# On the selection of domains and orbital pairs in local correlation treatments

Hans-Joachim Werner [*] and Klaus Pflüger

*Institut für Theoretische Chemie, Universität Stuttgart, Pfaffenwaldring 55, D-70569 Stuttgart, Germany*

**Abstract**

The accuracy of local correlation calculations can be controlled by the size of the virtual orbital subspaces (domains) used in the definition of the configuration spaces. Various strategies to define these domains are described and compared. The domains can be chosen on the basis of the localized occupied orbitals, as originally proposed by Pulay and co-workers. Extension of these domains leads to rapid convergence of the correlation energy towards the full canonical result. Alternatively, in local coupled cluster calculations energy criteria can be employed using the amplitudes of extended LMP2 calculations. Furthermore, orbital pairs can be classified according to distance, connectivity, or energy criteria and depending on the importance be treated at different computational levels. The accuracy and efficiency of the various approaches is compared.

[*] To whom correspondence should be addressed, email werner@theochem.uni-stuttgart.de

# 1 Introduction

Local correlation treatments, as originally proposed by Pulay [1] and first implemented by Pulay and Saebø [2–6] employ local orbital spaces to restrict the number of excited configurations in the wavefunction. The occupied orbitals are localized using standard procedures (e.g. Boys [7] or Pipek-Mezey [8] localization), while the virtual space is spanned by non-orthogonal projected atomic orbitals (PAOs). By fully exploiting the locality, our group has achieved linear scaling of the computational effort with molecular size for all standard single-reference methods [9–16]. This has extended the applicability of high-level methods such as local coupled-cluster with single and double excitations and perturbative treatment of triple excitations [LCCSD(T)] to molecules with 50-100 atoms and far over 1000 basis functions. This progress is based on two approximations: First, the excitations are restricted to subspaces of PAOs (domains). The size of these domains and therefore the number of excitations per electron pair is independent of the molecular size, while in conventional methods the number of excitations increases quadratically. Secondly, the correlation energy decreases quickly with the distance between two correlated localized orbitals, and therefore distant orbital pairs can be neglected or treated at a lower level than the pairs which describe the strong correlation of electrons which are close. More generally, based on distance or energy criteria, a hierarchy of *strong*, *close*, *weak*, and *distant* pairs can be defined. The number of pairs in each class increases linearly with molecular size, except for the remaining *very distant* pairs, which are neglected. This means that the total number of excitations and amplitudes increases linearly with molecular size, and this forms the basis for achieving linear scaling of all computational resources.

Using standard criteria for selecting the domains as proposed by Boughton and Pulay [17] and outlined in the next section, about 98-99% of the correlation energy obtained without local approximations and the same method and basis set are recovered. Most of the error is due to the use of finite domains. It has been demonstrated in previous work that this has very little effect on near equilibrium properties such as equilibrium geometries [18–20], harmonic vibrational frequencies [21–23], dipole moments and polarizabilities [24,25], or NMR chemical shifts [26]. However, energy differences like reaction energies or reaction barriers might be more affected if the electronic structure and the domains are different in the reactants and products, since then the absolute errors of the correlation energy may be different in reactants and products. Furthermore, the domains and thus the energy along a reaction path may not be contiguous [27]. The latter problems can be avoided by defining extended domains which are appropriate and fixed along the whole reaction path, as will be demonstrated in a forthcoming publication [28].

In the current work we discuss and compare different strategies to improve the

accuracy by domain extensions at fixed geometries. In principle, this allows to approach the non-local solution to arbitrary precision. However, the computational effort increases rather quickly with the domain sizes, in particular for LCCSD(T), and therefore it is important to find a best compromise between accuracy and cost. In contrast to conventional coupled-cluster calculations for small systems, the bottleneck of local coupled cluster calculations is usually the (direct) integral transformation, and the cost for this transformation depends strongly on the domain sizes. It is important to realize that in order to achieve low-order scaling in the integral transformation it is essential to choose the domains once in the beginning of a local calculation. Also the computation of the residuals in each coupled cluster iteration profits enormously if the domains are known in advance. Screening procedures applied during the coupled-cluster iterations have also been proposed [29, 30], and it has been shown that this leads to linear scaling of the number of amplitudes to be optimized. However, low-order scaling of the total computational effort has not been demonstrated for this method, and we believe that it cannot be achieved unless the domains are fixed in advance.

The transformation bottleneck has recently been much reduced by local density fitting (DF) approximations [20, 31–34]. In this method the 4-index two-electron integrals are approximated by products of 2-index and 3-index integrals. These are not only easier and faster to compute and transform, but also the scaling of the computational cost with basis set size per atom is reduced from $N_{AO}^4$ to $N_{AO}^3$. If large basis sets are used (triple or quadruple zeta) this reduces the transformation times by 1-2 orders of magnitude. The method is particularly fast for local second-order Møller-Plesset perturbation theory (DF-LMP2) [31]. This makes it possible to perform first an DF-LMP2 with large (or even full) domains, and to select the domains to be used in the subsequent DF-LCCSD(T) on the basis of the magnitude of LMP2 amplitudes or energy contributions. The price one has to pay is that an extra integral evaluation for the initial DF-LMP2 is necessary, but the cost for this is usually small as compared to the time for the subsequent DF-LCCSD(T). In the present paper it will be explored whether this method offers any advantages as compared to simpler methods in which the domains are determined only once.

## 2 Method

In the following sections, we summarize the criteria used to define the local orbital spaces and domains as implemented in the `MOLPRO` package of *ab initio* programs [35]. The keywords for thresholds are the same as used in this program.

*2.1 Orbital spaces*

In the following, we will denote occupied molecular orbitals by indices $i, j$, atomic orbitals (AOs) and projected atomic orbitals (PAOs) by indices $r, s$, and basis functions by indices $\mu, \nu$. The localized occupied orbitals (LMOs) are represented in a basis $\{\chi_\mu\}$ by a coefficient matrix $\mathbf{L}$. This is related to the occupied canonical orbitals, which are represented by the MO coefficient matrix $\mathbf{C}_o$, by a unitary transformation $\mathbf{W}$

$$|\phi_i^{\mathrm{LMO}}> = \sum_\mu |\chi_\mu > L_{\mu i} \qquad \mathbf{L} = \mathbf{C}_o \mathbf{W} . \tag{1}$$

In this work $\mathbf{W}$ is determined by Pipek-Mezey localization [8], which has the advantage that $\sigma$ and $\pi$ orbitals stay usually unmixed. Only the correlated orbitals are localized, in order to avoid artificial mixing with core orbitals.

In the case that large and diffuse basis sets are used (e.g. augmented correlation consistent basis sets), poor and non-physical Pipek-Mezey localization may arise due to large MO coefficients arising from near linear dependencies of the basis. We have encountered such problems, for instance in calculations for substituted benzenes (e.g. aniline). This problem can be avoided by excluding the contributions of the diffuse basis functions from the localization. Technically, this can simply be achieved by zeroing the corresponding rows and columns in the AO overlap matrix which is used in the Pipek-Mezey localization.

The virtual space is spanned by projected atomic orbitals $\{\phi_r^{\mathrm{PAO}}\}$, obtained by projecting a set of AOs against the occupied space:

$$|\phi_r^{\mathrm{AO}}> = \sum_\mu |\chi_\mu > C_{\mu r}^{\mathrm{AO}} , \tag{2}$$

$$|\phi_r^{\mathrm{PAO}}> = \left[ 1 - \sum_i^o |\phi_i^{\mathrm{LMO}}> < \phi_i^{\mathrm{LMO}}| \right] |\phi_r^{\mathrm{AO}}> = \sum_\mu |\chi_\mu > C_{\mu r}^{\mathrm{PAO}} . \tag{3}$$

Inserting the expansions yields for the coefficients of the PAOs

$$\mathbf{C}^{\mathrm{PAO}} = \left[ 1 - \mathbf{L}\mathbf{L}^\dagger \mathbf{S} \right] \mathbf{C}^{\mathrm{AO}} = \mathbf{P}\mathbf{C}^{\mathrm{AO}} . \tag{4}$$

If the basis functions are generally contracted and correspond to the atomic Hartree-Fock (HF) orbitals, as is for instance the case for the correlation consistent basis sets of Dunning [36], the matrix $\mathbf{C}^{\mathrm{AO}}$ is taken to be the unit matrix. For segmented basis sets $\mathbf{C}^{\mathrm{AO}}$ is often assumed to be a unit matrix as well, but it may be advantageous to generate atomic HF orbitals since then

inner-shell (core) orbitals can be eliminated. In the current work, we only use the correlation consistent basis sets and $\mathbf{C}^{AO} = \mathbf{1}$. Core orbitals are eliminated from the set of AOs, since the corresponding PAOs would be almost zero and are not well defined. The resulting set of PAOs are orthogonal to all occupied orbitals, but mutually non-orthogonal

$$< \phi_r^{PAO} | \phi_i^{LMO} > = 0 \ , \tag{5}$$

$$< \phi_r^{PAO} | \phi_s^{PAO} > = \left[ \mathbf{S}^{PAO} \right]_{rs} = \left[ \mathbf{C}^{PAO\dagger} \mathbf{S} \mathbf{C}^{PAO} \right]_{rs} \ , \tag{6}$$

where $S_{\mu\nu} = < \chi_\mu | \chi_\nu >$ is the overlap matrix of the basis functions. Due to the projection, the PAOs are linearly dependent, i.e., $\mathbf{S}^{PAO}$ has $m$ zero eigenvalues, where $m$ is the number of occupied orbitals minus the number of eliminated core orbitals. The number of redundant functions is smaller (or zero) if PAO subspaces are considered. Therefore, redundant functions are eliminated individually for each domain as described in the next section.

