PCCP

PAPER



Cite this: *Phys. Chem. Chem. Phys.*, 2014, **16**, 22167

Explicitly correlated PNO-MP2 and PNO-CCSD and their application to the S66 set and large molecular systems†

Gunnar Schmitz,*[a] Christof Hättig[a] and David P. Tew[b]

Received 6th August 2014,
Accepted 3rd September 2014

DOI: 10.1039/c4cp03502j

www.rsc.org/pccp

We present our current progress on the combination of explicit electron correlation with the pair natural orbital (PNO) representation. In particular we show cubic scaling PNO-MP2-F12, PNO-CCSD(2)$_{\overline{F12}}$ and PNO-CCSD[F12] implementations. The PNOs are constructed using a hybrid scheme, where the PNOs are generated in a truncated doubles space, spanned by orbital specific virtuals obtained using an iterative eigenvector algorithm. We demonstrate the performance of our implementation through calculations on a series of glycine chains. The accuracy of the local approximations is assessed using the S66 benchmark set, and we report for the first time explicitly correlated CCSD results for the whole set and improved estimates for the CCSD/CBS limits. For several dimers the PNO-CCSD[F12] calculations are more accurate than the current reference values. Additionally, we present pilot applications of our PNO-CCSD[F12] code to host–guest interactions in a cluster model for zeolite H-ZSM-5 and in a calix[4]arene–water complex.

1 Introduction

Although there has been an enormous development in computational technology, the investigation of large molecular systems using accurate wave function based methods is still a major challenge for theoretical chemistry. In this context, "large" refers to molecules with more than hundred atoms, which is still small in comparison to most proteins and enzymes. Density functional theory (DFT) is currently the method of choice for large molecules due to its low computational expense. However, the accuracy of DFT remains limited, particularly in the description of dispersion interactions and charge transfer excitations, and there is still no systematic route towards higher accuracy in the DFT framework. Wavefunction methods, in contrast, are systematically improvable and our goal is to develop low-cost, accurate alternatives to DFT to extend the arsenal of quantum chemical tools.

The applicability of correlated wavefunction methods to larger molecules is hampered by two problems: the steep scaling of the computational costs with the system size and the slow basis set convergence of the correlation energy. Even MP2, the simplest correlated wave function method, scales as $\mathcal{O}(N^5)$, where N is a measure for the system size. The "gold standard" of quantum chemistry CCSD(T) scales as $\mathcal{O}(N^7)$, which severely limits its applicability. The unfavorably high scaling is, however, unphysical and arises from the delocalized molecular orbital (MO) basis of the reference Hartree–Fock (HF) wave function. The heavy basis set requirements of wave function based methods, as compared to DFT, also hamper their applicability to large systems. This originates from the expansion of the many-electron wavefunction in Slater determinants built from one-electron functions (orbitals), which converges rather slowly in the regions of the electron cusps that are key features of the short-range correlation hole.

In the literature different approaches have been presented to address both issues. One route to a reduced scaling algorithm is to exploit the short-range nature of dynamic electron correlation in order to screen out negligible contributions and to truncate the wave function expansion. Ochsenfeld and coworkers have developed linear scaling MP2 implementations,[1,2] using an atomic orbital (AO) formulation of MP2, based on the work of Almlöf and Häser, together with refined integral estimates to screen negligible contributions at large interelectronic separations. An alternative route to linear scaling algorithms is to retain the MO formulation, but to localize the occupied orbitals through unitary transformation. In this localized basis, correlation of spatially distant electrons becomes negligible and a local virtual basis can be constructed using only virtuals spatially close to the occupied orbitals. Excitations from occupied to virtual orbitals are then restricted to these local domains. Several choices for sets of local virtual orbitals have been proposed: Schütz and Werner[3–5]

[a] *Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, D-44801 Bochum, Germany. E-mail: gunnar.schmitz@rub.de, christof.haettig@rub.de*
[b] *Center for Computational Chemistry, University of Bristol, Bristol BS8 1TS, UK*

† Electronic supplementary information (ESI) available. See DOI: 10.1039/c4cp03502j

successfully developed linear scaling algorithms based on the ansatz of Saebø and Pulay,[6] where non-orthogonal projected atomic orbitals (PAOs) are used as virtual orbitals. In combination with an *a priori* selection of atom domains for each local orbital one obtains a restricted set of virtuals. This PAO ansatz allows very efficient implementations, but has also a limited error control, since the PAO domains are predetermined, which leads to a biased treatment. In 2009 Neese and co-workers[7] have revived Meyer's pair natural orbitals (PNOs), where for every pair of electrons ij a basis of non-orthogonal virtuals is chosen by analyzing model amplitudes. The PNO representation is much more compact than the PAO representation and due to the analysis of model amplitudes the representation automatically adapts to the molecular system, which leads to a better error control. In fact the accuracy can be controlled by a single parameter, the PNO threshold T_{PNO}, which truncates the PNO expansion.[7] Helmich and Hättig[8] showed that a PNO representation can also be used in response theory for charge transfer and local excitations, which was problematic using the PAO approach. Yang *et al.*[9] realized that Pulay and Saebø's PAOs and Meyer's PNOs are two choices for a tensor factorization of the doubles amplitudes and introduced a third one, orbital specific virtuals (OSVs). In the order PAOs, OSVs and PNOs the representations become increasingly compact. PAOs are specific for a predetermined domain, OSVs are specific for an occupied orbital i and PNOs are specific for a given pair of occupied electrons ij. Later Krause *et al.*[10] and independently Hättig *et al.*[11] demonstrated that the OSV and PNO representation can be combined in a hybrid OSV–PNO approach to construct PNOs with quartic scaling costs. In this approach first the canonical virtual space is contracted into an OSV basis and then contracted into a PNO basis. As an improvement Schmitz *et al.*[12] developed a modified hybrid OSV–PNO algorithm, where the OSVs are generated iteratively and a shell based local density fitting (DF) scheme is applied. In this way the PNO generation scales only cubical with the system size. Riplinger *et al.*[13] also demonstrated how to combine the PAO and PNO approach in a hybrid way, where first virtuals are contracted in a large PAO basis and afterwards contracted in a PNO basis.

In addition to reducing the scaling with the system size, it is also beneficial to address the slow basis set convergence. Extrapolation techniques are often employed to accelerate basis set convergence towards the complete basis set (CBS) limit, and for correlation methods the most prominent form is the two-point X^{-3} extrapolation approach of Halkier *et al.*[14] Although the extrapolation strategy delivers satisfactory results, it does not improve the analytical quality of the underlying wave function. An alternative approach is the use of an explicitly correlated wave function ansatz in the correlation treatment, where the standard orbital basis expansion is augmented with pair-functions (geminals) that depend explicitly on the inter-electronic distance r_{ij}, resulting in a much improved description of the correlation hole. Modern F12 theory, which is an extension of the R12 methods of Kutzelnigg and Klopper,[15–17] is the most widely used explicitly correlated approach. Typically, a F12 calculation using a basis with cardinal number X is as accurate as a conventional calculation using a basis set with cardinal number $X + 2$, which means that reliable predictions can be made using DFT size basis sets. In our view, the combination of the hybrid OSV–PNO approach with F12 theory is an excellent route to low-scaling methods for accurate electronic structure predictions.

In this article, we build upon our earlier work, and present a revised implementation of PNO-MP2-F12 with reduced scaling and first implementations of the explicitly correlated coupled cluster models PNO-CCSD(2)$_{\overline{\mathrm{F12}}}$ and PNO-CCSD[F12]. The key aspects of our approach is the use of a hybrid OSV–PNO scheme, with an iterative generation of OSVs and the use of local density fitting (DF) for 4 index integrals, where orbital transformations are performed directly from the AO to the PNO basis. We present results of a series of benchmark calculations that reveal that our methods are accurate and have an early break even point with their conventional counterparts. We also demonstrate the utility of the current code through pilot applications, evaluating the interaction energies between weekly bound complexes. In particular we calculate interaction energies for the S66 set of weak molecular interaction reported by Hobza and co-workers.[38] The current CBS reference values for the S66 set are used to judge the accuracy of various methods and to fit empirical factors and corrections, for example those in ref. 43 and 44. We present revised reference values with significantly reduced basis set incompleteness errors.

2 Theory

2.1 Hybrid OSV–PNO approach in F12 theory

We have previously reported a hybrid OSV–PNO approach for MP2 theory without explicit correlation, using an iterative generation of OSVs in an $\mathcal{O}(N^3)$ scaling algorithm.[12] We now extend this technique to F12 theory, where the multiple sets of PNOs arising from the strong orthogonality projector require the construction of several different orbital-specific orbital sets.[11,18] Each set of PNOs, $d^{ij}_{p\bar{p}}$, and OSVs, $d^{ii}_{p\bar{p}}$, is an eigenvector of a corresponding pseudo density D_{pq}. In the following, PNOs are denoted with a bar, while OSVs and pre-PNOs are denoted with a tilde.

