# Local explicitly correlated second-order Møller–Plesset perturbation theory with pair natural orbitals

David P. Tew,[1,a)] Benjamin Helmich,[2,b)] and Christof Hättig[2,c)]
[1]*School of Chemistry, University of Bristol, Bristol BS8 1TS, United Kingdom*
[2]*Lehrstuhl für Theoretische Chemie, Ruhr-Universität Bochum, D-44780 Bochum, Germany*

(Received 9 June 2011; accepted 14 July 2011; published online 17 August 2011)

We explore using a pair natural orbital analysis of approximate first-order pair functions as means to truncate the space of both virtual and complementary auxiliary orbitals in the context of explicitly correlated F12 methods using localised occupied orbitals. We demonstrate that this offers an attractive procedure and that only 10–40 virtual orbitals per significant pair are required to obtain second-order valence correlation energies to within 1–2% of the basis set limit. Moreover, for this level of virtual truncation, only 10–40 complementary auxiliary orbitals per pair are required for an accurate resolution of the identity in the computation of the three- and four-electron integrals that arise in explicitly correlated methods. © *2011 American Institute of Physics*. [doi:10.1063/1.3624370]

## I. INTRODUCTION

Explicitly correlated coupled-cluster F12 methods have during the last decade evolved to become reliable tools for computing high accuracy ground state energies in a cost-effective way and have recently been implemented in several commercial and non-commercial program packages.[1–20] The introduction of an auxiliary basis[21,22] for the resolution of identities in the computation of three- and four-electron integrals, combined with efficient integral evaluation techniques,[23,24] Slater-type correlation factors,[25,26] and efficient approximations[6,8,14,18,19,27] to the full explicitly correlated coupled-cluster single and doubles model (CCSD-F12) has greatly reduced the computational overhead of coupled-cluster F12 methods. Nowadays, the computational costs of such calculations are only 20–40% larger than those for conventional calculations in a basis set of the same size, but due to a greatly enhanced convergence they have much smaller basis set errors. Recently, the CCSD-F12a model of Adler *et al.*[28,29] has been combined with a local correlation approach, where the excitations for a pair of localised occupied orbitals $i$, $j$ are restricted to projected atomic orbitals that are spatially close. This reduces the steep scaling of costs with the system size and further extends the applicability of these methods to larger systems. The CCSD(F12) method has also been combined with localisation techniques by Friedrich *et al.* in the context of the incremental scheme.[30]

In this article we explore an alternative approach to localisation through utilising pair natural orbitals (PNOs). A new interest in PNOs has recently been initiated by the work of Neese *et al.*[31,32] who presented efficient implementations of a PNO based coupled electron pair approximation (PNO-CEPA) and PNO-CCSD. Truncated pair natural orbital expansions have already, in the 1970s, been applied to drastically reduce the virtual orbital spaces in MP2, CI, and CEPA calculations in a way that led only to modest truncation errors in the correlation energy.[33–35] These methods were later dismissed because the transformation of four-index two-electron integrals to the PNO basis, and the subsequent storage of these integrals, was viewed as a severe bottleneck for large-scale applications. Recently, Neese and co-workers have demonstrated that this bottleneck can be avoided by applying the density-fitting approximation for the two-electron integrals.[31]

The PNO expansion of pair functions offers a way of introducing locality that avoids technical difficulties in selecting spatial domains and has the appealing feature that accuracy can be controlled through a single parameter, namely, the occupation number threshold for PNO truncation. Compared to the *a priori* truncation based on spatial domains, the truncation based on a PNO analysis of a correlated low-order trial wave function has the advantage that it introduces some additional optimisation of the virtual space and thereby also some self-adaptivity. This allows a larger reduction of the virtual space without sacrificing accuracy and also makes it possible to describe some non-local effects, without incommensurate expansion of the virtual space.

It is evident from the work of Werner and co-workers,[28,29,36] that the F12 geminal basis functions are highly suited local approaches, primarily because the geminals decay rapidly with increasing electron separation. In the domain-based approach the geminal basis can also be used to eliminate the domain error.[36] In other work, Höfener *et al.* have demonstrated that restricting double replacements to span only the F12 manifold recovers between 60% and 80% of the CC2 correlation energy.[37] In the context of PNO theory, therefore, it can reasonably be expected that the inclusion of geminal basis functions will permit a far more drastic truncation of the virtual space than is possible in standard calculations, without significant loss of accuracy. In this work we demonstrate that only 10–40 virtuals per pair are required to recover $>98\%$ of the basis set limit correlation energy, and that this is independent of the size of the molecule and the Hartree–Fock basis set.

[a)]Electronic mail: david.tew@bristol.ac.uk.
[b)]Electronic mail: benjamin.helmich@rub.de.
[c)]Electronic mail: christof.haettig@rub.de.

 © 2011 American Institute of Physics

Downloaded 11 Oct 2011 to 134.147.252.238. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

In R12 and F12 explicitly correlated methods, the three- and four-electron integrals that arise are evaluated through insertions of approximate resolutions of the identity (RI).[38,39] When small orbital basis sets are used, it is a standard practice to supplement the orbital basis with a complementary auxiliary (CA) basis in the RI insertion.[21,22] A large part of the additional expense of F12 calculations over their conventional analogues can be traced to the contractions over the CA indices. It is therefore of great interest to explore extensions of PNO theory to include the CA orbitals such that efficiency may be improved by systematically truncating the CA space. In this work we introduce complementary auxiliary pair natural orbitals (CAPNOs) and demonstrate their utility at the level of second-order Møller–Plesset theory. Our calculations indicate that although the atomic orbitals of the complementary auxiliary basis set (CABS) should sufficiently saturate the orbital space, the number of CAPNOs required for each pair for an accurate RI is remarkably small. In fact, for almost all occurrences of a RI insertion, the number of contributing CAPNOs is bounded by either the number of PNOs or the number of nearby occupied orbitals.

The size of the computational overhead in forming PNOs and CAPNOs means that—just as in the non-explicitly correlated case—the increase in efficiency in MP2 or MP2-F12 calculations can only be rather moderate, if no further approximations in the construction of the PNOs are made. However, once PNOs are implemented in an explicitly correlated coupled-cluster program, the computational savings will be considerable. The use of PNOs and CAPNOs will significantly reduce the computational overhead of CCSD-F12 and related approaches and will pave the way for the development of highly efficient PNO-CEPA-F12 or PNO-CCSD-F12 methods.

## II. THEORY

### A. Local MP2-F12 theory with pair natural orbitals

In PNO-MP2 theory, a different set of virtual orbitals is selected for the parameterisation of the first-order pair function for every occupied pair $ij$. In explicitly correlated PNO-MP2-F12 theory, the spin-free first-order pair correlation functions are, therefore, given by

$$u_{ij}(1,2) = \frac{1}{2}\sum_{\bar{a}\bar{b}} t^{ij}_{\bar{a}\bar{b}}\phi_{\bar{a}}(1)\phi_{\bar{b}}(2) + \hat{Q}_{ij} f(r_{12})\left(\frac{3}{8}\phi_i(1)\phi_j(2) + \frac{1}{8}\phi_j(1)\phi_i(2)\right). \quad (1)$$

Here, $i$, $j$ are localised occupied orbitals in the Hartree–Fock determinant and $\bar{a}, \bar{b}$ are pair natural virtual orbitals, which are yet to be defined, but have the general form

$$|\bar{a}\rangle = \sum_a |a\rangle d^{ij}_{a\bar{a}}, \quad (2)$$

where $a$ are the canonical Hartree–Fock virtual orbitals. We have chosen the fixed amplitude (SP) *ansatz*[40] where the coefficients of the geminal contribution are selected to satisfy the first-order $s$- and $p$-wave coalescence conditions. As usual, $f(r_{12})$ is a short-range correlation factor with the property $f(r_{12}) = r_{12} + \mathcal{O}(r_{12}^2)$. The strong orthogonality operator $\hat{Q}_{ij}$ is defined as

$$\hat{Q}_{ij} = (1-\hat{O}_1)(1-\hat{O}_2)(1-\hat{V}_1\hat{V}_2) \quad (3)$$

with

$$\hat{O} = \sum_i |i\rangle\langle i|, \quad (4)$$

$$\hat{V} = \sum_{\bar{a}} |\bar{a}\rangle\langle\bar{a}|. \quad (5)$$

Just as in Werner's local MP2-F12 theory,[36] $\hat{Q}_{ij}$ is different for every pair $ij$ because $\hat{V}$ projects onto the active virtual space, which in PNO-MP2-F12 theory is the set of pair natural virtual orbitals that survive truncation. In a closed-shell formulation, the amplitudes $t^{ij}_{\bar{a}\bar{b}}$ are determined through the residual equation

$$\sum_{\bar{c}} f^{\bar{c}}_{\bar{a}} t^{ij}_{\bar{c}\bar{b}} + \sum_{\bar{d}} f^{\bar{d}}_{\bar{b}} t^{ij}_{\bar{a}\bar{d}} - \sum_{k\bar{c}\bar{d}} S^{ijk}_{\bar{a}\bar{c}} S^{ijk}_{\bar{b}\bar{d}} t^{kj}_{\bar{c}\bar{d}} f^i_k - \sum_{l\bar{c}\bar{d}} S^{jil}_{\bar{a}\bar{c}} S^{jil}_{\bar{b}\bar{d}} t^{il}_{\bar{c}\bar{d}} f^j_l + C^{ij}_{\bar{a}\bar{b}} + g^{ij}_{\bar{a}\bar{b}} = 0, \quad (6)$$