## 3 Selection of domains using localized orbitals

### 3.1 Standard domains

For each correlated occupied orbital $\phi_i^{LMO}$ an *orbital domain* $[i]$ is generated according to the procedure proposed by Boughton and Pulay [17]. Some modifications were made in our program which are outlined in the following. The purpose of the method is to include all PAOs in the domain which arise from AOs (basis functions) that significantly contribute to the considered LMO. The PAOs arising from AOs at a given atom are considered as a group. Therefore, the first step is to select a set of atoms, and then the PAOs generated from all AOs at these atoms are included in the domain. First, for a given LMO, the atoms are ordered according to decreasing Löwdin charges

$$z_A^{(i)} = 2 \sum_{\mu \in A} \left[ \mathbf{S}^{1/2} \mathbf{L} \right]_{\mu i} \tag{7}$$

Here the sum runs over all basis functions at atom $A$. Mulliken gross charges could also be used, but we found the Löwdin charges to be somewhat less basis set dependent. All atoms with charges greater than a threshold CHGMAX (default 0.4) are immediately added to the domain list for orbital $i$. Furthermore, atoms with charge CHGMIN (default 0.01) are eliminated from the list and never included. For H-atoms, a separate threshold CHGMINH (default 0.03) is used. The elimination of atoms with small charges is particularly useful for diffuse basis sets, since then neighboring atoms often receive artificially large

charges due to orthogonalization tails. Further atoms are added step by step in the order of decreasing charges until the Boughton and Pulay criterion

$$T^{(i)} = 1 - |(\phi_i - \tilde{\phi}_i)^2| \geq \texttt{THRBP} \tag{8}$$

is fulfilled, where $\tilde{\phi}_i$ is an approximate orbital represented by just the basis functions at the already selected atoms

$$|\tilde{\phi}_i> = \sum_{A \in [i]} \sum_{\mu \in A} |\chi_\mu> \tilde{L}_{\mu i} \tag{9}$$

The coefficients $\tilde{L}_{\mu i}$ are determined by a simple least squares fitting procedure. [17]. The default value for $\texttt{THRBP}$ is 0.98. If larger basis sets are used, the BP criterion is fulfilled more quickly, and values of 0.985 (aug-cc-pVTZ) or 0.99 (aug-cc-pVQZ) may be needed to obtain sufficiently accurate domains.

In the following, we denote the PAO subspaces selected according to these criteria as *standard* or *primary* orbital domains.

## 3.2 Extended domains

In order to increase the accuracy, the standard orbital domains as defined in the previous section can be extended by adding the PAOs centered at neighboring atoms. In our program, two possibilities exist to select the additional atoms: either a distance criterion is used, or connectivity information is used. In the first case, all atoms are added to the atom domain list which are within the given distance $\texttt{REXT}$ from any primary atom selected by the standard procedure. In the second case, all atoms are added which are bound to any of the primary atoms. Two atoms A, B are considered to be bound if their distance is smaller than $1.2 \times (r_A + r_B)$, where $r_A$ and $r_B$ are the covalent radii. It is possible to add just the first shell of neighboring atoms ($\texttt{IEXT=1}$), or the first two shells ($\texttt{IEXT=2}$), and so on. For simple organic molecules containing just C, N, O, and H atoms, usually $\texttt{REXT} = 3$ Bohr is equivalent to $\texttt{IEXT=1}$, and $\texttt{REXT} = 5$ Bohr to $\texttt{IEXT=2}$. The use of connectivity has the advantage that the selection is independent of bond lengths, and works for atoms of different rows of the periodic system without change of the parameter. On the other hand, the use of distance criteria ($\texttt{REXT}$) may be advantageous in cases where atoms are close for sterical reasons, or if weak bonds are present.

## 3.3 Pair and triple domains

Domains for double and triple excitations are the union of the orbital domains from which the excitations are made, i.e., for doubles $[ij] = [i] \cup [j]$, and for triples $[ijk] = [i] \cup [j] \cup [k]$. If domains have been extended as described in section 3.2, the extended orbital domains are used by default. It is possible, however, to use the extended domains just for *strong* pairs, or for *strong* and *close* pairs. In these cases the keywords REXT or IEXT are replaced by REXTS, IEXTS or REXTC, IEXTC, respectively.

## 3.4 Domain merging

For aromatic molecules with high symmetry, for instance benzene, the localization may not be unique, i.e., some orbital rotation parameter is redundant. In benzene this corresponds to a rotation of the localized $\pi$-orbitals in the ring, which does not change the localization criterion. Such redundancies can cause problems in geometry optimizations, since the orbitals and domains may change discontinuously. Mathematically, redundancies lead to zero eigenvalues of the matrix of second-derivatives of the localization criterion with respect to orbital rotations, and this can in principle be used to detect redundancies. An easier and recommended way to eliminate redundancies is to merge the three $\pi$-orbital domains in benzene into one domain, which is then used for all three orbitals. The correlation energy is then invariant with respect to unitary transformations among the $\pi$-orbitals. A simple automatic way to detect critical cases without constructing and diagonalizing the second derivative matrix is to merge all orbital domains which overlap by more than one center, and to use the merged domains for all the contributing orbitals (option MERGEDOM=1). Normally, such overlapping domains only occur in conjugated systems. Note that domain merging is done with the standard domains, and any domain extensions apply to the merged domains.

## 3.5 Redundancy check

As mentioned in section 2.1, the PAOs in a domain may be linearly dependent. Redundancies are removed for each pair and triples domain individually by constructing and diagonalizing the PAO overlap matrix $[\mathbf{S}^{\mathrm{PAO}}]_{rs}$, $rs \in [ij]$ (eq. (6), unnormalized PAOs are used). For each eigenvalue which is smaller than a threshold THRLOC (default $10^{-6}$) one function is removed. Two different methods are possible: in the first case (IBASO=1), individual PAOs are eliminated. The advantage is that minimal domain sizes can be used in the integral transformation and LCCSD calculation, and the domains can be kept fixed

across different geometries in order to obtain strictly smooth potentials. However, there are various problems related to this: unless the small eigenvalues of $\mathbf{S}^{\mathrm{PAO}}$ are exactly zero (within numerical accuracy of the machine), the correlation energy will slightly depend on the choice of the deleted functions. In our algorithm the functions are selected in the order of decreasing coefficients in the corresponding eigenvector. This choice may be affected, however, by artifical mixing of degenerate eigenvectors. Furthermore, the energy may not be rotationally invariant if individual $p$- or $d$-functions are eliminated. This can be avoided by deleting whole shells of functions, but this may cause an additional loss of correlation energy. Other useful restrictions are not to eliminate $1s$ functions at H-atoms, and otherwise to eliminate only basis functions corresponding to atomic valence orbitals.

A better method (`IBASO=0`), which avoids all these problems, is to eliminate the linear combinations of PAOs which correspond to the eigenvectors with small eigenvalues. This can done by projection during the update of the LMP2 or LCCSD amplitudes (for details see Ref. [9]). A slight disadvantage of this method is that the LCCSD residual must be computed in the redundant basis, and is transformed to the non-redundant basis only at an intermediate stage. This may lead to some increase of the computational effort and disk space needed to store the amplitudes and transformed integrals.

### 3.6 Fixing domains

In geometry optimizations, the domains are automatically kept fixed once the geometry step size is smaller than a certain threshold (default 0.01 Bohr). In numerical frequency (Hessian) calculations the domains are always fixed. If eigenvectors of $\mathbf{S}^{\mathrm{PAO}}$ are eliminated, the number of deleted eigenvectors is kept fixed. In this way it is guaranteed that the potential energy surface is microscopically smooth and numerical artifacts in the finite difference calculation of the Hessian are avoided. This procedure is very similar to density functional theory (DFT), in which case the definition of the grid is frozen under the same conditions in order to ensure microscopically smooth potentials.

### 3.7 Selection of domains using energy thresholds

The procedure outlined in the previous sections to determine the domains was entirely based on the locality of the occupied orbitals. Previous applications, as well as results in the present paper, demonstrate that this works well for a large variety of applications. As an alternative, one could determine the domains for LCCSD calculations on the basis of an initial LMP2 calculation with full domains (i.e., including the full virtual space) using energy thresholds.