The appropriate PNOs for the virtual space (V-PNOs) in F12 theory are those that provide a compact representation of the double excitations that cannot be captured by the geminal basis. In ref. 11 it has been demonstrated that the eigenvectors of the pseudo (difference) density, built from difference amplitudes Δ^{ij}_{ab}, are a suitable choice for this purpose. Using a spin–orbital formalism,

$$\Delta^{ij}_{ab} = t^{ij}_{ab} - r^{ij}_{ab}, \tag{1}$$

where r^{ij}_{ab} is a short hand notation for $\langle ab|f_{12}|\widetilde{ij}\rangle$. The tilde in $|\widetilde{ij}\rangle$ indicates contraction with the geminal coefficients from the fixed amplitude (SP) approach[19] $|\widetilde{ij}\rangle = \hat{P}_{ij}\left(\frac{3}{8} + \frac{1}{8}\hat{S}_{ij}\right)|\widetilde{ij}\rangle$, where $\hat{S}_{ij}$ permutes the spatial components and $\hat{P}_{ij}$ is an antisymmetrisation operator defined by its action $\hat{P}_{pq}X_{pq} = X_{pq} - X_{qp}$. In eqn (1) the second term r^{ij}_{ab}, approximately removes from the MP2 doubles amplitudes t^{ij}_{ab} the contributions that can be spanned

 This journal is © the Owner Societies 2014

by the geminals. PNO-F12 theory requires additional auxiliary PNOs to express the strong orthogonality projectors ${}^{A}Q$ and ${}^{B}Q$ in a pair specific basis. In ref. 18, Tew and Hättig discussed various choices, which form the basis of our current implementation. For ${}^{B}\hat{Q}_{12}$ the recommendation is unchanged from our earlier work

$$ {}^{B}\hat{Q}_{12} = \bar{V}_1\bar{V}_2' + \bar{V}_1'\bar{V}_2 + \bar{V}_1'\bar{V}_2', \tag{2} $$

where for the sake of notational clarity the pair index ij is omitted. For ${}^{A}\hat{Q}_{12}$, we use approximation ${}^{A_3}\hat{Q}_{12}$ in ref. 18, which provides the most compact representation of the occupied–occupied and occupied-virtual product spaces and the number of auxiliary PNOs per pair quickly becomes independent of the system size,

$$ {}^{A_3}\hat{Q}_{12} = 1 - \bar{\bar{O}}_1\bar{\bar{O}}_2 - \bar{V}_1\bar{V}_2 - \bar{O}_1\bar{V}_2'' - \bar{V}_1''\bar{O}_2. \tag{3} $$

This is in contrast to our earlier work[18] where the $\mathcal{O}(N^4)$ construction of the densities and PNOs from the full MO basis were the dominant expense and the overhead of building additional PNO sets was not worthwhile. In the current hybrid OSV–PNO approach, use of ${}^{A_3}\hat{Q}_{12}$ is required to eliminate all $\mathcal{O}(N^4)$ steps. Above, the symbols V, V', V'' and O denote the virtual, the complementary virtual (CA), the union of virtual and CA and the occupied space, respectively.

For each term in the projectors, different types of auxiliary PNOs (X-PNOs) are required. The pseudo densities for each PNO generation are constructed as[18]

$$ \text{(OPNOs 1)} \quad \bar{\bar{K}}_{ij}\text{: } D^{ij}_{KL} = \sum_M r^{ij}_{KM} r^{ij}_{LM} \tag{4} $$

$$ \text{(OPNOs 2)} \quad \bar{K}_{ij}\text{: } D^{ij}_{KL} = \sum_{c''} r^{ij}_{Kc''} r^{ij}_{Lc''} \tag{5} $$

$$ \text{(CA-PNOs 1)} \quad \bar{a}_{ij}{}''\text{: } D^{ij}_{a''b''} = \sum_M r^{ij}_{a''M} r^{ij}_{b''M} \tag{6} $$

$$ \text{(CA-PNOs 2)} \quad \bar{a}_{ij}{}'\text{: } D^{ij}_{a'_{ij}b'_{ij}} = \sum_{\bar{c}_{ij}} r^{ij}_{a'_{ij}\bar{c}_{ij}} r^{ij}_{b'_{ij}\bar{c}_{ij}} \tag{7} $$

Both $\bar{\bar{K}}_{ij}$ and $\bar{K}_{ij}$ orbitals span the occupied orbital space, but $\bar{K}_{ij}$ occurs in the projector $\bar{O}\bar{V}''$, together with the $\bar{a}_{ij}{}''$ orbitals, which span the union of CA and virtual orbitals. This is accounted for in the PNO construction by using a density computed from integrals with one index in the occupied space and one in the CA plus virtual space, which is denoted by a double prime a''. Also the $\bar{a}'_{ij}$ orbitals span the union of CA and virtual orbitals, but include only the virtuals, which are discarded in the V-PNO representation *i.e.* they span the subspace of the union of CA and virtual space that is orthogonal to the selected V-PNOs. In this way, screened out virtuals are included in an extended CA space. The V-PNOs and X-PNOs in F12 theory can be computed in a pair specific pre-PNO basis in a similar manner to that used for conventional V-PNOs. The pre-PNO basis for pair ij is constructed by merging orbital-specific auxiliary orbital (OSX) sets for i and j. The OSVs and OSXs are analogues of their PNO counterparts and defined by the pseudo densities they diagonalize

$$ \text{(OSVs)} \quad \tilde{a}_i\text{: } D^{ii}_{ab} = \sum_c \Delta^{ii}_{ac}\Delta^{ii}_{bc} \tag{8} $$

$$ \text{(OSOs 1)} \quad \tilde{\bar{K}}_i\text{: } D^{ii}_{KL} = \sum_M r^{ii}_{KM} r^{ii}_{LM} \tag{9} $$

$$ \text{(OSOs 2)} \quad \tilde{K}_i\text{: } D^{ii}_{KL} = \sum_{p''} r^{ii}_{Kp''} r^{ii}_{Lp''} \tag{10} $$

$$ \text{(CAOSVs 1)} \quad \tilde{\bar{a}}_i{}''\text{: } D^{ii}_{a''b''} = \sum_{p''} r^{ii}_{a''p''} r^{ii}_{b''p''} \tag{11} $$

$$ \text{(CAOSVs 2)} \quad \tilde{a}_i{}''\text{: } D^{ii}_{a''b''} = \sum_{c''} r^{ii}_{a''c''} r^{ii}_{b''c''}, \tag{12} $$

where a'', b'' run over the virtual and CA orbitals, while p'' runs over the union of HF and CA basis. This large contraction space is needed to ensure an accurate approximation of r^{ij}_{KL}, $r^{ij}_{Kc''}$ and $r^{ij}_{a'b}$ in the OSX basis.[11] The occupation number thresholds for discarding unimportant X-PNOs and OSXs in F12 theory have been determined from ref. 18 and linked to the primary V-PNO occupation number threshold such that the truncation error in the auxiliary sets is always an order of magnitude smaller than the error from truncating the virtual set.

2.1.1 Iterative generation of OSVs and OSXs. As shown in ref. 12 a local density fitting scheme (DF) in combination with an iterative OSV generation that uses a modified block Davidson algorithm can be applied to obtain PNOs with quadratically scaling costs. In this approach Almlöfs "Laplace trick"[20] is used to factorize the orbital energy denominator by replacing it in the approximate MP2 doubles amplitudes by an equivalent Laplace transformed expression, which is evaluated numerically using n_L grid points and weights ω_z. Each quadrature point is then a factorisable expression, and the matrix vector product σ^k_a of the doubles amplitudes t^{ii}_{ac} with an arbitrary vector b^k_c can be written as

$$ \begin{aligned} \sigma^k_a &= \sum_c t^{ii}_{ac} b^k_c = -\sum_z^{n_L} \omega_z \sum_\nu C^z_{\nu a} \sum_{Q_i} (Q_i|\nu i^z) \\ &\quad \times \sum_{P_i} \left[{}^{(ii)}V^{-1}\right]_{Q_iP_i} \sum_\mu (P_i|\mu i^z) \sum_c C^z_{\mu c} b^k_c. \end{aligned} \tag{13} $$

The evaluation of arbitrary matrix vector products is needed for nearly all iterative diagonalization algorithms. In the expression P_i denotes an auxiliary function in a local DF basis, $(Q_i|\nu i)$ are 3 index DF-integrals with one index in the occupied and one index in the AO basis, V is defined as $V = (P|Q)$, $C^z_{\mu c}$ refers to grid point depended LMO coefficients $C^z_{\mu a} = C_{\mu a}e^{-\varepsilon_a t_z}$ and i^z to LMOs, which depend on the grid points $|i^z\rangle = \sum_j T_{ij}e^{\varepsilon_j t_z}|j\rangle$. Each of the matrix vector products in eqn (13) can be evaluated with operation counts that scale at most quadratically with the system size. The presented iterative OSV generation can be adapted for OSVs and OSXs in F12 theory. For OSXs with two

indices in the same orbital space (OSOs 1 and CAOSVs 2) one arrives with generalized dimensions p and q at

$$\begin{aligned}\sigma_a^k &= \sum_q r_{pq}^{ii} b_q^k \\ &= -\sum_\nu C_{\nu p} \sum_{Q_i} (Q_i|f_{12}|\nu i) \\ &\quad \times \sum_{P_i} \left[{}^{(ii)}V_{f_{12}}^{-1}\right]_{Q_i P_i} \sum_\mu (P_i|f_{12}|\mu i) \sum_q C_{\mu q} b_q^k,\end{aligned} \tag{14}$$

where now 3 index integrals over the correlation factor and its metric (for V) are used. The above expressions are appropriate for the iterative diagonalisation of the amplitudes, which is equivalent to diagonalising the pseudodensity for OSVs, OSOs 1 and CAOSVs 2. For OSOs 2 and CAOSVs 1, the amplitudes are not symmetric and the appropriate expressions for the iterative diagonalization of the pair density has to be written as

$$\begin{aligned}\sigma_p^k &= \sum_q D_{pq}^{ii} b_q^k \\ &= -\sum_{\nu^*} C_{\nu^* p} \sum_{Q_i} (Q_i|f_{12}|\nu^* i) \\ &\quad \times \sum_{P_i} \left[{}^{(ii)}V_{f_{12}}^{-1}\right]_{Q_i P_i} \sum_{\mu''} (P_i|f_{12}|\mu'' i) \sum_{b''} C_{\mu'' b''} \\ &\quad \times \sum_{\kappa''} C_{\kappa'' b''} \sum_{R_i} (R_i|f_{12}|\kappa'' i) \sum_{S_i} \left[{}^{(ii)}V_{f_{12}}^{-1}\right]_{R_i S_i} \\ &\quad \times \sum_{\lambda^*} (S_i|f_{12}|\lambda^* i) \sum_q C_{\lambda^* q} b_q^k,\end{aligned} \tag{15}$$

where the superscript * is used to denote that the AO basis (OSOs 1) as well as the AO + CA basis (OSOs 2 and CAOSVs 1 + 2) are used for different OSX types. For a fast convergence in the CAOSV generation the algorithm is started with the conventional OSVs of the same LMO augmented by a block of zeros for the CABS space. For the OSOs 1 + 2 the same strategy as in ref. 12 is used as start guess. As for CAOSVs 1 + 2 the conventional OSVs, which are required for subsidiary screening steps, are used as start guess for the (F12-)OSV construction.

2.2 Pre-PNO generation

The details of our algorithm for constructing the pre-PNOs from OSVs and the subsequent PNO formation have been discussed elsewhere.[11,12] Here we give a brief overview and highlight the deviations from our earlier work. After first obtaining OSVs and OSXs using the iterative AO-based scheme, the pair list is truncated using SOS–MP2 pair energy estimates in the conventional OSV basis, where we now exploit the local DF scheme.

