

# An efficient and near linear scaling pair natural orbital based local coupled cluster method

Christoph Riplinger, and Frank Neese

Citation: [The Journal of Chemical Physics](#) **138**, 034106 (2013); doi: 10.1063/1.4773581

View online: <http://dx.doi.org/10.1063/1.4773581>

View Table of Contents: <http://aip.scitation.org/toc/jcp/138/3>

Published by the [American Institute of Physics](#)

---

## Articles you may be interested in

[Natural triple excitations in local coupled cluster calculations with pair natural orbitals](#)

The Journal of Chemical Physics **139**, 134101 (2013); 10.1063/1.4821834

[Sparse maps—A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory](#)

The Journal of Chemical Physics **144**, 024109 (2016); 10.1063/1.4939030

[Efficient and accurate local approximations to coupled-electron pair approaches: An attempt to revive the pair natural orbital method](#)

The Journal of Chemical Physics **130**, 114108 (2009); 10.1063/1.3086717

[Efficient and accurate approximations to the local coupled cluster singles doubles method using a truncated pair natural orbital basis](#)

The Journal of Chemical Physics **131**, 064103 (2009); 10.1063/1.3173827

[A consistent and accurate ab initio parametrization of density functional dispersion correction \(DFT-D\) for the 94 elements H-Pu](#)

The Journal of Chemical Physics **132**, 154104 (2010); 10.1063/1.3382344

[Sparse maps—A systematic infrastructure for reduced-scaling electronic structure methods. I. An efficient and simple linear scaling local MP2 method that uses an intermediate basis of pair natural orbitals](#)

The Journal of Chemical Physics **143**, 034108 (2015); 10.1063/1.4926879

---



# SciLight

Sharp, quick summaries **illuminating**  
the latest physics research

Sign up for **FREE!**

AIP  
Publishing

# An efficient and near linear scaling pair natural orbital based local coupled cluster method

Christoph Riplinger and Frank Neese<sup>a)</sup>

Max Planck Institut für Chemische Energiekonversion, Stiftstr. 34-36, D-45470 Mülheim an der Ruhr, Germany

(Received 28 September 2012; accepted 13 December 2012; published online 17 January 2013)

In previous publications, it was shown that an efficient local coupled cluster method with single- and double excitations can be based on the concept of pair natural orbitals (PNOs) [F. Neese, A. Hansen, and D. G. Liakos, *J. Chem. Phys.* **131**, 064103 (2009)]. The resulting local pair natural orbital-coupled-cluster single double (LPNO-CCSD) method has since been proven to be highly reliable and efficient. For large molecules, the number of amplitudes to be determined is reduced by a factor of  $10^5$ – $10^6$  relative to a canonical CCSD calculation on the same system with the same basis set. In the original method, the PNOs were expanded in the set of canonical virtual orbitals and single excitations were not truncated. This led to a number of fifth order scaling steps that eventually rendered the method computationally expensive for large molecules (e.g., >100 atoms). In the present work, these limitations are overcome by a complete redesign of the LPNO-CCSD method. The new method is based on the combination of the concepts of PNOs and projected atomic orbitals (PAOs). Thus, each PNO is expanded in a set of PAOs that in turn belong to a given electron pair specific domain. In this way, it is possible to fully exploit locality while maintaining the extremely high compactness of the original LPNO-CCSD wavefunction. No terms are dropped from the CCSD equations and domains are chosen conservatively. The correlation energy loss due to the domains remains below <0.05%, which implies typically 15–20 but occasionally up to 30 atoms per domain on average. The new method has been given the acronym DLPNO-CCSD (“domain based LPNO-CCSD”). The method is nearly linear scaling with respect to system size. The original LPNO-CCSD method had three adjustable truncation thresholds that were chosen conservatively and do not need to be changed for actual applications. In the present treatment, no additional truncation parameters have been introduced. Any additional truncation is performed on the basis of the three original thresholds. There are no real-space cutoffs. Single excitations are truncated using singles-specific natural orbitals. Pairs are prescreened according to a multipole expansion of a pair correlation energy estimate based on local orbital specific virtual orbitals (LOSVs). Like its LPNO-CCSD predecessor, the method is completely of black box character and does not require any user adjustments. It is shown here that DLPNO-CCSD is as accurate as LPNO-CCSD while leading to computational savings exceeding one order of magnitude for larger systems. The largest calculations reported here featured >8800 basis functions and >450 atoms. In all larger test calculations done so far, the LPNO-CCSD step took less time than the preceding Hartree-Fock calculation, provided no approximations have been introduced in the latter. Thus, based on the present development reliable CCSD calculations on large molecules with unprecedented efficiency and accuracy are realized. © 2013 American Institute of Physics. [<http://dx.doi.org/10.1063/1.4773581>]

## I. INTRODUCTION

For a long time the accurate solution of the nonrelativistic  $N$ -electron molecular Schrödinger equation has been a dream of quantum chemistry. Impressive progress towards this goal has indeed been made in recent years. Based on the development of the powerful hierarchy of coupled cluster (CC)<sup>1</sup> methods in conjunction with efficient extrapolation<sup>2–6</sup> and explicit correlation (F12)<sup>7–9</sup> techniques, it has been possible to solve the nonrelativistic molecular Schrödinger equation to better than chemical accuracy (defined as 1 kcal/mol in relative energies). This accuracy can be routinely achieved for small molecules that are conceptually well described by a single Slater determinant. Fortunately, a very large fraction

of chemical problems falls under this category. However, at present the computational cost of these methods in rigorous form and their unfavorable scaling with system size prevents their application in mainstream computational chemistry. It is well known that the CC method including single- and double excitations (CCSD) method scales as  $O(N^6)$ , the approximate incorporation of triples to yield the gold-standard CCSD(T) method scales as  $O(N^7)$  while the full incorporation of triple excitations leads to  $O(N^8)$  scaling ( $N$  is a measure of system size). Although impressive progress has been made in parallel implementations of the methodology on modern supercomputers (see <http://www.nwchem-sw.org/index.php/Benchmarks> for parallel calculations employing > $10^5$  cores), it is obvious that one quickly hits a wall with respect to the required computational resources

<sup>a)</sup>Frank.Neese@cec.mpg.de.

(central processing unit (CPU) time, high-speed storage, and disk space). Even if Moore's law<sup>10</sup> can be satisfied for another two decades an increase of the required resources by a factor of 128 by only doubling the system size cannot be compensated for by ever more powerful hardware. In addition, such high-end hardware is only available to a small fraction of the computational chemistry community.

Thus, the conclusion appears to be inescapable that it is necessary to introduce some approximations in order to transform wavefunction based *ab initio* quantum chemistry into a major computational chemistry research tool for routine application to large molecules. In fact, the requirements of the efficiency that one needs to reach are very difficult to meet as the competing methodology based on density functional theory (DFT) comes roughly at the cost of a Hartree-Fock (HF) self-consistent field (SCF) calculation. Since some kind of SCF calculations appears to be a mandatory step before any CC calculation can take place, it becomes evident that it is necessary to organize the CC calculation such that it consumes either less or at least not much more time than the preceding SCF step. This is a highly ambitious goal since at the same time the approximations that are introduced must not spoil the inherent accuracy of the CC methodology.

It has been realized since the early days of wavefunction theory that the dynamic electron correlation that is covered by the CC methods is of short-range origin. Thus, local approximations suggest themselves and have therefore been developed in many flavors. Without attempting a comprehensive review of this large field, we mention the pioneering contributions of Pulay,<sup>11–14</sup> who suggested the concept of correlation domains and provided initial implementations of second through fourth order Möller-Plesset many body perturbation theory (MP2 to MP4) as well as configuration interaction with single- and double excitations (CISD). The method is based on projected atomic orbitals (PAOs) that span the virtual space but remain local.<sup>15</sup> These PAOs are necessarily linearly dependent and hence these linear dependencies must be removed. Since the atomic orbitals after projection remain largely local, they can be assigned to correlation domains. In general, each occupied orbital is assigned a correlation domain and the union of such domains is employed to describe higher excitations. Boughton and Pulay<sup>15</sup> have developed a method for automatic domain construction and Werner and Pflüger<sup>16</sup> have discussed the merits of various alternative domain construction schemes in detail. The PAO methodology has been perfectionized and efficiently implemented by Werner and co-workers<sup>16–37</sup> through extensive work during the past two decades. Schütz and Werner<sup>31</sup> have developed and implemented a local approximation to triple excitations thus leading to linear scaling LCCSD(T). The local equation of motion CCSD method (EOM-CCSD) has been developed for excited states by Werner and co-workers<sup>38</sup> while Schütz and co-workers<sup>39–43</sup> have obtained excellent results with a local version of the CC2 method. Schütz and co-workers<sup>44,45</sup> have also extended the PAO based methods to solids. An impressive recent development is the combination of the local CC methods with the emerging F12 techniques to deal with the problem of slow basis set convergence.<sup>17,18,20,21</sup> A second recent development is the orbital specific virtual (OSV) cor-

relation ansatz by Yang *et al.*<sup>46,47</sup> who have developed MP2 and CCSD versions of this method. This method is formulated in the language of tensor factorization theory and based on the idea to provide a set of correlating orbitals for each internal orbital.<sup>46,47</sup> Krause and Werner<sup>48</sup> have discussed the relative merits of the PAO, pair natural orbital (PNO), and OSV based methodologies as well as the combination of all of these techniques with F12 theory. At the present stage of development OSV-CCSD scales cubically with respect to system size. Crawford and Russ<sup>49,50</sup> explored the subject of coupled cluster linear response theory using PAO based coupled cluster methods and also investigated the smoothness of potential energy surfaces.

Alternative local correlation schemes are based on the fragment orbital method<sup>51–55</sup> and the incremental schemes.<sup>56–59</sup> These schemes consist of a series of smaller CC calculations that are combined to provide the total correlation energy of the system. Piecuch and co-workers<sup>60–62</sup> have developed and discussed a closely related scheme (the clusters-in-molecules method) and reported results for various flavors of completely renormalized coupled cluster methods.<sup>63</sup> Urban and co-workers<sup>64,65</sup> have reported results with an optimized virtual space method. Flocke and Bartlett<sup>66,67</sup> have investigated the use of natural bond orbitals in local coupled cluster calculations. Hoyau *et al.*<sup>68</sup> have developed local multireference CI (regional CI) schemes and have discussed several creative contraction schemes to render the wavefunction expansion more compact. Carter and co-workers<sup>69–73</sup> have developed local multireference CI schemes based on PAOs and have discussed several strategies for the efficient evaluation of the matrix elements. Several local schemes developed by Head-Gordon and co-workers<sup>74–82</sup> are also mentioned. Sherrill and Hohenstein<sup>83</sup> employed natural orbital concepts in the framework of symmetry adapted perturbation theory.

In our previous work, we were inspired by the pioneering contribution of Meyer,<sup>84–91</sup> Ahlrichs *et al.*,<sup>92,93</sup> as well as Taylor.<sup>93</sup> The idea in these contributions is based on the concept of PNOs (Meyer referred to them as pseudo-natural orbitals<sup>94</sup>), first discussed by Edmiston and Kraus.<sup>95,96</sup> It has been shown by Löwdin<sup>97</sup> early on that configuration interaction calculations on two-electron systems converge most rapidly when they are based on a set of NOs ordered according to decreasing occupation number. The NOs diagonalize the one-particle density. While in principle this requires the solution of the CI problem the approximation can be performed and extensive tests have shown that in practice approximate NOs perform almost as well as exact NOs. Ahlrichs and Driessler<sup>98</sup> have provided a self-consistent construction scheme for approximate NOs. The idea of the PNO based correlation methods consists of providing a separate set of NOs for each electron pair. This obviously leads to correlating orbitals that are orthonormal for a given electron pair but that are non-orthogonal for different pairs. Second, the total number of PNOs is much larger than the number of canonical virtual orbitals. Meyer<sup>91,94</sup> has shown that these limited non-orthogonalities only lead to minor complications in actual calculations. Based on the concept of PNOs, Meyer<sup>91,94</sup> has developed the coupled-electron pair approximation (CEPA) and has obtained impressively accurate results already in the

early 1970s that in hindsight were greatly ahead of their time. Taylor<sup>99</sup> as well as Fink and Staemmler<sup>100</sup> have generalized the CEPA method for multideterminantal reference wavefunctions. Kollmar<sup>101</sup> has developed an orbital invariant and stationary CEPA like functional that provides results that are essentially identical to CCSD.

In our previous work,<sup>6,102–106</sup> we have been intrigued by the attractive properties of the PNO based correlation calculations and have realized that it is possible to combine the PNO concept with local correlation ideas to arrive at accurate and efficient local correlation treatments. In fact, the PNOs constitute a set of correlating orbitals that is as local or non-local as the actual bonding situation of the electron pair in question demands and, provided that the PNOs for each pair are conservatively truncated, is the most compact space that recovers the correlation contribution of the electron pair. Based on these properties, it also behaves extremely well with respect to basis set extension as the number of correlating PNOs above threshold for each electron pair is approaching a constant as the basis set is approaching saturation.

Using these ideas we have developed the local pair natural orbital coupled electron pair (LPNO-CEPA) method.<sup>105</sup> By using essentially a single truncation parameter it was shown that more than 99.9% of the canonical correlation energy can be recovered. Extensive test calculations have established that the LPNO-CEPA and canonical CEPA methods can essentially be used interchangeably and that the LPNO-based methodology is robust and efficient. Subsequently, the generalization to the LPNO-CCSD method was achieved and it was shown that it behaves as well as the related LPNO-CEPA method.<sup>106</sup> The method has also been generalized to the open-shell case,<sup>104</sup> the parameterized CCSD method (pCCSD<sup>107</sup>) leading to LPNO-pCCSD<sup>102</sup> and it has also been efficiently parallelized.<sup>103</sup> The otherwise occasionally limiting construction of the singles Fock type matrix has been accelerated by the split-RI-J<sup>108</sup> and chain of spheres exchange (COSX)<sup>109,110</sup> approximations.<sup>111</sup> It has been used in correlation energy extrapolations and the study of weak intermolecular interactions.<sup>6,103,112</sup> A number of chemical applications have been reported by our group<sup>112–114</sup> as well as others.<sup>115–117</sup> The performance of OSV-CCSD was compared to LPNO-CCSD in the excellent paper by Yang *et al.*<sup>46</sup> and similar performance was found for the penicillin molecule with 42 atoms and about 1000 basis functions. This molecule will also be reinvestigated in the present work.

Tew, Hättig, and Helmich<sup>118,119</sup> have taken up the idea of the PNO correlation methods and have developed explicitly correlated MP2 and MP3 versions based on the same concept as well as a PNO based doubles correction (D<sup>120</sup>) to the configuration interaction with singles excited state method. Furthermore, Werner and Krause<sup>48</sup> have verified the original results and extended the LPNO-CCSD method to its F12 variant using a simulated PNO code. Rolik and Kallay<sup>121</sup> used concepts related to the cluster-in-molecules and natural orbital approaches to arrive at a local general-order coupled cluster implementation.