This would then be independent of the Boughton-Pulay procedure and the localization of the occupied orbitals, but would lead to higher scaling and high expense in the LMP2. A compromise is to perform the initial LMP2 with extended domains and use this to select the important parts of the domains using energy thresholds. The LMP2 correlation energy is given by

$$E_{\mathrm{corr}} = \sum_{i \geq j \in P} (2 - \delta_{ij}) \sum_{rs} \left[ 2K_{rs}^{ij} - K_{sr}^{ij} \right] T_{rs}^{ij} \tag{10}$$

where $K_{rs}^{ij} = (ri|sj)$ are the two-electron integrals in the LMO/PAO basis, $T_{rs}^{ij}$ are the LMP2 amplitudes, and $P$ represents the list of orbital pairs $(ij)$ which are included in the calculation. One can now further split this sum into contributions of centers $A$

$$E_{\mathrm{corr}} = \sum_{i \geq j \in P} \sum_{A} E_A^{ij} \tag{11}$$

with

$$E_A^{ij} = (1 - \tfrac{1}{2}\delta_{ij}) \sum_{r \in [A]} \sum_{s} (1 + \tau_{rs}) \left[ 2K_{rs}^{ij} - K_{sr}^{ij} \right] T_{rs}^{ij} \,, \tag{12}$$

where $\tau_{rs}$ permutes the indices $r$ and $s$, and $[A]$ denotes all PAOs at center $A$. Then all centers $A$ are included in the domain $[ij]$ for which $E_A^{ij}$ is above a certain threshold $T_s$. This results in pair-specific domains $[ij]$, which are not necessarily the union of the orbital domains $[i] = [ii]$ and $[j] = [jj]$. However, $[ij]$ must be (at least) the union of $[i]$ and $[j]$ in order to be able to form in the LCCSD intermediates like $C_{rs}^{ij} = T_{rs}^{ij} + t_r^i t_s^j$, where the singles amplitudes $t_r^i$ and $t_s^j$ are restricted to the domains $[i]$ and $[j]$, respectively. Therefore, we have adopted an alternative orbital-based approach: for each LMO $i$ and center $A$ the energies are accumulated. The energy contribution $E_A^{ij}$ is assigned to orbital $i$ if $A$ is closer to any atom contained in the primary (standard) domain $[i]$ than to any atom in the primary orbital domain $[j]$ (and vice versa). If both distances happen to be the same, the energy contribution is shared between orbitals $i$ and $j$. Finally, once the contributions of all pairs have been added, the atoms in the orbital domains $[i]$ are selected according to the energy threshold (the primary domains are always kept). Pair or triples domains can then be generated in the usual way by forming the union of the corresponding orbital domains.

In order to keep the linear scaling behaviour, extended domains may be used in the initial LMP2. In this case the summations over $r$ and $s$ in eqs. (10) and (12) are restricted to the extended domains.

## 4 Pair classes

The orbital pairs $(ij)$ are classified according to the closest distance $R^{(ij)}$ between atoms in the primary domains $[i]$ and $[j]$. This classification is independent of domain extensions. Furthermore, only atoms in the primary domains are considered for the pair classification if the atomic Löwdin charge is larger than CHGMIN_PAIRS (default value 0.2). This criterion was introduced in order to reduce the dependence of the pair selection on localization tails. The *strong pairs* $(0 \leq R^{(ij)} < \texttt{RCLOSE})$ contribute most to the correlation energy and are treated at highest level, e.g., LCCSD(T). *Close pairs* ($\texttt{RCLOSE} \leq R^{(ij)} < \texttt{RWEAK}$), *weak pairs* ($\texttt{RWEAK} \leq R^{(ij)} < \texttt{RDIST}$), and *distant pairs* ($\texttt{RDIST} \leq R^{(ij)} < \texttt{RVDIST}$) are normally optimized at the LMP2 level. It is also possible to determine the close pair amplitudes at the LMP4 level.

The close pair list as determined by RCLOSE has by default no effect on the LCCSD energy but affects the treatment of triple excitations (cf. section 5). If option KEEPCL=1 is set, the LMP2 amplitudes of the close pairs are included in the calculation of the LCCSD residuals of the strong pairs. This includes couplings between strong and close pairs, which are otherwise entirely neglected. As will be demonstrated in section 6, the inclusion of this coupling may significantly improve the results. The additional computational effort is much less than for the full treatment of all strong and close pairs, since for close pairs the residuals need not to be computed. In particular, the most expensive contributions, namely those of the integrals over four PAOs, are not needed for the close pairs.

The weak and distant pairs are treated exclusively at the LMP2 level. Optionally, the integrals for distant pairs are approximated by multipole approximations [10, 12] (option MULTP). Very distant pairs ($R^{(ij)} \geq \texttt{RVDIST}$) are neglected. Note that only the number of very distant pairs scales quadratically with molecular size. The pairs in each of the other classes scales linearly, independent of the choice of the distance criteria. The default values for the distance criteria are RCLOSE=1, RWEAK=3, RDIST=8, and RVDIST=15 (all distances in Bohr; note that RDIST has no effect unless the MULTP option is used). This means that in strong pairs the orbital domains $[i]$ and $[j]$ share at least one atom, while in close pairs the domains $[i]$ and $[j]$ are separated by at most one bond. Setting a distance criterion to zero means that all pairs up to the corresponding class are treated as strong pairs. For instance, RCLOSE=0 means that strong and close pairs are fully included in the LCCSD (in this case KEEPCL=1 has no effect).

As an alternative to the distance criteria, the classification can also be based on connectivity criteria, i.e., depending on the minimum number of bonds between the two orbital domains. The corresponding parameters are denoted

`ICLOSE`, `IWEAK`, `IDIST`, and `IVDIST` (default values 1,2,5,8, respectively). This means that strong pairs are separated by no bonds, close pairs by 1 bond, weak pairs by 2-4 bonds, distant pairs by 5-7 bonds, and very distant pairs by at least 8 bonds. As already discussed in the context of domain extensions, the advantage of using connectivity criteria is the independence of the bond lengths, while the advantage of distance criteria (default in our program) is that they are also effective in non-bonding situations.

## 5 Triple excitations

As for the double excitations, the triple excitations are restricted to domains (cf. section 3.3) and by a triples list $(ijk)$ of LMOs, so that the total number of triple excitations scales linearly with molecular size. As discussed in detail in Ref. [14], the triples list $(ijk)$ is defined by the condition the pairs $(ij)$, $(ik)$, or $(jk)$ must be either strong or close pairs. Additionally, at least one of these pairs must be strong. Secondly, the close pair amplitudes as determined in the initial LMP2 calculation are included in the triples calculation. This approximation turns out to be important for getting accurate triples corrections, but avoids the inclusion of all close pairs in the LCCSD. This is similar to the option `KEEPCL` (cf. section 4), which allows the inclusion of the close pair LMP2 amplitudes in the LCCSD residuals for the strong pairs.

In the local orbital basis the Fock matrix is non-diagonal, and therefore the perturbative treatment of triple excitations requires to solve a set of linear equations [13, 14]. This is computationally expensive and also makes it necessary to store all amplitudes of triple excitations. Therefore, two possible approximations have been introduced. In the (T0) approximation [13], the triple excitations arising from different LMO triples $(ijk)$ and $(i'j'k')$ are entirely decoupled. This means that couplings via off-diagonal elements $f_{ij}$ of the Fock matrix are neglected [13] (here $i, j$ refer to occupied orbitals only), but all couplings within each triple $(ijk)$ are fully included. The (T0) approximation is non-iterative and does not require to store the triples amplitudes.

A compromise between the full iterative (T) treatment and the simple (T0) approximation is the (T1) method. In this case the couplings $f_{ij}$ are included to first order. This corresponds to the first iteration of the (T) method, but a paging algorithm can be used to avoid storage of the triples amplitudes [14]. We found in numerous applications that the (T0) approximation works very well and yields results which are very close to the computationally much more demanding (T1) and (T) methods. In cases of doubt the (T1) method may be used to test the accuracy of the (T0) approximation.

A final note concerns the computational cost: despite linear scaling, the com-

putational effort for the triples can be substantial, in particular for large basis sets or with extended domains. It should be kept in mind that the CPU times scales with the fourth power of the domain sizes; therefore, calculations with extended domains may take one order of magnitude more time than with standard domains. In calculations for large molecules it may be helpful to compute the triples correction with a smaller basis set than used in the LCCSD. In recent calculations of barrier heights in enzymes [37] we found that this works well and hardly affects the results.

## 6  Applications and Discussion

### 6.1  Dependence of the correlation energy on the domain approximation

In order to demonstrate the effect of the domain approximation on correlation and reaction energies we have studied 50 chemical reactions involving 63 molecules [38]. Here we present only a representative subset of the results. The geometries of all molecules have been optimized at the MP2/aug-cc-pV(T+d)Z level. Tables 1-3 show the computed correlation energies of 22 molecules for LMP2, LCCSD, and LCCSD(T0). In each case, the results are compared to the full non-local calculations. For the $s$ and $p$ shells the aug-cc-pVTZ basis sets were used, while for $d$ and $f$ the standard cc-pVTZ sets were employed. For second-row atoms, additional hard $d$ functions were included [39]. The resulting basis is denoted aug[sp]-cc-pV(T+d)Z. The diffuse $s$ and $p$ functions were found to be particularly important for oxygen containing molecules; without them the Hartree-Fock errors of energy differences were often larger than the correlation errors. On the other hand, diffuse polarization functions had rather little effect, and therefore we found this basis to provide a suitable compromise between accuracy and cost. Similar calculations were also performed with larger basis sets and will be published elsewhere [38]. The Boughton-Pulay criterion for domain selection was chosen to be `THRBP=0.985`. The diffuse $s$ and $p$ functions were not included in the Pipek-Mezey localization (cf. section 2.1, options `CPLDEL=1`, `CPLMAXDL=1`).

