in homogeneous catalysis applications (cca. 175 atoms), while the bicarbonate or DAAO models of 565–601 atoms are in the typical size range of quantum systems in biochemical applications, for example, when modeling proteins with active centers in a quantum mechanics/molecular mechanics (QM/MM) framework. The Cbl radical and the $\text{FeC}_{72}\text{N}_2\text{H}_{100}$ complex pose a formidable challenge for local correlation methods due to the spin polarized transition metal atoms near their centers. This leads not only to high strong pair ratios of 22–26% but also to LISs that are somewhat more extended than those of the two protein models. It is therefore not completely surprising to see that the wall clock run times of the domain CCSD(T) calculations for the transition metal complexes somewhat exceed even those for PSII bicarbonate or the singlet DAAO species. Another reason for the unusually long CCSD(T) run time for the $\text{FeC}_{72}\text{N}_2\text{H}_{100}$ system is its moderate multireference character,¹¹⁵ which hinders the convergence of the CCSD iterations, resulting in up to 27 iterations performed in particular domains, compared to the usual 9–11 iterations.

Regarding the other contributions to the run time, it is reassuring that the formally cubic scaling orbital localization and the quadratically scaling pair energy evaluation take negligible time, even for the largest systems. However, the cubic-scaling local DF-HF computations can take time comparable to the LNO-CCSD(T) correlation energy computation for the large protein models, in accord with our experience with even larger proteins of 1000–2000 atoms and our restricted LNO-CCSD(T) implementation.^{52,53} Moreover, as observed on the singlet and triplet DAAO models, the time required for the integral transformation to the LNO basis increases by only up to 30% compared to the closed-shell algorithm, which increase is mainly

Table 7: Average (maximum) domain sizes, orbital space dimensions, DF-HF and correlation energies (in E_h), wall-clock times (in hours)^a, and memory requirements (in GB) for LNO-CCSD(T) computations of large molecules.

molecule	FeC ₇₂ N ₂ H ₁₀₀	Cbl radical	bicarbonate	DAAO	
atoms	175	179	565	601	
LMOs	205	250	788	837	838
SOMOs	4	1	2	0	2
AO basis	def2-TZVP	def2-TZVP	def2-TZVP	def2-TZVP	
basis functions	2939	3369	10 560	11 006	
auxiliary functions	7306	8379	26 064	27 071	
strong pairs [%]	26	22	6.8	5.9	5.9
atoms in ED	116 (171)	116 (169)	132 (294)	124 (268)	137 (353)
AOs in ED	2129 (2915)	2346 (3278)	2634 (5953)	2453 (5408)	2700 (6943)
LMOs in ED	54 (111)	55 (115)	55 (112)	50 (113)	51 (184)
PAOs in ED	1045 (1949)	1023 (1829)	975 (2037)	866 (1884)	905 (2932)
Occupied LNOs	32 (64)	31 (74)	29 (58)	26 (65)	27 (100)
Virtual LNOs	131 (268)	125 (272)	121 (250)	112 (269)	113 (449)
type of reference	ROHF	ROHF	QRO (UHF)	RHF	ROHF
DF-HF energy	-4156.159945	-5878.796625	-15197.85043 ^c	-14740.93979	-14740.90403
LNO-CCSD(T) energy	-13.3715	-17.6478	-56.0763	-59.1499	-59.1421
HF (1 iteration)	0.50	0.72	3.1 ^b	2.5 ^b	2.6 ^b
orbital localization	< 0.01	< 0.01	0.12	0.06	0.05
pair energies	0.03	0.04	0.17	0.09	0.19
LMP2 in EDs	0.74	0.97	1.1	0.67	1.93
integral trf. to LNOs	3.0	4.0	11	10	13
CCSD(T) in LISs	75	55	29	10	76
total LNO-CCSD(T)	85	67	57	33	106
memory req.	12	18	17	7.1	88

^a Using two 20-core, 2.4 GHz Intel Xeon Gold 6148 CPUs

^b Using the default local fitting domain size. The final iteration with larger fitting domains took about 3.5–4.8 times longer.

^c DF-HF energies calculated with semi-canonical QRO orbitals.

attributed to the larger domain sizes within the triplet system. As discussed in Section 3.8, an unrestricted implementation of the open-shell integral transformation would have twice the computational cost of the closed-shell analogue. However, due to the restricted intermediate basis and the distant spin polarization approximation, the computational costs of the most expensive integral transformation steps approach those of the closed-shell algorithm. This small overhead of the open-shell integral transformations means that, while this step is one of the rate determining parts of the closed-shell algorithm (30% of the singlet DAAO run time), its significance is much reduced in the open-shell case (12% of the triplet DAAO run time).

Looking at the average (maximum) ED sizes of 116–137 (169–353) atoms, at least for

the large protein models, the individual domain sizes are saturated, and we are getting close to the linear scaling regime of the ED and LIS computation parts of the LNO-CCSD(T) method. It is also clear that performing hundreds of MP2 energy computations for these domains is well beyond the capabilities of any conventional MP2 implementation. These ED MP2 computations are only feasible with a thoroughly optimized, Cholesky-decomposition- or Laplace-transform-based algorithm and extensive utilization of local approximations.^{65,66} In these (necessarily) large EDs, we can exploit the locality of the LMOs, use local AO and DF domains, and redundancy free MP1 amplitude expressions. As a result of these enhancements, the MP2 energy computation in the largest ED with over 180 LMOs and almost 3000 PAOs takes less than an hour. Turning to the average LIS sizes, the necessity and efficiency of the LNO compression is evident. The LISs contain somewhat over half as many occupied and around 6–8 times fewer virtual orbitals after the truncation of the ED basis dimensions. Without this compression of the correlated subspaces, the evaluation of hundreds of domain CCSD(T) energies would clearly be unfeasible.