Thus, all available evidence indicates that the LPNO based correlation methods are efficient and accurate and hold an enormous potential for chemical applications. The original

LPNO-CCSD method performed excellently for molecules with up to about 50-70 atoms for which it consumed only 2–4 times the time required for the preceding HF calculation. However, for large molecules, the method became demanding and some bottlenecks became apparent. These bottlenecks were related to the fact that the PNOs were expanded in terms of the canonical virtual orbitals. Furthermore, several integral classes with two external labels were still calculated in the canonical basis and transformed to the PNO basis. Last but not least the single excitations were not truncated. All this led to some fifth order scaling steps that ultimately limited the applicability of the method to molecules with more than, say, 100 or more atoms.

In the present work, an extensive redesign of the LPNO-CCSD (as well as LPNO-QCISD and LPNO-CEPA) methods is reported and evaluated through test calculations. The underlying idea is to expand the PNOs in a set of local orbitals that span the virtual space. In this way, one can take maximal advantage of locality while maintaining the compactness of the LPNO wavefunction. While several choices of local virtual orbitals would be feasible, we have chosen to employ PAOs for this purpose. Since we ultimately expand the wavefunction in PNOs, it is not critical that the correlation domains to which the PAOs are assigned remain as small as possible. In fact, we are choosing very conservative, large domains that recover more than 99.9% of the correlation energy. By extensively re-designing the integral transformation as well as each and every term in the CCSD residual equations, all bottlenecks of the previous implementation are avoided and a near linear scaling method is obtained that is dubbed DLPNO-CCSD and that retains all attractive features of the original LPNO-CCSD method.

## II. THEORY

### A. The CCSD equations

The theory of the CCSD method is, of course, well known and we only re-iterate the working closed-shell equations here as a basis for the discussion of the implementation. The reference wavefunction in CCSD theory is based on a single Slater determinant  $|\Psi_0\rangle$  (not necessarily the HF determinant). The molecular orbitals  $\{\psi\}$  that constitute this determinant are divided into internal (occupied) orbitals denoted with labels  $i, j, k, l$  as well as external (unoccupied) orbitals denoted with labels  $a, b, c, d$  throughout while  $p, q, r, s$  label orbitals from any subspace and  $\mu, \nu, \kappa, \tau$  refer to basis functions. Below, we will refer by  $\tilde{\mu}', \tilde{\nu}', \tilde{\kappa}', \tilde{\tau}'$  to the redundant PAO basis and use  $\tilde{\mu}, \tilde{\nu}, \tilde{\kappa}, \tilde{\tau}$  for the non-redundant PAOs. Likewise,  $\tilde{a}_{ij}, \tilde{b}_{ij}, \tilde{c}_{ij}, \tilde{d}_{ij}$  will be used to denote the expansion of the PNOs of pair  $ij$  in terms of non-redundant PAOs.

The CCSD wavefunction for a closed shell system is obtained by acting with the exponentially parameterized coupled cluster excitation operator onto the reference wavefunction

$$|\Psi_{CCSD}\rangle = \exp(T_1 + T_2)|\Psi_0\rangle \quad (1)$$

with  $T_1 = \sum_{ia} t_a^i E_i^a$  and  $T_2 = \frac{1}{2} \sum_{ijab} t_{ab}^{ij} E_i^a E_j^b$  contain the singles and doubles wavefunction amplitudes that have to be determined ( $E_i^a = a_{a\beta}^+ a_{i\beta}$  is the spin traced excitations operator and  $a_{i\sigma}$  and  $a_{a\sigma}^+$  are standard Fermion



destruction and creation operators for spin  $\sigma = \alpha, \beta$ ). The coupled cluster equations are solved through projection to yield the total energy

$$E = \langle \Psi_0 | e^{-T} H e^T | \Psi_0 \rangle \quad (2)$$

$$= E_0 + E_C = E_0 + 2 \sum_{ia} F_{ia} t_a^i + \sum_{ijab} [2(i a | j b) - (i b | j a)] (t_{ab}^{ij} + t_a^i t_b^j). \quad (3)$$

Here,  $T = T_1 + T_2$ ,  $E_0 = \langle \Psi_0 | H | \Psi_0 \rangle$ ,  $F_{pq} = h_{pq} + \sum_k 2(k k | p q) - (k p | k q)$  is the closed-shell Fock operator with  $h_{pq}$  being its one-electron part and  $(p q | r s)$  are two-electron integrals in (11|22) notation. The residual equations to be satisfied by the singles and doubles amplitudes are

$$R_a^i = \langle \Phi_i^a | e^{-T} H e^T | \Psi_0 \rangle = 0, \quad (4)$$

$$R_{ab}^{ij} = \langle \Phi_{ij}^{ab} | e^{-T} H e^T | \Psi_0 \rangle = 0. \quad (5)$$

$|\Phi_i^a\rangle = E_i^a |\Psi_0\rangle$  and  $|\Phi_{ij}^{ab}\rangle = E_j^b E_i^a |\Psi_0\rangle$  are singly and doubly excited configuration state functions consisting of two and four determinants, respectively.

Detailed equations for the CCSD residual have been discussed in many papers. In our implementation, we have been inspired by the CISD formulation of Meyer and Pulay<sup>122</sup> that is consistent with the CCSD formulation provided by Scuseria.<sup>123</sup> We furthermore found it useful to think of the CCSD intermediates in terms of dressed molecular integrals such that the CCSD residual takes the appearance of a CISD eigenvalue problem. The following quantities are involved:

$$\begin{aligned} R_a^i &= F_{ia} + \sum_b \tilde{F}_{ba} t_b^i - \sum_j \tilde{F}_{ij} t_a^j + G(\mathbf{t}_1)_{ia} \\ &+ \sum_{jb} (2\mathbf{t}^{ji} - \mathbf{t}^{ii+})_{ba} \tilde{F}_{bj} - \sum_{kjb} (2(ik|jb) - (ij|kb)) \tau_{ab}^{kj} \\ &+ \sum_{jbc} ((2(ib|ac) - (ic|ab)) \tau_{bc}^{ij} + (2(jb|ac) - (jc|ab)) \tau_{cb}^{ij}) \\ &+ \sum_{jb} (\tilde{F}_{jb} - 2F_{jb}) t_b^i t_a^j, \end{aligned} \quad (6)$$

$$\begin{aligned} R_{ab}^{ij} &= K_{ab}^{ij} + K(\tau^{ij})_{ab} + (\tilde{\mathbf{F}}^+ \tau^{ij} + \tau^{ij} \tilde{\mathbf{F}})_{ab} \\ &- \sum_k (\tilde{F}_{jk} \mathbf{t}^{ki+} + \tilde{F}_{ik} \mathbf{t}^{kj})_{ab} + \sum_{kl} (i\tilde{k}|jl) \tau_{ab}^{kl} \\ &+ \sum_k \left( (2\mathbf{t}^{ki+} - \mathbf{t}^{ki}) \left( \tilde{\mathbf{K}}^{jk+} - \frac{1}{2} \tilde{\mathbf{J}}^{jk} \right) - \frac{1}{2} \mathbf{t}^{ki} \tilde{\mathbf{J}}^{jk} - \tilde{\mathbf{J}}^{jk+} \mathbf{t}^{ki+} \right)_{ab} \\ &+ \sum_k \left( \left( \tilde{\mathbf{K}}^{ik} - \frac{1}{2} \tilde{\mathbf{J}}^{ik+} \right) (2\mathbf{t}^{kj} - \mathbf{t}^{kj+}) - \frac{1}{2} \tilde{\mathbf{J}}^{ik+} \mathbf{t}^{kj+} - \mathbf{t}^{kj} \tilde{\mathbf{J}}^{ik} \right)_{ab} \\ &- \sum_k ((jk|ia) t_b^k + (ik|jb) t_a^k) + \sum_c ((ia|cb) t_c^i + (ib|ac) t_c^j) \\ &- \sum_k ((\mathbf{K}^{ik} \mathbf{t}^j)_a t_b^k + (\mathbf{K}^{jk} \mathbf{t}^i)_b t_a^k + (\mathbf{J}^{ik} \mathbf{t}^j)_b t_a^k + (\mathbf{J}^{jk} \mathbf{t}^i)_a t_b^k), \end{aligned} \quad (7)$$

as usual,  $K_{ab}^{ij} = (ia|jb)$  and  $J_{ab}^{ij} = (ij|ab)$ .<sup>124</sup> The unique amplitudes obey the index restrictions  $i \leq j$  and all  $a, b$  (for  $i = j$  only  $a \leq b$  would be required but it is more convenient to keep a symmetric coefficient matrix  $t_{ab}^{ij}$  in this case<sup>122</sup>). The “dressed” quantities are<sup>106</sup>

$$\tau_{ab}^{ij} = t_{ab}^{ij} + t_a^i t_b^j, \quad (8)$$

$$\tilde{F}_{ij} = F_{ij} + \sum_k \langle (2\mathbf{K}^{jk} - \mathbf{K}^{jk+}) \mathbf{t}^{ik} \rangle, \quad (9)$$

$$\tilde{F}_{ab} = F_{ab} - \sum_{kl} (\mathbf{t}^{kl} (2\mathbf{K}^{lk} - \mathbf{K}^{lk+}))_{ab}, \quad (10)$$

$$\tilde{F}_{ai} = F_{ai} + \sum_k (2\mathbf{K}^{ik} - \mathbf{K}^{ik+}) \mathbf{t}_a^k, \quad (11)$$

$$\tilde{F}_{ij} = \tilde{F}_{ij} + G(\mathbf{t}_1)_{ij} + \sum_a F_{ia} t_a^j, \quad (12)$$

$$\tilde{F}_{ab} = \tilde{F}_{ab} + G(\mathbf{t}_1)_{ab} - \sum_i F_{ib} t_a^i, \quad (13)$$

$$G(\mathbf{t}_1)_{pq} = \sum_{jb} t_b^j (2(pq|jb) - (pj|qb)), \quad (14)$$

$$K(\tau^{ij})_{ab} = \sum_{cd} \left( (ac|bd) - \sum_k (kd|ac) t_b^k + (kc|bd) t_a^k \right) \tau_{cd}^{ij}, \quad (15)$$

$$(i\tilde{k}|jl) = (ik|jl) + \langle \mathbf{K}^{ij} \mathbf{t}^{kl+} \rangle + \sum_a (ki|la) t_a^j + (lj|ka) t_a^i, \quad (16)$$

$$\begin{aligned} \tilde{K}_{ab}^{ij} &= K_{ab}^{ij} + \sum_k \left( \mathbf{t}^{ik} \left( \mathbf{K}^{kj} - \frac{1}{2} \mathbf{K}^{kj+} \right) - \frac{1}{2} \mathbf{t}^{ki} \mathbf{K}^{kj} \right)_{ab} \\ &+ \sum_c (jc|ab) t_c^i - \sum_k (ik|jb) t_a^k, \end{aligned} \quad (17)$$

$$\tilde{J}_{ab}^{ij} = J_{ab}^{ij} - \frac{1}{2} \sum_k (\mathbf{K}^{kj} \mathbf{t}^{ki+})_{ab} + \sum_c (jb|ac) t_c^i - \sum_k (ij|ka) t_b^k. \quad (18)$$

As discussed before, the translation of the canonical equations into the framework of PNOs is straightforward.<sup>105,106</sup> The calculation of the individual terms will be discussed below. Our implementation includes the CEPA, variational CEPA,<sup>101</sup> coupled pair functional,<sup>125</sup> QCISD, CCSD, and pCCSD methods. For convenience, we will only describe the CCSD method, which is the most general of these methods. All other methods are obtained from CCSD through simple manipulation of neglect of certain terms as documented in the literature.<sup>106</sup>

## B. Initial guess and PNO construction

### 1. Domain construction

The DLPNO-CCSD method starts with the localization of the internal orbital space to yield localized internal orbitals  $|i_L\rangle$ . We usually employ the Foster-Boys criterion<sup>126</sup> but Pipek-Mezey<sup>127</sup> localization can also be used. Second, using the projector onto the localized occupied space

$\hat{P} = \sum_{i_L} |i_L\rangle\langle i_L|$  a redundant set of projected atomic orbitals that span the external space is obtained by defining  $|\tilde{\mu}'\rangle = (1 - \hat{P})|\mu\rangle$ . For convenience we will drop the subscript “L” from now on, remembering that the internal orbitals are always localized. In the next step of the procedure, each occupied orbital  $|i\rangle$  is assigned a correlation domain  $\{A\}_i$  of atoms  $A$  on which  $|i\rangle$  has a significant amplitude. We use the simplest possible criterion based on the gross Mulliken population  $p_i^A$  of each localized orbital  $|i\rangle$  on each atom  $A$ . The cutoff  $T_{CutMKN}$  has previously been introduced in order to create local fits to the PNOs.<sup>105</sup> The same cutoff is used here for the domain construction. Since it has been shown previously that with such a tight threshold the domains are large enough to accurately fit the PNOs (which are hence local), the same must be true for the PAOs into which the PNOs will be expanded. For electron pairs, the union of the individual orbital domains span the initial pair domain  $\{A\}_{ij} = \{A\}_i \cup \{A\}_j$ .

Owing to the projection, the PAOs do not remain strictly localized on the “parent” atoms. Thus, a given PAO on atom  $A$  will have contributions from basis functions on the neighboring atoms  $B$  according to  $|\tilde{\mu}'^A\rangle = \sum_{\mu^B} L_{\mu^B \tilde{\mu}'^A} |\mu^B\rangle$  where  $L$  is the expansion matrix for the PAOs in terms of basis functions. We found it is necessary to take this delocalization of the PAOs into account in the definition of the final pair domain. After extensive test calculations, we find that for a given atom  $A$  all atoms  $B$  should be included in this pair domain for which  $p^{AB} = \sum_{\tilde{\mu}'^B \mu^A} |L_{\mu^A \tilde{\mu}'^B}| > 0.1$ .

There is a second, extended domain for each pair  $ij$ , that consists of the union of the domains of orbitals  $i$  and  $j$  and that furthermore includes all atoms of the domains of all orbitals  $k$ , that fulfill the condition that none of the three pairs  $ij$ ,  $ik$ , and  $kj$  has been dropped from the calculation. This extended domain space is necessary in order to create a sufficiently large buffer space for the accurate calculation of the pair-pair interaction terms as described below.

Thus, in summary, the preparatory phase of a DLPNO-CCSD calculation consists of the localization of the occupied orbitals, the construction of orbital domains, and the construction of two sets of pair domains for each electron pair  $ij$ .

## 2. Electron pair prescreening

The second phase of the DLPNO-CCSD calculation consists of the initial guess for the cluster amplitudes and the PNO generation. This constitutes the main difference to the LPNO-CCSD method and hence it will be described in some detail.

In the LPNO-CCSD method, a semi-canonical RI-MP2 (SC-RI-MP2) calculation is performed in order to obtain pair densities from which the PNOs are calculated.<sup>105</sup> The term “semi-canonical” indicates in this context that the virtual part of the Fock matrix is diagonal while the internal part  $F_{ij}$  is not. Off-diagonal terms  $F_{ij}$  are ignored. In our experience, the semi-canonical treatment is a good approximation to the full MP2 energy. Since the complete MP2 energy is not used anywhere in the development, we had no strong incentive to proceed beyond this approximation. Throughout the calculation, the resolution of the identity (RI, also called density fitting)

approximation<sup>128</sup> is extensively used. The local MP2 method with use of density fitting is well known<sup>27</sup> and hence its semi-local simplification is straightforward.