Table 1 shows that with standard domains on the average 99.12% of the full MP2 energy for the same basis set is recovered. When the domains of the strong pairs are augmented by the PAOs at the next neighbors (`IEXTS=1`), the fraction increases to 99.84%, and if the domains of all pairs are augmented (`IEXT=1`) even to 99.92%. At the same time, the scatter becomes smaller, as indicated by the standard deviations of the average percentages. In general, it is found that the fraction is larger for molecules containing double bonds than for saturated molecules. This is somewhat surprising, since the orbitals in saturated molecules can be very well localized, and one would intuitively

assume that for such molecules the local approximations have a small effect. However, the opposite is true. We believe that this is due to the fact that unsaturated molecules, in particular those with conjugated bonds, have often planar structures or at least planar subgroups. This leads to smaller basis set superposition errors (BSSE) as compared to saturated molecules with three-dimensional structures. The BSSE is fully present in canonical calculations but small in local calculations [9, 40–42]. In addition, in molecules with conjugated double bonds the domains are often larger than for single bonds and may extend over 3-4 atoms. The worst case in the set is cyclohexane $C_6H_{12}$. For this molecule, only 98.1% of the correlation energy is recovered with standard domains. Furthermore, the convergence towards 100% with increasing domain size is slower than in most other cases. This particular case will be further investigated in section 6.4.

Tables 2 and 3 show similar results for LCCSD and LCCSD(T0), respectively. In these calculations, only strong pairs were treated at the LCCSD level, and the triples list $(ijk)$ was restricted as described in section 4. The weak pairs were treated by LMP2, and no coupling between weak and close pairs were included. We denote this approximation LCCSD(T0)|LMP2. Table 2 demonstrates that with LCCSD apparently a larger fraction of correlation energy (99.6% with standard domains) than with LMP2 (99.1%) is recovered. Clearly, this is due to error compensation. It is well known that MP2 overestimates long-range dispersion effects. Thus, the contributions of close and weak pairs, which are treated by LMP2, are overestimated relative to full LCCSD. With extended domains, this even leads to an overestimation of the total correlation energy relative to CCSD. On the other hand, it is also known that CCSD underestimates long-range correlation effects, and this error is largely corrected when perturbative triple excitations are included. Consequently, the overestimation of the correlation energy in LCCSD(T0)|LMP2 relative to full CCSD(T) is smaller than for LCCSD|LMP2; with standard domains, LCCSD(T0)|LMP2 yields an average of 99.4% of the full non-local correlation energy. We may also note that with full LCCSD and LCCSD(T0) (i.e. without weak pair approximations), 99.2% and 99.1%, respectively, of the correlation energy are obtained, consistent with the LMP2 result.

In contrast to LMP2, the scatter of the fraction of correlation energy obtained with the LCCSD and LCCSD(T0) methods does not decrease with increasing domain size. This indicates that part of this scatter is due to the weak pair approximation, which will be investigated in the next section.

13

## 6.2 Dependence of the correlation energy on the weak pair approximation

Table 4 demonstrates the effect of relaxing the weak pair approximations for LCCSD and LCCSD(T0). In all cases the strong pair domains were extended (`IEXTS=1`). In the columns denoted (a) all strong and close pairs were treated by LCCSD and LCCSD(T0); also, the triples list $(ijk)$ included all pairs $(ij)$, $(ik)$ and $(jk)$ which are either strong or close. In the columns denoted (b), the pair and triples list were the same as in Tables 1-3, i.e., only strong pairs were fully treated by LCCSD(T0), and the triples list was restricted such that at least one of the three pairs is a strong pair. However, in this case the close pair amplitudes, as optimized at the LMP2 level, were included in the computation of the LCCSD residual of the strong pairs (`KEEPCL=1`) and also in the (T0) calculation (which is the default). It is found that including the close pairs (case (a)) in the LCCSD eliminates most of the overshooting effect of the LCCSD|LMP2 approximation. Comparison with Tables 2 and 3 shows that the average fraction of correlation energy is reduced from 100.23% (`ICLOSE=1`) to 99.96% (`ICLOSE=2`) for LCCSD|LMP2, and from 100.09% to 99.77% for LCCSD(T0)|LMP2. Also, the standard deviations from the average value are now very small. Almost the same improvement is achieved in case (b) (`KEEPCL=1`). This demonstrates that the coupling of close and strong pairs has an effect which is not entirely negligible. It should be noted that the calculations in the latter case (b) are less expensive than in case (a), since the residuals and contributions of 4-external integrals are only needed for the strong pairs. In large molecules, this leads to substantial savings not only in CPU time, but also in the disk space requirements.

## 6.3 Reaction energies

More important than the errors of the individual correlation energies are errors of energy differences, for instance reaction energies. Small differences in the percentage of correlation energy recovered for reactants and products can lead to significant errors in computed energy differences. Here we present results for 33 reactions; a more extensive study will be presented elsewhere [38]. The reactions along with the conventional MP2, CCSD, and CCSD(T) reaction energies and experimental values are listed in Table 5. Experimental reaction enthalpies were computed from standard enthalpies of formation taken from [43, 44]. Zero point and thermal corrections were computed using DFT/BP86 with the SV(P) [45] basis set. These corrections were subtracted from the experimental reaction enthalpies to obtain reaction energies which can be directly compared to the computed energy differences. The errors introduced by the BP86 zero-point corrections were tested in several cases and found to be negligible when compared to other errors in the calculations. Table

14

5 also shows the minimum, maximum and average deviations of the computed and experimental values. The average errors of CCSD(T) with the aug[sp]-cc-pV(T+d)Z basis set used here are $1.34 \pm 1.34$ kcal/mol. Part of this error is attributed to basis set effects. Using aug-cc-pV(Q+d)Z reduces the error to $1.28 \pm 1.05$ kcal/mol. While in most cases the CCSD(T) values obtained with large basis sets approach the experimental ones closely, there are some exceptions. In particular, for the reactions involving $C_2H_3Cl$ and HNCO the deviations between the calculated and experimental values amount to more than 4 kcal/mol, even with the full aug-cc-pV(5+d)Z basis set. The enthalpies of formation of these molecules reported in the literature show a large scatter, and it can be concluded that the experimental data are not reliable. If the reactions involving these two molecules (reactions 12, 16, 30, 31) are excluded from the statistics, the average errors for the aug[sp]-cc-pV(T+d)Z and aug-cc-pV(Q+d)Z bases reduce to $0.9 \pm 0.7$ kcal/mol and $0.6 \pm 0.5$ kcal/mol, respectively.

In Tables 6 and 7 the minimum, maximum and average errors along with standard deviations of the LMP2 and LCCSD(T)|LMP2 reaction energies relative to (i) the conventional results and (ii) the experimental data are presented. The corresponding error statistics for the conventional MP2 and CCSD(T) methods, respectively, are shown in the last columns. Using standard domains, the average deviation of LMP2 from MP2 amounts to $1.02 \pm 0.77$ kcal/mol. With `IEXTS=1` or `IEXT=1` these errors are reduced to $0.45 \pm 0.57$ kcal/mol and $0.21 \pm 0.19$ kcal/mol, respectively. In all three cases the errors are small when compared to the errors relative to the experimental values. The latter are only slightly reduced by extending the domains, and these small improvements would normally not justify the increased computational effort.

A rather similar situation is seen in Table 7 for LCCSD(T0). Interestingly, in this case the average error of $0.6 \pm 0.5$ kcal/mol relative to canonical CCSD(T) obtained with standard domains is even smaller than the corresponding error of LMP2. This points to some error compensation: the reduction of correlation energy by the domain approximation is partly compensated by the overestimation of the LMP2 weak pair contributions and neglecting the couplings between strong and close pairs. When the domains are increased (`IEXTS=1` or `IEXT=1`) the errors slightly increase, since this error compensation becomes less balanced. On the other hand, if the coupling of strong and close pairs is included in the LCCSD residuals (`KEEPCL=1`), the errors are reduced to $0.33 \pm 0.30$ kcal/mol and are almost as small as the errors of LMP2 (which are independent of weak pair approximations). As in the case of LMP2, the latter improvement of the computational model does not lead to a significant improvement of the agreement with experimental values. We therefore believe that in most cases the standard local approximations (standard domains and treatment of only strong pairs by LCCSD) are sufficient with basis sets of this size. Only when highly accurate results are required and large basis sets

are used should the convergence of the local approximations be tested. It is then recommended to first check the domain approximation by comparing DF-MP2 and DF-LMP2 results (full DF-MP2 calculations should always be possible for cases where LCCSD(T0) is feasible). Secondly, the convergence of the LCCSD(T0) results with respect to the parameters `ICLOSE` and `IWEAK` can be tested. Such a systematic procedure was recently applied in accurate QM/MM calculations of barrier heights in enzymes [37].

## 6.4 A case study: the $C_6H_6 + 3\ H_2 \rightarrow C_6H_{12}$ reaction

In section 6.1 we discussed the surprisingly large errors in the correlation energy of $C_6H_{12}$ introduced by the domain approximation. Since the loss of correlation energy is much smaller in planar and aromatic molecules like benzene, a particular large error results for the reaction energy of $C_6H_6 + 3\ H_2 \rightarrow C_6H_{12}$. We have argued above that this effect is probably caused by BSSE effects, which are strongly reduced in local calculations [9, 40–42]. In this section we support this view by further calculations, in which the domain size and the basis set have been systematically varied.