where $f^q_p$ are the Fock matrix elements, $g^{ij}_{\bar{a}\bar{b}}$ are the usual electron repulsion integrals $(\bar{a}i|\bar{b}j)$, and $S^{ijk}_{\bar{a}\bar{c}}$ is the overlap matrix between PNOs $\bar{a}$ in pair $ij$ with the PNOs $\bar{c}$ in pair $kj$. Equation (6) differs from the non-explicitly correlated case only by the appearance of the matrix elements $C^{ij}_{\bar{a}\bar{b}}$, which describe the coupling to the double excitations into geminals. The correlation energy is computed through the MP2-F12 Lagrangian

$$E = \sum_{i\le j}(2-\delta_{ij})\sum_{\bar{a}\bar{b}} \tilde{t}^{ij}_{\bar{a}\bar{b}}\left(g^{\bar{a}\bar{b}}_{ij} + C^{\dagger\,\bar{a}\bar{b}}_{ij}\right) + \sum_{i\le j}(2-\delta_{ij})\left(B^{ij}_{ij} - X^{kj}_{ij} f^i_k - X^{il}_{ij} f^j_l + 2V^{ij}_{ij}\right), \quad (7)$$

where $\tilde{t}^{ij}_{ab} = 2t^{ij}_{ab} - t^{ji}_{ab}$ and the F12 intermediates have the usual definitions[3,41]

$$B^{ij}_{ij} = \langle\widetilde{ij}|f(r_{12})\hat{Q}_{ij}(\hat{F}_1+\hat{F}_2)\hat{Q}_{ij}f(r_{12})|\widetilde{ij}\rangle, \quad (8)$$

$$C^{ij}_{\bar{a}\bar{b}} = \langle\bar{a}\bar{b}|(\hat{F}_1+\hat{F}_2)\hat{Q}_{ij}f(r_{12})|\widetilde{ij}\rangle, \quad (9)$$

$$X^{ij}_{ij} = \langle\widetilde{ij}|f(r_{12})\hat{Q}_{ij}f(r_{12})|\widetilde{ij}\rangle, \quad (10)$$

$$V^{ij}_{ij} = \langle ij|r_{12}^{-1}\hat{Q}_{ij}f(r_{12})|\widetilde{ij}\rangle, \quad (11)$$

with

$$|ij\rangle = \phi_i(1)\phi_j(2), \quad (12)$$

$$\langle ij| = 2\phi_i(1)\phi_j(2) - \phi_j(1)\phi_i(2), \quad (13)$$

$$|\widetilde{ij}\rangle = \frac{3}{8}\phi_i(1)\phi_j(2) + \frac{1}{8}\phi_j(1)\phi_i(2), \quad (14)$$

Downloaded 11 Oct 2011 to 134.147.252.238. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

$$\langle \widetilde{ij}| = \frac{5}{8}\phi_i(1)\phi_j(2) - \frac{1}{8}\phi_j(1)\phi_i(2). \tag{15}$$

To obtain equations for the practical evaluation of the F12 intermediates, we follow standard F12 theory and approximate the $\hat{Q}_{ij}$ projector with the aid of a set of complementary auxiliary orbitals using one of following:

$$\hat{Q}_{ij}^{A} = 1 - \sum_{\bar{p}\bar{q}} |\bar{p}\bar{q}\rangle\langle\bar{p}\bar{q}| - \sum_{i\bar{q}'} |i\bar{q}'\rangle\langle i\bar{q}'| - \sum_{\bar{p}'j} |\bar{p}'j\rangle\langle\bar{p}'j|, \tag{16}$$

$$Q_{ij}^{B} = \sum_{\bar{a}\bar{q}'} |\bar{a}\bar{q}'\rangle\langle\bar{a}\bar{q}'| + \sum_{\bar{p}'\bar{b}} |\bar{p}'\bar{b}\rangle\langle\bar{p}'\bar{b}| + \sum_{\bar{p}'\bar{q}'} |\bar{p}'\bar{q}'\rangle\langle\bar{p}'\bar{q}'|, \tag{17}$$

where $\hat{Q}_{ij}^{B}$ is only used when $\hat{Q}_{ij}^{A}$ cannot be inserted to achieve the desired decomposition into sums of products of two-electron integrals. Here, the orbitals $\bar{p}$, $\bar{q}$ span the union of the frozen and active occupied orbitals and the PNOs of the pair $i$, $j$ that survive truncation. The orbitals $\bar{p}'$ are CAPNOs

$$|\bar{p}'\rangle = \sum_{a} |a\rangle d_{a\bar{p}'}^{ij} + \sum_{p''} |p''\rangle d_{p''\bar{p}'}^{ij}, \tag{18}$$

where $p''$ are the complementary auxiliary orbitals in standard F12 theory. The CAPNOs satisfy the condition $\langle\bar{p}'|\bar{a}\rangle = 0$ and the precise definition of the coefficients $d_{a\bar{p}'}^{ij}$ and $d_{p''\bar{p}'}^{ij}$ will be discussed in Sec. II C. In these choices for the strong orthogonality projector, the set of discarded PNO virtuals is transferred into the complementary space such that the accuracy in the RI is retained, provided that the space of CA orbitals is not truncated. However, it is then also possible to rank the CAPNOs and discard orbitals with negligible contribution such that the space of CAPNOs that needs to be considered becomes small and spatially localised if a localised basis is used for the occupied space.

Applying standard F12 methodology, the working equations for the F12 intermediates are obtained by combining the approximations for $\hat{Q}_{ij}$ with commutator manipulations. The resulting equations are completely analogous to standard F12 theory, except that the virtual orbitals are replaced by PNOs and the CA orbitals are replaced by CAPNOs. Since the F12 equations have been reported in many previous articles, we do not repeat them here but refer the reader to the literature.[41–44] In our present computer implementation we have made the choice that the CAPNOs are only truncated when they appear in the contraction indices of two two-electron integrals. When the RI results in an orbital transformation, no truncation is applied, e.g.,

$$\langle\bar{a}\bar{b}|(\hat{F}_1 + \hat{F}_2)\hat{Q}_{ij} f(r_{12})|\widetilde{ij}\rangle \approx f_{\bar{a}}^{\bar{p}'}\langle\bar{p}'\bar{b}|f(r_{12})|\widetilde{ij}\rangle + f_{\bar{b}}^{\bar{q}'}\langle\bar{a}\bar{q}'|f(r_{12})|\widetilde{ij}\rangle. \tag{19}$$

The reason for this choice is that orbital transformations are applied at the three-index level in our density-fitting driven implementation, and we do not yet have a local density fitting scheme that exploits locality for these quantities.

### B. Pair natural orbitals in the virtual space

The simplest definition of pair natural orbitals is the natural orbitals of the first-order density matrix, which in a closed-shell formalism is

$$D_{ab}^{ij} = \frac{1+\delta_{ij}}{N_{ij}} \sum_{c} \left(\tilde{t}_{ac}^{ij} t_{bc}^{ij} + \tilde{t}_{ac}^{ji} t_{bc}^{ji}\right), \tag{20}$$

$$N_{ij} = 1 + \sum_{ab} \tilde{t}_{ab}^{ij} t_{ab}^{ij}, \tag{21}$$

$$\tilde{t}_{ab}^{ij} = 2t_{ab}^{ij} - t_{ab}^{ji}, \tag{22}$$

$$t_{ab}^{ij} = -\frac{g_{ab}^{ij}}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j}. \tag{23}$$

Here, $g_{ab}^{ij}$ are the electron repulsion integrals $(ai|bj)$ for occupied $i$, $j$ and virtual $a$, $b$ orbitals in the canonical Hartree–Fock basis. For localised occupied orbitals, $\varepsilon_i$ and $\varepsilon_j$ are replaced by the diagonal elements of the Fock matrix $f_i^i$ and $f_j^j$. The set of virtuals $\bar{a}$ that diagonalise the density $\mathbf{D}^{ij}$ are the natural orbitals for each pair $ij$ (PNOs):

$$\sum_{ab} d_{a\bar{a}}^{ij} D_{ab}^{ij} d_{b\bar{b}}^{ij} = n_{\bar{a}}\delta_{\bar{a}\bar{b}}. \tag{24}$$

Ranking the PNOs according to their occupation numbers $1 > n_{\bar{a}} > 0$ provides a rapidly convergent series expansion for the MP2 energy:

$$E = \sum_{i\le j} (2-\delta_{ij}) \sum_{\bar{a}\bar{b}} \tilde{t}_{\bar{a}\bar{b}}^{ij} g_{ij}^{\bar{a}\bar{b}}. \tag{25}$$

Discarding PNOs with occupation numbers below a given threshold is a safe and systematic procedure for truncating the space of virtual orbitals, although for higher-order correlation methods, such as coupled-cluster approaches, modified definitions of PNOs may be considered.[31]

In F12 methods, there is considerable redundancy between the standard pair functions and the F12 geminals. Consequently, the above definition of PNOs is not optimal for these methods and an alternative is sought. The redundancy is the overlap of the geminal $(1-\hat{O}_1)(1-\hat{O}_2)f(r_{12})|\widetilde{ij}\rangle$ with the space of conventional doubles $|ab\rangle$, which are the amplitudes

$$r_{ab}^{ij} = \langle ab|f(r_{12})|\widetilde{ij}\rangle = \frac{3}{8}(ai|f(r_{12})|bj) + \frac{1}{8}(aj|f(r_{12})|bi). \tag{26}$$

We therefore substitute the MP2 amplitudes $t_{ab}^{ij}$ in Eqs. (20)–(23) with the difference amplitudes

$$t_{ab}^{ij} - r_{ab}^{ij}. \tag{27}$$

The difference amplitudes may be viewed as amplitudes for the part of the pair correlation function that cannot be captured by the F12 geminal. The geminal function $f(r_{12})|\widetilde{ij}\rangle$ is chosen to fit the short-range correlation hole and the difference amplitudes are weighted towards the longer range features of the correlation hole that require an anisotropic

Downloaded 11 Oct 2011 to 134.147.252.238. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

correlation factor. The pair natural orbitals that diagonalise the difference density provide a rapidly convergent series expansion for the MP2-F12 energy expression. In Sec. IV A we provide numerical results showing that the difference density definition for the PNOs is superior to the standard definition when PNOs are used in combination with F12 theory.