It is interesting to inspect the DAAO systems more closely due to the availability of very similar closed-shell and open-shell computations. Here, one immediately notices the surprisingly large maximum domain size of the triplet DAAO model in the last column of Table 7. Containing over 350 atoms, 100 occupied, and 449 virtual LNOs, the size of this domain is unprecedented in LNO-CCSD(T) calculations performed with the default threshold values. The central LMO of this domain is one of the SO LMOs of the system, depicted on the left panel of Figure 7. This orbital spreads across the entire FAD moiety and, to a small extent, even spills onto the peroxide molecule. The large spatial extent of this orbital means that it forms an unusually large number of strong pairs with other LMOs. This in turn leads to a large number of ED atoms and finally, to large LIS orbital dimensions.

Beyond the inherently delocalized electronic structure of the reduced FAD, another reason for the extreme spatial extent of this SO LMO is to be found in the restricted nature of the employed orbital localization scheme. During localization, the SOMOs are mixed only

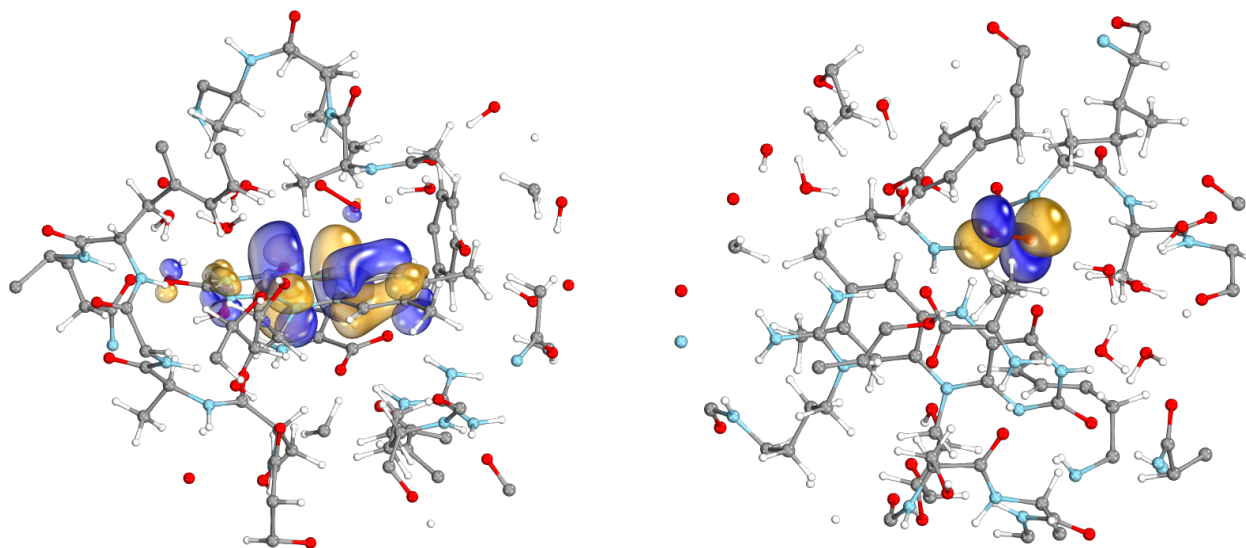


Figure 7: D-amino-acid oxidase model with its two singly occupied molecular orbitals depicted with blue and yellow surfaces. The left panel shows the highly delocalized orbital spreading over the entire flavin adenine dinucleotide, while the other singly occupied orbital localized almost exclusively on the peroxide radical can be seen on the right.

among themselves, in order to avoid the splitting of the spin up and spin down occupied orbitals. The triplet state of the DAAO model contains merely two SOMOs, but at the same time the singly occupied subspace spreads across the entire FAD moiety and the peroxide molecule. Due to the large spatial extent of the SO subspace and the small number of orbitals spanning it, it is not possible to localize the SO LMOs better in a restricted scheme. A (partial) remedy could be to employ a spin unrestricted localization scheme, where the two SOMOs would be allowed to mix with all other spin up orbitals. This could probably lead to orbitals more localized than the highly delocalized SOMO in question at the cost of introducing spin splitting to the LMOs, doubling the number of domains in all computations. Despite the complex electronic structure and extreme domain size, the LNO-CCSD(T) energy contribution of this delocalized central LMO can still be evaluated on a single CPU. This would not be possible without efficient local approximations, such as the compact NAF and LNO orbital spaces, the highly optimized CCSD(T) code featuring integral direct evaluation of four-center more-than-two-external ERIs, and the redundancy free Laplace-transform (T) formulation of LNO-CCSD(T). The broader significance of the present triplet DAAO LNO-

CCSD(T) computation is that it represents (i) the largest CCSD(T) computation in the literature with over 600 atoms and more than twice as many orbitals as in the def2-SVP level DLPNO-based PSII bicarbonate computation of Kumar et al.,⁵⁶ (ii) probably the most complicated system studied so far concerning the issues with delocalized SOMOs, illustrating that such problematic situations can also be handled with highly-optimized implementations.

It is also encouraging to see the comfortably manageable minimal memory requirements in the last line of Table 7, despite the complexity of the investigated systems and the employed triple- ζ quality basis set. Less than 20 GB of minimal memory is required for most systems in Table 7, while with about 90 GB, LNO-CCSD(T) was feasible even for the most challenging triplet DAAO system. Such low memory requirements are the result of the highly optimized integral evaluation and transformation routines and the tiled computation of the most memory intensive CCSD(T) terms, which removes the need to store any four-center ERI arrays with more than two external indices in the memory. Additionally to the small minimal memory requirements, the wall-clock run times of 1.5–4.5 days also indicate that LNO-CCSD(T) calculations for open-shell systems containing several hundred atoms, at least with reasonable triple- ζ quality basis sets and the default settings, have become routine tasks even on a single many-core CPU with easily accessible amount of memory.

Lastly, the reaction energies and energy differences obtained with the LNO-CCSD(T) method for the above discussed four large examples are collected in Table 8. Comparing the estimates of the energy differences obtained at the LMP2 and LNO-CCSD(T) levels of theory, one observes significant differences for all applications. The deviation of the relative energy estimates between these two methods is smallest for the DAAO oxidation and dAdoCbl formation processes, however, with 2.1–2.4 kcal/mol, this deviation is still clearly outside of the range of chemical accuracy typically defined to be 1 kcal/mol. Moreover, for the quintet–triplet energy differences of the two iron complexes, the disagreement between the two methods increases to 6.6–6.8 kcal/mol. These results are in accord with the observations emerging more and more often in the literature, enabled by large-scale CCSD(T) benchmarks,

namely that second order perturbation theory can often be insufficient to accurately describe such challenging extended systems and processes. This also verifies the need for higher order and more reliable methods.