Table 8 shows the convergence of the reaction energies as a function of the domain size for LMP2, LCCSD, and LCCSD(T0), using the cc-pVTZ basis set, `THRBP=0.98`, and `MERGEDOM=1`. We will first discuss the LMP2 results. With standard domains (`IEXT=0`) the LMP2 exothermicity is 5 kcal/mol less than for conventional MP2. There are several effects contributing to the over-stabilization of the reactants relative to the product: (i) as already mentioned above, the domain error is smaller in $C_6H_6$ than in $C_6H_{12}$; (ii) the domain merging procedure (cf. Section 3.4) increases the domains of the $\pi$-orbitals in benzene, but does not not affect $C_6H_{12}$; and (iii) there is no domain error for $H_2$ by construction. Furthermore, as will be shown below, the basis set errors of MP2 and LMP2 have opposite sign and are additive. Therefore, this reaction represents a worst-case scenario.

If the domains are extended by the next neighbors (`IEXT=1`) the LMP2 error is reduced to 1 kcal/mol, and for `IEXT=2` it amounts only to 0.2 kcal/mol. The dependence of the LCCSD|LMP2 and LCCSD(T0)|LMP2 reaction energies on the domain size is very similar as for LMP2. This supports our view that the domain approximation can be faithfully tested at the LMP2 level. For LCCSD|LMP2 and LCCSD(T0)|LMP2 the results with standard domains (`IEXT=0`) are apparently in better agreement with the conventional calculations than in the LMP2 case. Clearly, this is due to an error compensation between the domain approximation and the weak pair approximation, as already discussed in the previous section. If the domains are extended, the LCCSD and LCCSD(T0) reaction energies become too negative. This effect

16

is compensated, however, if close pairs are included in the LCCSD(T0) treatment (case [200] in Table 8; here we use the notation [`IWEAK ICLOSE KEEPCL`]). The [200] reaction energy with `IEXT=1` is again in close agreement with the canonical result. Almost the same improvement is achieved if only strong pairs are treated by LCCSD, but the close-pair LMP2 amplitudes are included in the calculation of the LCCSD residuals (`KEEPCL=1`). This is consistent with the findings in the previous section.

A more complete picture emerges if one considers the dependence of the reaction energy on the basis set, as shown in Fig. 1. The DF-MP2 and DF-LMP2 energies were computed with the cc-pV$n$Z basis sets up to quintuple zeta. An estimate for the MP2 complete basis set (CBS) limit was obtained by extrapolating the MP2/cc-pVQZ and MP2/cc-pV5Z correlation energies assuming $E_n = E_{CBS} + An^{-3}$. The figure shows that the basis set convergence is qualitatively different for MP2 and LMP2. While MP2 yields for small basis sets much too negative values and converges from below to the CBS limit, the opposite is true for LMP2 with standard domains (`IEXT=0`). In fact, for the cc-pVDZ basis ($n = 2$) the error of the LMP2 reaction energy relative to the CBS limit is smaller than that of MP2 and the basis set dependence of LMP2 is weaker than for MP2. It should be noted, however, that both methods seem not to converge to exactly the same limit; there is some intrinsic error in LMP2 due to the domain approximation, but this is much smaller than the apparent error observed for small basis sets. The exceptionally large difference between the LMP2 and MP2 results for small basis sets is due to the addition of the basis set errors of the two methods, which have different sign and origin. In MP2, the large BSSE in the product molecule $C_6H_{12}$ leads to too large an exothermicity. We believe that this is due to the fact that the basis functions at the neighbors of an atom with tetrahedral environment can to a certain extent mimic the effect of missing higher angular momentum functions in the basis. On the other hand, in LMP2 these BSSE effects are small, and the correlation energy of the product converges more slowly with basis set size than the energy of the reactants. If the domains are extended, some of the BSSE effects are introduced back into the LMP2 calculation, and the results approach the MP2 values. This leads to the interesting finding that in case that the domains of the strong and close pairs are extended (`IEXTC=1`) the CBS limit is already reached with the triple zeta basis set.

Finally, we tested the selection of domains using energy criteria, as described in section 3.7. Since, as shown above, the dependence of the LMP2 and LCCSD(T0) energies on the domain size is very similar, these tests were performed only at the LMP2 level. The results are shown in Table 9. As expected, the results converge towards the canonical ones with decreasing energy threshold. However, this convergence is less systematic than with the domain extension method. With selection threshold $10^{-3} E_h$ exactly the same domains and energies as with the standard Boughton-Pulay method are ob-

tained. If one considers the energy contributions of individual centers in the domains, one finds that the atoms in the standard domains have by far the largest contributions; those of the next important atoms are about one order of magnitude smaller. They vary somewhat from molecule to molecule, and therefore at certain thresholds more atoms may be added for the reactants or products, leading to unbalanced results. This is most clearly seen in Table 9 for the threshold $3.3 \times 10^{-4}$ $E_h$. Of course, for smaller thresholds these effects become less pronounced, but the domains needed to achieve a certain accuracy are not smaller than those determined by the standard procedure with extensions. Therefore, despite the appealing possibility to control the accuracy of a local calculation by an energy threshold, we do not consider this a useful procedure. Controlling the accuracy of the domain approximation is possible by another single parameter, namely the domain extension (IEXT), and this leads to much more systematic convergence. The weak pair approximation can also be controlled by an energy threshold, but again we do not find that this leads to a systematic improvement of the method.

## 7 Conclusions

We have reviewed the approximations made in the linear scaling local correlation methods developed in our group. The accuracy of local coupled cluster calculations depends on (i) the domain sizes and (ii) the definition of the strong pair list included in the LCCSD. These approximations can be controlled using connectivity or distance criteria. The domain sizes can be determined by a single parameter (IEXT or REXT), and we have shown that domain extensions lead to rapid and systematic convergence towards the canonical results. The list of strong pairs depends on parameters IWEAK and ICLOSE (or RWEAK and RCLOSE), and again convergence towards the canonical result with respect to these parameters has been demonstrated. It has been found that in most cases the domain and the weak pair approximations in LCCSD(T0)|LMP2 calculations compensate each other to a large extent. Therefore, rather accurate results are obtained with the standard approximation, in which the domains are determined by the Boughton-Pulay procedure and strong orbital pairs are determined by the condition that the domains of the two orbitals overlap. This may be considered a well defined new computational model, distinct from conventional CCSD(T). Extensive tests on 33 chemical reactions have shown that the errors introduced by these standard local approximations are usually small when compared to basis set errors and the deviations from experimental values. For the 33 reactions studied here, the average and maximum errors of the reaction energies relative to experimental values are very similar for CCSD(T) and LCCSD(T0). In cases in which high accuracy is required and large basis sets are used, it is recommended first to test the domain approximation at the

LMP2/MP2 level, and then to test the convergence of the LCCSD(T) with respect to the weak pair parameters.

It has also demonstrated that convergence of the results with basis set size may be qualitatively different for LMP2 and MP2, and faster convergence is found for LMP2. This effect is attributed to a reduction of basis set superposition errors in the local calculations. In a recent study on the influence of local approximations on vibrational frequencies [23] similar effects have been found, and it has been demonstrated that for some modes, which are sensitive to BSSE, the local calculations also yielded better basis set convergence and more accurate results than conventional calculations. Other studies have demonstrated that the effect of local approximations on equilibrium geometries [18–20], dipole moments and polarizabilities [24, 25], and NMR chemical shifts [26] are small. The combination of local and density fitting approximations makes it now possible perform DF-LCCSD(T0) calculations with large basis sets for molecules of chemical interest (50-100 atoms), as has recently been demonstrated in a QM/MM study of the enzymes chorismate mutase and para-hydroxybenzoate hydroxylase [37]. Further work to extend these methods to open-shell cases is in progress.