The calculation commences by an integral transformation that generates the three index integrals  $(i\tilde{\mu}'|K)$ ,  $(ia|K)$ , and  $(ij|K)$ , where  $K$  refers to a basis function in the auxiliary basis set that is used throughout the development. This step is very efficient and we saw no reason to perform it in a linear scaling fashion.

An outer loop over the electron pairs  $ij$  is performed. In order to obtain linear scaling and keeping with the design philosophy to not introduce any real space cutoffs, we have developed a prescreening algorithm that allows us to estimate the SC-MP2 pair correlation energy of weakly interacting electron pairs and hence bypass the more elaborate local SC-RI-MP2 estimate that would scale quadratically and have a reasonably large prefactor because our domains are large. The semi-canonical pair correlation energy of a given off-diagonal pair can be written as

$$\varepsilon_{ij}^{SC-MP2} = - \sum_{ab} \frac{4(ia|jb)(ia|jb) - 2(ia|jb)(ib|ja)}{\varepsilon_a + \varepsilon_b - F_{ii} - F_{jj}}, \quad (19)$$

where the  $\varepsilon$ 's are orbital energies and  $F_{ii}$  is a Fock matrix diagonal element. The first term is considered to be of Coulomb nature and the second of exchange nature. The latter is expected to fall off much more steeply than the former and hence can be ignored if attention is focused on the long range. One is then faced with the subject of estimating the integral  $(ia|jb)$  without calculating it exactly (which would introduce higher order scaling terms which we wish to avoid in the prescreening procedure).

The approximation that we pursue can be explained in two steps. In the first step, the RI approximation is introduced which results in the familiar expansion  $(ia|jb) \approx \sum_{\tilde{K}} (ia|\tilde{K})(jb|\tilde{K})$  where  $\tilde{K}$  refers to an auxiliary basis function that has been orthogonalized in the Coulomb metric using the  $V^{-1/2}$  matrix with  $V_{KL} = (K|L)$  being a two-index electron-electron repulsion integral. Instead of evaluating this expression in the canonical basis, the approximation is applied to a specific set of virtual orbitals in which there is a special set of virtual orbitals  $\{\tilde{a}_i\}$  for each occupied orbital. The set of correlating orbitals for each occupied orbital does asymptotically not grow with molecular size. These orbitals happen to be identical to the diagonal PNOs described below. Of course, having a truncated set of correlating orbitals for each occupied orbital is the essence of the OSV method.<sup>47</sup> However, we use this idea in a different context here and at the same time introduce local OSVs (LOSVs, *vide infra*). Truncation is performed with respect to a fixed “occupation number” cutoff  $T_{CutOSV}$  that will be justified below. As a result one only would have to calculate a much smaller set of integrals  $(i\tilde{a}_i|j\tilde{b}_j)$  and arrive at the following pair energy estimate:

$$\varepsilon_{ij}^{SC-OSV} \approx -4 \sum_{\tilde{a}_i \tilde{b}_j} \frac{(i\tilde{a}_i|j\tilde{b}_j)^2}{\varepsilon_{\tilde{a}_i} + \varepsilon_{\tilde{b}_j} - F_{ii} - F_{jj}}. \quad (20)$$

In order to obtain the OSV orbital energies  $\varepsilon_{\tilde{a}_i}$ , it is necessary to recanonicalize the OSVs after truncation has

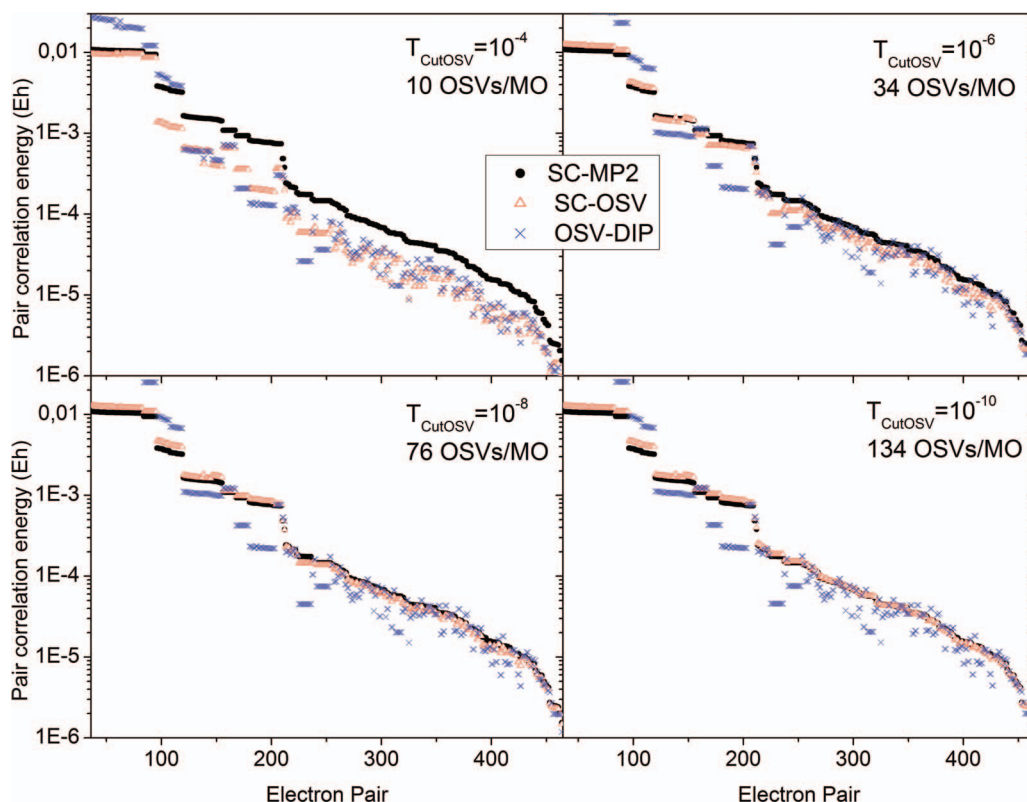


FIG. 1. Comparison of pair correlation energies for the stacked benzene dimer for four values of the OSV truncation threshold. Pairs are ordered to decreasing pair correlation energy (absolute values are plotted). Black solid dots = semi-canonical MP2 pair energy. Red open dots = OSV approximation to the semi-canonical MP2 pair correlation energy. Blue crosses = OSV based dipole-dipole approximation to the semi-canonical pair correlation energy.

been performed, which is a trivial procedure. While such a calculation is entirely feasible, even for large molecules, the generation of the OSVs scales as  $O(N^4)$  and the prescreening step would scale as  $O(N^3)$ , which seems inconsistent with the design goals of DLPNO method. In order to arrive at linear scaling for the OSV construction, we expand the OSVs not in terms of virtual orbitals but instead in terms of PAOs restricted to the domain of the diagonal  $ii$  pair (just like we will do for the PNOs below) to yield the LOSVs

$$|\tilde{a}_i\rangle = \sum_{\tilde{\mu}' \in \{i\}} d_{\tilde{\mu}'a}^i |\tilde{\mu}'\rangle. \quad (21)$$

Since the OSVs are identical to the diagonal PNOs<sup>46</sup> they are constructed exactly as outlined below. The computational effort of this step scales linearly.

Higher than quadratically scaling effort in the pair prescreening can also be avoided. Owing to the fact that during prescreening we are only interested in the weakly interacting tail of the correlation energy, it seems natural to assume that the charge distributions  $i\tilde{a}_i$  and  $j\tilde{b}_j$  are non-overlapping. Under this assumption one can then perform a bipolar expansion<sup>129</sup> truncated at the lowest order to arrive at the second step of our approximation scheme. It consists of a dipole-dipole estimate for the  $(i\tilde{a}_i|j\tilde{b}_j)$  integrals under which the OSV-dipole pair-energy (OSV-DIP) estimate becomes

$$\varepsilon_{ij}^{OSV-DIP} = -\frac{8}{R_{ij}^3} \sum_{ab} \frac{(\langle i|\mathbf{r}|\tilde{a}_i\rangle\langle j|\mathbf{r}|\tilde{b}_j\rangle)^2}{(\varepsilon_{\tilde{a}_i} + \varepsilon_{\tilde{b}_j} - F_{ii} - F_{jj})}, \quad (22)$$

where  $R_{ij}$  is the distance between the centers of the localized occupied orbitals  $i$  and  $j$ .

In order to demonstrate how the two approximations described work in a real life situation, we have chosen the stacked benzene-dimer from the “S22” set at the original geometry and with the aug-cc-pVDZ basis set.<sup>130</sup> In this structure there is, of course, an important dispersion interaction that one must describe correctly in a local correlation scheme. In Figure 1, the three sets of pair-correlation energies (semi-canonical RI-MP2, SC-OSV, and OSV-DIP) of this system are plotted for four different choices of  $T_{CutOSV}$ . For a very aggressive cutoff of  $10^{-4}$ , the agreement of either approximation with the correct result is rather modest although the correct asymptotic behavior is obtained. Already at a cutoff of  $10^{-6}$   $\varepsilon_{ij}^{SC-OSV}$  becomes a good approximation to  $\varepsilon_{ij}^{SC-MP2}$  which also demonstrates that the exchange part of the pair correlation energy is of very short range. As expected, the multipole estimate can be fairly far off the correct result for large pair correlation energies but it becomes accurate for small pair correlation energies of  $10^{-5}$  Eh and smaller, which constitute the dispersion tail of the correlation energy. At this point the calculation features an average of 34 LOSVs per occupied orbital. For even tighter cutoffs of  $10^{-8}$  and  $10^{-10}$ , the  $\varepsilon_{ij}^{SC-OSV}$  estimate becomes more and more accurate. The achievable accuracy is only limited by the neglect of the exchange term and deviations are only obvious for the large (intramolecular nearest neighbor orbital pairs) pair-correlation energies. The accuracy of the multipole estimate is limited by the low order expansion

and the assumption of non-overlapping charge distributions and cannot be further improved upon tightening  $T_{CutOSV}$ . Hence,  $10^{-6}$  is the appropriate value that we choose for our calculations.

On the basis of these results and additional test calculations, we decided to base our prescreening algorithm entirely on  $\varepsilon_{ij}^{OSV-DIP}$ . The quantities involved can be extremely rapidly calculated outside the pair loop. The actual pair prescreening requires quadratic effort with respect to molecular size but is so efficient that even for the largest calculations reported below the total time required for the quadratic part of the electron pair prescreening remained smaller than 3 s. Remembering that pairs with a pair correlation energy larger than  $T_{CutPairs}$  (default value  $10^{-4}$  Eh) are treated on the coupled cluster level, the following algorithm suggests itself: For a given pair,  $\varepsilon_{ij}^{OSV-DIP}$  is always calculated. If its absolute value is  $< \min(10^{-6}, 0.01T_{CutPairs})$ , the pair is dropped but the sum of all  $\varepsilon_{ij}^{OSV-DIP}$  of the dropped pairs is determined and serves as a correction for the truncation error. For all other pairs  $\varepsilon_{ij}^{SC-MP2}$  is calculated. If its value is  $\geq T_{CutPairs}$  the pair is kept, its PNOs are determined and it will be part of the coupled cluster iterations. Otherwise,  $\varepsilon_{ij}^{SC-MP2}$  is summed into a correction for the non-negligible pairs that will be added to the final energy as well. Furthermore, for the pairs that are kept in the CCSD iterations the difference between  $\varepsilon_{ij}^{SC-MP2}$  and its PNO equivalent  $\varepsilon_{ij}^{SC-PNO-MP2}$  serves as an estimate for the PNO truncation error. To give the reader some feeling for the size of these corrections, we quote the number for the benzene dimer calculation described above. For this system, the  $\varepsilon_{ij}^{OSV-DIP}$  error estimate is zero because no pair energy estimate falls below  $10^{-6}$  Eh. The  $\varepsilon_{ij}^{SC-MP2}$  intermediate pair energies sum to 6.99 mEh while the PNO truncation error estimate sums to 2.4 mEh relative to the sum of the  $\varepsilon_{ij}^{SC-PNO-MP2}$  of the kept pairs which in turn amounts to 1556.0 mEh. Thus, the total correction amounts to only 0.6% of the total MP2 correlation energy. It is, however, necessary to keep these corrections in order to achieve the high accuracy that we have already observed in the LPNO-CCSD method.

### 3. PNO construction

As described above, for the surviving pairs, the exchange operators  $K_{\tilde{\mu}'\tilde{\nu}'}^{ij}$  with  $\tilde{\mu}'\tilde{\nu}'$  confined to the pair domain  $ij$  are constructed. This involves an orthogonalization step using a pair specific, local Coulomb metric  $V_{KL}^{-1/2}(ij)$ .<sup>128</sup> Furthermore, the set of non-redundant PAOs for this pair domain is constructed to yield  $K_{\tilde{\mu}\tilde{\nu}}^{ij}$  with indices  $\tilde{\mu}$  and  $\tilde{\nu}$  confined to the delocalized pair domain described above. The extended pair domain defined above is only used at a later stage, as will be described below. In the process, the non-redundant, local PAOs are also recanonicalized by requiring that within the domain  $ij$  they diagonalize the Fock operator  $F_{\tilde{\mu}\tilde{\nu}}$ . As a result, the semi-canonical LMP2 amplitudes  $t_{\tilde{\mu}\tilde{\nu}}^{ij} = -K_{\tilde{\mu}\tilde{\nu}}^{ij}/(\varepsilon_{\tilde{\mu}} + \varepsilon_{\tilde{\nu}} - F_{ii} - F_{jj})$  with  $\varepsilon_{\tilde{\mu}} = F_{\tilde{\mu}\tilde{\mu}}$  and the SC-LMP2 pair correlation energies  $\varepsilon_{ij}^{SC-LMP2}$  are obtained for each pair  $ij$ . Since the number of electron pairs for which these steps are to be performed has been reduced to asymp-

totically linear through the pair prescreening, the procedure is linear scaling with respect to system size.

The calculation then proceeds, as in the LPNO-CCSD case, by constructing pair-specific densities

$$\mathbf{D}^{ij} = \tilde{\mathbf{t}}^{ij} \mathbf{t}^{ij+} + \tilde{\mathbf{t}}^{ij+} \mathbf{t}^{ij} \quad (23)$$

with  $\tilde{t}_{\tilde{\mu}\tilde{\nu}}^{ij} = \frac{1}{1+\delta_{ij}}(4t_{\tilde{\mu}\tilde{\nu}}^{ij} - 2t_{\tilde{\nu}\tilde{\mu}}^{ij})$ . The earlier LPNO-CCSD version used a normalization chosen to mimic the independent electron pair approximation (IEPA) amplitudes,<sup>105</sup> which amounts to a prefactor  $(1 + \delta_{ij})/(1 + \langle \tilde{\mathbf{t}}^{ij+} \mathbf{t}^{ij} \rangle)$ . Yang *et al.*<sup>46</sup> have commented in a footnote that this leads to slower convergence of the correlation energy in the PNO expansion. We have independently investigated this issue in Figure 2. It becomes apparent that the effect described by Yang *et al.*<sup>46</sup> is present but is insignificantly small (on average 1.4% per PNO included). Nevertheless, we have decided to change the normalization to keep future comparisons in the literature more transparent.