### C. Pair natural orbitals in the complementary space

For the projector $Q_{ij}^{A}$ in Eq. (16), we define CAPNOs through a density constructed from the geminal amplitudes $r_{p'k}^{ij}$, which replace $t_{ab}^{ij}$ in Eqs. (20)–(23), where $p'$ is the complementary space formed by the union of the standard CA orbitals $p''$ and the discarded virtual PNOs. The CAPNOs diagonalise the resulting CA density $D_{p'q'}^{A,ij}$ and form a rapidly convergent expansion for the CA space in $Q_{ij}^{A}$. This set of CAPNOs is, however, not suitable for use in the projector $Q_{ij}^{B}$ (Eq. (17)) and another set of CAPNOs is required. Here, we are faced with three possible definitions, namely, using the geminal amplitudes $r_{p'\bar{b}}^{ij}$, $r_{p'q'}^{ij}$, or $r_{p'q'''}^{ij}$ in place of $t_{ab}^{ij}$ in Eqs. (20)–(23), where $q'''$ is the union of $\bar{b}$ and $q'$. The choice $r_{p'\bar{b}}^{ij}$ will result in CAPNOs optimally suited to the $\bar{a}\bar{q}'$ and $\bar{p}'\bar{b}$ terms in $\hat{Q}_{ij}^{B}$, the choice $r_{p'q'}^{ij}$ will result in CAPNOs optimally suited to the $\bar{p}'\bar{q}'$ term and the choice $r_{p'q'''}^{ij}$ will be a compromise between the two.

In MP2-F12 theory, the energetically dominant contribution that involves $Q_{ij}^{B}$ is the exchange component of the matrix $B_{ij}^{ij}$. It has been noted by Werner *et al.* that neglecting the part in the exchange term that involves two CA indices (the hybrid approximation[43,45]) has very little impact on the accuracy for reaction energies. In CCSD(F12*) theory,[14,15] CCSD-F12x theory,[6,8] and CCSD(2)$_{\overline{\mathrm{F12}}}$ theory,[18,19] no further terms involving two CA indices appear and it, therefore, seems pragmatic to opt for CAPNOs that are suited to the $\bar{a}\bar{q}'$ and $\bar{p}'\bar{b}$ manifolds, namely, using amplitudes $r_{p'\bar{b}}^{ij}$. This is the definition adopted in this work. We observe, however, that in the full CCSD-F12 method[13,15,46,47] and in explicitly correlated triples theory,[48,49] intermediates requiring contractions over two CA indices appear and for these terms it may be advantageous to compute CAPNOs using $r_{p'q'}^{ij}$.

## III. COMPUTATIONAL DETAILS

The key quantity of interest in assessing the PNO-MP2-F12 method is the loss of accuracy as a function of the truncation of the PNO virtuals. Since we anticipate that the primary application of PNOs in F12 theory will be at the coupled cluster level, we choose to define accuracy in this MP2 study as the deviation of the correlation energy from the basis set limit, rather than from the experiment. For our assessment set we adopt the set of 106 molecules of H, C, N, O, and F atoms, compiled by Bakowies[50] and used in previous F12 assessment studies,[51–53] for which the basis set limit MP2 valence correlation energies are known to be within 99.95%.[52] Statistical error measures are reported for the basis set error of a set of 105 linearly independent reaction energies involving the 106 molecules. A list of the reactions is given in the supporting information.[54] We report all error measures in kJ/mol per valence electron.

All calculations have been performed using a developmental version of the TURBOMOLE program package.[55] The details of the MP2-F12 calculations are as follows: Occupied orbitals are localised using the Boys localisation procedure.[56] *Ansatz* 2 and approximation B is used as described in Ref. 41. As orbital basis sets we use the cc-pVXZ-F12 sets of Peterson with X = D, T, Q.[57] For the complementary auxiliary basis sets, we use the partner cc-pVXZ-F12 sets of Yousaf.[58] A Slater-type correlation factor approximated as a linear combination of six gaussians[26] with a Slater exponent of $1.0a_0^{-1}$ is used unless otherwise stated. Density fitting is used to approximate the two-electron integrals and we use the aug-cc-pwCVTZ, aug-cc-pwCVQZ, and aug-cc-pwCV5Z Coulomb fitting sets[59] in combination with the cc-pVDZ-F12, cc-pVTZ-F12, and cc-pVQZ-F12 orbital basis sets, respectively. For the Coulomb and exchange integrals in the Fock matrix, we use the aug-cc-pVTZ, aug-cc-pVQZ, and aug-cc-pV5Z RIJK basis sets, respectively.[60]

## IV. PERFORMANCE

### A. Local MP2-F12 using PNOs

Figure 1 is a plot of the mean absolute basis set error in computed PNO-MP2 and PNO-MP2-F12 correlation energies for the 106 molecules in our set. Figure 2 is the equivalent plot, but for the correlation contribution to 105 reaction energies. The basis set error per valence electron is plotted as a function of the average number of PNOs retained in the wave function expansion, where the average is over all orbital pairs in the set of 106 molecules. The 9 data points on each curve correspond to occupation number truncation thresholds of $1 \times 10^{-3}$ to $1 \times 10^{-10}$ in increments of 1/10, with a final point where all virtuals are included. No truncation of the CAPNOs was made in these PNO-MP2-F12 calculations.

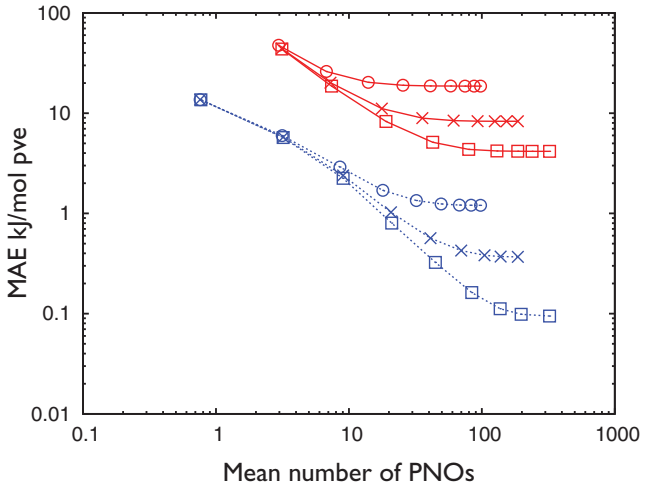



FIG. 1. Mean absolute correlation energy basis set errors for 106 molecules as a function of the average number of PNOs. The red solid lines are the results for the PNO-MP2 method and the blue dotted lines are for PNO-MP2-F12. The orbital basis sets are ○ cc-pVDZ-F12, × cc-pVTZ-F12, and □ cc-pVQZ-F12.

Downloaded 11 Oct 2011 to 134.147.252.238. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

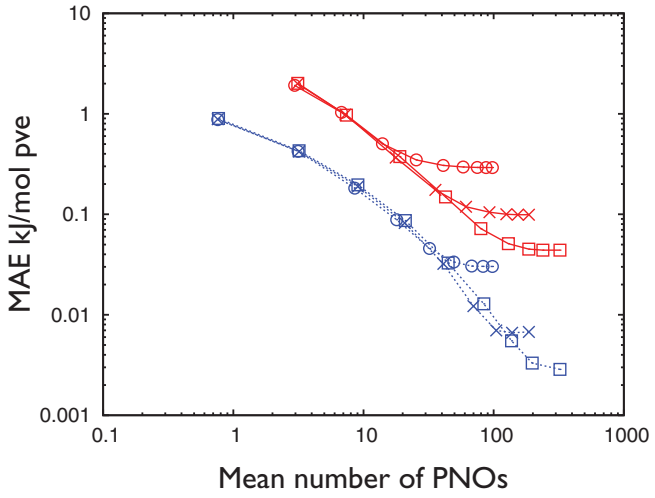



FIG. 2. Mean absolute correlation energy basis set errors for 105 reaction energies as a function of the average number of PNOs. The red solid lines are the results for the PNO-MP2 method and the blue dotted lines are for PNO-MP2-F12. The orbital basis sets are ○ cc-pVDZ-F12, × cc-pVTZ-F12, and □ cc-pVQZ-F12.

From Fig. 1 it is evident that the accelerated basis set convergence observed for standard MP2-F12 theory carries through to PNO-MP2-F12 theory. Although the factor of improvement of F12 over orbital-only energies is not as large for truncated PNO virtual space as for untruncated expansions, it is nonetheless considerable. For relative energies, Fig. 2 shows that on average four to five times fewer virtuals are required for PNO-MP2-F12 than for PNO-MP2 to achieve a given basis set error in the correlation energy. At the CCSD level of theory, which involves matrix operations that scale as $O^2V^4$ and $O^3V^3$, where $O$ is the number of occupied and $V$ is the number of virtual orbitals, combining F12 geminals with PNO truncations will constitute a considerable saving in computational effort. It is remarkable that for both PNO-MP2 and PNO-MP2-F12, the curves for the cc-pVDZ-F12, cc-pVTZ-F12, and cc-pVQZ-F12 basis sets lie almost perfectly on top of each other in the limit of severe PNO truncation. This feature is much more pronounced for relative energies than for total energies. It is also noteworthy that for the PNO-MP2 method, the slope of the curves in this region is between −1.0 and −1.1, which is very close to the value of −1 that would be expected from an $X^{-3}$ energy convergence in correlation consistent basis sets.[61] However, preliminary investigations indicate that the behaviour is not sufficiently regular for a simple $N^{-1}$ extrapolation to be useful, where $N$ is the average number of PNOs per significant pair.