## Acknowledgments

**Tables**

Table 1
LMP2 energies as function of the domain size

| Molecule | MP2 | LMP2 | | |
|---|---|---|---|---|
| | | IEXTS=0 | IEXTS=1 | IEXT=1 |
| $C_2H_2$ | -0.311213 | -0.309642 (99.50) | -0.311136 (99.98) | -0.311136 (99.98) |
| $C_2H_4$ | -0.336334 | -0.332675 (98.91) | -0.336096 (99.93) | -0.336162 (99.95) |
| $C_2H_6$ | -0.370764 | -0.366111 (98.74) | -0.370071 (99.81) | -0.370405 (99.90) |
| $H_2CO$ | -0.397511 | -0.395534 (99.50) | -0.397459 (99.99) | -0.397464 (99.99) |
| $CH_3NH_2$ | -0.406442 | -0.401554 (98.80) | -0.405616 (99.80) | -0.406101 (99.92) |
| $CH_3OH$ | -0.430764 | -0.426533 (99.02) | -0.430076 (99.84) | -0.430465 (99.93) |
| $H_2O_2$ | -0.500681 | -0.497630 (99.39) | -0.500313 (99.93) | -0.500538 (99.97) |
| $CH_3CN$ | -0.519120 | -0.515115 (99.23) | -0.518504 (99.88) | -0.518656 (99.91) |
| $C_2H_3Cl$ | -0.517689 | -0.513164 (99.13) | -0.517021 (99.87) | -0.517266 (99.92) |
| $H_2CCO$ | -0.541430 | -0.537787 (99.33) | -0.541102 (99.94) | -0.541140 (99.95) |
| HNCO | -0.581119 | -0.578557 (99.56) | -0.580987 (99.98) | -0.581011 (99.98) |
| $CH_3CHO$ | -0.570019 | -0.565149 (99.15) | -0.569271 (99.87) | -0.569518 (99.91) |
| $C_2H_4O$ | -0.577946 | -0.572522 (99.06) | -0.577399 (99.91) | -0.577651 (99.95) |
| $HCONH_2$ | -0.606922 | -0.602279 (99.24) | -0.606527 (99.93) | -0.606637 (99.95) |
| $C_2H_5OH$ | -0.604622 | -0.597066 (98.75) | -0.602717 (99.68) | -0.603745 (99.86) |
| HCOOH | -0.633158 | -0.629535 (99.43) | -0.632855 (99.95) | -0.632966 (99.97) |
| $C_2H_3CN$ | -0.661225 | -0.655179 (99.09) | -0.660227 (99.85) | -0.660501 (99.89) |
| $COCl_2$ | -0.766275 | -0.762349 (99.49) | -0.765222 (99.86) | -0.765833 (99.94) |
| $HCONHCH_3$ | -0.780409 | -0.773522 (99.12) | -0.779097 (99.83) | -0.779622 (99.90) |
| $HCOOCH_3$ | -0.803457 | -0.796554 (99.14) | -0.801773 (99.79) | -0.802562 (99.89) |
| $NH_2CONH_2$ | -0.813438 | -0.806032 (99.09) | -0.812265 (99.86) | -0.812730 (99.91) |
| $C_2H_4(OH)_2$ | -0.839214 | -0.828236 (98.69) | -0.835590 (99.57) | -0.837641 (99.81) |
| $CH_3NO_2$ | -0.858234 | -0.852897 (99.38) | -0.857021 (99.86) | -0.857576 (99.92) |
| $C_6H_{12}$ | -1.045942 | -1.026304 (98.12) | -1.038709 (99.31) | -1.042408 (99.66) |
| Average[a] | | (99.12 ± 0.33) | (99.84 ± 0.14) | (99.92 ± 0.07) |

a) Average fraction of correlation energy relative to MP2 in percent.

The error bounds are standard deviations.

Table 2
LCCSD energies as function of the domain size

| | CCSD | LCCSD\|LMP2$^a$ | | |
|---|---|---|---|---|
| Molecule | | IEXTS=0 | IEXTS=1 | IEXT=1 |
| $C_2H_2$ | -0.322948 | -0.321504 ( 99.55) | -0.322874 ( 99.98) | -0.322876 ( 99.98) |
| $C_2H_4$ | -0.360599 | -0.357982 ( 99.27) | -0.361004 (100.11) | -0.361077 (100.13) |
| $C_2H_6$ | -0.401843 | -0.399422 ( 99.40) | -0.402999 (100.29) | -0.403326 (100.37) |
| $H_2CO$ | -0.407090 | -0.405201 ( 99.54) | -0.406972 ( 99.97) | -0.406979 ( 99.97) |
| $CH_3NH_2$ | -0.430262 | -0.427842 ( 99.44) | -0.431448 (100.28) | -0.431923 (100.39) |
| $CH_3OH$ | -0.448749 | -0.446227 ( 99.44) | -0.449411 (100.15) | -0.449793 (100.23) |
| $H_2O_2$ | -0.507312 | -0.505233 ( 99.59) | -0.507763 (100.09) | -0.507982 (100.13) |
| $CH_3CN$ | -0.532250 | -0.531361 ( 99.83) | -0.534496 (100.42) | -0.534649 (100.45) |
| $C_2H_3Cl$ | -0.550382 | -0.548913 ( 99.73) | -0.552350 (100.36) | -0.552597 (100.40) |
| $H_2CCO$ | -0.549266 | -0.546455 ( 99.49) | -0.549481 (100.04) | -0.549517 (100.05) |
| HNCO | -0.577266 | -0.575304 ( 99.66) | -0.577494 (100.04) | -0.577526 (100.04) |
| $CH_3CHO$ | -0.589670 | -0.587092 ( 99.56) | -0.590885 (100.21) | -0.591132 (100.25) |
| $C_2H_4O$ | -0.595732 | -0.593008 ( 99.54) | -0.597507 (100.30) | -0.597749 (100.34) |
| $HCONH_2$ | -0.616527 | -0.613322 ( 99.48) | -0.617170 (100.10) | -0.617299 (100.13) |
| $C_2H_5OH$ | -0.632951 | -0.629493 ( 99.45) | -0.634604 (100.26) | -0.635620 (100.42) |
| HCOOH | -0.637004 | -0.634419 ( 99.59) | -0.637514 (100.08) | -0.637621 (100.10) |
| $C_2H_3CN$ | -0.676157 | -0.675902 ( 99.96) | -0.680494 (100.64) | -0.680765 (100.68) |
| $COCl_2$ | -0.789483 | -0.791779 (100.29) | -0.794310 (100.61) | -0.794924 (100.69) |
| $HCONHCH_3$ | -0.800478 | -0.796843 ( 99.55) | -0.801969 (100.19) | -0.802481 (100.25) |
| $HCOOCH_3$ | -0.818812 | -0.815630 ( 99.61) | -0.820413 (100.20) | -0.821204 (100.29) |
| $NH_2CONH_2$ | -0.825674 | -0.821861 ( 99.54) | -0.827515 (100.22) | -0.827967 (100.28) |
| $C_2H_4(OH)_2$ | -0.864727 | -0.859758 ( 99.43) | -0.866409 (100.19) | -0.868437 (100.43) |
| $CH_3NO_2$ | -0.865569 | -0.863578 ( 99.77) | -0.867420 (100.21) | -0.867976 (100.28) |
| $C_6H_{12}$ | -1.106198 | -1.103528 ( 99.76) | -1.114741 (100.77) | -1.118381 (101.10) |
| Average$^b$ | | $(99.60 \pm 0.21)$ | $(100.24 \pm 0.20)$ | $(100.31 \pm 0.25)$ |

a) LCCSD\|LMP2 using `IWEAK=2, ICLOSE=1`

b) Average fraction of correlation energy relative to CCSD in percent.

The error bounds are standard deviations.

Table 3
LCCSD(T0) energies as function of the domain size

| | CCSD(T) | LCCSD(T0)|LMP2[a] | | |
| --- | --- | --- | --- | --- |
| Molecule | | IEXTS=0 | IEXTS=1 | IEXT=1 |
| $C_2H_2$ | -0.339615 | -0.338003 (99.53) | -0.339443 ( 99.95) | -0.339443 ( 99.95) |
| $C_2H_4$ | -0.375751 | -0.372737 (99.20) | -0.376013 (100.07) | -0.376084 (100.09) |
| $C_2H_6$ | -0.415408 | -0.412527 (99.31) | -0.416392 (100.24) | -0.416752 (100.32) |
| $H_2CO$ | -0.423993 | -0.421630 (99.44) | -0.423569 ( 99.90) | -0.423575 ( 99.90) |
| $CH_3NH_2$ | -0.445378 | -0.442216 (99.29) | -0.446252 (100.20) | -0.446784 (100.32) |
| $CH_3OH$ | -0.463960 | -0.460634 (99.28) | -0.464238 (100.06) | -0.464668 (100.15) |
| $H_2O_2$ | -0.526470 | -0.523453 (99.43) | -0.526395 ( 99.99) | -0.526646 (100.03) |
| $CH_3CN$ | -0.557876 | -0.556557 (99.76) | -0.559697 (100.33) | -0.559869 (100.36) |
| $C_2H_3CL$ | -0.575719 | -0.573305 (99.58) | -0.577096 (100.24) | -0.577361 (100.29) |
| $H_2CCO$ | -0.575956 | -0.572202 (99.35) | -0.575464 ( 99.91) | -0.575506 ( 99.92) |
| HNCO | -0.606660 | -0.603389 (99.46) | -0.605784 ( 99.86) | -0.605810 ( 99.86) |
| $CH_3CHO$ | -0.614158 | -0.610674 (99.43) | -0.614795 (100.10) | -0.615068 (100.15) |
| $C_2H_4O$ | -0.620367 | -0.616660 (99.40) | -0.621591 (100.20) | -0.621858 (100.24) |
| $HCONH_2$ | -0.642937 | -0.638640 (99.33) | -0.642931 (100.00) | -0.643051 (100.02) |
| $C_2H_5OH$ | -0.655761 | -0.650568 (99.21) | -0.656257 (100.08) | -0.657350 (100.24) |
| HCOOH | -0.663816 | -0.660095 (99.44) | -0.663530 ( 99.96) | -0.663651 ( 99.98) |
| $C_2H_3CN$ | -0.711718 | -0.709890 (99.74) | -0.714840 (100.44) | -0.715133 (100.48) |
| $COCl_2$ | -0.828566 | -0.828447 (99.99) | -0.831335 (100.33) | -0.831975 (100.41) |
| $HCONHCH_3$ | -0.834752 | -0.828825 (99.29) | -0.834419 ( 99.96) | -0.834960 (100.02) |
| $HCOOCH_3$ | -0.853381 | -0.847756 (99.34) | -0.853069 ( 99.96) | -0.853901 (100.06) |
| $NH_2CONH_2$ | -0.860759 | -0.854820 (99.31) | -0.861162 (100.05) | -0.861651 (100.10) |
| $C_2H_4(OH)_2$ | -0.896933 | -0.888724 (99.08) | -0.896253 ( 99.92) | -0.898407 (100.16) |
| $CH_3NO_2$ | -0.907362 | -0.902179 (99.43) | -0.906443 ( 99.90) | -0.907023 ( 99.96) |
| $C_6H_{12}$ | -1.151504 | -1.144232 (99.37) | -1.156460 (100.43) | -1.160306 (100.76) |
| Average[b] | | (99.42 ± 0.19) | (100.09 ± 0.17) | (100.16 ± 0.21) |

a) LCCSD(T0)|LMP2 using IWEAK=2, ICLOSE=1

b) Average fraction of correlation energy relative to CCSD(T) in percent.