In general for two matrices A and B we define  $\langle \mathbf{AB} \rangle = \sum_{pq} A_{pq} B_{qp}$ . In full analogy to the previously described LPNO-CCSD method, the diagonalization of the pair density  $\mathbf{D}^{ij}$  yields the PNOs and their occupation numbers

$$\mathbf{D}^{ij} \mathbf{d}^{ij} = n^{ij} \mathbf{d}^{ij}. \quad (24)$$

Thus, each PNO for a given pair  $ij$  is expanded in terms of the non-redundant PAOs of the domain  $ij$  as

$$|\tilde{a}_{ij}\rangle = \sum_{\tilde{\mu} \in ij} d_{\tilde{\mu}}^{ij} |\tilde{\mu}\rangle. \quad (25)$$

Only the PNOs with  $n_a^{ij} > T_{CutPNO}$  are kept and are then recanonicalized with respect to the Fock operator, as in the LPNO-CCSD case.<sup>105,106</sup> The next step of the procedure consists of transforming the exchange operator  $K_{\tilde{\mu}\tilde{\nu}}^{ij}$  into the PNO basis to yield  $K_{\tilde{a}\tilde{b}}^{ij}$  from which the final PNO basis guess amplitudes  $t_{\tilde{a}\tilde{b}}^{ij}$  are obtained. From these PNO amplitudes, the semi-canonical PNO-MP2 pair correlation energy ( $\varepsilon_{ij}^{SC-PNO-MP2}$ ) is re-calculated and used as described above. The more elaborate PNO construction scheme used in LPNO-CCSD based on ‘‘Meyer style PNOs’’ is not taken over to the DLPNO-CCSD method since the results of MP2-PNOs and Meyer-PNOs are very similar.

### 4. Single excitation treatment

The single excitations were not truncated in the LPNO-CCSD method.<sup>105,106</sup> While this choice was sound, it also introduced some higher order scaling steps that are avoided in DLPNO-CCSD. Singles are hence expanded in a set of PNOs that coincide with the PNOs of the diagonal pairs. However, based on extensive test calculations we found that significant accuracy can be gained by truncating them even more conservatively than the doubles excitations. Following these test calculations, we find that  $0.03T_{CutPNO}$  is a choice that leads to virtually no loss of accuracy.



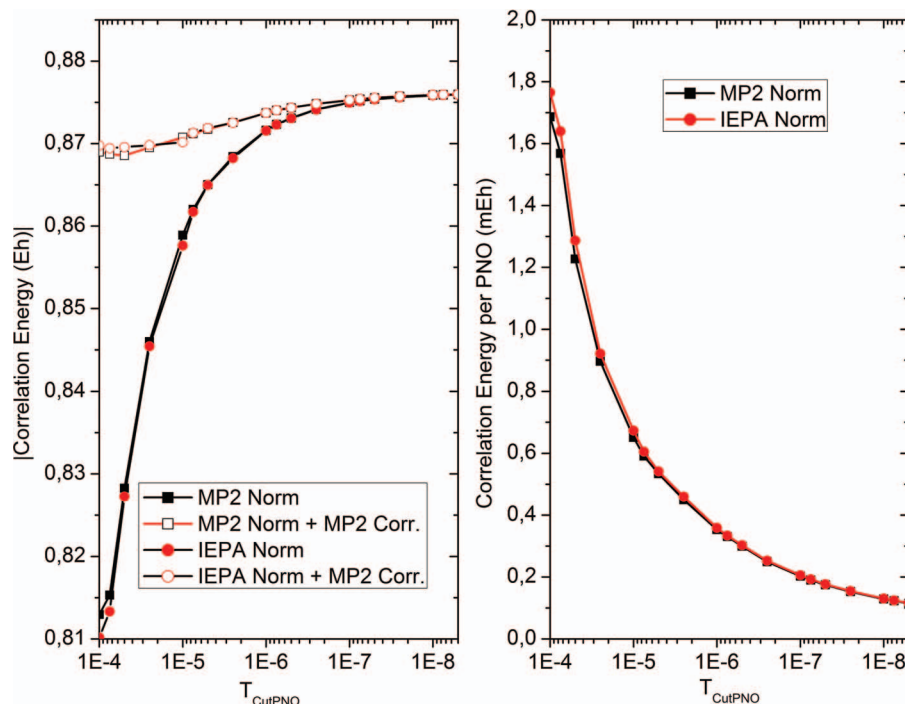


FIG. 2. Comparison of the convergence rate for PNOs normalized with the “IEPA norm” or “MP2 norm” as defined in the text. Calculations were carried out with the cc-pVDZ basis set on the cyclohexane molecule.

### C. Integral transformations and pair-pair interaction treatment

As pointed out by Werner,<sup>48</sup> among the PAO based, OSV and PNO based local CC methods, the PNO method requires the most extensive integral transformation. Hence, it is important to organize this step carefully. In the present work, the original LPNO-CCSD machinery was completely redesigned in order to take advantage of the localized nature of the PNOs, PAOs, and basis functions.

The most complicated integral transformation generates a linear scaling set of integrals  $(\tilde{\mu}'\tilde{\nu}'|K)$ . To this end, we keep track of which indices  $\tilde{\mu}'$  and  $\tilde{\nu}'$  that are ever needed together with a given aux-function  $K$  and ensure that only those integrals are generated and stored. While requiring extensive bookkeeping, the transformation is realized in a linear scaling fashion.

The second integral transformation generates all required integrals over PNOs. To this end, the relevant subset of integrals  $(\tilde{\mu}'\tilde{\nu}'|K)$ ,  $(k\tilde{\nu}'|K)$ , and  $(kl|K)$  are determined for a given pair  $ij$  and read from disk. Here, the indices  $k$  and  $l$  refer to internal orbital indices of pairs that interact with pair  $ij$  in the sense described above. The various sets of three index integrals are transformed into the PNO basis of pair  $ij$  and subsequently the necessary four index integrals  $(\tilde{a}_{ij}\tilde{c}_{ij}|\tilde{b}_{ij}\tilde{d}_{ij})$ ,  $(k\tilde{a}_{ij}|\tilde{b}_{ij}\tilde{c}_{ij})$ ,  $(i\tilde{a}_{ij}|j\tilde{b}_{ij})$ ,  $(ij|\tilde{a}_{ij}\tilde{b}_{ij})$ ,  $(ik|j\tilde{a}_{ij})/(jk|i\tilde{a}_{ij})$ ,  $(ij|k\tilde{a}_{ij})$  and  $(ik|jl)$ , are generated and stored. The number of such integrals is linear scaling with respect to system size.

Special attention must be given to the two-virtual integrals that occur in the semi-joint pair-pair interactions. In the initial LPNO-CEPA and CCSD versions, it was attempted to project the integrals from the PNO set of one pair into the

PNO set of another pair. This is clearly not a good approximation and led to large errors. Therefore, in the present LPNO-CEPA and LPNO-CCSD versions, these integrals were generated in the canonical virtual basis and then transformed into the respective PNO basis as needed. This was organized in a step outside the actual coupled cluster iterations.

However, the generation and storage of the canonical four index Coulomb and exchange operators is out of the question in the framework of the DLPNO-CCSD method that aims at (near) linear scaling. Hence, we have developed a third, fairly elaborate, integral transformation that generates the integrals that are required for the computation of the pair-pair interaction terms and pre-stores the amplitude independent parts on disk. To illustrate the process, consider the term occurring in the doubles residual

$$R_{\tilde{a}_{ij}\tilde{b}_{ij}}^{ij} \leftarrow \sum_{\tilde{c}_{kj}\tilde{d}_{kj}} \left( K_{\tilde{a}_{ij}\tilde{c}_{kj}}^{ik} - \frac{1}{2} J_{\tilde{c}_{kj}\tilde{a}_{ij}}^{ik} \right) \left( 2t_{\tilde{c}_{kj}\tilde{d}_{kj}}^{kj} - t_{\tilde{d}_{kj}\tilde{c}_{kj}}^{kj} \right) S_{\tilde{d}_{kj},\tilde{b}_{ij}}^{kj,ij}. \quad (26)$$

This contribution obviously involves the integrals  $K_{\tilde{a}_{ij}\tilde{c}_{kj}}^{ik}$  and  $J_{\tilde{c}_{kj}\tilde{a}_{ij}}^{ik}$  in mixed PNO bases. These integrals are generated from the corresponding integrals  $K_{\tilde{\mu}'\tilde{\nu}'}^{ik}$  and  $J_{\tilde{\mu}'\tilde{\nu}'}^{ik}$  that are generated on the fly from the various pre-stored three index integrals and then transformed to the corresponding PNO basis. Note that in only this step, the indices  $\tilde{\mu}'$  and  $\tilde{\nu}'$  belong to the extended domains defined above. Owing to the domain construction algorithm described above, the extended domains are so large that the resulting integrals over PNOs are accurately represented and hence no inaccurate projection is involved while the computational effort to generate and manipulate the integrals still remains linear scaling with system size.

For each pair  $ij$  the so generated integrals over PNOs ( $K_{\tilde{a}_{ij}\tilde{c}_{kj}}^{ik}$  and  $J_{\tilde{c}_{kj}\tilde{a}_{ij}}^{ik}$  for all interacting  $k$  in the sense defined above) are stored in a “pair interaction file.” The size of this file scales linearly with respect to system size and since the PNO spaces are highly compact it never grows out of proportion.

During the CEPA iterations the pair interaction file is read and the relevant integrals are contracted with the CEPA amplitudes. In the QCISD and CCSD cases, the situation is slightly more complex because the integrals must first be dressed with cluster amplitudes. As in the LPNO-CCSD case, these nonlinear terms are generated through projection rather than transformation. To make the process more transparent, consider the following term that occurs in the doubles residual:

$$R_{ab}^{ij} \leftarrow - \sum_{kc} \tilde{J}_{ac}^{kj} t_{bc}^{ki}, \quad (27)$$

which becomes in the PNO case

$$R_{\tilde{a}_{ij}\tilde{b}_{ij}}^{ij} \leftarrow - \sum_{k,\tilde{c}_{ki},\tilde{d}_{ki}} \tilde{J}_{\tilde{a}_{ij}\tilde{c}_{ki}}^{kj} t_{\tilde{d}_{ki}\tilde{b}_{ij}}^{ki} S_{\tilde{d}_{ki}\tilde{b}_{ij}}^{ki,ij}. \quad (28)$$

This requires the dressed integral  $\tilde{J}_{\tilde{a}_{ij}\tilde{c}_{ki}}^{kj}$ , which in turn is generated as

$$\tilde{J}_{\tilde{a}_{ij}\tilde{c}_{ki}}^{kj} = J_{\tilde{a}_{ij}\tilde{c}_{ki}}^{kj} - \frac{1}{2} \sum_{l,\tilde{b}_{lj},\tilde{e}_{lj},\tilde{d}_{lk},\tilde{f}_{lk}} S_{\tilde{a}_{ij}\tilde{b}_{lj}}^{ij,lj} K_{\tilde{b}_{lj}\tilde{e}_{lj}}^{lj} S_{\tilde{e}_{lj}\tilde{d}_{lk}}^{lj,lk} \tau_{\tilde{f}_{lk}\tilde{d}_{lk}}^{lk} S_{\tilde{f}_{lk}\tilde{c}_{ki}}^{lk,ki} + \dots \quad (29)$$

The PNO-PNO overlap matrices in Eqs. (26)–(29) serve to project between PNO bases of different pairs. It is noteworthy that the calculation of the PNO pair/pair overlaps is slightly more complicated than in the original LPNO-CCSD case due to the non-orthogonality of the (redundant) PAO space

$$S_{\tilde{a}_{ij},\tilde{b}_{kl}}^{ij,kl} = \sum_{\tilde{\mu}' \in ij} \sum_{\tilde{\nu}' \in kl} d_{\tilde{\mu}'\tilde{a}}^{ij} S_{\tilde{\mu}'\tilde{\nu}'} d_{\tilde{\nu}'\tilde{b}}^{kl}. \quad (30)$$

These overlaps are generated and stored on disk in a compressed fashion with linear scaling effort and storage requirements given that the number of interacting pairs is asymptotically linear scaling. Again, since the PNO spaces are so compact, the size of the file containing the overlap integrals never grows out of proportion.

#### D. Solution of the DLPNO-CCSD equations

The CCSD equations are solved in the usual fashion by employing a direct-inversion in iterative subspace algorithm.<sup>131,132</sup> All these steps proceed as detailed before in the description of the LPNO-CCSD method.<sup>105,106</sup> The main deviation is that interacting pairs (and singles) are restricted to those combinations that feature overlapping domains. Due to this restriction the entire residual construction proceeds in an essentially linear scaling fashion. The only nonlinear term is the calculation of the singles Fock matrix. However, it is well known how to linearize Fock matrix construction algorithms.<sup>133–138</sup> In addition, the contribution of this term is small and essentially no loss of accuracy occurs upon invoking the RIJCOSX approximation.<sup>111</sup> In the largest calculations that we have performed to date with more than 8800

basis functions, this term consumed about 40%–50% of the sigma-vector construction time, which in turn contributes typically about 20% to the total calculation time. Hence, further optimization of this term will likely lead to modest improvements in the wall clock times. Wherever possible matrix multiplications have been used in the formation of the residual. However, since these matrices are generally of low dimension this is not nearly as important as in the canonical case where BLAS level 3 operations are mandatory in order to obtain high performance.

#### E. Summary

The preceding paragraphs have provided a detailed description of the DLPNO-CCSD method. Relative to its predecessor all steps except for the essentially trivial initial generation of the integrals ( $ij|K$ ) and ( $i\tilde{\mu}'|K$ ) are generated in a linear scaling fashion but using very tight thresholds. In fact, relative to the LPNO-CCSD method no additional thresholds were introduced. The few additional truncations have been adjusted according to the previous thresholds  $T_{CutPNO}$ ,  $T_{CutPairs}$ , and  $T_{CutMKN}$  of which  $T_{CutPNO}$  is the single most important parameter. Thus, in DLPNO-CCSD, singles are truncated according to  $0.03T_{CutPNO}$  and pairs are being prescreened according to  $0.01T_{CutPairs}$  using a multipole based algorithm. Initial orbital domains are constructed according to  $T_{CutMKN}$  and pair domains according to a fixed delocalization parameter with the value 0.1 as described above. Furthermore, extended domains are used to provide an accurate buffer space for the calculation of pair-pair interactions that, nevertheless, remain local. The truncation  $T_{CutOSV} = 10^{-6}$  controls the number of LOSVs that are used in the pair prescreening process. It is important to point out that all truncations have been chosen conservatively. They are not meant to be changed or re-investigated in actual applications of the DLPNO-CCSD method.

### III. RESULTS

A number of calculations have been performed on a range of medium sized to large molecules that are shown in Figure 3.

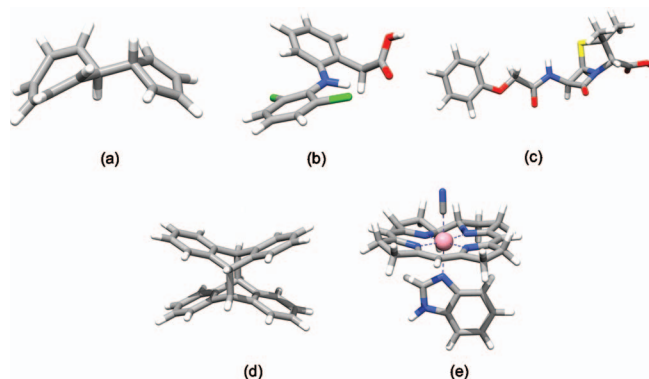


FIG. 3. Structures of some of the molecules studied in this work: (a) cyclopentadienyltropenyle, (b) diclophenac, (c) penicillin, (d) anthracene dimer and (e) vitamin B12.