Table 8: Contributions to reaction energies and spin-state gaps in kcal/mol for the four largest examples evaluated with the def2-TZVP basis. ΔE^{HF} – DF-HF, ΔE^{LMP2} – LMP2 correlation contribution, $\Delta E^{\text{LNO-CCSD(T)}}$ – LNO-CCSD(T) correlation contribution, $\Delta E_{\text{total}}^{\text{LNO-CCSD(T)}}$ – total LNO-CCSD(T).

	ΔE^{HF}	ΔE^{LMP2}	$\Delta E^{\text{LNO-CCSD(T)}}$	$\Delta E_{\text{total}}^{\text{LNO-CCSD(T)}}$
FeC ₇₂ N ₂ H ₁₀₀ ⁵ A – ³ A	57.56	-10.57	-17.39	40.17
bicarbonate ⁵ A – ³ A	52.67	-12.07	-18.70	33.97
Cbl + Ado → dAdoCbl	50.41	-102.47	-104.82	-54.41
DAAO oxidation	22.44	7.01	4.88	27.32

8 Summary and conclusions

A high-spin open-shell extension of the asymptotically linear-scaling local natural orbital (LNO) based CCSD(T) method is presented. The efficiency of the open-shell algorithm approaches that of our closed-shell LNO-CCSD(T) method⁵² for large molecules due to the utilization of restricted open-shell Hartree–Fock (ROHF) or Kohn–Sham (ROKS) reference determinants, the use of restricted open-shell intermediate basis sets, and a novel approximation of long-range spin-polarization effects at the CCSD(T) level. For compact molecules, where the domain CCSD(T) computations are rate-determining, the open-shell LNO-CCSD(T) method requires at least 2–3 times more operations and data than the closed-shell variant, just as for alternative local CCSD(T) approaches. The presented method is identical to its closed-shell counterpart when closed-shell systems are considered, enabling the combined use of these algorithms to efficiently compute consistent energy differences between open- and closed-shell species.

The proposed approach efficiently extends and generalizes the optimized algorithms developed for our other local correlation methods to the open-shell LNO-CCSD(T) case,^{52,53,62,65,66}

for example, the algorithms for domain constructions, auxiliary basis compression, integral transformation, and LNO construction. Its implementation is integral-direct, memory- and disk use economic, and OpenMP-parallel, although the scaling of the current version above 10–20 cores should be improved. The local approximations are defined without relying on empirical parameters, such as real-space cutoffs of fragment definition, and adapt to the complexity of the wave function by construction. Additionally, a highly accurate open-shell local MP2 correlation energy is also obtained for free as part of the LNO-CCSD(T) calculation. The open-shell LNO-CCSD(T) code will be made available open-access for academic use in a forthcoming release of the MRCC program suite.^{102,103}

The effect of the local approximations and the corresponding thresholds are profiled on both energy differences and correlation energies. Systematically tightening these thresholds, satisfactory convergence towards the exact CCSD(T) energies is observed. The errors caused by the local approximations using the default threshold values are benchmarked on diverse test sets of radical stabilization energies, vertical ionization potentials, and singlet–triplet energy gaps. For systems where the exact CCSD(T) reference is accessible, mean absolute errors are found to be below 0.25 kcal/mol, while maximum errors do not exceed 0.7 kcal/mol and are below 0.5 kcal/mol for all but the most challenging systems.

The performance of the open-shell LNO-CCSD(T) algorithm is demonstrated on systems of 175–601 atoms using triple- ζ quality basis sets. Among these, a triplet state of the O₂ reduction process via a D-amino acid oxidase model of 601 atoms is especially challenging, not only due to its unprecedented size with over 11,000 basis functions, but also because of a singly-occupied molecular orbital delocalized over about 20 atoms. The complicated triplet electronic structure of the 565-atom model of the bicarbonate protein in photosystem II is also investigated using about twice as many basis functions as in a previous local CCSD study.⁴⁷ Each of these calculations is feasible using less than 100 GB of memory, on a single node of 20–40 cores within cca. 2.5–4.5 days of wall clock run time. These results demonstrate that the presented LNO-CCSD(T) algorithm extends the applicability range of open-shell CCSD(T)

calculations up to biochemical systems of 500-600 atoms using reasonable triple- ζ basis sets and commodity hardware. Having the ROHF/ROKS/QRO reference at hand, which can become relatively costly in this size range the open-shell LNO-CCSD(T) method should scale to even larger systems/basis sets, which approaches the capabilities of closed-shell LNO-CCSD(T), currently standing at about 1000-2000 atoms and 45,000 basis functions.^{52,53}

Appendix: Derivation of virtual density matrix fragments

In this Appendix, the orbital notation of the main text is simplified using $i, j, k \dots$ and $a, b, c \dots$ indices for the occupied and virtual (semi-)canonical orbitals, respectively. Again, lowercase letters stand for spin up, uppercase letters label spin down orbitals, while MP1 amplitudes are denoted by t . Moreover, expressions only for the spin up density matrix fragments are shown since their spin down counterparts can be obtained by mirroring all spins of the expressions.