The error bounds are standard deviations.

Table 4
Effect of weak pair approximations on LCCSD energies

| Molecule | LCCSD\|LMP2 | | LCCSD(T0)\|LMP2 | |
|---|---|---|---|---|
| | (a) | (b) | (a) | (b) |
| $C_2H_2$ | -0.322862 ( 99.97) | -0.322841 ( 99.97) | -0.339431 (99.95) | -0.339405 (99.94) |
| $C_2H_4$ | -0.360441 ( 99.96) | -0.360361 ( 99.93) | -0.375356 (99.89) | -0.375287 (99.88) |
| $C_2H_6$ | -0.401508 ( 99.92) | -0.401746 ( 99.98) | -0.414718 (99.83) | -0.415006 (99.90) |
| $H_2CO$ | -0.407040 ( 99.99) | -0.406998 ( 99.98) | -0.423625 (99.91) | -0.423584 (99.90) |
| $CH_3NH_2$ | -0.429943 ( 99.93) | -0.430137 ( 99.97) | -0.444575 (99.82) | -0.444809 (99.87) |
| $CH_3OH$ | -0.448461 ( 99.94) | -0.448373 ( 99.92) | -0.463191 (99.83) | -0.463086 (99.81) |
| $H_2O_2$ | -0.507154 ( 99.97) | -0.507118 ( 99.96) | -0.525715 (99.86) | -0.525644 (99.84) |
| $CH_3CN$ | -0.531795 ( 99.91) | -0.532502 (100.05) | -0.556546 (99.76) | -0.557374 (99.91) |
| $C_2H_3Cl$ | -0.550124 ( 99.95) | -0.550473 (100.02) | -0.574476 (99.78) | -0.574917 (99.86) |
| $H_2CCO$ | -0.548965 ( 99.95) | -0.548767 ( 99.91) | -0.574765 (99.79) | -0.574523 (99.75) |
| HNCO | -0.577120 ( 99.97) | -0.576955 ( 99.95) | -0.605271 (99.77) | -0.605067 (99.74) |
| $CH_3CHO$ | -0.589216 ( 99.92) | -0.589481 ( 99.97) | -0.612831 (99.78) | -0.613129 (99.83) |
| $C_2H_4O$ | -0.595454 ( 99.95) | -0.596062 (100.06) | -0.619246 (99.82) | -0.619968 (99.94) |
| $HCONH_2$ | -0.616266 ( 99.96) | -0.616385 ( 99.98) | -0.641809 (99.82) | -0.641941 (99.85) |
| $C_2H_5OH$ | -0.632406 ( 99.91) | -0.632795 ( 99.98) | -0.653743 (99.69) | -0.654199 (99.76) |
| HCOOH | -0.636813 ( 99.97) | -0.636740 ( 99.96) | -0.662644 (99.82) | -0.662562 (99.81) |
| $C_2H_3CN$ | -0.676049 ( 99.98) | -0.677009 (100.13) | -0.709633 (99.71) | -0.710760 (99.87) |
| $COCl_2$ | -0.789971 (100.06) | -0.790742 (100.16) | -0.826156 (99.71) | -0.827006 (99.81) |
| $HCONHCH_3$ | -0.800101 ( 99.95) | -0.800713 (100.03) | -0.832190 (99.69) | -0.832899 (99.78) |
| $HCOOCH_3$ | -0.818387 ( 99.95) | -0.818739 ( 99.99) | -0.850621 (99.68) | -0.851027 (99.72) |
| $NH_2CONH_2$ | -0.825267 ( 99.95) | -0.825889 (100.03) | -0.858491 (99.74) | -0.859191 (99.82) |
| $C_2H_4(OH)_2$ | -0.863853 ( 99.90) | -0.864383 ( 99.96) | -0.893263 (99.59) | -0.893875 (99.66) |
| $CH_3NO_2$ | -0.865443 ( 99.99) | -0.865899 (100.04) | -0.903933 (99.62) | -0.904484 (99.68) |
| $C_6H_{12}$ | -1.106354 (100.01) | -1.109230 (100.27) | -1.146770 (99.59) | -1.150110 (99.88) |
| Average[c] | (99.96 ± 0.03) | (100.01 ± 0.08) | (99.77 ± 0.09) | (99.83 ± 0.08) |

(a) LCCSD|LMP2, `IWEAK=2, ICLOSE=0, IEXTS=1`

(b) LCCSD|LMP2, `IWEAK=2, ICLOSE=1, IEXTS=1, KEEPCL=1`

c) Average fraction of correlation energy in percent.

The error bounds are standard deviations.

Table 5
Reference values for reaction energies$^a$ (in kcal/mol)

| Reaction | MP2 | CCSD | CCSD(T) | Experiment |
|---|---|---|---|---|
| $C_2H_2 + H_2 \rightarrow C_2H_4$ | 0.41 | -2.64 | -1.69 | -48.06 |
| $CO + H_2 \rightarrow H_2CO$ | -0.60 | -0.22 | 0.06 | -4.73 |
| $H_2O_2 + H_2 \rightarrow 2\ H_2O$ | -5.38 | -2.68 | -0.93 | -86.27 |
| $C_2H_6 + H_2 \rightarrow 2\ CH_4$ | 0.08 | -0.90 | -0.37 | -18.16 |
| $C_2H_4 + H_2 \rightarrow C_2H_6$ | -1.59 | -1.05 | -0.06 | -39.45 |
| $H_2CO + H_2 \rightarrow CH_3OH$ | -0.43 | -0.89 | 0.18 | -29.31 |
| $CH_3CHO + H_2 \rightarrow C_2H_5OH$ | -2.09 | -2.72 | -1.66 | -23.34 |
| $CO + H_2O \rightarrow CO_2 + H_2$ | -1.50 | 4.99 | 2.73 | -7.40 |
| $C_2H_2 + H_2O \rightarrow CH_3CHO$ | 2.13 | 0.31 | 0.53 | -39.10 |
| $C_2H_4 + H_2O \rightarrow C_2H_5OH$ | -0.36 | 0.23 | 0.56 | -14.38 |
| $C_2H_4O + H_2O \rightarrow C_2H_4(OH)_2$ | 2.54 | 0.84 | 1.22 | -26.73 |
| $CO_2 + NH_3 \rightarrow HNCO + H_2O$ | -4.13 | -3.91 | -4.20 | 24.86 |
| $CO + NH_3 \rightarrow HCONH_2$ | -0.85 | 2.23 | 1.56 | -9.57 |
| $HCOOH + NH_3 \rightarrow HCONH_2 + H_2O$ | -1.19 | -0.41 | -0.28 | -0.35 |
| $NH_3 + 4\ H_2O_2 \rightarrow HNO_3 + 5\ H_2O$ | -16.62 | 2.17 | 0.63 | -176.83 |
| $HNCO + NH_3 \rightarrow NH_2CONH_2$ | 6.62 | 4.05 | 5.49 | -24.74 |
| $CO + H_2O_2 \rightarrow CO_2 + H_2O$ | -6.87 | 2.31 | 1.81 | -93.68 |
| $CH_4 + 4\ H_2O_2 \rightarrow CO_2 + 6\ H_2O$ | -20.68 | -3.65 | -0.65 | -288.43 |
| $C_2H_4 + H_2O_2 \rightarrow C_2H_4O + H_2O$ | -5.94 | -0.86 | 0.08 | -50.90 |
| $C_2H_5OH + H_2O_2 \rightarrow CH_3CHO + 2\ H_2O$ | -3.29 | 0.03 | 0.73 | -62.94 |
| $C_2H_4 + H_2O_2 \rightarrow C_2H_4(OH)_2$ | -3.40 | -0.02 | 1.30 | -77.62 |
| $H_2CCO + H_2CO \rightarrow C_2H_4O + CO$ | 0.08 | -1.73 | -0.72 | -3.74 |
| $C_2H_2 + HCN \rightarrow C_2H_3CN$ | 1.92 | 1.86 | 1.54 | -45.95 |
| $CO + CH_3OH \rightarrow HCOOCH_3$ | -0.70 | 2.53 | 1.27 | -13.21 |
| $HCOOH + CH_3OH \rightarrow HCOOCH_3 + H_2O$ | -1.04 | -0.12 | -0.57 | -3.99 |
| $cis\text{-}CH_3CH{=}CHCH_3 \rightarrow trans\text{-}CH_3CH{=}CHCH_3$ | -0.91 | -1.02 | -0.92 | -0.96 |
| $CS_2 + 2\ H_2O \rightarrow CO_2 + 2\ H_2S$ | 6.03 | -3.80 | 0.08 | -11.85 |
| $SO_2 + CO_2 \rightarrow SO_3 + CO$ | 2.72 | -0.83 | 0.69 | 44.88 |
| $SO_2 + H_2O_2 \rightarrow SO_3 + H_2O$ | -4.15 | 1.48 | 2.49 | -48.80 |
| $C_2H_2 + HCl \rightarrow C_2H_3Cl$ | -2.67 | -2.87 | -3.24 | -27.56 |
| $C_2H_4 + Cl_2 \rightarrow C_2H_3Cl + HCl$ | -7.26 | -4.52 | -4.67 | -24.51 |
| $CO + Cl_2 \rightarrow COCl_2$ | -4.49 | 3.52 | 1.19 | -27.47 |
| $COCl_2 + 2\ NH_3 \rightarrow NH_2CONH_2 + 2\ HCl$ | 1.29 | -2.67 | -0.28 | -24.81 |
| Min. Error | 0.08 | 0.02 | 0.06 | |
| Max. Error | 20.68 | 4.99 | 5.49 | |
| Av. Error | 3.64 | 1.94 | 1.34 | |
| $\sigma$ | 4.83 | 1.41 | 1.34 | |