In Tables I and II we report further statistical measures for the basis set errors displayed in Fig. 2. Using the PNO-MP2-F12 method, an average of 20 PNOs (maximum of 40) is sufficient to achieve a mean absolute error (MAE) of 0.1 kJ/mol per valence electron, and to keep the maximum error (MAE) below 0.4 kJ/mol per valence electron for our test set. For such calculations there is no need to use a basis set larger than cc-pVDZ-F12. For higher accuracy, using the cc-pVTZ-F12 basis, an average of 70 PNOs (maximum of 120) reduces MAE and MAX errors to 0.01 and 0.05 kJ/mol per valence electron, respectively.

Figure 3 compares PNO-MP2-F12 results using two definitions of PNOs, those obtained from the MP2 amplitudes equation (23) and from the difference amplitudes equation (27). No CAPNO truncation was applied in these calculations. Examining the cc-pVQZ-F12 curve for the simpler approach, one sees that upon truncation the mean absolute error passes through the untruncated cc-pVTZ-F12 and cc-pVDZ-F12 data points. This indicates that PNOs have been discarded

TABLE I. Statistical measures for the basis set error in correlation contributions to 105 reaction energies computed using the PNO-MP2 method with various PNO truncation thresholds. Energies are in kJ/mol per valence electron.

| | | $-\log_{10}$ (Occupation number thresholds) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Basis | | ∞ | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 |
| cc-pVDZ-F12 | MAE[a] | 0.292 | 0.292 | 0.293 | 0.296 | 0.307 | 0.348 | 0.503 | 1.033 | 1.918 |
| | RMS[b] | 0.371 | 0.371 | 0.372 | 0.376 | 0.390 | 0.442 | 0.648 | 1.346 | 2.537 |
| | MAX[c] | 1.152 | 1.152 | 1.154 | 1.162 | 1.207 | 1.366 | 2.064 | 3.941 | –7.418 |
| | PNO[d] | 98 | 87 | 75 | 58 | 41 | 26 | 14 | 7 | 3 |
| | NMX[e] | 161 | 152 | 133 | 99 | 76 | 49 | 29 | 15 | 7 |
| cc-pVTZ-F12 | MAE | 0.099 | 0.099 | 0.100 | 0.105 | 0.119 | 0.175 | 0.367 | 0.984 | 1.994 |
| | RMS | 0.130 | 0.131 | 0.132 | 0.136 | 0.154 | 0.227 | 0.488 | 1.286 | 2.639 |
| | MAX | 0.398 | 0.399 | 0.401 | 0.416 | 0.487 | 0.744 | 1.532 | −3.780 | 9.148 |
| | PNO | 186 | 153 | 125 | 93 | 61 | 36 | 18 | 7 | 3 |
| | NMX | 299 | 271 | 219 | 161 | 114 | 69 | 38 | 17 | 7 |
| cc-pVQZ-F12 | MAE | 0.044 | 0.044 | 0.045 | 0.051 | 0.072 | 0.149 | 0.375 | 0.971 | 2.009 |
| | RMS | 0.058 | 0.059 | 0.060 | 0.066 | 0.091 | 0.194 | 0.489 | 1.265 | 2.642 |
| | MAX | 0.155 | 0.156 | 0.160 | 0.184 | 0.274 | 0.589 | 1.485 | −3.547 | 9.144 |
| | PNO | 323 | 239 | 186 | 130 | 80 | 43 | 19 | 7 | 3 |
| | NMX | 520 | 418 | 307 | 230 | 151 | 85 | 41 | 18 | 7 |

[a]Mean absolute error.
[b]Root mean squared error.
[c]Maximum error.
[d]Mean number of PNOs that survive truncation.
[e]Maximum number of PNOs that survive truncation.

Downloaded 11 Oct 2011 to 134.147.252.238. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

TABLE II. Statistical measures for the basis set error in correlation contributions to 105 reaction energies computed using the PNO-MP2-F12 method with various PNO truncation thresholds. Energies are in kJ/mol per valence electron.

| | | $-\log_{10}$ (Occupation number thresholds) | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Basis | | $\infty$ | 10 | 9 | 8 | 7 | 6 | 5 | 4 | 3 |
| cc-pVDZ-F12 | MAE[a] | 0.030 | 0.030 | 0.031 | 0.034 | 0.046 | 0.088 | 0.182 | 0.423 | 0.875 |
| | RMS[b] | 0.039 | 0.039 | 0.040 | 0.044 | 0.061 | 0.116 | 0.240 | 0.572 | 1.185 |
| | MAX[c] | −0.115 | −0.115 | −0.118 | −0.133 | 0.182 | −0.386 | −0.762 | −2.131 | 4.268 |
| | PNO[d] | 98 | 83 | 68 | 49 | 32 | 18 | 9 | 3 | 1 |
| | NMX[e] | 161 | 148 | 123 | 87 | 59 | 36 | 21 | 10 | 5 |
| cc-pVTZ-F12 | MAE | 0.007 | 0.007 | 0.007 | 0.012 | 0.032 | 0.082 | 0.189 | 0.421 | 0.883 |
| | RMS | 0.009 | 0.009 | 0.009 | 0.016 | 0.041 | 0.105 | 0.246 | 0.569 | 1.198 |
| | MAX | −0.027 | 0.026 | 0.027 | −0.049 | −0.137 | −0.356 | −0.852 | −2.181 | 4.316 |
| | PNO | 186 | 139 | 105 | 70 | 41 | 21 | 9 | 3 | 1 |
| | NMX | 299 | 238 | 179 | 123 | 78 | 43 | 23 | 10 | 5 |
| cc-pVQZ-F12 | MAE | 0.003 | 0.003 | 0.006 | 0.013 | 0.033 | 0.087 | 0.196 | 0.429 | 0.898 |
| | RMS | 0.004 | 0.004 | 0.007 | 0.016 | 0.042 | 0.111 | 0.257 | 0.579 | 1.206 |
| | MAX | 0.021 | 0.018 | 0.016 | −0.044 | −0.133 | −0.351 | −0.889 | −2.204 | 4.325 |
| | PNO | 323 | 198 | 137 | 84 | 45 | 21 | 9 | 3 | 1 |
| | NMX | 520 | 327 | 236 | 152 | 87 | 45 | 23 | 10 | 5 |

[a]Mean absolute error.
[b]Root mean squared error.
[c]Maximum error.
[d]Mean number of PNOs that survive truncation.
[e]Maximum number of PNOs that survive truncation.

that are important in MP2-F12 calculations, whereas PNOs that are unimportant, presumably those that fit the short-range Coulomb hole, have been kept. Using the difference density approach, the trend of the curves with truncation is much closer to that observed for the PNO-MP2 method and fewer virtuals are required for a given accuracy. PNOs obtained from the difference amplitudes are clearly superior and this definition has been used in all other PNO-MP2-F12 calculations reported in this work.

It has been observed for MP2-F12 calculations that the dependence on the Slater exponent $\gamma$ in the correlation factor is stronger for smaller basis sets than for larger basis sets and that the optimal $\gamma$ increases with increasing basis size.[26] Figure 4 displays the $\gamma$ dependence in PNO-MP2-F12 calculations as a function of PNO truncation. Again, no CAPNO truncation was applied in these calculations. From Fig. 4 it can be seen that the $\gamma$ dependence in PNO-MP2-F12 mirrors the trends observed for MP2-F12. In absolute terms, the basis set errors depend more strongly on $\gamma$ for small PNO sets than for large PNO sets. The $\gamma$ that gives lowest errors is predictably $1.0\ a_0^{-1}$ in the limit of no truncation, but upon truncation, the optimum value increases to around $1.2\ a_0^{-1}$, at least for the

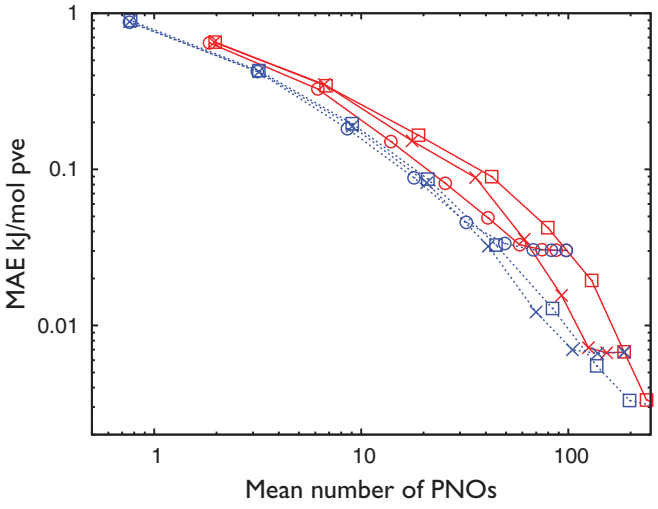



FIG. 3. Mean absolute correlation energy basis set errors for 105 reaction energies as a function of the average number of PNOs. The red solid lines are the PNO-MP2-F12 results using PNOs obtained from the MP2 amplitudes equation (23). The blue dotted lines use PNOs obtained from difference amplitudes equation (27). The orbital basis sets are ○ cc-pVDZ-F12, × cc-pVTZ-F12, and □ cc-pVQZ-F12.