## A. Accuracy tests

### 1. Convergence with respect to truncation parameters

The convergence of the LPNO-CCSD results with respect to the three truncation parameters has been studied for the molecule cyclopentadienyltropenyle (see Figure 3). This system is large enough such that the domains do not include the entire molecule and small enough for canonical CCSD calculations to be performed as reference calculations.

The behavior of DLPNO-CCSD with respect to the most important truncation parameter  $T_{\text{CutPNO}}$  is very similar to the behavior of LPNO-CCSD (Figure 4). Smooth convergence from below is observed. At the default threshold of  $3.33 \times 10^{-7}$  about 99.8% of the correlation energy is recovered which increases to 99.95% with inclusion of the MP2 correction. In our opinion, the small residual deviations from the canonical results observed in DLPNO-CCSD are entirely acceptable for chemical applications.

The convergence with respect to  $T_{\text{CutPairs}}$  follows the established pattern and is very close to what has been obtained for LPNO-CCSD (Figure 5). At the default value of  $10^{-4}$  Eh, the remaining error is virtually zero after inclusion of the MP2 correction. There is a noticeable tendency for over-correction at less tight values of  $T_{\text{CutPairs}}$ . In our experience, the default value is, however, already on the safe side.

The convergence with respect to  $T_{\text{CutMKN}}$  in Figure 6 is also very similar to the LPNO-CCSD case with the default value of  $10^{-3}$  leading well into the converged regime. This is an important observation since  $T_{\text{CutMKN}}$  takes on a more important role in DLPNO-CCSD. In LPNO-CCSD, this parameter was only used to establish a fitting domain of auxiliary basis functions. In DLPNO-CCSD, it additionally controls the sizes of the domains that contribute their PAOs to the PNOs of the given pair. Since the fitting domains in LPNO-CCSD were already large enough to accurately fit the PNOs it appears to

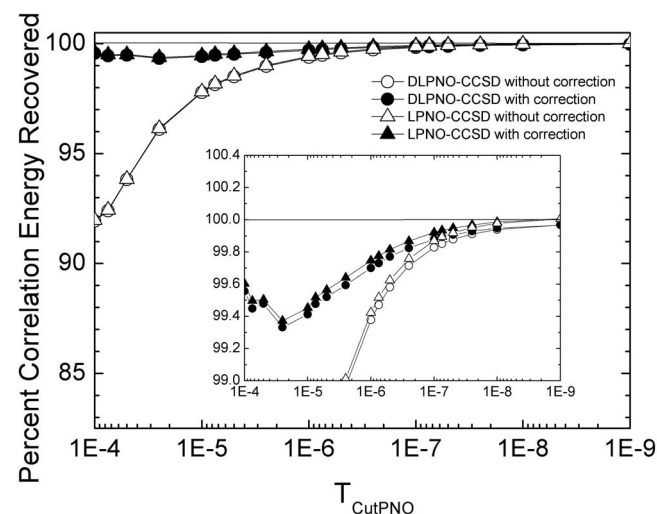


FIG. 4. The convergence of the DLPNO-CCSD (circles) and LPNO-CCSD (triangles) method with respect to variations of the  $T_{\text{CutPNO}}$  parameter. Calculations were carried out on the cyclopentadienyltropenyle molecule with the def2-TZVP basis set. Open symbols = results without perturbative correction for the PNO error, full symbols = results with perturbative correction of the PNO and pair truncation error. Thresholds: ( $T_{\text{CutPairs}} = 0$ ;  $T_{\text{CutMKN}} = 10^{-3}$ ).

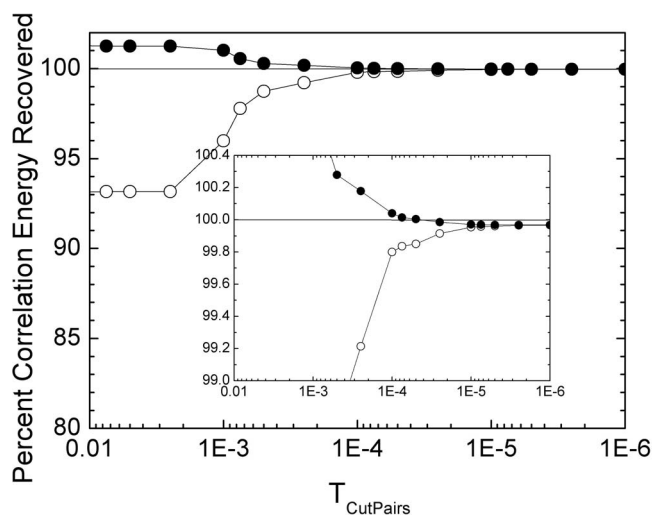


FIG. 5. The convergence of the DLPNO-CCSD method with respect to variations of the  $T_{\text{CutPairs}}$  parameter. Calculations were carried out on the cyclopentadienyltropenyle molecule with the def2-TZVP basis set. Full circles = results with perturbative correction for the PNO and local errors, open circles = results without perturbative correction of the PNO and local errors. Thresholds: ( $T_{\text{CutPNO}} = 1 \times 10^{-9}$ ;  $T_{\text{CutMKN}} = 10^{-3}$ ).

be reasonable that the PAOs of the same domain are sufficient to accurately expand the PNOs. The only slight extension is the delocalization that has been described above.

### 2. Accuracy of absolute energies

The results for absolute correlation energies of a number of medium sized molecules that are identical to the ones studied previously,<sup>106</sup> are collected in Table I. Relative to CCSD these numbers contain a small error due to the RI approximation, that is, however, insignificant relative to the local and PNO approximations. It is evident from the table that both, LPNO-CCSD and DLPNO-CCSD recover more than 99.95% of the correlation energy for each system. The mean absolute

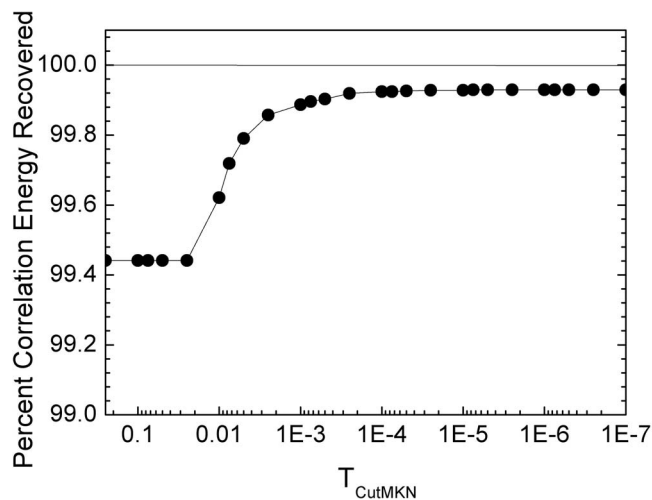


FIG. 6. The convergence of the DLPNO-CCSD method with respect to variations of the  $T_{\text{CutMKN}}$  parameter. Calculations were carried out on the cyclopentadienyltropenyle molecule with the def2-TZVP basis set. Thresholds: ( $T_{\text{CutPNO}} = 3.33 \times 10^{-7}$ ;  $T_{\text{CutPairs}} = 10^{-4}$ ).

TABLE I. Error analysis of DLPNO-CCSD for the absolute energies of a series of medium sized reference molecules with the def2-TZVP basis set. All molecules have an average pair domain size &lt;90% of the whole system size.

	E <sub>corr</sub> [Eh]	Error [kcal/mol]		%E <sub>corr</sub> (CCSD)	
	CCSD	LPNO-CCSD	DLPNO-CCSD	LPNO-CCSD	DLPNO-CCSD
Heptatriyne	−1.068458329	0.38	0.12	99.94	99.98
2,3-dimethylbut-2-ene	−1.072641372	0.36	0.29	99.95	99.96
Octamethylcyclobutane	−2.183565069	0.59	0.98	99.96	99.93
<i>p</i> -xylene	−1.321960090	0.58	0.41	99.93	99.95
[2,2] <i>para</i> -cyclophane	−2.613418615	−0.09	−1.25	100.01	100.08
Cyclopentadienyltropenyle	−1.943415908	0.65	0.44	99.95	99.96
<i>n</i> -octane	−1.463003832	0.43	0.58	99.95	99.94
2,2,3,3-tetramethylbutane	−1.481895747	0.49	0.50	99.95	99.95
<i>n</i> -hexane H <sub>ttt</sub> (elongated)	−1.103845100	0.29	0.44	99.96	99.94
<i>n</i> -hexane H <sub>x</sub> +g <sub>x</sub> + (folded)	−1.107709465	0.17	0.16	99.98	99.98
Phe-Gly-Gly (folded) <sup>a</sup>	−3.061001482	0.94	1.20	99.95	99.94
Phe-Gly-Gly (elongated) <sup>a</sup>	−3.052492950	1.64	1.62	99.91	99.92
<i>n</i> -pentane	−0.924815332	0.23	0.19	99.96	99.97
Dimethyl peroxide	−0.851572103	0.25	0.32	99.95	99.94
Vinyl acetate	−1.125124521	0.36	0.43	99.95	99.94
Pentan-2,4-dione	−1.305652949	0.25	0.22	99.97	99.97
Hexanoic acid	−1.523329340	0.30	0.40	99.97	99.96
Methyl pivalate	−1.526873610	0.33	0.24	99.97	99.97
	MAD [kcal/mol]	0.46	0.54		
	MD [kcal/mol]	0.45	0.41	Mean [%]	99.96
	MAX [kcal/mol]	1.64	1.62		

<sup>a</sup>Calculated with the def2-SVP basis set.

deviations of 0.46 and 0.48 kcal/mol for the chosen test set are virtually identical and are both well within the chemical accuracy limit of 1 kcal/mol. The error is obviously an extensive property and will grow linearly with system size. However, our experience with the LPNO-CCSD method indicates that these deviations are insignificant in actual chemical applications. As discussed above, a very slight overshooting of the correlation energy is occasionally observed and attributed to the MP2 correction for the pair and PNO errors. In our opinion, this is a very small price to pay since the MP2 correction captures the remaining pair approximation and PNO errors very well.

### 3. Accuracy of relative energies

A collection of results for reaction energies (see Ref. 106) is presented in Table II. These reactions cover a number of problematic isomerization reactions in which the reactants and products have significantly different structures and hence the domain structure changes strongly between reactants and products. A point in case is the isomerization from *n*-octane to 2,2,3,3-tetramethylbutane. There is only a single case for which the errors exceed 1 kcal/mol in both LPNO-CCSD and DLPNO-CCSD which is the reaction 2 *p*-xylene → [2,2]*para*cyclophane + 2 H<sub>2</sub>. We believe that this case represents a worst case scenario for local correlation methods. Nevertheless, the MADs of about 0.3 and 0.4 kcal/mol for LPNO-CCSD and DLPNO-CCSD show that both methods behave excellently for chemical reactions and that the typical error is smaller than other uncertainties including errors in geometries, treatment of environment effects, and basis set incompleteness.

### 4. Potential energy surfaces

In order to investigate whether any artificial bumps or discontinuities occur in potential energy surfaces, we have re-investigated some systems that were treated earlier in the development of the LPNO-CEPA and LPNO-CCSD methods using the same geometries and basis sets.<sup>105,106</sup> The first system to be studied is the CO dissociation from ketene which was identified by Crawford<sup>49</sup> to cause a number of discontinuities in PAO based local correlation methods. Mata and Werner<sup>139</sup> demonstrated how domain merging could overcome these problems. As is evident from Figure 7, neither LPNO-CCSD nor DLPNO-CCSD have any requirement for domain merging and there are no discernible kinks, humps or jumps in the potential energy surface. Furthermore, the two methods are virtually indistinguishable. This also holds for the dissociation region of the molecule, where the domain structure changes. We take this as a first indication that the truncations have been done conservatively enough to ensure smoothness in potential energy surfaces.

The second test is performed on the biphenyl rotational barrier. This barrier to rotation around the central C–C bond is less than 1.5 kcal/mol high. Thus, the barrier is not only small, but there will also be changes in the extension of the delocalized  $\pi$ -system over the course of the rotation coordinate that might lead to artifacts if domains behave inconsistently. Hence, this system might be a sensitive test for a local correlation treatment. Again, Figure 8 shows that the DLPNO-CCSD method behaves perfectly well and is essentially indistinguishable from the LPNO-CCSD curve, which in turn follows the canonical CCSD curve to an accuracy of better than 0.05 kcal/mol. For completeness, we have also shown MP2



TABLE II. Error analysis of DLPNO-CCSD for the reaction energies of a series of medium sized reference molecules with the def2-TZVP basis set. At least one of the reaction partners of each reaction has an average pair domain size <90% of the whole system size.

Reaction	$E_{\text{reac}}$ [kcal/mol]	Error [kcal/mol]	
	CCSD	LPNO-CCSD	DLPNO-CCSD
Heptahexane $\rightarrow$ heptatriyne	-18.02	-0.76	-0.57
2,2,3-dimethylbut-2-ene $\rightarrow$ octamethylcyclobutane	-17.66	-0.13	0.39
2 p-xylene $\rightarrow$ [2,2]paracyclophane + 2 H <sub>2</sub>	63.21	-1.25	-2.08
C <sub>12</sub> H <sub>12</sub> (D <sub>6h</sub> cage) $\rightarrow$ C <sub>12</sub> H <sub>12</sub> (cyclopentadienyletropyne)	24.95	-0.43	-0.45
n-octane $\rightarrow$ 2,2,3,3-tetramethylbutane	-0.45	0.07	-0.09
n-hexane isomerization (H <sub>ttt</sub> $\rightarrow$ H <sub>xx</sub> +g <sub>xx</sub> )	5.37	-0.12	0.41
Phe-Gly-Gly isomerization (folded $\rightarrow$ elongated) <sup>a</sup>	1.51	0.70	-0.29
Neo-pentane $\rightarrow$ n-pentane	3.49	-0.05	-0.08
Ethandiol $\rightarrow$ dimethylperoxide	64.15	0.22	0.21
Dihydrofuran-2(3H)-one $\rightarrow$ vinyl acetate	15.61	-0.01	0.06
Tetrahydropyran-2-one $\rightarrow$ pentan-2,4-dione	6.39	-0.14	-0.10
Hexanoic acid $\rightarrow$ methyl pivalate	8.08	0.02	-0.17
	MAD [kcal/mol]	0.33	0.41
	MD [kcal/mol]	-0.16	-0.23
	MAX [kcal/mol]	-1.25	-2.08

<sup>a</sup>Calculated with the def2-SVP basis set.

and B3LYP rotational curves that show significant deviations from the coupled cluster results.

The final test, shown in Figure 9, deals with the rotational potential energy surface around the central C–C bond in ethane-1,2-biphenyl. This is an interesting case as it involves stacked rings for a dihedral angle of 0° and essentially non-interacting rings at 180°. As one moves through the rotational surface one thus covers all intermediate interaction regimes and it is clear that there must be significant changes to the domain structure over the course of a full rotation. Hence, it is most gratifying to once more observe that LPNO-CCSD and DLPNO-CCSD behave in an essentially indistinguishable way and both follow the canonical CCSD curve to an accuracy

of about 0.3 kcal/mol. Again, MP2 and B3LYP show much larger deviations.