For all surviving pairs ij a pre-PNO basis is constructed by merging and orthogonalizing the OSV sets of i and j. Linear dependencies in this merged basis are removed. During the pre-PNO generation (non-unitary) transformation matrices from the OSV/OSX to the pre-PNO basis are constructed and saved on disk. This avoids the need for saving the pre-PNO coefficients in the canonical or AO basis. If they are needed they are constructed on the fly. With a proper algorithm design this leads to a reduced I/O. The pseudo densities for the PNO generation are then built in the reduced doubles space spanned by the pre-PNOs.

2.3 Integral evaluation

All 4-index integrals are built from local density fitting (DF) 3-index intermediates. To reduce the costs for the evaluation of 3-index precursors, integral screening techniques are introduced, which exploit the short range nature of the operators $f_{12}, f_{12}g_{12}, f_{12}^2$ and $f_{12}^2 r_{12}^2$. Our screening algorithm follows the work of Adler *et al.*,[21] but introduces a prescreening step with a cheaper estimate, which only uses the most diffuse basis function in the contracted GTO. The bounds for these kinds of integrals can be found in the ESI.†

2.3.1 4-index integral construction. All 4-index integrals required for the OSV and PNO generation are built using local DF from 3-index precursors with an index in the AO basis and one occupied index. The integrals over f_{12}, $f_{12}g_{12}$, f_{12}^2 are positive (or negative) definite and density fitting proceeds using the natural fit metric. Manby's robust density fitting formulation[22] is used for the $f_{12}^2 r_{12}^2$ integrals since they are not positive or negative definite in our implementation.[23] The procedure for standard DF is shown in pseudocode 1 and for robust DF in 2. In both cases the 3-index AO integrals are directly transformed to the pre-PNO basis. The required pre-PNO coefficients in the AO basis are obtained by contracting the OSV to pre-PNO transformation matrices with the AO coefficients for the OSVs. Then a local DF matrix ${}^{(ij)}V_{P_{ij}Q_{ij}}^{(\hat{o}_{12})}$ is constructed for the operator $\hat{o}_{12}$. Since the number of pre-PNOs is small compared to the dimension of the pair-specific fitting basis sets it is more efficient to solve the linear equations for the local DF intermediates *via* Cholesky decomposition of ${}^{(ij)}V_{P_{ij}Q_{ij}}^{(\hat{o}_{12})}$ instead of building the inverse of this matrix.

2.4 Explicitly correlated PNO-CCSD

Our PNO implementation of the standard CCSD working equations follows that of Neese *et al.*[24] and we introduce the same approximations with one exception: we do not neglect the contribution to the doubles residual, which involves the term $\sum_k J_{\bar{a}\bar{c}}^{k\bar{d}} t_{\bar{b}}^k$. The implementation of the terms in the singles residual is only slightly changed from that reported for PNO-CC2 in ref. 25 and 26; we no longer use MO precursors but compute all PNO target integrals directly from integrals in the AO basis. In the following, we give the details of the additional terms required for PNO-CCSD(2)$_{\overline{\text{F12}}}$[27] and PNO-CCSD[F12].[28] The choice of intermediates follows that of ref. 28. Several aspects of our implementation have already been reported and are not repeated here. In particular, terms common to PNO-MP2-F12 and PNO-MP3(F12) have been reported in ref. 11 and 18. For the PNO-MP2-F12 intermediates, our current implementation differs slightly from ref. 11 and 18 in that ${}^{A_3}\hat{Q}_{12}$ has replaced ${}^{A_2}\hat{Q}_{12}$ and the AO based local DF algorithms 1 and 2 are used to construct the 4-index integrals in the PNO basis.

 This journal is © the Owner Societies 2014

Algorithm 1 Local DF scheme to build 4-index integrals in a pre-PNO basis $X^{ij}_{\tilde{r}_{ij}\tilde{s}_{ij}} = \langle \tilde{r}_{ij}\tilde{s}_{ij}|\hat{o}_{12}|ij\rangle$

$C_{\mu^*\tilde{r}_{ij}} = \sum_r C_{\mu^* r} d_{r\tilde{r}_{ij}}$

$C_{\nu^\times\tilde{s}_{ij}} = \sum_s C_{\nu^\times s} d_{s\tilde{s}_{ij}}$

Get sub-matrix $V^{(\hat{o}_{12})}_{P_{ij}Q_{ij}}$ of full matrix $V^{(\hat{o}_{12})}_{PQ} = (P|\hat{o}_{12}|Q)$

Get 3-index integrals $(Q_{ij}|\hat{o}_{12}|\mu^\star i)$ and $(Q_{ij}|\hat{o}_{12}|\nu^\times j)$

Transform AO-index to *e.g.* pre-PNO basis:

$\left(Q_{ij}|\hat{o}_{12}|\tilde{r}_{ij}i\right) = \sum_{\mu^*} \left(Q_{ij}|\hat{o}_{12}|\mu^* i\right) C_{\mu^*\tilde{r}_{ij}}$

$\left(Q_{ij}|\hat{o}_{12}|\tilde{s}_{ij}j\right) = \sum_{\nu^\times} \left(Q_{ij}|\hat{o}_{12}|\nu^\times j\right) C_{\nu^\times\tilde{s}_{ij}}$

Cholesky decomposition: $V^{(\hat{o}_{12})}_{P_{ij}Q_{ij}} = \sum_{R_{ij}} I^{(\hat{o}_{12})}_{P_{ij}R_{ij}} I^{(\hat{o}_{12})}_{Q_{ij}R_{ij}}$

Solve $\sum_{Q_{ij}} V^{(\hat{o}_{12})}_{P_{ij}Q_{ij}} B_{Q_{ij}\tilde{r}_{ij}i} = \left(P_{ij}|\hat{o}_{12}|\tilde{r}_{ij}i\right)$ using $I^{(\hat{o}_{12})}_{P_{ij}R_{ij}}$

Calculate 4-index integrals in pre-PNO basis:

$X^{ij}_{\tilde{r}_{ij}\tilde{s}_{ij}} = \sum_{Q_{ij}} B_{Q_{ij}\tilde{r}_{ij}i} \left(Q_{ij}|\hat{o}_{12}|\tilde{s}_{ij}j\right)$

Algorithm 2 Local robust DF scheme to build 4-index integrals in a pre-PNO basis $X^{ij}_{\tilde{r}_{ij}\tilde{s}_{ij}} = \langle \tilde{r}_{ij}\tilde{s}_{ij}|\hat{o}_{12}|ij\rangle$

$C_{\mu^*\tilde{r}_{ij}} = \sum_r C_{\mu^* r} d_{r\tilde{r}_{ij}}$

$C_{\nu^\times\tilde{s}_{ij}} = \sum_s C_{\nu^\times s} d_{s\tilde{s}_{ij}}$

Get sub-matrix $V^{(\hat{g}_{12})}_{P_{ij}Q_{ij}}$ of full matrix $V^{(\hat{g}_{12})}_{PQ} = (P|g_{12}|Q)$

Get sub-matrix $V^{(\hat{o}_{12})}_{P_{ij}Q_{ij}}$ of full matrix $V^{(\hat{o}_{12})}_{PQ} = (P|\hat{o}_{12}|Q)$

Get integrals $(Q_{ij}|\hat{o}_{12}|\mu^\star i)$, $(Q_{ij}|\hat{o}_{12}|\nu^\times j)$, $(Q_{ij}|g_{12}|\mu^\star i)$ and $(Q_{ij}|g_{12}|\nu^\times j)$ and transform to pre-PNO basis as 1

Cholesky decomposition: $V^{(\hat{g}_{12})}_{P_{ij}Q_{ij}} = \sum_{R_{ij}} I^{(\hat{g}_{12})}_{P_{ij}R_{ij}} I^{(\hat{g}_{12})}_{Q_{ij}R_{ij}}$

Solve $\sum_{Q_{ij}} V^{(\hat{g}_{12})}_{P_{ij}Q_{ij}} G_{Q_{ij}\tilde{r}_{ij}i} = \left(P_{ij}|g_{12}|\tilde{r}_{ij}i\right)$ using $I^{(\hat{g}_{12})}_{P_{ij}R_{ij}}$

Solve $\sum_{Q_{ij}} V^{(\hat{g}_{12})}_{P_{ij}Q_{ij}} G_{Q_{ij}\tilde{s}_{ij}j} = \left(P_{ij}|g_{12}|\tilde{s}_{ij}j\right)$ using $I^{(\hat{g}_{12})}_{P_{ij}R_{ij}}$

Construct robust DF intermediates:

$B^{(\hat{o}_{12})}_{Q_{ij}\tilde{r}_{ij}i} = \left(Q_{ij}|\hat{o}_{12}|\tilde{r}_{ij}i\right) - \frac{1}{2}\sum_{P_{ij}} V^{(\hat{o}_{12})}_{Q_{ij}P_{ij}} G_{P_{ij}\tilde{r}_{ij}i}$

$B^{(\hat{o}_{12})}_{Q_{ij}\tilde{s}_{ij}j} = \left(Q_{ij}|\hat{o}_{12}|\tilde{s}_{ij}j\right) - \frac{1}{2}\sum_{P_{ij}} V^{(\hat{o}_{12})}_{Q_{ij}P_{ij}} G_{P_{ij}\tilde{s}_{ij}j}$

Calculate 4-index integrals in pre-PNO basis:

$X^{ij}_{\tilde{r}_{ij}\tilde{s}_{ij}} = \sum_{Q_{ij}} B^{(\hat{o}_{12})}_{Q_{ij}\tilde{r}_{ij}i} B^{(\hat{o}_{12})}_{Q_{ij}\tilde{s}_{ij}j}$

2.4.1 PNO-CCSD(2)$_{\overline{F12}}$. In the CCSD(2)$_{\overline{F12}}$ approximation the CCSD equations are solved without F12 contributions, but the energy is evaluated with a modified Lagrangian, which is obtained from the conventional CCSD Lagrangian plus the most important explicitly correlated contributions:[27,28]

$$L_{(2)\overline{F12}} = E_{\mathrm{CCSD}} + \Delta E^{\mathrm{unc}}_{\mathrm{MP2-F12}} + 2\sum_{\bar{a}>\bar{b},i>j}\left(\mathscr{C}^{ij}_{\bar{a}\bar{b}} + \mathscr{V}^{ij}_{\bar{a}\bar{b}}\right) t^{ij}_{\bar{a}\bar{b}} \tag{16}$$