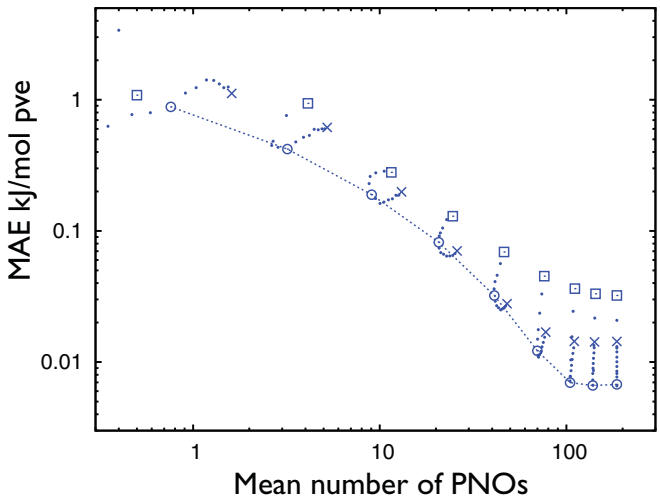



FIG. 4. Mean absolute correlation energy basis set errors for 105 reaction energies computed using the cc-pVTZ-F12 basis and various $\gamma$ exponents and occupation number thresholds, plotted as a function of the average number of PNOs. $\gamma = 0.5, 0.6, \ldots, 1.8a_0^{-1}$ with $\gamma = 0.5$ □, $\gamma = 1.8$ ×, and the dashed line connects points ○ with $\gamma = 1.0a_0^{-1}$.

Downloaded 11 Oct 2011 to 134.147.252.238. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

valence electrons of the molecules in our test set. In the limit of full truncation, $\gamma$ values of less than 1.0 $a_0^{-1}$ are advantageous, but the errors are too large for this to be of practical use.

### B. Local MP2-F12 using CAPNOs

In Fig. 5 we plot the basis set errors for the set of 105 reaction energies as a function of the number of CAPNOs included in the evaluation of the MP2-F12 intermediates. The results plotted are for PNO-MP2-F12 calculations using the cc-pVDZ-F12 orbital basis and a PNO truncation at occupation number threshold $1 \times 10^{-6}$. Two sets of calculations are presented, one where the CAPNOs for $\hat{Q}_{ij}^{A}$ are truncated, but those used for $\hat{Q}_{ij}^{B}$ are not, and vice versa. As before, the data points correspond to occupation number thresholds of $1 \times 10^{-3}$ to $1 \times 10^{-10}$ in increments of 1/10, with a final point where all virtuals are included. It is remarkable that for type-A CAPNOs even an occupation number threshold of $1 \times 10^{-10}$ discards all but a maximum of 40 orbitals (20 on average) for the molecules in our set, with a completely negligible loss of accuracy in the RI. In fact, our data suggests that for more severe truncation, with an average of only 10 CAPNOs, error cancellation leads to even smaller basis set errors. The basis set errors for type-B CAPNOs are significantly smaller than for type-A, since only a part of the exchange contribution to the energy is approximated using $\hat{Q}_{ij}^{B}$ in MP2-F12 theory. Type-B CAPNO truncation at $1 \times 10^{-8}$ leaves an average of 25 and a maximum of 50 orbitals with a similar impressive level of accuracy as for type-A CAPNOs. This constitutes an enormous saving compared to the CABS RI used in standard F12 theory.

### C. Large molecules

The largest molecule in the test set is $C_2F_4$, which has 18 valence orbitals. To investigate whether the conclusions of Secs. IV A and IV B are sensitive to the size of the molecule under consideration, we have performed PNO-MP2-F12 calculations on ethylenedioxy-tetrathiafulvalene (EDO-TTF), $C_8H_6O_2S_4$ (37 valence orbitals), leflunomide, $C_{12}H_9N_2O_2F_3$ (50 valence orbitals), and the interaction of methanol with a cluster model for zeolite H-ZSM-5, $CH_3OH\cdots HAl(OH)(OSiH_3)_3$ (32 valence orbitals). For these systems, MP2-F12/cc-pVXZ-F12 energies with X = D, T, and Q have been published previously in Ref. 41. The geometries were taken from the supporting information of Ref. 41.

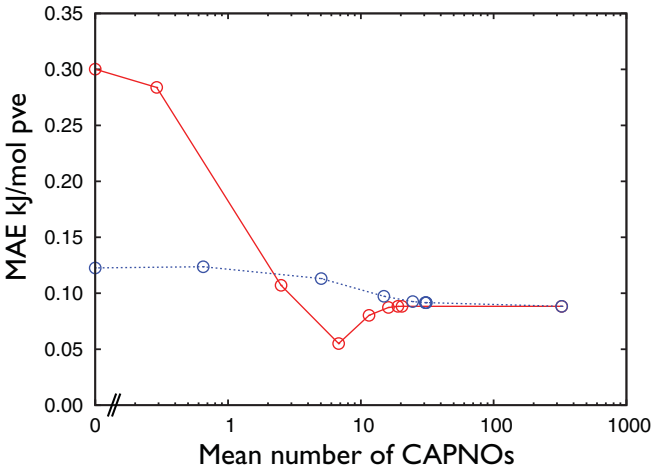



FIG. 5. Mean absolute correlation energy basis set errors for 105 reaction energies as a function of the average number of CAPNOs. The red solid line plots truncation of type-A CAPNOs and the blue dotted line truncation of type-B CAPNOs. The cc-pVDZ-F12 orbital basis set was used with PNO truncation at occupation numbers less than $1 \times 10^{-6}$.

In Table III we report valence correlation energies of EDO-TTF computed using the PNO-MP2-F12 method for a range of choices of thresholds. Specifically, MP2-F12 pair energies are discarded if the conventional MP2 pair contribution is less than $\epsilon_{ij}$ (the strong pair approximation[62]) and PNOs and CAPNOs are discarded if the occupation numbers are less than $n_{\mathrm{occ}}$. Similar results for leflunomide and the cluster model are reported in Tables IV and V, respectively.

Taking leflunomide as an example, we see that with the cc-pVDZ-F12 basis and a loose threshold $n_{\mathrm{occ}} = 1 \times 10^{-6}$, at most 40 PNOs survive truncation for any given pair. For a tighter threshold of $n_{\mathrm{occ}} = 1 \times 10^{-8}$, at most 120 survive. This is in accord with the values reported in Table II for the smaller molecules in our test set. Concerning the CAPNOs, the number that survive with $n_{\mathrm{occ}} = 1 \times 10^{-6}$ is also very similar to the values displayed in Fig. 5. Furthermore, the loss of accuracy due to PNO and CAPNO truncation is comparable to that observed for the test set. Using a tight $n_{\mathrm{occ}} = 1 \times 10^{-8}$ threshold, the correlation energy is within 0.2% of the untruncated value, provided that the pair energy threshold is also tight $\epsilon_{ij} = 1 \times 10^{-5}$. With a looser $n_{\mathrm{occ}} = 1 \times 10^{-6}$ threshold, ~0.5% of the correlation energy is lost. The results for leflunomide and the cluster model both display a slight increase in the computed correlation energy with tight thresholds. This results from the CABS RI error, and is also seen in Fig. 5.

The PNO and CAPNO truncation errors appear to be similar in magnitude and this raises the question of whether there is a fortuitous cancellation of errors. In Table VI we report interaction energies for methanol on the model zeolite cluster, without structural relaxation (for structures, see Ref. 41). The untruncated, counterpoise corrected cc-pVDZ-F12 value is $-13.07$ mE$_h$, which is 97.3% of the basis set limit value of $-13.43$ mE$_h$. With tight thresholds of $\epsilon_{ij} = 1 \times 10^{-6}$ and $n_{\mathrm{occ}} = 1 \times 10^{-8}$, an interaction energy of $-13.01$ mE$_h$ is obtained, a loss in accuracy of only 0.4%. This should be compared with the total energy of the complex in Table V, where the loss of accuracy was less than 0.1%. Although the RI error due to truncation is at least 0.9 mE$_h$ in the complex, the combined PNO and CAPNO truncation is only 0.06 mE$_h$ in the interaction energy. This accuracy is impressive considering that a maximum of only 62 out of a possible 413 PNOs survive truncation and that the level of CAPNO truncation is similarly high. The accuracy of the interaction energy deteriorates by further 0.6% when using a loose threshold of $n_{\mathrm{occ}} = 1 \times 10^{-6}$. It is critical, however, for these interaction energies to have a tight threshold for $\epsilon_{ij}$.

To summarise, our results demonstrate the expected result that the level of truncation afforded for each significant

Downloaded 11 Oct 2011 to 134.147.252.238. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

TABLE III. PNO-MP2-F12 valence correlation energies of EDO-TTF with various pair energy $\epsilon_{ij}$ and occupation number $n_{\mathrm{occ}}$ thresholds. The basis set limit is −2.758(2) $E_{\mathrm{h}}$, see Ref. 41.