The spin up virtual second-order density matrix reads as

$$D_{ab} = \frac{1}{2} \left(\sum_{cij} t_{ij}^{ac} t_{ij}^{bc} + \sum_{CiJ} t_{iJ}^{aC} t_{iJ}^{bC} + \sum_{CIj} t_{Ij}^{aC} t_{Ij}^{bC} \right). \quad (28)$$

After transforming one of the occupied orbitals, say the i and I indices, to the LMO basis, one has three different choices to continue the derivation. If we first restrict the summations to the central LMO index, we arrive at the virtual density matrix expression used in the presented open-shell LNO-CCSD(T) algorithm (eq 19). The other two forms can be obtained by recognizing that the last two terms of eq 28 can be combined together in two ways. One can either switch the order of the occupied indices of the last term of eq 28 and rename them

to $I \rightarrow J, j \rightarrow i$, yielding:

$$\begin{aligned}
D_{ab} &= \frac{1}{2} \left(\sum_{cij} t_{ij}^{ac} t_{ij}^{bc} + \sum_{CiJ} t_{iJ}^{aC} t_{iJ}^{bC} + \sum_{CIj} t_{Ij}^{aC} t_{Ij}^{bC} \right) = \\
& \frac{1}{2} \left(\sum_{cij} t_{ij}^{ac} t_{ij}^{bc} + \sum_{CiJ} t_{iJ}^{aC} t_{iJ}^{bC} + \sum_{CIj} t_{iJ}^{aC} t_{iJ}^{bC} \right) = \frac{1}{2} \sum_{cij} t_{ij}^{ac} t_{ij}^{bc} + \sum_{CiJ} t_{iJ}^{aC} t_{iJ}^{bC}.
\end{aligned} \tag{29}$$

Here, we exploited that exchanging the occupied indices of the antisymmetric MP1 amplitudes introduces two sign changes that cancel each other ($t_{Ij}^{aC} = -t_{jI}^{aC}$). Transforming the i index to the LMO basis and restricting its summation to the central LMO, one obtains a virtual density matrix expression different from the first one:

$$D_{ab}^{\mathcal{I}} = \frac{1}{2} \sum_{cj} t_{i'j}^{ac} t_{i'j}^{bc} + \sum_{CJ} t_{i'J}^{aC} t_{i'J}^{bC}. \tag{30}$$

However, in this form the central LMO with spin down occupation does not contribute to the spin up virtual density matrix fragment at all. For the last form, one can perform the exchanging and renaming of the occupied indices on the second term of eq 28, yielding:

$$\begin{aligned}
D_{ab} &= \frac{1}{2} \left(\sum_{cij} t_{ij}^{ac} t_{ij}^{bc} + \sum_{CiJ} t_{iJ}^{aC} t_{iJ}^{bC} + \sum_{CIj} t_{Ij}^{aC} t_{Ij}^{bC} \right) = \\
& \frac{1}{2} \left(\sum_{cij} t_{ij}^{ac} t_{ij}^{bc} + \sum_{CIj} t_{Ij}^{aC} t_{Ij}^{bC} + \sum_{CIj} t_{Ij}^{aC} t_{Ij}^{bC} \right) = \frac{1}{2} \sum_{cij} t_{ij}^{ac} t_{ij}^{bc} + \sum_{CIj} t_{Ij}^{aC} t_{Ij}^{bC}.
\end{aligned} \tag{31}$$

If this form were to be used after restricting the summations to the central LMO index, both the central LMO with spin up and spin down occupation would contribute to the spin up density matrix. However, the contribution of the spin down central LMO would be weighted twice as high as that of the spin up central LMO.

Supporting Information Available

See the Supporting Information for reference DF-CCSD(T) energies of the RSE30, IP21, and AC12 test sets.

Acknowledgement

The authors are grateful for the financial support from the ERC Starting Grant No. 101076972, “aCCuracy”, the National Research, Development, and Innovation Office (NKFIH, Grant No. FK142489 and KKP126451), the János Bolyai Research Scholarship of the Hungarian Academy of Sciences, ÚNKP-23-5-BME-408 New National Excellence Program of the Ministry for Culture and Innovation sourced from the NKFIH fund, and the computing time granted on the Hungarian HPC Infrastructure at NIIF Institute, Hungary. The research reported in this paper is part of project BME-EGA-02, implemented with the support provided by the Ministry of Innovation and Technology of Hungary from the National Research, Development and Innovation Fund, financed under the TKP2021 funding scheme.

References

- (1) Zhang, J.; Head-Gordon, M. Electronic structures and reaction dynamics of open-shell species. *Phys. Chem. Chem. Phys.* **2009**, *11*, 4699.
- (2) Krylov, A. I. *Reviews in Computational Chemistry*; John Wiley & Sons, Ltd, 2017; Chapter 4, p 151.
- (3) Stanton, J. F.; Gauss, J. *Advances in Chemical Physics*; John Wiley & Sons, Ltd, 2003; p 101.
- (4) Bally, T.; Borden, W. T. *Reviews in Computational Chemistry*; John Wiley & Sons, Ltd, 1999; p 1.

- (5) Møller, C.; Plesset, M. S. Note on an Approximation Treatment for Many-Electron Systems. *Phys. Rev.* **1934**, *46*, 618.
- (6) Bartlett, R. J.; Musiał, M. Coupled-cluster theory in quantum chemistry. *Rev. Mod. Phys.* **2007**, *79*, 291.
- (7) Crawford, T. D.; Schaefer III, H. F. An Introduction to Coupled Cluster Theory for Computational Chemists. *Rev. Comp. Chem.* **1999**, *14*, 33.
- (8) 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.
- (9) Epifanovsky, E.; Zuev, D.; Feng, X.; Khistyayev, K.; Shao, Y.; Krylov, A. I. General implementation of the resolution-of-the-identity and Cholesky representations of electron repulsion integrals within coupled-cluster and equation-of-motion methods: Theory and benchmarks. *J. Chem. Phys.* **2013**, *139*, 134105.
- (10) DePrince, A. E.; Sherrill, C. D. Accuracy and Efficiency of Coupled-Cluster Theory Using Density Fitting/Cholesky Decomposition, Frozen Natural Orbitals, and a t_1 -Transformed Hamiltonian. *J. Chem. Theory Comput.* **2013**, *9*, 2687.
- (11) Peng, C.; Calvin, J. A.; Valeev, E. F. Coupled-cluster singles, doubles and perturbative triples with density fitting approximation for massively parallel heterogeneous platforms. *Int. J. Quantum Chem.* **2019**, *119*, e25894.
- (12) Shen, T.; Zhu, Z.; Zhang, I. Y.; Scheffler, M. Massive-parallel Implementation of the Resolution-of-Identity Coupled-cluster Approaches in the Numeric Atom-centered Orbital Framework for Molecular Systems. *J. Chem. Theory Comput.* **2019**, *15*, 4721.
- (13) Gyevi-Nagy, L.; Kállay, M.; Nagy, P. R. Integral-direct and parallel implementation of the CCSD(T) method: Algorithmic developments and large-scale applications. *J. Chem. Theory Comput.* **2020**, *16*, 366.