Here $\Delta E^{\mathrm{unc}}_{\mathrm{MP2-F12}}$ is the explicitly correlated part of the MP2-F12 energy without coupling to $t^{ij}_{\bar{a}\bar{b}}$. The coupling is included *via* $\mathscr{C}^{ij}_{\bar{a}\bar{b}}$ and $\mathscr{V}^{ij}_{\bar{a}\bar{b}}$. The matrix elements $\mathscr{C}^{ij}_{\bar{a}\bar{b}} = \langle\bar{a}\bar{b}|\left(\hat{F}_1+\hat{F}_2\right)^{\mathrm{B}}\hat{Q}_{12}f_{12}|\widetilde{ij}\rangle$ are available from PNO-MP2-F12, while $\mathscr{V}^{ij}_{\bar{a}\bar{b}}$ is an additional intermediate, given in terms of PNOs and X-PNOs as

$$\mathscr{V}^{ij}_{\bar{a}\bar{b}} = v^{ij}_{\bar{a}\bar{b}} - \sum_{\bar{K}>\bar{L}} r^{ij}_{\bar{K}\bar{L}} g^{\bar{K}\bar{L}}_{\bar{a}\bar{b}} - \sum_{\bar{L}\bar{c}''} r^{ij}_{\bar{c}''\bar{L}} g^{\bar{c}''L}_{\bar{a}\bar{b}} - \sum_{\bar{c}>\bar{d}} r^{ij}_{\bar{c}\bar{d}} g^{\bar{c}\bar{d}}_{\bar{a}\bar{b}}, \tag{17}$$

with the usual definition of intermediates $r^{ij}_{\bar{p}\bar{q}} = \langle\bar{p}\bar{q}|f_{12}|\widetilde{ij}\rangle$, $g^{\overline{rs}}_{\bar{p}\bar{q}} = \langle\bar{p}\bar{q}||\overline{rs}\rangle$, $v^{ij}_{\bar{p}\bar{q}} = \langle\bar{p}\bar{q}|f_{12}g_{12}|\widetilde{ij}\rangle$. The intermediate $\mathscr{V}^{ij}_{\bar{a}\bar{b}}$ also appears in PNO-MP3(F12) and we refer the reader to ref. 11 for implementation details.

2.4.2 PNO-CCSD[F12]. For the PNO-CCSD[F12] model the leading additional terms to the singles and doubles residuals are added to the PNO-CCSD residuals during the solution of the cluster equations:

$$\Omega^i_{a,[\mathrm{F12}]} = \Omega^i_{a,\mathrm{CCSD}} + \mathscr{C}^i_a + \mathscr{V}^i_a + \mathscr{U}^i_a \tag{18}$$

$$\Omega^{ij}_{\bar{a}\bar{b},[\mathrm{F12}]} = \Omega^{ij}_{\bar{a}\bar{b},\mathrm{CCSD}} + \mathscr{C}^{ij}_{\bar{a}\bar{b}} + \mathscr{V}^{ij}_{\bar{a}\bar{b}} + \mathscr{U}^{ij}_{\bar{a}\bar{b}} \tag{19}$$

Starting from the PNO-CCSD Lagrangian the expression for the PNO-CCSD[F12] Lagrangian becomes[28]

$$L_{[\mathrm{F12}]} = E_{\mathrm{CCSD}} + \Delta E^{\mathrm{unc}}_{\mathrm{MP2-F12}} + \sum_{ai}\left(\mathscr{V}^i_a + \mathscr{U}^i_a\right) t^i_a + \sum_{\bar{a}>\bar{b},i>j}\left(\mathscr{C}^{ij}_{\bar{a}\bar{b}} + \mathscr{V}^{ij}_{\bar{a}\bar{b}} + \mathscr{U}^{ij}_{\bar{a}\bar{b}}\right) t^{ij}_{\bar{a}\bar{b}}, \tag{20}$$

where E_{CCSD} is the usual PNO-CCSD energy expression evaluated with the amplitudes obtained as solutions of eqn (18) and (19). Using V-PNOs and X-PNOs for a size independent strong orthogonality operator, the intermediates from F12 theory $\mathscr{C}^i_a$, $\mathscr{U}^i_a$, $\mathscr{V}^i_a$, $\mathscr{U}^{ij}_{\bar{a}\bar{b}}$ become

$$\mathscr{C}^i_a = \sum_j \sum_{\bar{a}} d^{ij}_{a\bar{a}} \hat{P}_{ij} \sum_{\bar{c}'} r^{ij}_{\bar{a}\bar{c}'} \left(\sum_{c'} F_{jc'} d^{ij}_{c'\bar{c}'}\right), \tag{21}$$

$$\mathscr{U}^i_a = -\sum_{kl}\sum_{\bar{a}} d^{kl}_{a\bar{a}} \sum_{\bar{d}'} r^{kl}_{\bar{a}\bar{d}'} g^{lk}_{\bar{d}'i}, \tag{22}$$

$$\mathscr{V}^i_a = \sum_l \mathscr{V}^{il}_{al} = \sum_l \left[v^{il}_{al} - \sum_{\bar{K}>\bar{L}} r^{il}_{\bar{K}\bar{L}} g^{\bar{K}\bar{L}}_{al} - \sum_{\bar{L}\bar{c}''} r^{il}_{\bar{c}''\bar{L}} g^{\bar{c}''\bar{L}}_{al} - \sum_{\bar{c}>\bar{d}} r^{il}_{\bar{c}\bar{d}} g^{\bar{c}\bar{d}}_{al} \right], \tag{23}$$

and

$$\mathscr{W}^{ij}_{\bar{a}\bar{b}} = \hat{P}_{\bar{a}\bar{b}}\hat{P}_{ij}\sum_{k\bar{c}'}\left(K^{ik}_{\bar{a}_{ij}\bar{c}'_{jk}} - J^{ik}_{\bar{a}_{ij}\bar{c}'_{jk}}\right) r_{\bar{c}'_{kj}\bar{b}_{kj}}{}^{kj} S^{kj,ij}_{\bar{b}_{kj}\bar{b}_{ij}}, \tag{24}$$

The terms $\mathscr{V}^{ij}_{\bar{a}\bar{b}}$, $\mathscr{W}^{ij}_{\bar{a}\bar{b}}$ and $\mathscr{C}^{ij}_{\bar{a}\bar{b}}$ are common to PNO-MP3(F12). Their implementation has been reported in ref. 11. Implementation of the additional contributions $\mathscr{C}^{i}_{a}$, $\mathscr{W}^{i}_{a}$ and $\mathscr{V}^{i}_{a}$ required only minor changes and generalizations to the existing routines for PNO-CC2: $\mathscr{C}^{i}_{a}$ needed a generalization of Ω^{I}_{ai} to treat a Fock matrix in the CA space and F12 amplitudes in the V-PNO and CAPNO-2 basis. For $\mathscr{W}^{i}_{a}$ the same strategy as that for Ω^{II}_{ai} can be used, but with generalization to the V-PNO and CAPNO-2 index spaces. The intermediate $\mathscr{V}^{i}_{a}$ is more involved. The first term v^{il}_{al} is precomputed using DF techniques and the remaining terms are constructed using the following scheme, where $\bar{p}$ and $\bar{q}$ denote generalized PNO spaces (occupied, virtual, CA):

$$C_{Q_{i_u},\bar{q}j} = \sum_{P_{i_u}} (j\bar{q}|P_{i_u}) \left[{}^{i_u}V^{-1}\right]_{P_{i_u}Q_{i_u}} \tag{25}$$

$$Y_{Q_{i_u},\bar{p}ij} = \sum_{\bar{q}} r^{ij}_{\bar{p}\bar{q}} C_{Q_{i_u},\bar{q}j} \tag{26}$$

$$\Gamma_{Q_{i_u},\nu^* i} = \sum_{j\in\mathscr{J}(i)}\sum_{\bar{p}} Y_{Q_{i_u},\bar{p}ij} C_{\nu^*\bar{p}} \tag{27}$$

$$\mathscr{V}^{i}_{\mu} += \sum_{Q_{i_u},\nu^*} \Gamma_{Q_{i_u},\nu^* i}(Q_{i_u}|\mu\nu^*) \tag{28}$$

The index i_u indicates the unified DF basis,[25] and the star denotes that for the contributions of $p = c''$ and $q = \bar{L}$, the union of the AO basis and AO-CABS basis is used.

3 Computational details

All calculations were performed using a development version of the TURBOMOLE package.[29] For canonical DF-MP2-F12, CCSD(2)$_{\overline{F12}}$ and CCSD[F12] calculations the ricc2[28,30,31] module was employed, whereas the local development program pnoccsd was used for PNO-MP2-F12, PNO-CCSD(2)$_{\overline{F12}}$ and PNO-CCSD[F12] calculations. For the orbital basis the cc-pVDZ-F12,[32] cc-pVDZ, aug-cc-pVDZ and aug-cc-pVTZ[33] were chosen, which will be abbreviated as DZ-F12, DZ, aDZ and aTZ. For DF and RI, the appropriate auxiliary basis sets were taken from ref. 34 and 35. In one case we also used the aug-cc-pwCVDZ basis[36] in combination with the auxiliary basis from ref. 37 for the DF and RI, which will be abbreviated as wCDZ.

4 Results and discussion

4.1 Impact of the integral screening

Integral screening was tested using a set of molecules composed of an adenine–thymine base pair, cyclohexane, dibutylphenol, norbornane, $(Gly)_1$, $(Gly)_2$, $(Gly)_4$ and a $(GaAs)_4$ cluster as an example for an inorganic molecule with occupied d-shells. For these molecules, four sets of PNO-MP2-F12 calculations were performed to determine the errors due to F12 screening Δ_{F12} compared to standard Schwarz screening Δ_{con}. Δ_{F12} was computed as the difference between energies with and without F12 screening, but with default levels of Schwarz screening. Δ_{con} was computed as the difference between energies evaluated using the default threshold for the convergence of the density to 10^{-7} and the tighter threshold of 10^{-13}, and with no F12 screening. The default threshold for F12 screening is the same as that for Schwarz screening. The results are shown in Table 1 and reveal that the error due to the Schwarz screening is in most cases one order of magnitude larger than the error due to the F12 screening. There are some exceptions where this trend flips, but nevertheless the errors due to the screening are at least in the range of 10^{-6}% and therefore negligible at this level of convergence for the one-electron density, *i.e.* the HF-SCF wavefunction.