| Basis | $-\log(\epsilon_{ij})$ | $-\log(n_{\mathrm{occ}})$ | $N_{ij}$[a] | PNO[b] | CAPNO-A[c] | CAPNO-B[d] | $E_{\mathrm{corr}}$ | % |
|---|---|---|---|---|---|---|---|---|
| cc-pVDZ-F12 | ∞ | ∞ | 703 | 443 | 1056 | 1056 | −2.7220 | 98.7 |
| cc-pVDZ-F12 | 6 | 8 | 670 | 41 (93) | 14 (29) | 34 (90) | −2.7177 | 98.5 |
| cc-pVDZ-F12 | 5 | 8 | 551 | 48 (93) | 16 (29) | 41 (90) | −2.7169 | 98.5 |
| cc-pVDZ-F12 | 4 | 8 | 413 | 57 (93) | 19 (29) | 50 (90) | −2.7084 | 98.2 |
| cc-pVDZ-F12 | 6 | 6 | 603 | 11 (37) | 4 (10) | 6 (28) | −2.7071 | 98.2 |
| cc-pVDZ-F12 | 5 | 6 | 551 | 12 (37) | 4 (10) | 7 (28) | −2.7066 | 98.1 |
| cc-pVDZ-F12 | 4 | 6 | 413 | 15 (37) | 5 (10) | 9 (28) | −2.6977 | 97.8 |
| cc-pVTZ-F12 | ∞ | ∞ | 703 | 819 | 2115 | 2115 | −2.7493 | 99.7 |
| cc-pVTZ-F12 | 6 | 8 | 670 | 49 (132) | 11 (24) | 38 (115) | −2.7460 | 99.6 |
| cc-pVTZ-F12 | 5 | 8 | 553 | 58 (132) | 13 (24) | 46 (115) | −2.7453 | 99.5 |
| cc-pVTZ-F12 | 4 | 8 | 413 | 70 (132) | 16 (24) | 115 (115) | −2.7365 | 99.2 |
| cc-pVTZ-F12 | 6 | 6 | 606 | 12 (39) | 3 (9) | 6 (27) | −2.7321 | 99.1 |
| cc-pVTZ-F12 | 5 | 6 | 553 | 13 (39) | 3 (9) | 7 (27) | −2.7316 | 99.0 |
| cc-pVTZ-F12 | 4 | 6 | 413 | 16 (39) | 4 (9) | 9 (27) | −2.7226 | 98.7 |

[a]Number of pairs that survive truncation.
[b]Mean number of PNOs that survive truncation (maximum in parentheses).
[c]Mean number of type-A CAPNOs that survive truncation (maximum in parentheses).
[d]Mean number of type-B CAPNOs that survive truncation (maximum in parentheses).

pair is not sensitive to the size of molecule. This, coupled with the fact that the number of significant pairs will eventually grow only linearly with the size of the molecule, is very encouraging for the applicability of our approach to larger chemical problems in the framework of coupled-cluster theory.

## V. DISCUSSION

The high degree of truncation possible for the CAPNOs can be understood from the close relationship of pair natural orbitals with a singular value decomposition of the amplitudes.[63] In a spin-orbital formalism, if the amplitudes have singular values $\omega$

$$t_{ab}^{ij} = \sum_{\bar{c}} L_{\bar{c}a}^{ij} \omega_{\bar{c}} L_{\bar{c}b}^{ij}, \tag{28}$$

then

$$D_{ab}^{ij} = \frac{1}{1 + \sum_{\bar{c}} \omega_{\bar{c}}^2} \sum_{\bar{c}} L_{\bar{c}a}^{ij} \omega_{\bar{c}}^2 L_{\bar{c}b}^{ij}. \tag{29}$$

The eigenvectors that diagonalise the pair density are the eigenvectors of the amplitude matrix. In the case of type-A CAPNOs, the number of non-zero singular values of the amplitudes $r_{p'k}^{ij}$ is bounded by the number of occupied orbitals. Similarly, the number of type-B CAPNOs that contribute to $r_{p'\bar{b}}^{ij}$ is bounded by the number of PNOs for the pair $ij$. This clearly explains the high level of truncation observed in

TABLE IV. PNO-MP2-F12 valence correlation energies of leflunomide with various pair energy $\epsilon_{ij}$ and occupation number $n_{\mathrm{occ}}$ thresholds. The basis set limit is −3.9615(20) $E_{\mathrm{h}}$, see Ref. 41.

| Basis | $-\log(\epsilon_{ij})$ | $-\log(n_{\mathrm{occ}})$ | $N_{ij}$[a] | PNO[b] | CAPNO-A[c] | CAPNO-B[d] | $E_{\mathrm{corr}}$ | % |
|---|---|---|---|---|---|---|---|---|
| cc-pVDZ-F12 | ∞ | ∞ | 1275 | 582 | 1452 | 1452 | −3.9105 | 98.7 |
| cc-pVDZ-F12 | 6 | 8 | 973 | 35 (120) | 12 (32) | 30 (119) | −3.9139 | 98.8 |
| cc-pVDZ-F12 | 5 | 8 | 769 | 42 (120) | 15 (32) | 38 (119) | −3.9127 | 98.8 |
| cc-pVDZ-F12 | 4 | 8 | 510 | 54 (120) | 18 (32) | 51 (119) | −3.8977 | 98.4 |
| cc-pVDZ-F12 | 6 | 6 | 859 | 10 (40) | 4 (11) | 6 (28) | −3.8986 | 98.4 |
| cc-pVDZ-F12 | 5 | 6 | 769 | 11 (40) | 4 (11) | 7 (28) | −3.8977 | 98.4 |
| cc-pVDZ-F12 | 4 | 6 | 510 | 14 (40) | 6 (11) | 10 (28) | −3.8823 | 98.0 |
| cc-pVTZ-F12 | ∞ | ∞ | 1275 | 1100 | 1794 | 1794 | −3.9500 | 99.7 |
| cc-pVTZ-F12 | 6 | 8 | 974 | 41 (154) | 10 (28) | 33 (142) | −3.9459 | 99.6 |
| cc-pVTZ-F12 | 5 | 8 | 774 | 50 (154) | 12 (28) | 42 (142) | −3.9447 | 99.6 |
| cc-pVTZ-F12 | 4 | 8 | 515 | 65 (154) | 15 (28) | 57 (142) | −3.9301 | 99.2 |
| cc-pVTZ-F12 | 6 | 6 | 860 | 11 (44) | 3 (10) | 6 (27) | −3.9271 | 99.1 |
| cc-pVTZ-F12 | 5 | 6 | 774 | 12 (44) | 3 (10) | 7 (27) | −3.9263 | 99.1 |
| cc-pVTZ-F12 | 4 | 6 | 515 | 15 (44) | 5 (10) | 10 (27) | −3.9114 | 98.7 |

[a]Number of pairs that survive truncation.
[b]Mean number of PNOs that survive truncation (maximum in parentheses).
[c]Mean number of type-A CAPNOs that survive truncation (maximum in parentheses).
[d]Mean number of type-B CAPNOs that survive truncation (maximum in parentheses).

Downloaded 11 Oct 2011 to 134.147.252.238. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

TABLE V. PNO-MP2-F12 valence correlation energies of the methanol complexed to a cluster model for zeolite H-ZSM-5, with various pair energy $\epsilon_{ij}$ and occupation number $n_{\mathrm{occ}}$ thresholds. The basis set limit is −2.128(1) $E_{\mathrm{h}}$, see Ref. 41.

| Basis | $-\log(\epsilon_{ij})$ | $-\log(n_{\mathrm{occ}})$ | $N_{ij}$[a] | PNO[b] | CAPNO-A[c] | CAPNO-B[d] | $E_{\mathrm{corr}}$ | % |
|---|---|---|---|---|---|---|---|---|
| cc-pVDZ-F12 | ∞ | ∞ | 528 | 413 | 990 | 990 | −2.1005 | 98.7 |
| cc-pVDZ-F12 | 6 | 8 | 528 | 25 (62) | 8 (21) | 20 (69) | −2.1014 | 98.8 |
| cc-pVDZ-F12 | 5 | 8 | 445 | 29 (62) | 9 (21) | 24 (69) | −2.1008 | 98.7 |
| cc-pVDZ-F12 | 4 | 8 | 237 | 43 (62) | 14 (21) | 39 (69) | −2.0920 | 98.3 |
| cc-pVDZ-F12 | 6 | 6 | 519 | 8 (29) | 2 (9) | 4 (24) | −2.0961 | 98.5 |
| cc-pVDZ-F12 | 5 | 6 | 445 | 9 (29) | 2 (9) | 5 (24) | −2.0957 | 98.5 |
| cc-pVDZ-F12 | 4 | 6 | 237 | 13 (29) | 4 (9) | 9 (24) | −2.0868 | 98.1 |
| cc-pVTZ-F12 | ∞ | ∞ | 528 | 778 | 1365 | 1365 | −2.1206 | 99.7 |
| cc-pVTZ-F12 | 6 | 8 | 528 | 31 (91) | 7 (17) | 23 (88) | −2.1194 | 99.6 |
| cc-pVTZ-F12 | 5 | 8 | 445 | 35 (91) | 8 (17) | 27 (88) | −2.1188 | 99.6 |
| cc-pVTZ-F12 | 4 | 8 | 237 | 57 (91) | 11 (17) | 46 (88) | −2.1106 | 99.2 |
| cc-pVTZ-F12 | 6 | 6 | 528 | 9 (32) | 2 (7) | 4 (26) | −2.1118 | 99.2 |
| cc-pVTZ-F12 | 5 | 6 | 445 | 10 (32) | 2 (7) | 5 (26) | −2.1114 | 99.2 |
| cc-pVTZ-F12 | 4 | 6 | 237 | 14 (32) | 4 (7) | 9 (26) | −2.1033 | 98.8 |

[a]Number of pairs that survive truncation.
[b]Mean number of PNOs that survive truncation (maximum in parentheses).
[c]Mean number of type-A CAPNOs that survive truncation (maximum in parentheses).
[d]Mean number of type-B CAPNOs that survive truncation (maximum in parentheses).

Fig. 5 and the fact that fewer type-A CAPNOs are required than type-B. Note that the upper bound to the number of non-zero singular values is independent of the definition of the occupied orbitals and that truncation of the CAPNOs is also possible without loss of accuracy when using (semi-)canonical orbitals. Using localised orbitals further decreases the number of significant CAPNOs through the zero overlap of orbitals $i$, $j$ with occupied orbitals $k$ in the projection operator that are spatially distant.