- (14) Kobayashi, R.; Rendell, A. P. A direct coupled cluster algorithm for massively parallel computers. *Chem. Phys. Lett.* **1997**, *265*, 1.
- (15) Pitoňák, M.; Aquilante, F.; Hobza, P.; Neogrády, P.; Noga, J.; Urban, M. Parallelized implementation of the CCSD(T) method in MOLCAS using optimized virtual orbitals space and Cholesky decomposed two-electron integrals. *Collect. Czech. Chem. Commun.* **2011**, *76*, 713.
- (16) Asadchev, A.; Gordon, M. S. Fast and Flexible Coupled Cluster Implementation. *J. Chem. Theory Comput.* **2013**, *9*, 3385.
- (17) Solomonik, E.; Matthews, D.; Hammond, J. R.; Stanton, J. F.; Demmel, J. A massively parallel tensor contraction framework for coupled-cluster computations. *J. Parallel Distr. Comput.* **2014**, *74*, 3176.
- (18) Löwdin, P.-O. Quantum theory of many-particle systems. I. Physical interpretations by means of density matrices, natural spin-orbitals, and convergence problems in the method of configurational interaction. *Phys. Rev.* **1955**, *97*, 1474.
- (19) Taube, A. G.; Bartlett, R. J. Frozen Natural Orbital Coupled-Cluster Theory: Forces and Application to Decomposition of Nitroethane. *J. Chem. Phys.* **2008**, *128*, 164101.
- (20) Gyevi-Nagy, L.; Kállay, M.; Nagy, P. R. Accurate reduced-cost CCSD(T) energies: parallel implementation, benchmarks, and large-scale applications. *J. Chem. Theory Comput.* **2021**, *17*, 860.
- (21) Nagy, P. R.; Gyevi-Nagy, L.; Kállay, M. Basis set truncation corrections for improved frozen natural orbital CCSD(T) energies. *Mol. Phys.* **2021**, *119*, e1963495.
- (22) Kutzelnigg, W.; Klopper, W. Wave functions with terms linear in the interelectronic coordinates to take care of the correlation cusp. I. General theory. *J. Chem. Phys.* **1991**, *94*, 1985.
- (23) Klopper, W.; Manby, F. R.; S.Ten-no,; Valeev, E. F. R12 methods in explicitly corre-

- lated molecular electronic structure theory. *Int. Rev. Phys. Chem.* **2006**, *25*, 427.
- (24) Hättig, C.; Klopper, W.; Köhn, A.; Tew, D. P. Explicitly Correlated Electrons in Molecules. *Chem. Rev.* **2012**, *112*, 4.
- (25) Kállay, M.; Horváth, R. A.; Gyevi-Nagy, L.; Nagy, P. R. Size-consistent explicitly correlated triple excitation correction. *J. Chem. Phys.* **2021**, *155*, 034107.
- (26) Kállay, M.; Horváth, R. A.; Gyevi-Nagy, L.; Nagy, P. R. Basis set limit CCSD(T) energies for extended molecules via a reduced-cost explicitly correlated approach. *J. Chem. Theory Comput.* **2023**, *19*, 174.
- (27) Zaleśny, R.; Papadopoulos, M.; Mezey, P.; Leszczynski, J. *Linear-Scaling Techniques in Computational Chemistry and Physics: Methods and Applications*; Challenges and Advances in Computational Chemistry and Physics; Springer: Netherlands, 2011.
- (28) Herbert, J. M. Fantasy versus reality in fragment-based quantum chemistry. *J. Chem. Phys.* **2019**, *151*, 170901.
- (29) Gordon, M., Ed. *Fragmentation: Toward Accurate Calculations on Complex Molecular Systems*; Wiley: New York, 2017.
- (30) Pulay, P. Localizability of dynamic electron correlation. *Chem. Phys. Lett.* **1983**, *100*, 151.
- (31) Pulay, P.; Saebø, S. Orbital-invariant formulation and second-order gradient evaluation in Møller–Plesset perturbation theory. *Theor. Chem. Acc.* **1986**, *69*, 357.
- (32) Collins, M. A.; Bettens, R. P. A. Energy-Based Molecular Fragmentation Methods. *Chem. Rev.* **2015**, *115*, 5607.
- (33) Raghavachari, K.; Saha, A. Accurate Composite and Fragment-Based Quantum Chemical Models for Large Molecules. *Chem. Rev.* **2015**, *115*, 5643.
- (34) Li, W.; Piecuch, P.; Gour, J. R.; Li, S. Local correlation calculations using standard