4.2 Performance of PNO-MP2-F12, PNO-CCSD(2)$_{\overline{F12}}$ and PNO-CCSD[F12]

The performance of our PNO-MP2-F12 implementation in comparison to the canonical DF-MP2-F12 implementation presented in ref. 23 and the previous quartic scaling PNO-MP2-F12 method is tested on glycine chains $(Gly)_n$ with n = 1, 2, 4, 8, 16, 32 in the DZ-F12 basis. In Fig. 1 the wall time is plotted on a logarithmic scale against the number of orbital basis functions for $T_{PNO} = 10^{-7}$. Only for $(Gly)_1$ the canonical implementation outperforms the cubic PNO-MP2-F12 code. Afterwards the benefit of the new implementation is substantial. The break even point with the quartic scaling version reported in ref. 11 is a bit later, but with $(Gly)_4$ still at a reasonable system size. Table 2 shows power laws of the form $10^{b}\cdot N^{n}_{bas}$, which are fitted to the wall clock time. The exponent n shows the effective scaling of the implementations

Table 1 Error in the total correlation energy due to F12 screening Δ_{F12} and Schwarz screening Δ_{con} in μE_h and % in different basis sets. For the Schwarz (and F12-) screening $T^{def.}_{schwarz}$ was used as threshold, while for the reference with (numerical) inactive screening $T^{tight}_{schwarz}$ was applied

Molecules	Basis	$T^{def.}_{schwarz}$	$T^{tight}_{schwarz}$	$\Delta_{F12}\mu E_h$	Δ_{F12} (%)	$\Delta_{con}\mu E_h$	Δ_{con} (%)
Adenine–thymine	DZ-F12	1.6×10^{-12}	1.6×10^{-18}	-1.55×10^{-2}	4.04×10^{-7}	1.54×10^{-2}	-4.02×10^{-7}
Cyclohexane	DZ-F12	3.7×10^{-12}	3.7×10^{-18}	2.90×10^{-3}	-2.53×10^{-7}	-1.15×10^{-2}	1.00×10^{-6}
Dibutylphenol	DZ-F12	1.6×10^{-12}	1.6×10^{-18}	9.70×10^{-3}	-3.39×10^{-7}	-7.09×10^{-2}	2.48×10^{-6}
Norbornane	DZ-F12	3.3×10^{-12}	3.3×10^{-18}	4.70×10^{-3}	-3.57×10^{-7}	-2.81×10^{-2}	2.13×10^{-6}
$(Gly)_1$	aTZ	2.8×10^{-12}	2.8×10^{-18}	9.00×10^{-4}	-6.67×10^{-8}	-2.04×10^{-2}	1.51×10^{-6}
$(Gly)_2$	aTZ	1.6×10^{-12}	1.6×10^{-18}	2.05×10^{-2}	-8.72×10^{-7}	-7.04×10^{-2}	2.99×10^{-6}
$(Gly)_1$	DZ-F12	5.4×10^{-12}	5.4×10^{-18}	6.34×10^{-2}	-4.76×10^{-6}	1.90×10^{-3}	-1.43×10^{-7}
$(Gly)_2$	DZ-F12	3.1×10^{-12}	3.1×10^{-18}	1.40×10^{-3}	-6.05×10^{-8}	-9.20×10^{-3}	3.97×10^{-7}
$(Gly)_4$	aTZ	8.6×10^{-13}	8.6×10^{-19}	-1.33×10^{-2}	3.04×10^{-7}	-2.20×10^{-3}	5.03×10^{-8}
$(GaAs)_4$	wCDZ	2.8×10^{-12}	2.8×10^{-18}	-9.40×10^{-3}	1.92×10^{-7}	-2.18×10^{-2}	4.46×10^{-7}

This journal is © the Owner Societies 2014

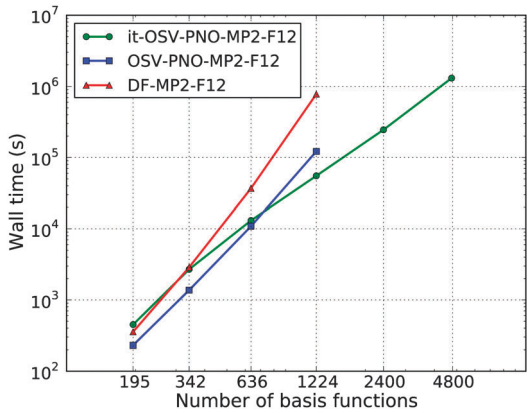


Fig. 1 Wall time (s) for calculations on glycine chains $(Gly)_n$ with n = 1, 2, 4, 8, 16, 32 in the DZ-F12 basis set using canonical DF-MP2-F12 and PNO-MP2-F12 with a quartic and cubic algorithm. The PNO threshold is $T_{PNO} = 10^{-7}$.

Table 2 Fitted power laws for the wall clock time for the canonical DF-MP2-F12 implementation and the cubic and quartic scaling PNO-MP2-F12 implementation

T_{PNO}	$\mathcal{O}(N^3)$	$\mathcal{O}(N^4)$	DF-MP2-F12
10^{-6}	$10^{-2.89}N_{bas}^{2.42}$	$10^{-5.43}N_{bas}^{3.37}$	$10^{-6.47}N_{bas}^{3.93}$
10^{-7}	$10^{-2.81}N_{bas}^{2.44}$	$10^{-5.48}N_{bas}^{3.41}$	$10^{-6.47}N_{bas}^{3.93}$
10^{-8}	$10^{-2.81}N_{bas}^{2.47}$	$10^{-5.59}N_{bas}^{3.48}$	$10^{-6.47}N_{bas}^{3.93}$

and reveals an effective sub-cubic scaling for the new code, but it is fair to mention that also the effective scaling of the old code is a bit below quartic scaling in the studied size regime. Nevertheless the scaling of the new implementation is significantly reduced in comparison to the previous one, which opens up new possibilities. For PNO-CCSD[F12] and PNO-CCSD(2)$_{\overline{F12}}$ analogous calculations were carried out and compared to the canonical CCSD[F12] and CCSD implementations in the TURBOMOLE package. Fig. 2 shows the timings of these methods using a T_{PNO} of 10^{-7} and the DZ-F12 basis. Fig. 2 reveals that even for $(Gly)_1$ PNO-CCSD[F12] and PNO-CCSD(2)$_{\overline{F12}}$ are faster than both canonical CCSD and CCSD[F12]. The benefits of the

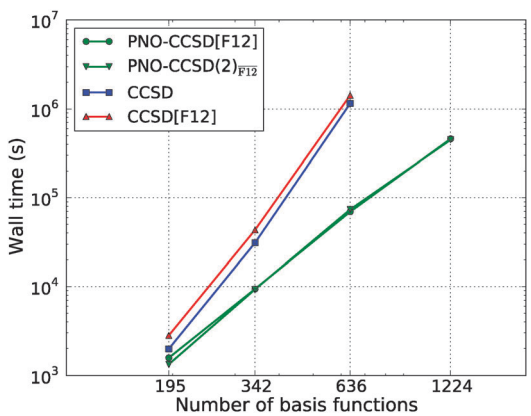


Fig. 2 Wall time (s) for calculations on glycine chains $(Gly)_n$ with n = 1, 2, 4, 8 in the DZ-F12 basis set using canonical CCSD, CCSD[F12], PNO-CCSD[F12] and PNO-CCSD(2)$_{\overline{F12}}$ with $T_{PNO} = 10^{-7}$.

Table 3 Fitted power laws for the canonical CCSD[F12] implementation and the cubic scaling PNO-CCSD[F12] and PNO-CCSD(2)$_{\overline{F12}}$ implementation

T_{PNO}	PNO-CCSD[F12]	PNO-CCSD(2)$_{\overline{F12}}$	CCSD[F12]
10^{-6}	$10^{-3.93}N_{bas}^{3.06}$	$10^{-3.96}N_{bas}^{3.05}$	$10^{-8.67}N_{bas}^{5.26}$
10^{-7}	$10^{-3.89}N_{bas}^{3.10}$	$10^{-4.12}N_{bas}^{3.18}$	$10^{-8.67}N_{bas}^{5.26}$
10^{-8}	$10^{-3.94}N_{bas}^{3.18}$	$10^{-4.00}N_{bas}^{3.20}$	$10^{-8.67}N_{bas}^{5.26}$

PNO approximation are really substantial, and in combination with the enhanced basis set convergence PNO-CCSD[F12] it is extremely promising as a fast and accurate computational approach. The curves for PNO-CCSD[F12] and PNO-CCSD(2)$_{\overline{F12}}$ are very close. CCSD(2)$_{\overline{F12}}$ is slightly faster for a given basis set and truncation threshold, but a better measure of performance is the cost-accuracy ratio, which is discussed in the next section. The results of power law fits to the wall clock time are shown in Table 3. Cubic scaling is not quite reached within the sequence of increasingly larger glycine chains from $(Gly)_1$ to $(Gly)_8$ in the DZ-F12 basis, but it is observed in the smaller DZ basis within this sequence with a fitted power law of $10^{-1.95}N_{bas}^{2.55}$. It is likely that cubic scaling would be observed also for the DZ-F12 basis if $(Gly)_{16}$ were included in the fit.

4.3 S66 benchmarks

To measure the accuracy of our PNO-MP2-F12, PNO-CCSD(2)$_{\overline{F12}}$ and PNO-CCSD[F12] implementations we used the S66 set[38] and performed calculations in the aDZ and aTZ basis sets. For aDZ it was feasible to carry out canonical reference calculations for the whole set, while for aTZ only the S22 subset was computed using the canonical code. These reference data enable us to demonstrate that the PNO truncation error is very small and we use the PNO-CCSD[F12]/aTZ level of theory to improve upon previous estimates for the CCSD/CBS limits of the S66 set,[38] which are given in the ESI.† To the best of our knowledge this is the first time that calculations for the whole S66 set have been performed using F12 theory. The largest previous calculation we have found are CCSD(T)-F12b[39] on a small subset of the S66 molecules, using the incremental scheme.[40]

4.3.1 Analysis of the accuracy. In the first analysis we focus on the relative error in the correlation energy caused by the PNO approximation

$$\Delta_{rel} = \frac{E_c^{PNO} - E_c^{can.}}{E_c^{can.}} \times 100, \qquad (29)$$

which is listed in Tables 4–6 for different thresholds for PNO-MP2-F12, PNO-CCSD(2)$_{\overline{F12}}$ and PNO-CCSD[F12], respectively. In our analysis we included the dimer as well as the monomer structures with and without counterpoise correction. Additionally Fig. 3 shows the distribution of Δ_{rel} in the aDZ and aTZ basis for PNO-CCSD[F12]. It is clear from this graphic that the behaviour for these two basis sets is very similar and the errors converge rapidly to zero as the truncation threshold is tightened. The PNO-MP2-F12 and PNO-CCSD[F12] methods show very similar relative errors, which indicates that the approximations used for the higher-order terms in the CCSD amplitude equations are not a significant source of error, and that our previous conclusions for the accuracy of PNO-MP2-F12 and PNO-MP3(F12)[11,41] are transferable to PNO-CCSD[F12].