An alternative view of the above is obtained by considering an explicit RI insertion, for example,

$$\langle ij|f(r_{12})\hat{O}_1 f(r_{12})|ij\rangle = \sum_{k\mu}\langle ij|f(r_{12})|k\mu\rangle\langle k\mu|f(r_{12})|ij\rangle, \tag{30}$$

TABLE VI. The counterpoise corrected correlation contribution to the interaction energy of methanol with a cluster model for zeolite H-ZSM-5, with various pair energy $\epsilon_{ij}$ and occupation number $n_{\mathrm{occ}}$ thresholds. The basis set limit is −13.43(1) m$E_{\mathrm{h}}$, see Ref. 41.

| Basis | $-\log(\epsilon_{ij})$ | $-\log(n_{\mathrm{occ}})$ | $\Delta E$ | CP | $\Delta E_{\mathrm{CP}}$ | % |
|---|---|---|---|---|---|---|
| cc-pVDZ-F12 | ∞ | ∞ | −13.38 | 0.32 | −13.07 | 97.3 |
| cc-pVDZ-F12 | 6 | 8 | −13.41 | 0.40 | −13.01 | 96.9 |
| cc-pVDZ-F12 | 5 | 8 | −13.00 | 0.40 | −12.60 | 93.9 |
| cc-pVDZ-F12 | 4 | 8 | −9.56 | 0.53 | −9.03 | 67.2 |
| cc-pVDZ-F12 | 6 | 6 | −13.20 | 0.28 | −12.93 | 96.2 |
| cc-pVDZ-F12 | 5 | 6 | −12.92 | 0.28 | −12.64 | 94.1 |
| cc-pVDZ-F12 | 4 | 6 | −9.47 | 0.41 | −9.05 | 67.4 |
| cc-pVTZ-F12 | ∞ | ∞ | −13.53 | 0.19 | −13.34 | 99.3 |
| cc-pVTZ-F12 | 6 | 8 | −13.32 | 0.16 | −13.16 | 98.0 |
| cc-pVTZ-F12 | 5 | 8 | −12.95 | 0.16 | −12.78 | 95.2 |
| cc-pVTZ-F12 | 4 | 8 | −9.70 | 0.16 | −9.54 | 71.0 |
| cc-pVTZ-F12 | 6 | 6 | −13.00 | 0.18 | −12.82 | 95.4 |
| cc-pVTZ-F12 | 5 | 6 | −12.75 | 0.18 | −12.57 | 93.6 |
| cc-pVTZ-F12 | 4 | 6 | −9.49 | 0.18 | −9.31 | 69.3 |

where $\mu$ is a complete set of orthonormal orbitals. For each orbital $k$, there is a specific linear combination of the complete set of orbitals, $\tilde{\mu}^k$ such that

$$\langle ij|f(r_{12})\hat{O}_1 f(r_{12})|ij\rangle = \sum_k \langle ij|f(r_{12})|k\tilde{\mu}^k\rangle\langle k\tilde{\mu}^k|f(r_{12})|ij\rangle, \tag{31}$$

where

$$\phi_{\tilde{\mu}^k}(2) = N\int d1\,\phi_i(1)\phi_j(2)f(r_{12})\phi_k(1) \tag{32}$$

and $N$ is a normalisation constant (note that we have not imposed orthogonality on the orbitals $\langle\tilde{\mu}^k|\tilde{\mu}^l\rangle \neq \delta_{kl}$). This viewpoint makes clear the importance of the diagonal structure of the F12 terms. It also emphasises that even though the number of CAPNOs required is small, the extent of the CA basis must be sufficiently large for the RI to be accurate. In our approach, by moving discarded PNOs into the CA space, we ensure that the extent of the underlying CA space is not curtailed before the construction of CAPNOs.

In the calculations reported in Secs. IV A–IV C, we have applied an occupation number threshold to the CAPNOs in the same way as for the virtual PNOs. However, since the number of CAPNOs is bounded, one may as well keep all CAPNOs with numerically non-vanishing singular values and do away with the CAPNO occupation number threshold. In this way there will be a vanishing truncation errors for type-A CAPNOs, and the type-B CAPNOs in the so called C and D (or ring) terms in the coupled-cluster expressions.

One important consideration remains to be discussed. When selecting the density used to define the type-B CAPNOs, we opted to use the amplitudes $r^{ij}_{p'\bar{b}}$ and to ignore the amplitudes $r^{ij}_{p'q'}$. The number of non-vanishing singular values for the amplitudes $r^{ij}_{p'q'}$ is, in principle, much larger than for $r^{ij}_{p'\bar{b}}$. Furthermore, using the CAPNOs defined using $r^{ij}_{p'\bar{b}}$ to

Downloaded 11 Oct 2011 to 134.147.252.238. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

expand $r^{ij}_{p'q'}$ is guaranteed on symmetry grounds to miss vectors that are important. The justification for our choice is that these amplitudes only appear in the MP2-F12 and CCSD(F12*) equations in a single term, which is known to be small and have negligible impact on relative energies in the canonical methods. The question remains: how large an error does our choice introduce and would a better choice be to adopt the hybrid approximation and discard these terms completely. We have therefore investigated three possibilities: (i) defining type-B CAPNOs using the $r^{ij}_{p'\bar{b}}$ amplitudes and keeping all terms in the Lagrangian, (ii) defining type-B CAPNOs using the $r^{ij}_{p'\bar{b}}$ amplitudes and using the hybrid approximation to discard terms that depend on $r^{ij}_{p'q'}$, and (iii) defining type-B CAPNOs using $r^{ij}_{p'q'''}$. The mean absolute basis set reaction energy errors as a function of the mean number of type-B CAPNOs are plotted in Fig. 6. The type-A CAPNOs are not truncated and for the PNOs, $n_{\mathrm{occ}} = 1 \times 10^{-6}$ and $n_{\mathrm{occ}} = 1 \times 10^{-8}$ are used with the cc-pVDZ-F12 and cc-pVTZ-F12 basis sets, respectively.

From Fig. 6 we see the error in assuming that the hybrid approximation is only 10% of the basis set error in the reaction energies, which affirms its validity in the present context. Concerning the alternative definition of the type-B CAPNOs, we observe the expected result that the number of CAPNOs decreases slowly as a function of the occupation number threshold, since the number of non-zero singular values is not bound by the number of PNOs. Interestingly, the improvement in accuracy over the original definition is only marginal and the results, in fact, worsen at more severe CAPNO occupation number truncation thresholds. Our results indicate that the best choice in terms of accuracy is to use a density constructed from $r^{ij}_{p'\bar{b}}$ and to not assume the hybrid approximation. In future developments one could instead consider different strategies to compute the contributions from the pair-specific projectors $\sum_{\bar{p}'\bar{q}'} |\bar{p}'\bar{q}'\rangle\langle\bar{p}'\bar{q}'|$ to the MP2-F12 intermediates.

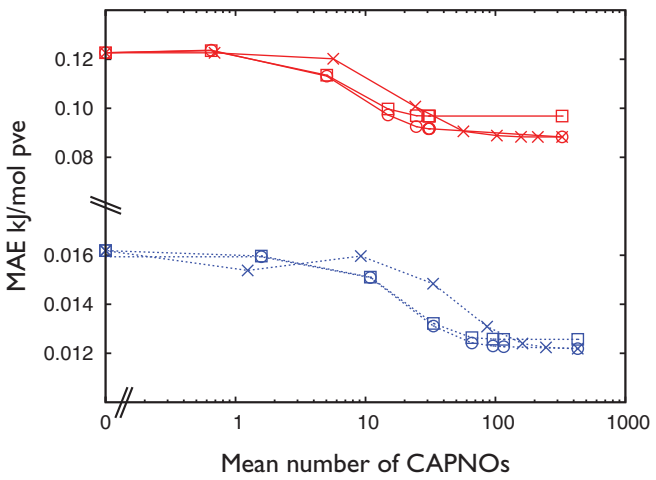



FIG. 6. Mean absolute correlation energy basis set errors for 105 reaction energies as a function of the average number of type-B CAPNOs with $-\log(n_{\mathrm{occ}}) = 3, 4, \ldots, 10, \infty$. ○ and □ are for CAPNOs constructed using $r^{ij}_{p'\bar{b}}$ but, □ is with the hybrid approximation. × are for CAPNOs constructed using $r^{ij}_{p'q'''}$. Calculations using an cc-pVDZ-F12 are indicated by a red solid line and cc-pVTZ-F12 with a blue dotted line.

## VI. CONCLUSION

The pair natural orbital approach has two appealing features as compared to domain-based localisation approaches and also to the more recent work of Yang *et al.*[64] that utilises tensor decompositions into orbital contributions (rather than pair contributions). Firstly, in the PNO approach, the number of virtuals retained adapts in a near optimal manner according to the complexity of the pair correlation function being represented. Secondly, in our formulation we are able to minimise the redundancy between the geminal and the orbital pair functions when truncating the virtual space. One disadvantage, however, is that the computational overhead of forming the PNOs remains significant, but a local density fitting scheme such as the Mulliken-based approach of Neese *et al.*[31] can be used to reduce this.

Our investigations have demonstrated that a pair natural orbital analysis of the first-order correlation function is a very attractive means by which to truncate the space of double excitations in explicitly correlated methods, when using localised occupied orbitals. We find that, as expected, far more severe truncation of the virtual space is possible when F12 geminal functions are included in the wave function expansion, compared to non-explicitly correlated PNO methods. For the reaction energies in our test set, it was only necessary to retain an average of 40 virtual orbitals per significant pair, using the cc-pVTZ-F12 basis, to ensure a root mean squared error of only 2 kJ/mol compared to the basis set limit (with a maximum error of 6 kJ/mol).