Since all calculations have been performed on the basis of the standard thresholds we are confident that the introduction of domains into the DLPNO-CCSD method has not caused any artifacts in calculated potential energy surfaces.

## B. Scaling behavior

The scaling behavior of the DLPNO-CCSD method relative to LPNO-CCSD and canonical CCSD has been investigated on linear alkane chains in standard geometries and in conjunction with the def2-SVP basis set. This clearly is a “best case” scenario for any local correlation method since the onset of linear scaling is reached earlier than in true three-dimensional systems. We nevertheless believe that it is still

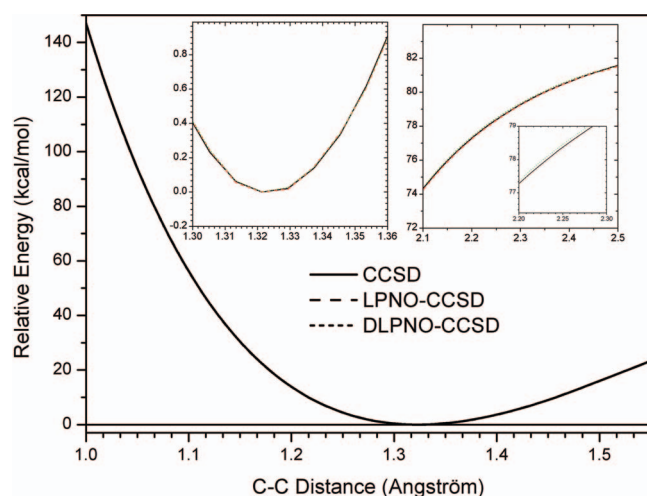


FIG. 7. Scanned potential energy surface for the dissociation of CO from ketene. All calculations were done with the def2-SVP basis set. Solid line = CCSD, dashed line = LPNO-CCSD, short dashed line = DLPNO-CCSD. The insets show magnifications of the calculated curve. At the level of plot resolution no difference is discernible between LPNO-CCSD and DLPNO-CCSD. Default thresholds were used ( $T_{\text{CutPNO}} = 3.33 \times 10^{-7}$ ;  $T_{\text{CutPairs}} = 10^{-4}$  Eh,  $T_{\text{CutMKN}} = 10^{-3}$ ).

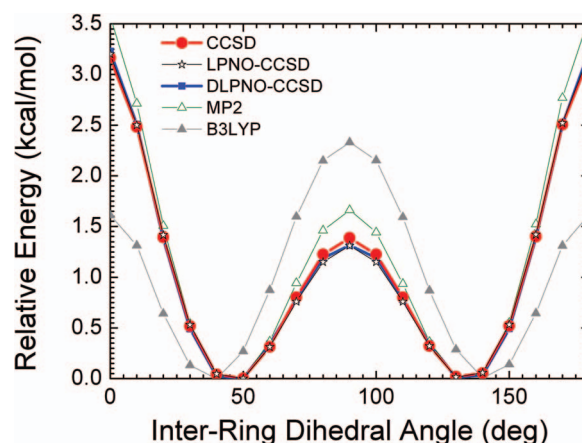


FIG. 8. Rotational barrier of biphenyl. The structures come from a relaxed surface scan with the PBE functional and the TZVP basis set together with van der Waals corrections. All calculations shown were done with the SV(P) basis set. Circles = CCSD, stars = LPNO-CCSD, quadrats = DLPNO-CCSD, open triangles = MP2, filled triangles = B3LYP. Default thresholds were used ( $T_{\text{CutPNO}} = 3.33 \times 10^{-7}$ ,  $T_{\text{CutPairs}} = 10^{-4}$  Eh,  $T_{\text{CutMKN}} = 10^{-3}$ ).

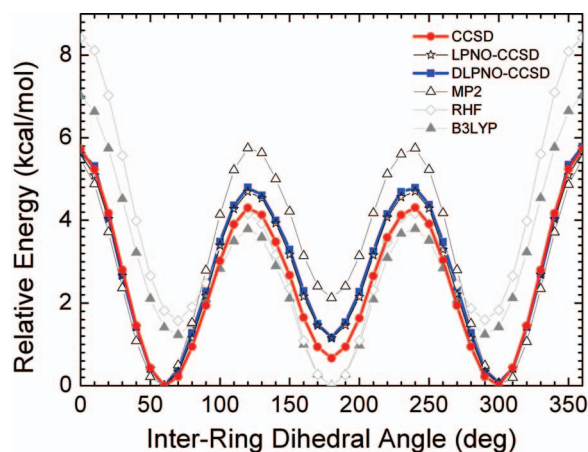


FIG. 9. Rotational barrier of ethane-1,2-diphenyl. The structures come from a relaxed surface scan with the PBE functional and the TZVP basis set together with van der Waals corrections. All calculations shown were done with the def2-SV(P) basis set. Circles = CCSD, stars = LPNO-CCSD, quadrats = DLPNO-CCSD, open triangles = MP2, filled triangles = B3LYP, open diamond = RHF. Default thresholds were used ( $T_{\text{CutPNO}} = 3.33 \times 10^{-7}$ ,  $T_{\text{CutPairs}} = 10^{-4}$  Eh,  $T_{\text{CutMKN}} = 10^{-3}$ ).

appropriate since all conclusions will carry over to real three-dimensional systems.

It is obvious from Figure 9 that LPNO-CCSD is very much more efficient than canonical CCSD and there is virtually no crossover region. LPNO-CCSD remains efficient up to about 100 atoms after which the nonlinear terms take over and the calculations start to become expensive. It is most gratifying to observe that DLPNO-CCSD shows a performance that is very similar to LPNO-CCSD for smaller systems but outperforms it by a large margin for systems with  $> 80$  atoms. Again, there virtually is no crossover point and DLPNO-CCSD is always more efficient than LPNO-CCSD. In order to achieve this, we have taken an approach that balances the prefactor and scaling behavior of the calculations. This concerns the most expensive step in the integral transformation, which consists of the generation of the  $\mathbf{J}^{ik}$ ,  $\mathbf{K}^{ik}$  operators in the extended domains, as required for the accurate calculation of pair-pair interactions. The most cumbersome terms arise from the Coulomb operators as within the density fitting approximation this requires the potential large set of integrals ( $\tilde{\mu}'\tilde{\nu}'|\bar{K}$ ). In order to achieve optimal performance, for smaller molecules ( $< 60$  atoms based on test calculations), we employ an integral transformation that generates all of those integrals (thus sacrificing linear scaling), while for molecules larger than 60 atoms we use a linear scaling local transformation algorithm. The result is that calculations on smaller molecules are almost twice as fast as they would be if only a linear scaling integral transformation would be used throughout.

As is evident from Figure 10, DLPNO-CCSD shows essentially linear scaling with respect to systems size between 30 and 220 atoms. Thus, future improvements notwithstanding, DLPNO-CCSD is applicable to pretty much any molecule for which one is able to perform a self-consistent field calculation while the resource usage for the DLPNO-

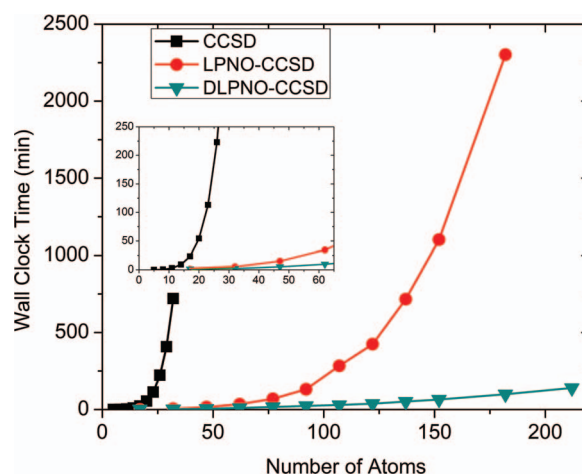


FIG. 10. Scaling behavior of the CCSD, LPNO-CCSD, and DLPNO-CCSD methods. The calculations were carried out on linear hydrocarbon chains  $\text{H}_3\text{C}-(\text{CH}_2)_{n-2}-\text{CH}_3$  (in standardized geometries with all angles =  $109.4712^\circ$ ,  $\text{RCH} = 109$  pm and  $\text{RCC} = 155$  pm) and default thresholds. The insets show magnifications of the calculated curve. Default thresholds were used ( $T_{\text{CutPNO}} = 3.33 \times 10^{-7}$ ,  $T_{\text{CutPairs}} = 10^{-4}$  Eh,  $T_{\text{CutMKN}} = 10^{-3}$ ).

CCSD part is not growing out of proportion. Thus, the design objectives of DLPNO-CCSD have been met.

### C. Representative timings

In this section, we provide some representative timings for DLPNO-CCSD and LPNO-CCSD calculations on larger molecules for a set of molecules that more closely resemble real-life applications of the method. For this reason, we have also chosen the def2-TZVP basis set<sup>140</sup> that is of triple zeta quality and features 2d1f and 1p polarization sets on the heavy atoms and hydrogens, respectively. With basis sets of this size, sufficiently accurate results are usually obtained while CC results with double-zeta basis sets are clearly not suitable for chemical applications. This is alleviated by explicitly correlated methods. However, we note in passing that the basis set flagged “cc-pVDZ-F12” for F12 calculations<sup>141</sup> contains as many as five sets of p-functions and hence is almost of the same size than a regular triple- $\zeta$  basis set.

Representative timings for the molecules in Figure 3 are collected in Tables III and IV. The systems feature 30-452 atoms with 660-8878 basis functions (Table IV). The average number of atoms in each pair domain is typically about 20 but can reach up to 30 in more dense and electronically delocalized systems such as the anthracene dimer. It is clear that the number of PAOs in each domain (400-650) is too demanding in order for pure PAO based CCSD calculations to be performed in an efficient manner. It is therefore gratifying to observe that only 20-36 PNOs are required on average to accurately describe the pair correlations. It is this compactness of the PNO space that allows such calculations to be performed efficiently.

Molecules with up to about 100 atoms complete the entire DLPNO-CCSD calculation in a few hours of computing time on a single core while the larger calculations may take a few days to complete. It is most gratifying that the present code is able to perform a calculation on the

TABLE III. Wall clock times (s) for DLPNO calculations on a series of large systems with def2-TZVP. All calculations were carried out on a single core of a Linux computer (operating system Linux 2.6.32) with 4 octo-core Intel XEON 2.67 GHz CPUs E7- 8837 that have 24 MB level 2 cache and 256 GB main memory. (n.d. = not determined. The reference determinants for these calculations came from RI-BP86 calculations for the large Vancomycin and  $C_{150}H_{302}$  calculations. All other reference wavefunctions were Hartree-Fock determinants.)

	Diclophenac (def2-TZVP)	Penicillin (def2-TZVP)	Anthracene-dimer (def2-TZVP)	Vitamin B12 (def2-TZVP)	Vancomycin (def2-SVP)	Vancomycin (def2-TZVP)	$C_{150}H_{302}$ (def2-SVP)	$C_{150}H_{302}$ (def2-TZVPP)
HF time (RIJCOSX approximation)	2154	3183	6026	20 187	n.d.	n.d.	n.d.	n.d.
HF time (no approximation)	20 391	27 585	78 952	211 593	n.d.	n.d.	n.d.	n.d.
DLPNO-CCSD total wall clock time [s]	13 323	14 545	66 710	219 974	52 514	436 861	76 957	1 164 034
Initial Fock matrix (RIJCOSX) (%)	2	2	1	0	2	2	4	12
$(ij K)$ , $(i\tilde{\mu}' K)$ , $(\tilde{\mu}'\tilde{\nu}' K)$ (%)	4	10	4	1	11	16	43	58
OSV, SL-MP2, and PNO generation (%)	10	9	12	5	11	6	9	4
PNO-integral transformation (%)	14	9	29	5	6	3	11	2
$(ij \tilde{\mu}\tilde{\nu})$ , $(i\tilde{\mu} j\tilde{\nu})$ (%)	21	26	24	75	26	49	8	12
PP overlap (%)	3	2	4	1	2	1	0	0
Precomputation of pair-pair interaction (%)	2	1	2	0	0	0	0	0
# Iterations	11	11	9	14	11	11	7	7
Sigma-vector (%)	44	40	24	11	37	22	18	11

Vancomycin molecule (Figure 11) with 176 atoms and the def2-TZVP basis set which amounts to more than 3600 basis functions in about 5 days on a single core. With the def2-SVP basis set the same calculation takes only about 12–14 h on a single core. The largest molecule treated ( $C_{150}H_{302}$ ) required only 24 h on a single core to complete with the def2-SVP basis set. Our calculation on  $C_{150}H_{302}$  with the accurate def2-TZVPP basis set featured 8878 basis functions and more than 20 000 auxiliary basis functions. To the best of our knowledge it represents the largest coupled cluster cal-

culation performed to date. While the effort for such a calculation is high ( $\sim 13$  days on a single processor and use of about 800 GB of disk space), the result shows that giant coupled cluster calculations with 10 000 basis functions and hundreds of atoms will soon be routine on powerful parallel computers. This calculation is dominated by the initial generation of the three index integrals  $(i\tilde{\mu}|K)$  showing that further progress is possible for such calculations with fully linearized code. Such calculations were far outside of the feasible range in the previous LPNO-CCSD implementation and would

TABLE IV. System size for the DLPNO-CCSD calculations on a series of large systems.