Table 4 Average $\bar{\Delta}_{rel}$, standard deviation Δ_{STD} and maximal Δ_{max} relative error in total correlation energy for PNO-MP2-F12 in the aDZ and aTZ basis sets

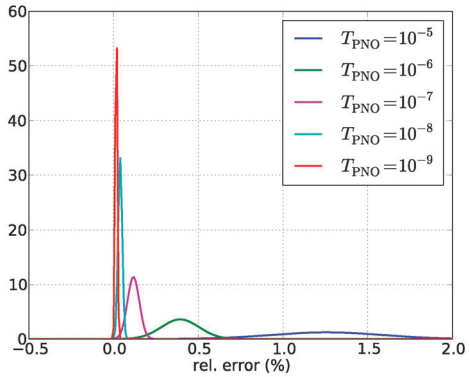
T_{PNO}	aDZ $\bar{\Delta}_{rel}$	aDZ Δ_{STD}	aDZ Δ_{max}	aTZ $\bar{\Delta}_{rel}$	aTZ Δ_{STD}	aTZ Δ_{max}
10^{-5}	1.551	0.465	2.280	1.740	0.450	2.651
10^{-6}	0.483	0.152	0.714	0.560	0.167	0.890
10^{-7}	0.158	0.050	0.234	0.163	0.056	0.277
10^{-8}	0.052	0.016	0.086	0.047	0.021	0.089
10^{-9}	0.020	0.006	0.040	0.013	0.006	0.027

Table 5 Average $\bar{\Delta}_{rel}$, standard deviation Δ_{STD} and maximal Δ_{max} relative error in total correlation energy for PNO-CCSD(2)$_{\overline{F12}}$ in the aDZ and aTZ basis sets

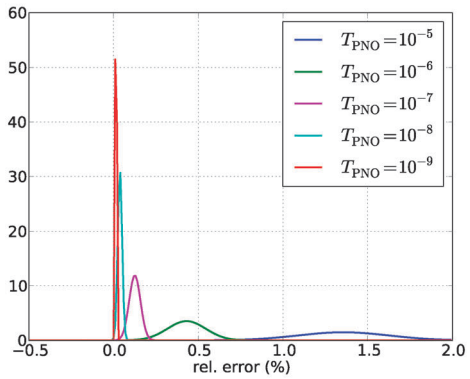
T_{PNO}	aDZ $\bar{\Delta}_{rel}$	aDZ Δ_{STD}	aDZ Δ_{max}	aTZ $\bar{\Delta}_{rel}$	aTZ Δ_{STD}	aTZ Δ_{max}
10^{-5}	1.955	0.363	2.447	2.806	0.309	3.493
10^{-6}	0.683	0.183	0.943	1.180	0.145	1.526
10^{-7}	0.255	0.073	0.369	0.528	0.054	0.652
10^{-8}	0.099	0.024	0.024	0.175	0.030	0.240
10^{-9}	0.037	0.006	0.051	0.059	0.016	0.089

Table 6 Average $\bar{\Delta}_{rel}$, standard deviation Δ_{STD} and maximal Δ_{max} relative error in total correlation energy for PNO-CCSD[F12] in the aDZ and aTZ basis sets

T_{PNO}	aDZ $\bar{\Delta}_{rel}$	aDZ Δ_{STD}	aDZ Δ_{max}	aTZ $\bar{\Delta}_{rel}$	aTZ Δ_{STD}	aTZ Δ_{max}
10^{-5}	1.260	0.326	1.746	1.386	0.302	2.548
10^{-6}	0.395	0.112	0.577	0.441	0.120	0.682
10^{-7}	0.118	0.180	0.035	0.129	0.036	0.212
10^{-8}	0.039	0.012	0.071	0.039	0.014	0.069
10^{-9}	0.016	0.006	0.040	0.015	0.004	0.026



(a) aDZ



(b) aTZ

Fig. 3 Distribution of error in the total correlation energy for PNO-CCSD[F12] for different basis sets aXZ (X = D, T).

In contrast, the relative errors of PNO-CCSD(2)$_{\overline{F12}}$ are 1.5–4 times larger than for PNO-CCSD[F12] and the PNO truncation clearly leads to inaccurate doubles amplitudes. This can be explained: in our approach the V-PNOs are chosen to span the doubles space complementary to the F12 contribution, but the PNO-CCSD(2)$_{\overline{F12}}$ method requires V-PNOs that span the conventional doubles amplitude space, because it proceeds by first solving the conventional CCSD amplitudes without coupling to F12. Due to the non-linearity of the CCSD equations, the amplitudes obtained by solving standard CCSD in the space of F12 optimised V-PNOs differ substantially from those obtained when including the F12 terms in the amplitude equations (as for CCSD[F12]). These deviations are not accounted for in the non-iterative F12 correction terms in CCSD(2)$_{\overline{F12}}$. We have performed PNO-CCSD(2)$_{\overline{F12}}$ calculations using conventional V-PNOs and indeed the truncation errors are then close to those observed for PNO-MP2-F12 and PNO-CCSD[F12]. However, the number of V-PNOs retained is then much larger than when using the difference density approach making these PNO-CCSD(2)$_{\overline{F12}}$ calculations much more expensive and uncompetitive with PNO-CCSD[F12].

Having established that the truncation error for total energies reduces with tighter truncation thresholds, we now turn to the more important task of examining the truncation error for energy differences, and characterising its magnitude relative to the inherent basis set incompleteness error (BSIE) for the canonical calculation. In Fig. 4 we present, for the S22 set, the mean absolute deviation between the PNO-MP2-F12, PNO-CCSD(2)$_{\overline{F12}}$ and CCSD[F12] interaction energies and their CBS limit values. Curves from both aDZ and aTZ calculations are shown with T_{PNO} decreasing from 10^{-5} to 10^{-9}. The CBS limits were computed by extrapolating the canonical aDZ and aTZ values for each method using the optimized extrapolation scheme (with optimized exponents) of Hill *et al.*[42] The mean BSIEs for the canonical aDZ and aTZ calculations are indicated by horizontal lines and these estimates differ only slightly from those previously computed at the MP2.5(F12) level of theory.[11] Fig. 4 highlights that tightening the T_{PNO} threshold beyond 10^{-8} for the aDZ basis and 10^{-9} for the aTZ basis does not significantly improve accuracy since the basis set is almost exhausted, it only increases costs. In Fig. 5 we present a bar chart of the absolute PNO errors to the canonical limit in the aTZ basis. For all thresholds the errors are largest for PNO-MP2-F12. Both Fig. 4 and 5 highlight that although the convergence of the PNO-CCSD(2)$_{\overline{F12}}$ method for total energies with PNO truncation is erratic, the BSIE errors of the interaction energies are nevertheless similar to PNO-CCSD[F12], indicating a high level of error cancellation.

This journal is © the Owner Societies 2014

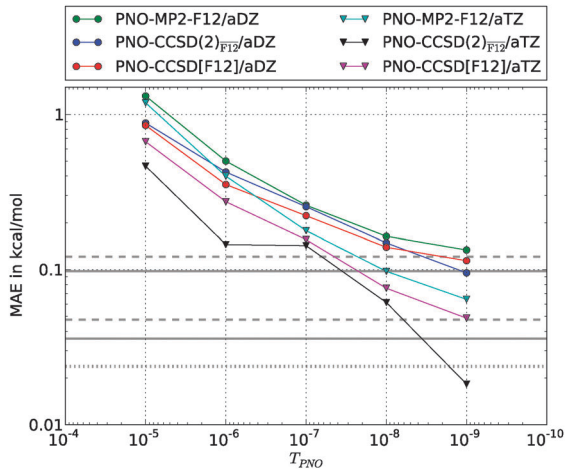


Fig. 4 Remaining PNO truncation error and BSIE for different PNO thresholds. The BSIE limit is sketched with grey lines (dashed MP2-F12, solid CCSD[F12], dotted CCSD(2)$_{\overline{F12}}$).

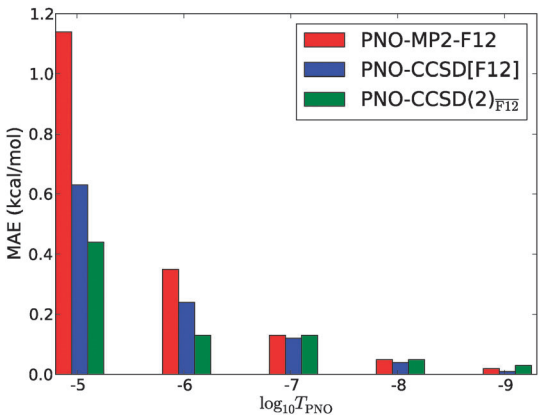


Fig. 5 MAE of the PNO truncation error for the counterpoise corrected interaction energies in the S22 subset using the aTZ basis.

4.3.2 Improved CCSD/CBS limits for the S66 set. For the S66 set, Hobza and co-workers use the following scheme to estimate the energy from a post-MP2 method MCOR at the basis set limit:[38]

$$E(\mathrm{MCOR/CBS}) = E(\mathrm{HF}) + E^{\mathrm{corr}}(\mathrm{MP2/CBS}) + \Delta\mathrm{MCOR}, \quad (30)$$

where E(HF) is the Hartree–Fock energy in the aQZ basis, E^{corr}(MP2/CBS) the extrapolated MP2 CBS limit using Helgaker's extrapolation[14] with the aTZ and aQZ basis and a higher order correction ΔMCOR = E(MCOR) − E(MP2) in the aDZ basis. The PNO representation lifts the basis set limitations. aTZ basis sets can be used for the MCOR = CCSD part, and the combination with F12 theory delivers energies of almost quintuple-ζ quality. The PNO-CCSD[F12] and PNO-CCSD(2)$_{\overline{F12}}$ calculations in the aTZ basis, reported in the previous section, are used here to obtain improved CCSD/CBS estimates for the entire S66 set. Such PNO-CCSD-F12 calculations are both simpler than Hobza's composite approach and closer to the CCSD/CBS limit.