We have also extended the ideas behind the pair natural orbital analysis to introduce complementary auxiliary pair natural orbitals, CAPNOs, which can be truncated to reduce the number of orbitals in the resolution of the identity used to approximate the three- and four-electron integrals in F12 methods. We find that separate definitions of CAPNOs are required for the two approximations to the strong orthogonality operator used in F12 theory. For the first, more accurate approximation, the number of CAPNOs that have non-negligible contribution for each pair is bounded by the number of nearby occupied orbitals. For the second approximation, our calculations indicate that it is neither necessary nor desirable to define CAPNOs suited to the terms that depend quadratically on the CA orbitals, choosing instead to focus on the terms with linear CA dependence. Consequently, the number of contributing CAPNOs is bounded by the number of PNOs for that pair. Therefore, although the atomic CA basis should be sufficiently saturated up to an angular momentum three times the highest angular momentum of the occupied orbitals, the number of CAPNOs per pair required for an accurate RI is small and does not increase with the system size.

The enormous reduction of both the virtual and complementary auxiliary spaces afforded by the PNO approach provides an excellent platform for the development of efficient explicitly correlated coupled-cluster programs that aim at applications to larger molecules.

Downloaded 11 Oct 2011 to 134.147.252.238. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions

## ACKNOWLEDGMENTS

This work was supported by the Deutsche Forschungsgemeinschaft. D.P.T thanks the Royal Society for support from the University Research Fellow scheme.

[1] D. P. Tew, W. Klopper, C. Neiss, and C. Hättig, Phys. Chem. Chem. Phys. **9**, 1921 (2007).
[2] D. P. Tew, W. Klopper, C. Neiss, and C. Hättig, Phys. Chem. Chem. Phys. **10**, 6325 (2008).
[3] D. P. Tew, C. Hättig, R. A. Bachorz, and W. Klopper, in *Recent Progress in Coupled Cluster Methods – Theory and Applications*, edited by P. Čársky, J. Paldus, and J. Pittner (Springer, Dordrecht/Heidelberg/London/New York, 2010), pp. 535–572, ISBN 978-90-481-2884-6.
[4] D. P. Tew, W. Klopper, and C. Hättig, Chem. Phys. Lett. **452**, 326 (2008).
[5] D. P. Tew, C. Hättig, R. A. Bachorz, and W. Klopper, in *Recent Progress in Coupled Cluster Methods: Theory and Applications*, edited by P. Cársky, J. Pittner, and J. Paldus (Springer, Dordrecht, Heidelberg, London, New York, 2009), pp. 1–37.
[6] T. B. Adler, G. Knizia, and H.-J. Werner, J. Chem. Phys. **127**, 221106 (2007).
[7] H.-J. Werner, T. B. Adler, G. Knizia, and F. R. Manby, in *Recent Progress in Coupled Cluster Methods: Theory and Applications*, edited by P. Cársky, J. Pittner, and J. Paldus (Springer, Dordrecht, Heidelberg, London, New York, 2009), pp. 1–47.
[8] G. Knizia, T. B. Adler, and H.-J. Werner, J. Chem. Phys. **130**, 054104 (2009).
[9] D. Bokhan, S. Ten-no, and J. Noga, Phys. Chem. Chem. Phys. **10**, 3320 (2008).
[10] J. Noga, S. Kedžuch, J. Šimunek, and S. Ten-no, J. Chem. Phys. **128**, 174103 (2008).
[11] J. Noga, S. Kedžuch, J. Šimunek, and S. Ten-no, J. Chem. Phys. **130**, 029901E (2009).
[12] D. Bokhan, S. Bernadotte, and S. Ten-no, Chem. Phys. Lett. **469**, 214 (2009).
[13] A. Köhn, G. W. Richings, and D. P. Tew, J. Chem. Phys. **129**, 201103 (2008).
[14] C. Hättig, D. P. Tew, and A. Köhn, J. Chem. Phys. **132**, 231102 (2010).
[15] A. Köhn and D. P. Tew, J. Chem. Phys. **133**, 174117 (2010).
[16] T. Shiozaki, E. F. Valeev, and S. Hirata, J. Chem. Phys. **131**, 044118 (2009).
[17] T. Shiozaki, M. Kamiya, S. Hirata, and E. F. Valeev, J. Chem. Phys. **130**, 054101 (2009).
[18] M. Torheyden and E. F. Valeev, Phys. Chem. Chem. Phys. **10**, 3410 (2008).
[19] E. F. Valeev and T. D. Crawford, J. Chem. Phys. **128**, 244113 (2008).
[20] E. F. Valeev, Phys. Chem. Chem. Phys. **10**, 106 (2008).
[21] W. Klopper and C. C. M. Samson, J. Chem. Phys. **116**, 6397 (2002).
[22] E. F. Valeev, Chem. Phys. Lett. **395**, 190 (2004).
[23] F. R. Manby, J. Chem. Phys. **119**, 4607 (2003).
[24] S. Ten-no and F. R. Manby, J. Chem. Phys. **119**, 5358 (2003).
[25] S. Ten-no, Chem. Phys. Lett. **398**, 56 (2004).
[26] D. P. Tew and W. Klopper, J. Chem. Phys. **123**, 074101 (2005).
[27] H. Fliegl, W. Klopper, and C. Hättig, J. Chem. Phys. **122**, 084107 (2005).
[28] T. B. Adler, H.-J. Werner, and F. R. Manby, J. Chem. Phys. **130**, 054106 (2009).
[29] T. B. Adler and H.-J. Werner, J. Chem. Phys. **130**, 241101 (2009).
[30] J. Friedrich, D. P. Tew, W. Klopper, and M. Dolg, J. Chem. Phys. **132**, 164114 (2010).
[31] F. Neese, F. Wennmohs, and A. Hansen, J. Chem. Phys. **130**, 114108 (2009).
[32] F. Neese, A. Hansen, and D. G. Liakos, J. Chem. Phys. **131**, 064103 (2009).
[33] W. Meyer, Int. J. Quantum Chem. Symp. **5**, 341 (1971).
[34] W. Meyer, J. Chem. Phys. **58**, 1017 (1975).
[35] R. Ahlrichs, F. Driessler, H. Lischka, V. Staemmler, and W. Kutzelnigg, J. Chem. Phys. **62**, 1225 (1975).
[36] H.-J. Werner, J. Chem. Phys. **129**, 101103 (2008).
[37] S. Höfener, D. P. Tew, W. Klopper, and T. Helgaker, Chem. Phys. **356**, 25 (2009).
[38] W. Kutzelnigg, Theor. Chem. Acc. **68**, 445 (1985).
[39] W. Kutzelnigg and W. Klopper, J. Chem. Phys. **94**, 1985 (1991).
[40] S. Ten-no, J. Chem. Phys. **126**, 014108 (2007).
[41] R. A. Bachorz, F. A. Bischoff, A. Glöß, C. Hättig, S. Höfener, W. Klopper, and D. P. Tew, J. Comput. Chem. **32**, 2492 (2011).
[42] S. Kedzuch, M. Milko, and J. Noga, Int. J. Quantum Chem. **105**, 929 (2005).
[43] H.-J. Werner, T. B. Adler, and F. Manby, J. Chem. Phys. **126**, 164102 (2007).
[44] W. Klopper, F. R. Manby, S. Ten-no, and E. F. Valeev, Int. Rev. Phys. Chem. **25**, 427 (2006).
[45] W. Klopper, J. Chem. Phys. **120**, 10890 (2004).
[46] T. Shiozaki, M. Kamiya, S. Hirata, and E. F. Valeev, Phys. Chem. Chem. Phys. **10**, 3358 (2008).
[47] T. Shiozaki, M. Kamiya, S. Hirata, and E. F. Valeev, J. Chem. Phys. **129**, 071101 (2008).
[48] A. Köhn, J. Chem. Phys. **130**, 131101 (2009).
[49] A. Köhn, J. Chem. Phys. **133**, 174118 (2010).
[50] D. Bakowies, J. Chem. Phys. **127**, 084105 (2007).
[51] W. Klopper, B. Ruscic, D. P. Tew, F. A. Bischoff, and S. Wolfsegger, Chem. Phys. **356**, 14 (2010).
[52] F. A. Bischoff, S. Wolfsegger, D. P. Tew, and W. Klopper, Mol. Phys. **107**, 963 (2009).
[53] W. Klopper, R. A. Bachorz, C. Hättig, and D. P. Tew, Theor. Chem. Acc. **126**, 289 (2010).
[54] See supplementary material at http://dx.doi.org/10.1063/1.3624370 for the definitions of the 105 reactions and the basis set limit correlation contributions to the reaction energies.
[55] See http://www.turbomole.com for turbomole development version (2011).
[56] S. F. Boys, in *In Quantum Theory of Atoms, Molecules and the Solid State*, edited by P. O. Löwdin (Academic, New York, 1966), p. 253.
[57] K. A. Peterson, T. B. Adler, and H.-J. Werner, J. Chem. Phys. **128**, 084102 (2008).
[58] K. E. Yousaf and K. A. Peterson, J. Chem. Phys. **129**, 184108 (2008).
[59] C. Hättig, Phys. Chem. Chem. Phys. **7**, 59 (2005).
[60] F. Weigend, J. Comput. Chem. **29**, 167 (2008).
[61] T. Helgaker, W. Klopper, H. Koch, and J. Noga, J. Chem. Phys. **106**, 9639 (1997).
[62] S. Saebo and P. Pulay, Annu. Rev. Phys. Chem. **44**, 213 (1993).
[63] F. Bell, D. Lambrecht, and M. Head-Gordon, Mol. Phys. **108**, 27592773 (2010).
[64] J. Yang, Y. Kurashige, F. R. Manby, and G. K. L. Chan, J. Chem. Phys. **134**, 044123 (2011).

Downloaded 11 Oct 2011 to 134.147.252.238. Redistribution subject to AIP license or copyright; see http://jcp.aip.org/about/rights_and_permissions