- and renormalized coupled-cluster approaches. *J. Chem. Phys.* **2009**, *131*, 114109.
- (35) Li, W.; Ni, Z.; Li, S. Cluster-in-molecule local correlation method for post-Hartree–Fock calculations of large systems. *Mol. Phys.* **2016**, *114*, 1447.
- (36) Guo, Y.; Becker, U.; Neese, F. 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.* **2018**, *148*, 124117.
- (37) Ziólkowski, M.; Jansík, B.; Kjærgaard, T.; Jørgensen, P. Linear scaling coupled cluster method with correlation energy based error control. *J. Chem. Phys.* **2010**, *133*, 014107.
- (38) Eriksen, J. J.; Baudin, P.; Ettenhuber, P.; Kristensen, K.; Kjærgaard, T.; Jørgensen, P. Linear-Scaling Coupled Cluster with Perturbative Triple Excitations: The Divide–Expand–Consolidate CCSD(T) Model. *J. Chem. Theory Comput.* **2015**, *11*, 2984.
- (39) Li, W.; Li, S. Divide-and-conquer local correlation approach to the correlation energy of large molecules. *J. Chem. Phys.* **2004**, *121*, 6649.
- (40) Kobayashi, M.; Nakai, H. Divide-and-conquer-based linear-scaling approach for traditional and renormalized coupled cluster methods with single, double, and noniterative triple excitations. *J. Chem. Phys.* **2009**, *131*, 114108.
- (41) Nakano, M.; Yoshikawa, T.; Hirata, S.; Seino, J.; Nakai, H. Computerized implementation of higher-order electron-correlation methods and their linear-scaling divide-and-conquer extensions. *J. Comput. Chem.* **2017**, *38*, 2520.
- (42) Stoll, H. Correlation energy of diamond. *Phys. Rev. B* **1992**, *46*, 6700.
- (43) Friedrich, J.; Hanrath, M.; Dolg, M. Using symmetry in the framework of the incremental scheme: Molecular applications. *Chem. Phys.* **2008**, *346*, 266.
- (44) Friedrich, J.; Dolg, M. Fully Automated Incremental Evaluation of MP2 and CCSD(T) Energies: Application to Water Clusters. *J. Chem. Theory Comput.* **2009**, *5*, 287.

- (45) Fiedler, B.; Schmitz, G.; Hättig, C.; Friedrich, J. Combining Accuracy and Efficiency: An Incremental Focal-Point Method Based on Pair Natural Orbitals. *J. Chem. Theory Comput.* **2017**, *13*, 6023.
- (46) Kurashige, Y.; Yang, J.; Chan, G. K.-L.; Manby, F. R. Optimization of orbital-specific virtuals in local Møller–Plesset perturbation theory. *J. Chem. Phys.* **2012**, *136*, 124106.
- (47) Saitow, M.; Becker, U.; Riplinger, C.; Valeev, E. F.; Neese, F. A new near-linear scaling, efficient and accurate, open-shell domain-based local pair natural orbital coupled cluster singles and doubles theory. *J. Chem. Phys.* **2017**, *146*, 164105.
- (48) 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*, 024116.
- (49) Ma, Q.; Werner, H.-J. Explicitly correlated local coupled-cluster methods using pair natural orbitals. *Wiley Interdiscip. Rev.: Comput. Mol. Sci.* **2018**, *8*, e1371.
- (50) Ma, Q.; Werner, H.-J. 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.* **2021**, *17*, 902.
- (51) Schmitz, G.; Hättig, C. Perturbative triples correction for local pair natural orbital based explicitly correlated CCSD(F12*) using Laplace transformation techniques. *J. Chem. Phys.* **2016**, *145*, 234107.
- (52) Nagy, P. R.; Samu, G.; Kállay, M. Optimization of the linear-scaling local natural orbital CCSD(T) method: Improved algorithm and benchmark applications. *J. Chem. Theory Comput.* **2018**, *14*, 4193.
- (53) Nagy, P. R.; Kállay, M. Approaching the basis set limit of CCSD(T) energies for

- large molecules with local natural orbital coupled-cluster methods. *J. Chem. Theory Comput.* **2019**, *15*, 5275.
- (54) Guo, Y.; Riplinger, C.; Becker, U.; Liakos, D. G.; Minenkov, Y.; Cavallo, L.; Neese, F. 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.* **2018**, *148*, 011101.
- (55) 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*, 214102.
- (56) Kumar, A.; Neese, F.; Valeev, E. F. Explicitly correlated coupled cluster method for accurate treatment of open-shell molecules with hundreds of atoms. *J. Chem. Phys.* **2020**, *153*, 094105.
- (57) Ma, Q.; Werner, H.-J. Scalable Electron Correlation Methods. 7. Local Open-Shell Coupled-Cluster Methods Using Pair Natural Orbitals: PNO-RCCSD and PNO-UCCSD. *J. Chem. Theory Comput.* **2020**, *16*, 3135.
- (58) Rolik, Z.; Kállay, M. A general-order local coupled-cluster method based on the cluster-in-molecule approach. *J. Chem. Phys.* **2011**, *135*, 104111.
- (59) Rolik, Z.; Szegedy, L.; Ladjánszki, I.; Ladóczki, B.; Kállay, M. An efficient linear-scaling CCSD(T) method based on local natural orbitals. *J. Chem. Phys.* **2013**, *139*, 094105.
- (60) Kállay, M. Linear-scaling implementation of the direct random-phase approximation. *J. Chem. Phys.* **2015**, *142*, 204105.
- (61) Mester, D.; Nagy, P. R.; Kállay, M. Reduced-scaling correlation methods for the excited states of large molecules: Implementation and benchmarks for the second-order algebraic-diagrammatic construction approach. *J. Chem. Theory Comput.* **2019**, *15*,

6111.

- (62) Nagy, P. R.; Kállay, M. Optimization of the linear-scaling local natural orbital CCSD(T) method: Redundancy-free triples correction using Laplace transform. *J. Chem. Phys.* **2017**, *146*, 214106.
- (63) Anacker, T.; Tew, D. P.; Friedrich, J. First UHF Implementation of the Incremental Scheme for Open-Shell Systems. *J. Chem. Theory Comput.* **2016**, *12*, 65.
- (64) Zhang, J.; Dolg, M. Third-Order Incremental Dual-Basis Set Zero-Buffer Approach for Large High-Spin Open-Shell Systems. *J. Chem. Theory Comput.* **2015**, *11*, 962.
- (65) Szabó, P. B.; Csóka, J.; Kállay, M.; Nagy, P. R. Linear scaling open-shell MP2 approach: algorithm, benchmarks, and large-scale applications. *J. Chem. Theory Comput.* **2021**, *17*, 2886.
- (66) Nagy, P. R.; Samu, G.; Kállay, M. An integral-direct linear-scaling second-order Møller–Plesset approach. *J. Chem. Theory Comput.* **2016**, *12*, 4897.
- (67) Al-Hamdani, Y. S.; Nagy, P. R.; Barton, D.; Kállay, M.; Brandenburg, J. G.; Tkatchenko, A. Interactions between large molecules pose a puzzle for reference quantum mechanical methods. *Nat. Commun.* **2021**, *12*, 3927.
- (68) Hégyely, B.; Nagy, P. R.; Ferenczy, G. G.; Kállay, M. Exact density functional and wave function embedding schemes based on orbital localization. *J. Chem. Phys.* **2016**, *145*, 064107.
- (69) Hégyely, B.; Nagy, P. R.; Kállay, M. Dual basis set approach for density functional and wave function embedding schemes. *J. Chem. Theory Comput.* **2018**, *14*, 4600.
- (70) Lauderdale, W. J.; Stanton, J. F.; Gauss, J.; Watts, J. D.; Bartlett, R. J. Many-body perturbation theory with a restricted open-shell Hartree–Fock reference. *Chem. Phys. Lett.* **1991**, *187*, 21.
- (71) Knowles, P. J.; Andrews, J. S.; Amos, R. D.; Handy, N. C.; Pople, J. A. Restricted