Comparing our PNO-MP2-F12/aTZ correlation energies, using a T_{PNO} of 10^{-8}, with Hobza's reference MP2 CBS limits E^{corr}(MP2/CBS), we find a good agreement with a RMSD of 0.08 kcal mol^{-1}. Comparing our PNO-CCSD[F12] and PNO-CCSD(2)$_{\overline{F12}}$ correlation energies with Hobza's E^{corr}(CCSD/CBS) values, we find larger deviations (RMSDs of 0.22 and 0.18 kcal mol^{-1}, respectively). Many of the complexes beyond the S22 subset show larger deviations. In the block S24–S29 we see for instance deviations up to 0.5 kcal mol^{-1}. These relatively large deviations occur mainly for examples with large singles contributions, where the correlation energy is often overestimated due to missing interference effects in the composite scheme. In Table 7 we give some numerical examples for both good agreement and larger differences. To verify that we reached convergence we extrapolated from the aDZ and aTZ results to the CBS limit. Although the extrapolation scheme of Hill *et al.* assumes a relatively slow convergence, the RMSD between the aTZ results and the extrapolated CBS limits is with 0.035 kcal mol^{-1} much smaller than the RMSD between the extrapolated CBS limits and previous results from the literature, which is with 0.20 kcal mol^{-1} substantially. Our new CBS/limits are given in the ESI.† There we combined the correlation energy with the HF energy in the aTZ basis plus the CABS correction $E_{\mathrm{aTZ}}^{\mathrm{HF+CABS}}$ and with the HF energy in the aQZ basis $E_{\mathrm{aQZ}}^{\mathrm{HF}}$. When going from $E_{\mathrm{aTZ}}^{\mathrm{HF+CABS}}$ to $E_{\mathrm{aQZ}}^{\mathrm{HF}}$ there is only a negligible change in energy. The mean absolute deviation between the two data sets is 0.005 kcal mol^{-1} and the maximal and minimal absolute deviation is 0.021 and 4.78 × 10^{-4} kcal mol^{-1}, respectively. Calculating the HF energy in the aQZ basis here only increases the computational costs, but not the accuracy. For a measure of the BSIE we give the difference of PNO-CCSD[F12] and PNO-CCSD(2)$_{\overline{F12}}$ in the aTZ basis. In a complete basis set and a complete PNO set this difference should vanish. For the test set we get a mean absolute deviation for both methods of 0.052 kcal mol^{-1}. The differences are mainly very small, but there are also indications that for some dimers calculations in larger basis sets could be interesting. The uracil–uracil π–π (no. 26) is

Table 7 Contribution of the CCSD correlation energy to the whole interaction energy in kcal mol^{-1} for PNO-CCSD[F12] in the aDZ and aTZ basis sets, the extrapolation $\Delta E_{\mathrm{int}}^{\mathrm{lim}}$ from aDZ and aTZ and the S66 reference in kcal mol^{-1}

Dimer	aDZ	aTZ	$\Delta E_{\mathrm{int}}^{\mathrm{lim}}$	Lit.	($\Delta E_{\mathrm{int}}^{\mathrm{lim}}$-aTZ)	($\Delta E_{\mathrm{int}}^{\mathrm{lim}}$-Lit.)
02 Water–MeOH	−1.537	−1.591	−1.622	−1.632	−0.031	0.010
11 $MeNH_2$–peptide	−3.192	−3.245	−3.273	−3.403	−0.030	0.130
12 $MeNH_2$–water	−2.258	−2.325	−2.361	−2.389	−0.038	0.028
13 Peptide–MeOH	−2.491	−2.548	−2.578	−2.631	−0.032	0.053
26 Uracil–uracil π–π	−7.894	−7.950	−7.985	−8.495	−0.035	0.510
27 Benzene–pyridine π–π	−5.579	−5.581	−5.584	−5.952	−0.003	0.368
28 Benzene–uracil π–π	−7.171	−7.199	−7.219	−7.718	−0.019	0.500

such an example, where the difference is −0.244 kcal mol^{-1}. The new results show where the current CCSD/CBS and also CCSD(T)/CBS reference limits have to be improved. In the ESI† we give a list of all dimers, where we observe an absolute deviation of larger than 0.2 kcal mol^{-1} and the new benchmark data for the full S66 test set.

4.4 Large scale application

4.4.1 Large basis set. The PNO-CCSD[F12] model was applied to some medium sized systems. One of these systems is a cluster model for zeolite H-ZSM-5 described by Svelle *et al.* in ref. 45, where the model was denoted as the "4T" cluster. In the current study the interaction energy of a methanol molecule with the 4T cluster was calculated on the PNO-CCSD[F12]/aDZ, PNO-CCSD[F12]/aTZ and PNO-CCSD[F12]/aQZ levels of theory. Canonical CCSD[F12] calculations using the very large and accurate aQZ basis are not feasible in a reasonable amount of time. The aQZ dimer calculation, which was the most expensive part, took 3 days on a machine using two 2.93 GHz Intel Xenon X5670 CPUs and 48 GiB RAM, and 6 threads and a partially OpenMP parallelized code.

The counterpoise corrected and uncorrected interaction energies are listed in Table 8. For the aQZ basis the results with and without counterpoise correction are very close, which is a strong indication that the results are close to the CBS limit. The counterpoise corrected values for aTZ and aQZ are very close, as expected from the enhanced basis set convergence with explicit electron correlation. From the results we can conclude that the CCSD/CBS limit (at the given geometry) is 20.6 ± 0.1 kcal mol^{-1}.

4.5 Large molecular system

As another example we chose a water–calix[4]arene complex (C4A–H_2O), shown in Fig. 6, and calculated the counterpoise corrected interaction energy at the PNO-MP2-F12/aDZ and PNO-CCSD[F12]/aDZ levels. This cluster has already been studied by Hontama *et al.*[46] and we thank the authors for providing us with their DF-MP2/aDZ optimized structure. For this complex we calculated the binding energy as

$$E_{\text{int}} = E^{\text{C4A}-\text{H}_2\text{O}}_{\text{C4A}-\text{H}_2\text{O}}(\text{C4A}-\text{H}_2\text{O}) - E^{\text{C4A}}_{\text{C4A}}(\text{C4A}) - E^{\text{H}_2\text{O}}_{\text{H}_2\text{O}}(\text{H}_2\text{O}), \quad (31)$$

where the superscripts denote the basis and the subscript the geometry and the molecule is given in parentheses. We additionally computed the counterpoise corrected interaction energies as

$$E^{\text{CP}}_{\text{int}} = E^{\text{C4A}-\text{H}_2\text{O}}_{\text{C4A}-\text{H}_2\text{O}}(\text{C4A}-\text{H}_2\text{O}) - E^{\text{C4A}-\text{H}_2\text{O}}_{\text{C4A}-\text{H}_2\text{O}}(\text{C4A}) - E^{\text{C4A}-\text{H}_2\text{O}}_{\text{C4A}-\text{H}_2\text{O}}(\text{H}_2\text{O}) + E^{\text{H}_2\text{O}}_{\text{relax}} + E^{\text{C4A}}_{\text{relax}}, \quad (32)$$

including the relaxation $E^{\text{H}_2\text{O}}_{\text{relax}}$ and $E^{\text{C4A}}_{\text{relax}}$ due to the change in geometry

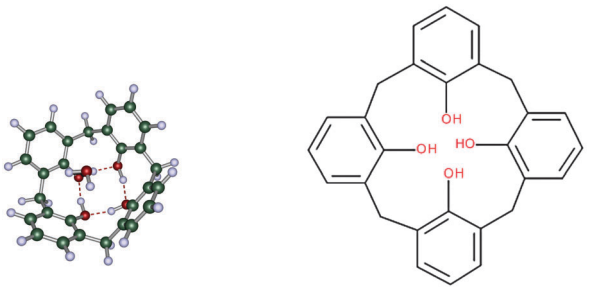
$$E^{\text{H}_2\text{O}}_{\text{relax}} = E^{\text{H}_2\text{O}}_{\text{C4A}-\text{H}_2\text{O}}(\text{H}_2\text{O}) - E^{\text{H}_2\text{O}}_{\text{H}_2\text{O}}(\text{H}_2\text{O}) \quad (33)$$

$$E^{\text{C4A}}_{\text{relax}} = E^{\text{C4A}}_{\text{C4A}-\text{H}_2\text{O}}(\text{C4A}) - E^{\text{C4A}}_{\text{C4A}}(\text{C4A}) \quad (34)$$

Our results, together with the values of Hontama *et al.*,[46] are listed in Table 9. The comparison shows a near perfect agreement of our uncorrected PNO-MP2-F12/aDZ results with the uncorrected DF-MP2/aQZ value of Hontama *et al.*, which underlines the expected enhancement in basis set convergence. Counterpoise corrected reference calculations in the aQZ basis are not available, but we expect a similar agreement.

The binding energies vary greatly with the level of theory applied. The PNO-CCSD[F12] binding energy is 46% lower than the PNO-MP2-F12 value, and the SCS-PNO-MP2-F12 result is close to PNO-CCSD[F12] – both 30% lower than PNO-MP2-F12, which is consistent with the usual overestimation of dispersion interactions by MP2 theory.

Further investigation of the complex using the TPSS-D3/aTZ level of theory revealed an additional minimum, Fig. 7, where one hydrogen bond of the host opens to form a hydrogen bond with the guest. Computed interaction energies are listed in Table 10. Due to the large change in geometry, we only give here



(a) Optimized complex (b) Structural formula of the host

Fig. 6 Structure of the calix[4]arene–water complex. Hydrogen bonds are indicated using red dashed lines.