	Diclophenac (def2-TZVP)	Penicillin (def2-TZVP)	Anthracene dimer (def2-TZVP)	Vitamin B <sub>12</sub> (def2-TZVP)	Vancomycin (def2-SVP)	Vancomycin (def2-TZVP)	$C_{150}H_{302}$ (def2-SVP)	$C_{150}H_{302}$ (def2-TZVPP)
Atoms	30	42	48	62	176	176	452	452
Basis functions	667	858	988	1261	1797	3593	3610	8879
Correlated electrons	98	128	132	178	542	542	902	902
Total number of electron pairs	1225	2080	2211	4005	36 856	36 856	101 926	101 926
Electron pairs skipped due to prescreening	95 (7.8%)	736 (35.4%)	148 (7.0%)	520 (12.9%)	26 172 (71.0%)	25 480 (69.1%)	93 726 (91.9%)	93 528 (91.9%)
Electron pairs treated by semicanonical local MP2	1130	1344	2063	3485	10 684	11 376	8200	8398
Surviving electron pairs used in CCSD iterations (% of total no. of pairs)	575 (46.9)	716 (34.4)	965 (43.6)	1203 (30.0)	3160 (8.6)	3476 (9.4)	4025 (4.0)	4269 (4.0)
Correction for prescreened pairs (mEh)	0.1	0.2	0.1	2.9	3.5	3.6	1.3	1.5
Correction for pairs < $T_{\text{CutPairs}}$ (mEH)	13.2	14.3	27.8	44.9	122.2	126.6	73.4	84.4
PNO error correction (mEh)	10.1	11.9	16.6	19.6	21.0	55.2	20.8	71.1
Correlation energy including all corrections (mEh)	3286.2	4333.8	4430.5	6055.7	14 923.7	18 575.9	21 883.9	29 010.8
Average numbers per pair domain								
Atoms	18	17	31	23	20	19	20	21
PAOs	435	380	713	539	200	404	132	434
PNOs	33	30	33	30	20	28	20	34

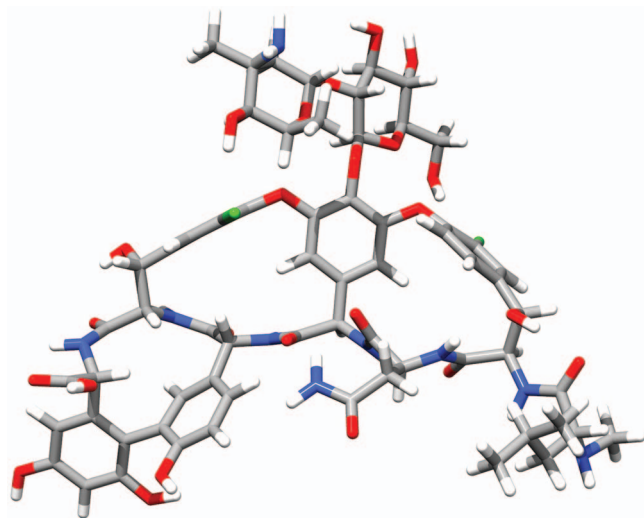


FIG. 11. Structure of Vancomycin (176 atoms).

probably be very challenging, if not impossible, for other local CC methods. It is instructive to note that the two giant DLPNO-CCSD calculations on Vancomycin and  $C_{150}H_{302}$  would have featured  $3.8 \times 10^{11}$  and  $6.9 \times 10^{12}$  canonical amplitudes, respectively. In the PNO treatment, this was reduced to  $4.4 \times 10^6$  and  $6.9 \times 10^6$  amplitudes, respectively. Thus, the PNO treatment leads to a wavefunction, which is  $10^5$ – $10^6$  times more compact than the canonical representation, thus, demonstrating the tremendous advantages of the PNO based methodology.

It should be noticed from Table III that the time required to complete a DLPNO-CCSD calculation is less than the time required to finish the preceding Hartree-Fock calculation, provided the latter is done without approximations. Only after invoking the RIJCOSX approximation to the HF method is the DLPNO-CCSD step slower. Furthermore, all “giant” calculations have been done on the basis of DFT orbitals since the generation of HF orbitals was considered to be too costly. Thus, one of the important and most ambitious design objectives has been met and the numbers show that the DLPNO-CCSD energy can be obtained in reasonable turnaround times for pretty much any system for which the HF calculation is still feasible.

For all calculations, a maximum core memory of 8–32 GB was used (Vancomycin and  $C_{150}H_{302}$  with the large basis sets 200 GB were allowed). However, the availability of large amounts of core memory is not of critical importance because all integral transformations have been organized in batches according to available memory. The overhead associated with multiple batches is very small. The computer time distributes over a number of significant steps to varying extents: (a) the initial integral transformation (typically 25% and due to poorer scaling up to 50% for the largest calculations), (b) the initial guess including the PNO generation (5%–15%), (c) the PNO integral transformation (10%–30%, more time is required for delocalized systems that have larger domains on average), (d) the precomputation of extended domain Coulomb and exchange operators used in the calculation of the pair-pair interactions (10%–50%), and finally (e) the

solution of the CCSD equations (20%–40%). For the largest calculations, the time required to form the singles Fock matrix becomes a significant part of the sigma-vector construction time owing to its quadratic scaling. It is noteworthy that for delocalized systems, such as the Anthracene dimer, the domain sizes become rather large and hence owing to the larger number of PAOs in each domain together with the reduced propensity for pair prescreening the calculations become also more expensive. However, the PNO space retains its compactness even in these cases.

Yang *et al.*<sup>46,47</sup> have compared their OSV-CCSD for the Penicillin molecule with our previous LPNO-CCSD calculation together with the def2-TZVPP basis set (1009 basis functions in the paper of Yang *et al.*,<sup>46,47</sup> 1007 basis functions in our calculation). With DLPNO-CCSD this calculation completes in 337 min elapsed time and gives a final correlation energy of  $-4.3779$  Eh (including a MP2 correction of 27 mEh). The computation time is comparable to the medium accuracy calculation of Yang *et al.*<sup>46,47</sup> and about three times faster than their most accurate calculation. However, the correlation energy is significantly higher (Yang *et al.*<sup>46,47</sup> report values around  $-4.40$  Eh). However, based on Table V of Yang *et al.*,<sup>46,47</sup> the OSV method appears to converge to the correct result from below while in DLPNO one converges to the correct correlation energy from above. With the previous LPNO-CCSD code, we calculate a correlation energy of  $-4.3781$  Eh. The deviation of just 0.2 mEh excludes the possibility that the additional approximations in DLPNO-CCSD significantly influence the result. Since we wanted more conclusive insight into the discrepancy in the results, we have undertaken the expensive canonical CCSD calculation on penicillin with the def2-TZVPP basis set (featuring  $1.7 \times 10^9$  cluster amplitudes) This resulted in a correlation energy of  $-4.3799$  Eh (we note in passing that the singles Fock operator was also formed by the RIJCOSX approximation in these calculations. This leads to deviations from the fully canonical result of less than 0.1 mEh). The RI/density fitting error in LPNO-CCSD and DLPNO-CCSD was estimated to be approximately +0.5 mEh (i.e., to higher correlation energies) from a comparison of MP2 and RI-MP2 calculations using the same basis set. Thus, the LPNO-CCSD and DLPNO-CCSD results are within 2 mEh of the canonical result while the most accurate LCCSD and OSV-CCSD results deviate by about 19 mEh from this number. Since slight differences in the basis sets used may contribute to the discrepancy in the results between Yang *et al.*<sup>46,47</sup> and the present ones, we also report the DLPNO-CCSD value with the cc-pVTZ basis set (976 basis functions) which is  $-4.3860$  Eh. Calculation times are virtually identical to the def2-TZVPP calculation.

Anticipating the efficient parallelization of the code, it is evident that all of the calculations reported here, including Vancomycin with almost 180 atoms in conjunction with a fairly accurate def2-TZVP basis set, can be performed in a routine fashion in a few hours elapsed time if a few dozen cores are devoted. We anticipate excellent scaling behavior with respect to the number of CPUs on modern computer architectures. This clearly brings the DLPNO-CCSD method into a regime where—at fixed geometry—it can be used in a routine fashion for mainstream computer chemistry



applications in essentially the same way that density functional theory is used today.

#### IV. DISCUSSION

In the present work, the LPNO-CCSD method was extensively re-designed to take advantage of locality. Through this development the near linear scaling and highly efficient DLPNO-CCSD method is obtained. Relative to LPNO-CCSD a few additional approximations obviously needed to be introduced. We have, however, designed these approximations such, that the closest possible match to the LPNO-CCSD method is obtained and the method remains entirely in the spirit of the original development. Thus, no real-space cutoffs or complex pair hierarchies were introduced and all thresholds were taken as conservatively as possible in order to guarantee that the method is of black box character and remains robust and reliable. Even higher efficiency could obviously be obtained through more aggressive truncation and the method would still be accurate enough for most purposes. However, we choose to not do this in the interest of utmost robustness. We have also made a dedicated effort during the development to keep the underlying computer code as straightforward as possible which, in our opinion, is necessary in order to facilitate future developments.

Relative to the LPNO-CCSD method, the additional truncations concern the initial screening of negligible pairs, the pair domain extension as well as the truncation of the single excitations. Since our cutoffs are conservative, the pair domains are large and accurate. They contain about 15–20 but up to 30 atoms on average. With a standard triple-zeta basis set, this means that the domains contain 300–715 PAOs. This would probably be prohibitive for a PAO based LCCSD calculation but is manageable in the PNO case owing to the compactness of the PNOs, which condense the correlation space to about 20–40 PNOs on average. Thus, DLPNO-CCSD retains the excellent behavior of the LPNO-CCSD method with respect to basis set extension.

Test calculations reveal that the loss of correlation energy due to the additional approximations in DLPNO-CCSD is very small (<0.05% on average). All chemically relevant tests on energy differences show that the DLPNO-CCSD method is nearly identical to LPNO-CCSD and canonical CCSD. Calculations on medium size to large molecules reveal that DLPNO-CCSD is significantly faster than the parent LPNO-CCSD method, which in turn already is an efficient local correlation method. Since all resource usage (disk, memory, and CPU) is either perfectly or near linear scaling with respect to system size, CCSD calculations with a few hundred atoms and several thousand basis functions are clearly within reach and have been demonstrated in this work.

In our opinion, the present development combines all the advantages of the PAO and PNO based correlation methods. Relative to the PAO methods, the present development benefits from the much greater compactness of the wavefunction. This makes it possible to use significantly larger and more reliable domains and reduces the prefactor of the calculations to the extent that larger calculations than before can be tackled and higher accuracy be obtained. Relative to the previous

PNO treatments, the present treatment achieves near linear scaling while retaining the compactness of the PNO approach even for large molecules and large basis sets. By dropping the requirement that the PNOs are expanded in the canonical virtual space, we were also able to calculate several terms with higher accuracy than before. The accuracy is gained by using a subset of the PAOs as buffer space rather than employing the less precise projection of the PNOs of one pair into the PNO space of another pair.

Relative to the OSV methods, the PNO methods have the significant advantage that the expansion space for intermediate and weak pairs is automatically small while in the OSV approach the expansion spaces of the constituting internal orbitals are used. Thus, the PNO wavefunction is significantly more compact than the OSV one. How the methods compare in terms of actual computer timings remains to be seen. Relative to fragment and incremental approaches, the PNO methods have the advantage that the correlation energy is calculated in a single run and is not pieced together by separate calculations which may involve significant overhead due to redundant computational steps. Furthermore, in the LPNO approaches there is no chemically motivated division into molecular fragments necessary that may occasionally lead to ambiguities or inconsistencies.

Overall, we believe that the present development constitutes an important step forward in the development of local correlation methods. Obvious future developments include the extensions of the method to open shells and its efficient parallelization. These as well as other subjects are under development in our research group.

#### ACKNOWLEDGMENTS

The authors gratefully acknowledge financial support of this work by the Max Planck Society as well as stimulating discussions on the subject of this paper with Professor Marcel Nooijen.

- <sup>1</sup>R. J. Bartlett, *Rev. Mod. Phys.* **79**, 291 (2007).
- <sup>2</sup>A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper, H. Koch, J. Olson, and A. K. Wilson, *Chem. Phys. Lett.* **286**, 243 (1998).
- <sup>3</sup>D. G. Truhlar, *Chem. Phys. Lett.* **294**, 45 (1998).
- <sup>4</sup>A. Tajti *et al.*, *J. Chem. Phys.* **121**, 11599 (2004).
- <sup>5</sup>A. Karton, E. Rabinovich, J. M. L. Martin, and B. Ruscic, *J. Chem. Phys.* **125**, 144108 (2006).
- <sup>6</sup>D. G. Liakos and F. Neese, *J. Phys. Chem. A* **116**, 4801 (2012).
- <sup>7</sup>T. Helgaker, W. Klopper, and D. P. Tew, *Mol. Phys.* **106**, 2107 (2008).
- <sup>8</sup>E. F. Valeev, and T. D. Crawford, *J. Chem. Phys.* **128**, 244113 (2008).
- <sup>9</sup>T. B. Adler, G. Knizia, and H. Werner, *J. Chem. Phys.* **127**, 221106 (2007).
- <sup>10</sup>G. E. Moore, Moore's law is the observation that over the history of computing hardware, the number of transistors on integrated circuits doubles approximately every two years. The period often quoted as "18 months" is due to Intel executive David House, who predicted the period for a doubling in chip performance (being a combination of the effect of more transistors and their being faster), 1965; see [http://en.wikipedia.org/wiki/Moore's\\_Law](http://en.wikipedia.org/wiki/Moore's_Law).
- <sup>11</sup>S. Saebo and P. Pulay, *J. Chem. Phys.* **88**, 1884 (1988).
- <sup>12</sup>S. Saebo and P. Pulay, *J. Chem. Phys.* **86**, 914 (1987).
- <sup>13</sup>S. Saebo and P. Pulay, *Chem. Phys. Lett.* **113**, 13 (1985).
- <sup>14</sup>P. Pulay, *Chem. Phys. Lett.* **100**, 151 (1983).
- <sup>15</sup>J. W. Boughton and P. Pulay, *J. Comput. Chem.* **14**, 736 (1993).
- <sup>16</sup>H.-J. Werner and K. Pflüger, *Annu. Rep. Comp. Chem.* **2**, 53 (2006).
- <sup>17</sup>T. B. Adler, H.-J. Werner, and F. R. Manby, *J. Chem. Phys.* **130**, 054106 (2009).