- Møller–Plesset theory for open-shell molecules. *Chem. Phys. Lett.* **1991**, *186*, 130.
- (72) Földes, T.; Madarász, Á.; Révész, Á.; Dobi, Z.; Varga, S.; Hamza, A.; Nagy, P. R.; Pihko, P. M.; Pápai, I. Stereocontrol in Diphenylprolinol Silyl Ether Catalyzed Michael Additions: Steric Shielding or Curtin–Hammett Scenario? *J. Am. Chem. Soc.* **2017**, *139*, 17052.
- (73) Rahman, S.; Wineman-Fisher, V.; Nagy, P. R.; Al-Hamdani, Y.; Tkatchenko, A.; Varma, S. Methyl-Induced Polarization Destabilizes the Noncovalent Interactions of N-Methylated Lysines. *Chem. Eur. J.* **2021**, *27*, 11005.
- (74) Nagy, P. R.; Gyevi-Nagy, L.; Lőrincz, B. D.; Kállay, M. Pursuing the basis set limit of CCSD(T) non-covalent interaction energies for medium-sized complexes: case study on the S66 compilation. *Mol. Phys.* **2023**, *121*, e2109526.
- (75) Ott, A.; Nagy, P. R.; Benkő, Z. Stability of Carbocyclic Phosphinyl Radicals: Effect of Ring Size, Delocalization, and Sterics. *Inorg. Chem.* **2022**, *61*, 16266.
- (76) Sylvetsky, N.; Banerjee, A.; Alonso, M.; Martin, J. M. L. Performance of Localized Coupled Cluster Methods in a Moderately Strong Correlation Regime: Hückel–Möbius Interconversions in Expanded Porphyrins. *J. Chem. Theory Comput.* **2020**, *16*, 3641.
- (77) Efremenko, I.; Martin, J. M. L. Coupled Cluster Benchmark of New DFT and Local Correlation Methods: Mechanisms of Hydroarylation and Oxidative Coupling Catalyzed by Ru(II, III) Chloride Carbonyls. *J. Phys. Chem. A* **2021**, *125*, 8987.
- (78) Paulechka, E.; Kazakov, A. Efficient Estimation of Formation Enthalpies for Closed-Shell Organic Compounds with Local Coupled-Cluster Methods. *J. Chem. Theory Comput.* **2018**, *14*, 5920.
- (79) Semidalas, E.; Martin, J. M. L. The MOBH35 metal-organic barrier heights reconsidered: performance of local-orbital coupled cluster approaches in different static correlation regimes. *J. Chem. Theory Comput.* **2022**, *18*, 883.

- (80) Sandler, I.; Sharma, S.; Chan, B.; Ho, J. Accurate Quantum Chemical Prediction of Gas-Phase Anion Binding Affinities and Their Structure-Binding Relationships. *J. Phys. Chem. A* **2021**, *125*, 9838.
- (81) Santra, G.; Semidalas, E.; Mehta, N.; Karton, A.; Martin, J. M. L. S66x8 noncovalent interactions revisited: new benchmark and performance of composite localized coupled-cluster methods. *Phys. Chem. Chem. Phys.* **2022**, *24*, 25555.
- (82) Santra, G.; Martin, J. M. Performance of Localized-Orbital Coupled-Cluster Approaches for the Conformational Energies of Longer n-Alkane Chains. *J. Phys. Chem. A* **2022**, *126*, 9375.
- (83) Karton, A.; Chan, B. Performance of local G4(MP2) composite ab initio procedures for fullerene isomerization energies. *Comput. Theor. Chem.* **2022**, *1217*, 113874.
- (84) Wineman-Fisher, V.; Al-Hamdani, Y.; Nagy, P. R.; Tkatchenko, A.; Varma, S. Improved description of ligand polarization enhances transferability of ion–ligand interactions. *J. Chem. Phys.* **2020**, *153*, 094115.
- (85) Wineman-Fisher, V.; Delgado, J. M.; Nagy, P. R.; Jakobsson, E.; Pandit, S. A.; Varma, S. Transferable interactions of Li⁺ and Mg²⁺ ions in polarizable models. *J. Chem. Phys.* **2020**, *153*, 104113.
- (86) Stanton, J. F.; Gauss, J.; Watts, J. D.; Bartlett, R. J. A direct product decomposition approach for symmetry exploitation in many-body methods. I. Energy calculations. *J. Chem. Phys.* **1991**, *94*, 4334.
- (87) 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.
- (88) Polly, R.; Werner, H.-J.; Manby, F. R.; Knowles, P. J. Fast Hartree–Fock theory using local fitting approximations. *Mol. Phys.* **2004**, *102*, 2311.