Table 8 Counterpoise corrected $E^{\text{CP}}_{\text{int}}$ and uncorrected E_{int} binding energies for the 4T cluster using a $T_{\text{PNO}} = 10^{-8}$

Basis	$E^{\text{CP}}_{\text{int}}$ (kcal mol^{-1})	E_{int} (kcal mol^{-1})
aDZ	−20.10	−22.02
aTZ	−20.41	−21.15
aQZ	−20.53	−20.70

Table 9 Counterpoise corrected $E^{\text{CP}}_{\text{int}}$ and uncorrected E_{int} binding energies for the endo calix[4]arene–water complex using a $T_{\text{PNO}} = 10^{-7}$. The DF-MP2 results are from ref. 46

Method	E_{int} (kcal mol^{-1})	$E^{\text{CP}}_{\text{int}}$ (kcal mol^{-1})
PNO-MP2-F12/aDZ	−8.50	−7.73
PNO-CCSD[F12]/aDZ	−5.00	−4.14
SCS-PNO-MP2-F12/aDZ	−5.70	−4.95
RHF	2.27	4.26
DF-MP2/aTZ	−10.29	−7.97
DF-MP2/aQZ	−8.94	—

This journal is © the Owner Societies 2014

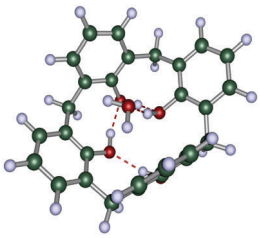


Fig. 7 Structure of the calix[4]arene–water complex optimized at the TPSS-D3/aTZ level of theory. Hydrogen bonds are indicated using red dashed lines.

Table 10 Counterpoise corrected E_{int}^{CP} and uncorrected E_{int} binding energies for the TPSS-D3 optimized endo calix[4]arene–water complex in the aDZ basis using a $T_{PNO} = 10^{-7}$

Method	E_{int}^{CP} (kcal mol^{-1})
PNO-MP2-F12/aDZ	−13.72
PNO-CCSD[F12]/aDZ	−10.72
SCS-PNO-MP2-F12/aDZ	−11.05
RHF	−3.79

the counterpoise corrected interaction energies including the relaxation since the uncorrected values are not meaningful. The host–guest interaction in the TPSS-D3 optimized structure is stronger due to the presence of the hydrogen bond and is predicted to be stable even at the RHF level of theory.

5 Conclusion and outlook

We have presented low scaling PNO-MP2-F12, PNO-CCSD(2)$_{\overline{F12}}$ and PNO-CCSD[F12] implementations with early break even points to their canonical counterparts. For this purpose we extended our hybrid OSV–PNO approach with an iterative generation of OSVs to F12 theory, which required minor changes in the existing algorithm for the PNO generation. The terms for the singles residuals are based on our PNO-CC2[25,26] implementation and extended to the F12 contributions in CCSD[F12]. For the conventional contributions to the doubles residuals we followed Neese *et al.*,[24] and for the F12 contributions we could use the previously implemented intermediates from PNO-MP3(F12).[11] Furthermore we adapted the integral evaluation to exploit the short range nature of the operators f_{12}, $f_{12}g_{12}$, f_{12}^2 and $f_{12}^2 r_{12}^2$, screening small contributions.

We find that for the iterative CCSD[F12] approximation to CCSD-F12 the PNO truncation errors with the very compact set of F12-PNOs converge as fast as for MP2-F12, while for the perturbative CCSD(2)$_{\overline{F12}}$ approximation the larger set of conventional PNOs has to be used to maintain this convergence.

The accuracy of the implemented methods is demonstrated using calculations on the S66 set of weakly bound complexes with the aDZ and aTZ basis sets. The PNO error smoothly decreases with the PNO truncation threshold T_{PNO} and our best calculations for the S66 set are closer to the CBS limit than previous reference values. An advantage of the PNO methods is that they reduce the need for complicated composite extrapolation schemes. The improvement in accuracy arises because larger basis sets can be used in the reduced scaling implementation. In this manner we were also able to calculate the interaction energy of a methanol in a model for zeolite H-ZSM-5 in the aQZ basis, which is not feasible with canonical implementations in practical time. Our calculations on a calix[4]arene–water complex demonstrate that PNO-CCSD[F12] is a practical tool for the evaluation of interaction energies in large van der Waals complexes, and we find that MP2 overestimates the binding energy for this complex by nearly a factor of two. Our investigations on this complex revealed another minimum with a significantly stronger host–guest interaction than that of the previously studied structure. The success of low-scaling PNO-CC-F12 methods for modeling weak interactions is enormously encouraging. Since host–guest interactions are present in various fields from biochemistry to inorganic chemistry, with many industrial applications such as the storage of small molecules, we see great potential in PNO-CC-F12 methods to contribute to a better understanding of such processes.

Acknowledgements

We thank S. Xantheas for providing the structure of the calix[4]-arene–water complex. Financial support from the Deutsche Forschungsgemeinschaft through grant no. HA/2588/7 is gratefully acknowledged. DPT thanks the Royal Society for support through a University Research Fellowship.

References

1 D. S. Lambrecht, B. Doser and C. Ochsenfeld, *J. Chem. Phys.*, 2005, **123**, 184102.
2 B. Doser, D. S. Lambrecht and C. Ochsenfeld, *Phys. Chem. Chem. Phys.*, 2008, **10**, 3335–3344.
3 M. Schütz, G. Hetzer and H.-J. Werner, *J. Chem. Phys.*, 1999, **111**, 5691.
4 M. Schütz and H.-J. Werner, *J. Chem. Phys.*, 2001, **114**, 661.
5 M. Schütz and H.-J. Werner, *Chem. Phys. Lett.*, 2000, **318**, 370.
6 S. Saebø and P. Pulay, *Chem. Phys. Lett.*, 1985, **113**, 13.
7 F. Neese, F. Wennmohs and A. Hansen, *J. Chem. Phys.*, 2009, **130**, 114108.
8 B. Helmich and C. Hättig, *J. Chem. Phys.*, 2011, **135**, 214106.
9 J. Yang, G. K.-L. Chan, F. R. Manby, M. Schütz and H.-J. Werner, *J. Chem. Phys.*, 2012, **136**, 144105.
10 C. Krause and H.-J. Werner, *Phys. Chem. Chem. Phys.*, 2012, **14**, 7591–7604.
11 C. Hättig, D. P. Tew and B. Helmich, *J. Chem. Phys.*, 2012, **136**, 204105.
12 G. Schmitz, B. Helmich and C. Hättig, *Mol. Phys.*, 2013, **111**, 2463–2476.
13 C. Riplinger and F. Neese, *J. Chem. Phys.*, 2013, **138**, 034106.
14 A. Halkier, T. Helgaker, P. Jørgensen, W. Klopper and J. Olsen, *Chem. Phys. Lett.*, 1999, **302**, 437–446.
15 W. Kutzelnigg, *Theor. Chim. Acta*, 1985, **68**, 445–469.
16 W. Klopper and W. Kutzelnigg, *J. Phys. Chem.*, 1990, **94**, 5625–5630.

17 V. Termath, W. Klopper and W. Kutzelnigg, *J. Chem. Phys.*, 1991, **94**, 2002.
18 D. P. Tew and C. Hättig, *Int. J. Quantum Chem.*, 2013, **113**, 224–229.
19 S. Ten-no, *J. Chem. Phys.*, 2004, **121**, 117–129.
20 M. Häser and J. Almlöf, *J. Chem. Phys.*, 1992, **96**, 489.
21 T. B. Adler, H.-J. Werner and F. R. Manby, *J. Chem. Phys.*, 2009, **130**, 054106.
22 F. R. Manby, *J. Chem. Phys.*, 2003, **119**, 4607–4613.
23 R. A. Bachorz, F. A. Bischoff, A. Glöß, C. Hättig, S. Höfener, W. Klopper and D. P. Tew, *J. Comput. Chem.*, 2011, **32**, 2492–2513.
24 F. Neese, A. Hansen and D. G. Liakos, *J. Chem. Phys.*, 2009, **131**, 064103.
25 B. Helmich and C. Hättig, *J. Chem. Phys.*, 2013, **139**, 084114.
26 B. Helmich, PhD thesis, Ruhr-Universität Bochum, 2014.
27 M. Torheyden and E. F. Valeev, *Phys. Chem. Chem. Phys.*, 2008, **10**, 3410–3420.
28 C. Hättig, D. P. Tew and A. Köhn, *J. Chem. Phys.*, 2010, **132**, 231102.
29 TURBOMOLE V6.5 2013, a development of University of Karlsruhe and Forschungszentrum Karlsruhe GmbH, 1989–2007, TURBOMOLE GmbH, since 2007; available from http://www.turbomole.com.
30 C. Hättig and F. Weigend, *J. Chem. Phys.*, 2000, **113**, 5154.
31 R. A. Bachorz, F. A. Bischoff, A. Glöß, C. Hättig, S. Höfener, W. Klopper and D. P. Tew, *J. Comput. Chem.*, 2011, **32**, 2492–2513.
32 K. A. Peterson, T. B. Adler and H.-J. Werner, *J. Chem. Phys.*, 2008, **128**, 084102.
33 T. H. J. Dunning, *J. Chem. Phys.*, 1989, **90**, 1007.
34 F. Weigend, A. Köhn and C. Hättig, *J. Chem. Phys.*, 2002, **116**, 3175–3183.
35 K. E. Yousaf and K. A. Peterson, *J. Chem. Phys.*, 2008, **129**, 184108.
36 K. Peterson and T. J. Dunning, *J. Chem. Phys.*, 2002, **117**, 10548.
37 C. Hättig, G. Schmitz and J. Koßmann, *Phys. Chem. Chem. Phys.*, 2012, **14**, 6549–6555.
38 J. Rezac, K. E. Riley and P. Hobza, *J. Chem. Theory Comput.*, 2011, **7**, 2427–2438.
39 T. B. Adler, G. Knizia and H.-J. Werner, *J. Chem. Phys.*, 2007, **127**, 221106.
40 J. Zhang and M. Dolg, *J. Chem. Phys.*, 2014, **140**, 044114.
41 D. P. Tew, B. Helmich and C. Hättig, *J. Chem. Phys.*, 2011, **135**, 074107.
42 J. G. Hill, K. A. Peterson, G. Knizia and H.-J. Werner, *J. Chem. Phys.*, 2009, **131**, 194105.
43 G. A. DiLabio and M. Koleini, *J. Chem. Phys.*, 2014, **140**, 18A542.
44 M. Goldey and M. Head-Gordon, *J. Phys. Chem. B*, 2014, **118**, 6519–6525.
45 S. Svelle, C. Tuma, X. Rozanska, T. Kerber and J. Sauer, *J. Am. Chem. Soc.*, 2009, **131**, 816–825.
46 N. Hontama, Y. Inokuchi, T. Ebata, C. Dedonder-Lardeux, C. Jouvet and S. S. Xantheas, *J. Phys. Chem. A*, 2010, **114**, 2967–2972.

 This journal is © the Owner Societies 2014