- <sup>18</sup>H.-J. Werner, *J. Chem. Phys.* **129**, 101103 (2008).
- <sup>19</sup>R. A. Mata, H.-J. Werner, and M. Schütz, *J. Chem. Phys.* **128**, 144106 (2008).
- <sup>20</sup>H. J. Werner and F. R. Manby, *J. Chem. Phys.* **124**, 054114 (2006).
- <sup>21</sup>H.-J. Werner and F. R. Manby, *J. Chem. Phys.* **124**, 054114 (2006).
- <sup>22</sup>T. Hrenar, G. Rauhut, and H. J. Werner, *J. Phys. Chem. A* **110**, 2060 (2006).
- <sup>23</sup>J. G. Hill, J. A. Platt, and H. J. Werner, *Phys. Chem. Chem. Phys.* **8**, 4072 (2006).
- <sup>24</sup>M. Schütz, H. J. Werner, R. Lindh, and F. R. Manby, *J. Chem. Phys.* **121**, 737 (2004).
- <sup>25</sup>R. Polly, H. J. Werner, F. R. Manby, and P. J. Knowles, *Mol. Phys.* **102**, 2311 (2004).
- <sup>26</sup>T. Korona, K. Pflüger, and H. J. Werner, *Phys. Chem. Chem. Phys.* **6**, 2059 (2004).
- <sup>27</sup>H.-J. Werner, F. R. Manby, and P. J. Knowles, *J. Chem. Phys.* **118**, 8149 (2003).
- <sup>28</sup>G. Rauhut and H. J. Werner, *Phys. Chem. Chem. Phys.* **5**, 791 (2003).
- <sup>29</sup>M. Schütz and H. J. Werner, *J. Chem. Phys.* **114**, 661 (2001).
- <sup>30</sup>G. Rauhut and H. J. Werner, *Phys. Chem. Chem. Phys.* **3**, 4853 (2001).
- <sup>31</sup>M. Schütz and H. J. Werner, *Chem. Phys. Lett.* **318**, 370 (2000).
- <sup>32</sup>G. Hetzer, M. Schütz, H. Stoll, and H. J. Werner, *J. Chem. Phys.* **113**, 9443 (2000).
- <sup>33</sup>M. Schütz, G. Hetzer, and H. J. Werner, *J. Chem. Phys.* **111**, 5691 (1999).
- <sup>34</sup>G. Hetzer, P. Pulay, and H. J. Werner, *Chem. Phys. Lett.* **290**, 143 (1998).
- <sup>35</sup>A. El Azhary, G. Rauhut, P. Pulay, and H. J. Werner, *J. Chem. Phys.* **108**, 5185 (1998).
- <sup>36</sup>H. J. Werner and M. Schütz, *J. Chem. Phys.* **135**, 144116 (2011).
- <sup>37</sup>T. B. Adler and H. J. Werner, *J. Chem. Phys.* **135**, 144117 (2011).
- <sup>38</sup>T. Korona and H. J. Werner, *J. Chem. Phys.* **118**, 3006 (2003).
- <sup>39</sup>D. Kats, T. Korona, and M. Schütz, *J. Chem. Phys.* **127**, 064107 (2007).
- <sup>40</sup>D. Kats, T. Korona, and M. Schütz, *J. Chem. Phys.* **125**, 104106 (2006).
- <sup>41</sup>D. Kats and M. Schütz, *Z. Phys. Chem.* **224**, 601 (2010).
- <sup>42</sup>K. Freundorfer, D. Kats, T. Korona, and M. Schütz, *J. Chem. Phys.* **133**, 244110 (2010).
- <sup>43</sup>D. Kats and M. Schütz, *J. Chem. Phys.* **131**, 124117 (2009).
- <sup>44</sup>D. Usvyat, L. Maschio, C. Pisani, and M. Schütz, *Z. Phys. Chem.* **224**, 441 (2010).
- <sup>45</sup>C. Pisani, M. Schütz, S. Casassa, D. Usvyat, L. Maschio, M. Lorenz, and A. Erba, *Phys. Chem. Chem. Phys.* **14**, 7615 (2012).
- <sup>46</sup>J. Yang, G. K. L. Chan, F. R. Manby, M. Schütz, and H. J. Werner, *J. Chem. Phys.* **136**, 144105 (2012).
- <sup>47</sup>Y. Kurashige, J. Yang, G. K. L. Chan, and F. R. Manby, *J. Chem. Phys.* **136**, 124106 (2012).
- <sup>48</sup>C. Krause and H. J. Werner, *Phys. Chem. Chem. Phys.* **14**, 7591 (2012).
- <sup>49</sup>N. J. Russ and T. D. Crawford, *J. Chem. Phys.* **121**, 691 (2004).
- <sup>50</sup>N. J. Russ and T. D. Crawford, *Chem. Phys. Lett.* **400**, 104 (2004).
- <sup>51</sup>T. Nagata, D. G. Fedorov, H. Li, and K. Kitaura, *J. Chem. Phys.* **136**, 204112 (2012).
- <sup>52</sup>G. D. Fletcher, D. G. Fedorov, S. R. Pruitt, T. L. Windus, and M. S. Gordon, *J. Chem. Theo. Comp.* **8**, 75 (2012).
- <sup>53</sup>C. H. Choi and D. G. Fedorov, *Chem. Phys. Lett.* **543**, 159 (2012).
- <sup>54</sup>T. Nagata, D. G. Fedorov, K. Ishimura, and K. Kitaura, *J. Chem. Phys.* **135**, 044110 (2011).
- <sup>55</sup>D. G. Fedorov, K. Ishimura, T. Ishida, K. Kitaura, P. Pulay, and S. Nagase, *J. Comput. Chem.* **28**, 1476 (2007).
- <sup>56</sup>K. Doll, M. Dolg, P. Fulde, and H. Stoll, *Phys. Rev. B* **52**, 4842 (1995).
- <sup>57</sup>H. Stoll, *Ann. Phys.* **508**, 355 (1996).
- <sup>58</sup>H. Stoll, *J. Chem. Phys.* **97**, 8449 (1992).
- <sup>59</sup>J. Friedrich, M. Hanrath, and M. Dolg, *J. Chem. Phys.* **126**, 154110 (2007).
- <sup>60</sup>W. Li and P. Piecuch, *J. Phys. Chem. A* **114**, 8644 (2010).
- <sup>61</sup>W. Li, P. Piecuch, J. R. Gour, and S. H. Li, *J. Chem. Phys.* **131**, 114109 (2009).
- <sup>62</sup>W. Li, P. Piecuch, and J. R. Gour, in *Advances in the Theory of Atomic and Molecular Systems: Conceptual and Computational Advances in Quantum Chemistry*, edited by P. Piecuch *et al.* (Springer, Heidelberg, 2009), p. 131.
- <sup>63</sup>P. M. Kozłowski *et al.*, *J. Chem. Theory Comput.* **8**, 1870 (2012).
- <sup>64</sup>M. Pitonak, F. Holka, P. Neogady, and M. Urban, *J. Mol. Struct.: THEOCHEM* **768**, 79 (2006).
- <sup>65</sup>M. Pitonak, P. Neogady, V. Kello, and M. Urban, *Mol. Phys.* **104**, 2277 (2006).
- <sup>66</sup>N. Flocke and R. J. Bartlett, *J. Chem. Phys.* **118**, 5326 (2003).
- <sup>67</sup>N. Flocke and R. J. Bartlett, *Chem. Phys. Lett.* **367**, 80 (2003).
- <sup>68</sup>S. Hoyau, D. Maynau, and J. P. Malrieu, *J. Chem. Phys.* **134**, 054125 (2011).
- <sup>69</sup>D. Walter, A. Venkatnathan, and E. A. Carter, *J. Chem. Phys.* **118**, 8127 (2003).
- <sup>70</sup>D. Walter, A. B. Szilva, K. Niedfeldt, and E. A. Carter, *J. Chem. Phys.* **117**, 1982 (2002).
- <sup>71</sup>D. Walter and E. A. Carter, *Chem. Phys. Lett.* **346**, 177 (2001).
- <sup>72</sup>G. Reynolds and E. A. Carter, *Chem. Phys. Lett.* **265**, 660 (1997).
- <sup>73</sup>G. Reynolds, T. J. Martinez, and E. A. Carter, *J. Chem. Phys.* **105**, 6455 (1996).
- <sup>74</sup>J. E. Subotnik, A. Sodt, and M. Head-Gordon, *J. Chem. Phys.* **128**, 034103 (2008).
- <sup>75</sup>J. E. Subotnik, A. Sodt, and M. Head-Gordon, *J. Chem. Phys.* **125**, 074116 (2006).
- <sup>76</sup>J. E. Subotnik and M. Head-Gordon, *J. Chem. Phys.* **122**, 034109 (2005).
- <sup>77</sup>J. E. Subotnik and M. Head-Gordon, *J. Chem. Phys.* **123**, 064108 (2005).
- <sup>78</sup>P. E. Maslen, A. D. Dutoi, M. S. Lee, Y. H. Shao, and M. Head-Gordon, *Mol. Phys.* **103**, 425 (2005).
- <sup>79</sup>M. Head-Gordon, T. Van Voorhis, G. J. O. Beran, and B. Dunietz, in *Proceedings of the International Conference on Computational Science - ICCS 2003, Part IV* (Springer, Heidelberg, 2003), p. 96.
- <sup>80</sup>P. E. Maslen, M. S. Lee, and M. Head-Gordon, *Chem. Phys. Lett.* **319**, 205 (2000).
- <sup>81</sup>P. E. Maslen and M. Head-Gordon, *Chem. Phys. Lett.* **283**, 102 (1998).
- <sup>82</sup>P. E. Maslen and M. Head-Gordon, *J. Chem. Phys.* **109**, 7093 (1998).
- <sup>83</sup>E. G. Hohenstein and C. D. Sherrill, *J. Chem. Phys.* **133**, 104107 (2010).
- <sup>84</sup>P. Rosmus and W. Meyer, *J. Chem. Phys.* **69**, 2745 (1978).
- <sup>85</sup>P. Botschwina and W. Meyer, *Chem. Phys.* **20**, 43 (1977).
- <sup>86</sup>P. Botschwina and W. Meyer, *J. Chem. Phys.* **67**, 2390 (1977).
- <sup>87</sup>H. J. Werner and W. Meyer, *Mol. Phys.* **31**, 855 (1976).
- <sup>88</sup>W. Meyer and P. Rosmus, *J. Chem. Phys.* **63**, 2356 (1975).
- <sup>89</sup>W. Meyer, *Theor. Chim. Acta* **35**, 277 (1974).
- <sup>90</sup>W. Meyer, *J. Chem. Phys.* **58**, 1017 (1973).
- <sup>91</sup>W. Meyer, *Int. J. Quantum Chem.* **5**(S5), 341 (1971).
- <sup>92</sup>R. Ahlrichs, F. Driessler, H. Lischka, V. Staemmler, and W. Kutzelnigg, *J. Chem. Phys.* **62**, 1235 (1975).
- <sup>93</sup>P. R. Taylor, G. B. Bacskay, N. S. Hush, and A. C. Hurley, *Chem. Phys. Lett.* **41**, 444 (1976).
- <sup>94</sup>W. Meyer, in *Methods of Electronic Structure Theory*, edited by H. F. Schaefer III (Plenum, New York, 1977), p. 413.
- <sup>95</sup>C. Edmiston, *J. Chem. Phys.* **45**, 1833 (1966).
- <sup>96</sup>C. Edmiston and M. Krauss, *J. Chem. Phys.* **42**, 1119 (1965).
- <sup>97</sup>P. O. Löwdin, *Phys. Rev.* **97**, 1474 (1955).
- <sup>98</sup>R. Ahlrichs and F. Driessler, *Theor. Chim. Acta* **36**, 275 (1975).
- <sup>99</sup>P. R. Taylor, *J. Chem. Phys.* **74**, 1256 (1981).
- <sup>100</sup>R. Fink and V. Staemmler, *Theor. Chim. Acta* **87**, 129 (1993).
- <sup>101</sup>C. Kollmar and F. Neese, *J. Chem. Phys.* **135**(8), 084102 (2011).
- <sup>102</sup>L. M. J. Huntington, A. Hansen, F. Neese, and M. Nooijen, *J. Chem. Phys.* **136**, 064101 (2012).
- <sup>103</sup>D. G. Liakos, A. Hansen, and F. Neese, *J. Chem. Theory Comput.* **7**, 76 (2011).
- <sup>104</sup>A. Hansen, D. G. Liakos, and F. Neese, *J. Chem. Phys.* **135**, 214102 (2011).
- <sup>105</sup>F. Neese, F. Wennmohs, and A. Hansen, *J. Chem. Phys.* **130**, 114108 (2009).
- <sup>106</sup>F. Neese, A. Hansen, and D. G. Liakos, *J. Chem. Phys.* **131**, 064103 (2009).
- <sup>107</sup>L. M. J. Huntington and M. Nooijen, *J. Chem. Phys.* **133**, 184109 (2010).
- <sup>108</sup>F. Neese, *J. Comput. Chem.* **24**, 1740 (2003).
- <sup>109</sup>R. Izsak and F. Neese, *J. Chem. Phys.* **135**, 144105 (2011).
- <sup>110</sup>F. Neese, F. Wennmohs, A. Hansen, and U. Becker, *Chem. Phys.* **356**, 98 (2009).
- <sup>111</sup>R. Izsak, A. Hansen, and F. Neese, *Mol. Phys.* **110**, 2413 (2012).
- <sup>112</sup>J. Antony, S. Grimme, D. G. Liakos, and F. Neese, *J. Phys. Chem. A* **115**, 11210 (2011).
- <sup>113</sup>A. Anoop, W. Thiel, and F. Neese, *J. Chem. Theory Comput.* **6**, 3137 (2010).
- <sup>114</sup>D. G. Liakos and F. Neese, *J. Chem. Theory Comput.* **7**, 1511 (2011).
- <sup>115</sup>S. S. Zade, N. Zamoshchik, A. R. Reddy, G. Fridman-Marueli, D. Sheberla, and M. Bendikov, *J. Am. Chem. Soc.* **133**, 10803 (2011).
- <sup>116</sup>A. Kubas, S. Brase, and K. Fink, *Chem. Eur. J.* **18**, 8377 (2012).
- <sup>117</sup>M. Ashtari and N. M. Cann, *J. Chromatogr. A* **1218**, 6331 (2011).
- <sup>118</sup>C. Hättig, D. P. Tew, and B. Helmich, *J. Chem. Phys.* **136**, 204105 (2012).
- <sup>119</sup>D. P. Tew, B. Helmich, and C. Hättig, *J. Chem. Phys.* **135**, 074107 (2011).

- <sup>120</sup>B. Helmich and C. Hättig, *J. Chem. Phys.* **135**, 214106 (2011).
- <sup>121</sup>Z. Rolik and M. Kallay, *J. Chem. Phys.* **135**, 104111 (2011).
- <sup>122</sup>P. Pulay, S. Saebo, and W. Meyer, *J. Chem. Phys.* **81**, 1901 (1984).
- <sup>123</sup>G. E. Scuseria, C. L. Janssen, and H. F. Schaefer, *J. Chem. Phys.* **89**, 7382 (1988).
- <sup>124</sup>H.-J. Werner, in *Ab Initio Methods in Quantum Chemistry, Advances in Chemical Physics Vol. LXIX*, edited by K. P. Lawley (Wiley, New York, 1987), Vol. 2.
- <sup>125</sup>R. Ahlrichs, P. Scharf, and C. Ehrhardt, *J. Chem. Phys.* **82**, 890 (1985).
- <sup>126</sup>S. F. Boys, *Rev. Mod. Phys.* **32**, 296 (1960).
- <sup>127</sup>J. Pipek and P. G. Mezey, *J. Chem. Phys.* **90**, 4916 (1989).
- <sup>128</sup>O. Vahtras, J. Almlöf, and M. W. Feyereisen, *Chem. Phys. Lett.* **213**, 514 (1993).
- <sup>129</sup>L. S. Salmon, F. W. Birss, and K. Ruedenberg, *J. Chem. Phys.* **49**, 4293 (1968).
- <sup>130</sup>P. Jurecka, J. Sponer, J. Cerny, and P. Hobza, *Phys. Chem. Chem. Phys.* **8**, 1985 (2006).
- <sup>131</sup>P. Pulay, *J. Comput. Chem.* **3**, 556 (1982).
- <sup>132</sup>P. Pulay, *Chem. Phys. Lett.* **73**, 393 (1980).
- <sup>133</sup>G. E. Scuseria, *J. Phys. Chem. A* **103**, 4782 (1999).
- <sup>134</sup>C. Ochsenfeld, C. A. White, and M. Head-Gordon, *J. Chem. Phys.* **109**, 1663 (1998).
- <sup>135</sup>E. Schwegler, M. Challacombe, and M. Head-Gordon, *J. Chem. Phys.* **106**, 9708 (1997).
- <sup>136</sup>C. A. White, B. G. Johnson, P. M. W. Gill, and M. Head-Gordon, *Chem. Phys. Lett.* **253**, 268 (1996).
- <sup>137</sup>R. E. Stratmann, G. E. Scuseria, and M. J. Frisch, *Chem. Phys. Lett.* **257**, 213 (1996).
- <sup>138</sup>M. C. Strain, G. E. Scuseria, and M. J. Frisch, *Science* **271**, 51 (1996).
- <sup>139</sup>R. A. Mata and H. J. Werner, *J. Chem. Phys.* **125**, 184110 (2006).
- <sup>140</sup>F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.* **7**, 3297 (2005).
- <sup>141</sup>K. A. Peterson, T. B. Adler, and H. J. Werner, *J. Chem. Phys.* **128**, 084102 (2008).