- (89) Csóka, J.; Kállay, M. Speeding up density fitting Hartree–Fock calculations with multipole approximations. *Mol. Phys.* **2020**, *118*, e1769213.
- (90) Csóka, J.; Kállay, M. Speeding up Hartree–Fock and Kohn–Sham calculations with first-order corrections. *J. Chem. Phys.* **2021**, *154*, 164114.
- (91) Foster, J. M.; Boys, S. F. Canonical Configurational Interaction Procedure. *Rev. Mod. Phys.* **1960**, *32*, 300.
- (92) Pipek, J.; Mezey, P. A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions. *J. Chem. Phys.* **1989**, *90*, 4916.
- (93) Knizia, G. Intrinsic Atomic Orbitals: An Unbiased Bridge between Quantum Theory and Chemical Concepts. *J. Chem. Theory Comput.* **2013**, *9*, 4834.
- (94) Jansík, B.; Høst, S.; Kristensen, K.; Jørgensen, P. Local orbitals by minimizing powers of the orbital variance. *J. Chem. Phys.* **2011**, *134*, 194104.
- (95) Boughton, J. W.; Pulay, P. Comparison of the Boys and Pipek–Mezey Localizations in the Local Correlation Approach and Automatic Virtual Basis Selection. *J. Comput. Chem.* **1993**, *14*, 736.
- (96) Nagy, P. R.; Surján, P. R.; Szabados, Á. Mayer’s orthogonalization: relation to the Gram–Schmidt and Löwdin’s symmetrical scheme. *Theor. Chem. Acc.* **2012**, *131*, 1109.
- (97) Tóth, Z.; Nagy, P. R.; Jeszenszki, P.; Szabados, Á. Novel orthogonalization and biorthogonalization algorithms. *Theor. Chem. Acc.* **2015**, *134*, 100.
- (98) Boys, S. F.; Cook, G. B.; Reeves, C. M.; Shavitt, I. Automatic Fundamental Calculations of Molecular Structure. *Nature* **1956**, *178*, 1207.
- (99) Whitten, J. L. Coulombic potential energy integrals and approximations. *J. Chem. Phys.* **1973**, *58*, 4496.

- (100) Samu, G.; Kállay, M. Efficient evaluation of three-center Coulomb integrals. *J. Chem. Phys.* **2017**, *146*, 204101.
- (101) Kállay, M. A systematic way for the cost reduction of density fitting methods. *J. Chem. Phys.* **2014**, *141*, 244113.
- (102) Kállay, M.; Nagy, P. R.; Mester, D.; Rolik, Z.; Samu, G.; Csontos, J.; Csóka, J.; Szabó, P. B.; Gyevi-Nagy, L.; Hégyel, B.; Ladjánszki, I.; Szegedy, L.; Ladóczki, B.; Petrov, K.; Farkas, M.; Mezei, P. D.; Ganyecz, Á. The MRCC program system: Accurate quantum chemistry from water to proteins. *J. Chem. Phys.* **2020**, *152*, 074107.
- (103) Kállay, M.; Nagy, P. R.; Mester, D.; Gyevi-Nagy, L.; Csóka, J.; Szabó, P. B.; Rolik, Z.; Samu, G.; Csontos, J.; Hégyel, B.; Ganyecz, Á.; Ladjánszki, I.; Szegedy, L.; Ladóczki, B.; Petrov, K.; Farkas, M.; Mezei, P. D.; Horváth, R. A. MRCC, a quantum chemical program suite. See <https://www.mrcc.hu/> **Accessed August 1, 2023**,
- (104) Boys, S. F. Construction of Some Molecular Orbitals to Be Approximately Invariant for Changes from One Molecule to Another. *Rev. Mod. Phys.* **1960**, *32*, 296.
- (105) Weigend, F.; Ahlrichs, R. Balanced basis sets of split valence, triple zeta valence and quadruple zeta valence quality for H to Rn: Design and assessment of accuracy integrals over Gaussian functions. *Phys. Chem. Chem. Phys.* **2005**, *7*, 3297.
- (106) Dunning Jr., 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.
- (107) Dunning Jr., T. H.; Peterson, K. A.; Wilson, A. K. Gaussian basis sets for use in correlated molecular calculations. X. The atoms aluminum through argon revisited. *J. Chem. Phys.* **2001**, *114*, 9244.
- (108) Radoń, M. Benchmarking quantum chemistry methods for spin-state energetics of iron complexes against quantitative experimental data. *Phys. Chem. Chem. Phys.* **2019**, *21*, 4854–4870.

- (109) Weigend, F.; Köhn, A.; Hättig, C. Efficient use of the correlation consistent basis sets in resolution of the identity MP2 calculations. *J. Chem. Phys.* **2002**, *116*, 3175.
- (110) Karton, A.; Martin, J. M. L. Comment on: “Estimating the Hartree–Fock limit from finite basis set calculations”. *Theor. Chem. Acc.* **2006**, *115*, 330.
- (111) Helgaker, T.; Klopper, W.; Koch, H.; Noga, J. Basis-set convergence of correlated calculations on water. *J. Chem. Phys.* **1997**, *106*, 9639.
- (112) Liu, Y. Linear Scaling High-spin Open-shell Local Correlation Methods. Ph.D. thesis, Institut für Theoretische Chemie der Universität Stuttgart, 2011.
- (113) Goerigk, L.; Grimme, S. A general database for main group thermochemistry, kinetics, and noncovalent interactions – Assessment of common and reparameterized (meta-)GGA density functionals. *J. Chem. Theory Comput.* **2010**, *6*, 107.
- (114) Ghafarian Shirazi, R.; Neese, F.; Pantazis, D. A. Accurate Spin-State Energetics for Aryl Carbenes. *J. Chem. Theory Comput.* **2018**, *14*, 4733.
- (115) Guo, Y.; Sivalingam, K.; Valeev, E. F.; Neese, F. SparseMaps—A systematic infrastructure for reduced-scaling electronic structure methods. III. Linear-scaling multireference domain-based pair natural orbital N-electron valence perturbation theory. *J. Chem. Phys.* **2016**, *144*, 094111.
- (116) Wick, C. R.; Smith, D. M. Modeling the Reactions Catalyzed by Coenzyme B12 Dependent Enzymes: Accuracy and Cost-Quality Balance. *J. Phys. Chem. A* **2018**, *122*, 1747.
- (117) Kiss, D. J.; Ferenczy, G. G. A detailed mechanism of the oxidative half-reaction of D-amino acid oxidase: another route for flavin oxidation. *Org. Biomol. Chem.* **2019**, *17*, 7973.

TOC Graphic