a) Basis aug[sp]-cc-pV(T+d)Z; the MP2, CCSD, and CCSD(T)

   values are the errors relative to the experimental ones.

Table 6

Errors of LMP2 reaction energies[a] (in kcal/mol)

| Error | Relative to MP2 | | | Relative to Experiment | | | |
|---|---|---|---|---|---|---|---|
| | IEXT=0 | IEXTS=1 | IEXT=1 | IEXT=0 | IEXTS=1 | IEXT=1 | MP2 |
| Minimum | 0.04 | 0.01 | 0.01 | 0.28 | 0.03 | 0.05 | 0.08 |
| Maximum | 2.68 | 1.93 | 0.80 | 22.61 | 21.55 | 21.00 | 20.68 |
| Average | 1.02 | 0.45 | 0.21 | 3.85 | 3.70 | 3.64 | 3.64 |
| $\sigma$ | 0.77 | 0.47 | 0.19 | 4.86 | 4.60 | 4.47 | 4.83 |

a) Basis aug[sp]-cc-pV(T+d)Z

Table 7
Errors of LCCSD(T0) reaction energies[a] (in kcal/mol)

| Error | Relative to CCSD(T) | | | | Relative to Experiment | | | | |
|---|---|---|---|---|---|---|---|---|---|
| | IEXT=0 | IEXT=1 | IEXTS=1 | IEXTS=1[b] | IEXT=0 | IEXT=1 | IEXTS=1 | IEXTS=1[b] | CCSD(T) |
| Minimum | 0.01 | 0.04 | 0.02 | 0.04 | 0.14 | 0.01 | 0.01 | 0.00 | 0.06 |
| Maximum | 1.73 | 2.54 | 2.35 | 1.51 | 5.73 | 4.52 | 4.53 | 5.31 | 5.49 |
| Average | 0.60 | 0.77 | 0.62 | 0.33 | 1.49 | 1.50 | 1.45 | 1.49 | 1.34 |
| $\sigma$ | 0.51 | 0.60 | 0.53 | 0.30 | 1.38 | 1.33 | 1.36 | 1.33 | 1.34 |

a) Basis aug[sp]-cc-pV(T+d)Z, IWEAK=2, ICLOSE=1

b) Using KEEPCL=1, see text.

Table 8
Reaction energies for $C_6H_6 + 3\ H_2 \rightarrow C_6H_{12}$ as function of domain size and weak pair approximations (in kcal/mol without ZPC)

| IEXT | LMP2 | LCCSD | | | LCCSD(T0) | | |
|------|------|-------|-------|-------|-------|-------|-------|
| | | $[210]^b$ | [200] | [211] | [210] | [200] | [211] |
| 0 | -60.2 | -75.7 | -70.9 | -72.2 | -70.3 | -65.0 | -66.5 |
| 1 | -64.3 | -79.4 | -74.4 | -75.7 | -74.5 | -69.0 | -70.5 |
| 2 | -65.1 | -80.2 | | | -75.4 | | |
| 3 | -65.3 | | | | | | |
| Full | -65.3 | | -74.0 | | | -69.5 | |

a) All calculation using density fitting and cc-pVTZ basis sets.

b) The notation [IWEAK ICLOSE KEEPCL] is used.

Table 9
LMP2 energies[a] as function of the domain selection threshold

| THRSEL[b] | $E(C_6H_6)$ | $E(C_6H_{12})$ | $\Delta E^c$ | $N_{av}^d(C_6H_6)$ | $N_{av}(C_6H_{12})$ |
|---|---|---|---|---|---|
| $1.0 \times 10^{-3}$ | -231.716592 | -235.306494 | -60.2 | 123 | 87 |
| $3.3 \times 10^{-4}$ | -231.723150 | -235.306494 | -56.1 | 167 | 87 |
| $1.0 \times 10^{-4}$ | -231.723932 | -235.320147 | -64.2 | 184 | 193 |
| $3.3 \times 10^{-5}$ | -231.725079 | -235.322537 | -65.0 | 235 | 255 |
| $1.0 \times 10^{-5}$ | -231.725079 | -235.323101 | -65.3 | 235 | 274 |
| Full | -231.726506 | -235.324493 | -65.3 | 263 | 348 |

a) All values calculated using density fitting and cc-pVTZ basis sets.

b) Energy threshold for domain selection (see text) using

   LMP2 (`IEXTS=2`) as reference.

c) Reaction energy for $C_6H_6 + 3\ H_2 \rightarrow C_6H_{12}$ in kcal/mol

   (without ZPC).

d) Average domain sizes

**Figure Captions**

Fig. 1: Dependence of the MP2 and LMP2 reaction energies for $C_6H_6 + 3\ H_2$ $\rightarrow C_6H_{12}$ on the basis set. The cc-pV$n$Z basis sets have been used for $n = 2-5$.

# References

[1] P. Pulay, Localizability of dynamic electron correlation, Chem. Phys. Lett. 100 (1983) 151–154.

[2] S. Saebø, P. Pulay, Local configuration-interaction: An efficient approach for larger molecules, Chem. Phys. Lett. 113 (1985) 13–18.

[3] P. Pulay, S. Saebø, Orbital-invariant formulation and second-order gradient evaluation in Møller-Plesset perturbation theory, Theor. Chim. Acta 69 (1986) 357–368.

[4] S. Saebø, P. Pulay, Fourth-order Møller-Plesset perturbation theory in the local correlation treatment. I. Method, J. Chem. Phys. 86 (1987) 914–922.

[5] S. Saebø, P. Pulay, The local correlation treatment. II. Implementation and tests, J. Chem. Phys. 88 (1988) 1884–1890.

[6] S. Saebø, P. Pulay, Local treatment of electron correlation, Annu. Rev. Phys. Chem. 44 (1993) 213–236.

[7] S. F. Boys, in: P. O. Löwdin (Ed.), Quantum Theory of Atoms, Molecules, and the Solid State, Academic Press, New York, 1966, p. 253.

[8] J. Pipek, P. G. Mezey, A fast intrinsic localization procedure applicable for ab initio and semiempirical linear combination of atomic orbital wave functions, J. Chem. Phys. 90 (1989) 4916–4926.

[9] C. Hampel, H.-J. Werner, Local treatment of electron correlation in coupled cluster theory, J. Chem. Phys. 104 (1996) 6286–6297.

[10] G. Hetzer, P. Pulay, H.-J. Werner, Multipole approximation of distant pair energies in local MP2 calculations, Chem. Phys. Lett. 290 (1998) 143–149.

[11] M. Schütz, G. Hetzer, H.-J. Werner, Low-order scaling local electron correlation methods. I. Linear scaling local MP2, J. Chem. Phys. 111 (1999) 5691–5705.

[12] G. Hetzer, M. Schütz, H. Stoll, H.-J. Werner, Low-order scaling local correlation methods II: Splitting the coulomb operator in linear scaling local second-order Møller-Plesset perturbation theory, J. Chem. Phys. 113 (2000) 9443–9455.

[13] M. Schütz, H.-J. Werner, Local perturbative triples correction (T) with linear cost scaling, Chem. Phys. Lett. 318 (2000) 370–378.

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

[15] M. Schütz, H.-J. Werner, Low-order scaling local electron correlation methods. IV. Linear scaling local coupled-cluster (LCCSD), J. Chem. Phys. 114 (2001) 661–681.

[16] M. Schütz, Linear scaling local connected triples beyond local (T): Local CCSDT-1b with O(N) scaling, J. Chem. Phys. 116 (2002) 8772–8785.

[17] J. W. Boughton, P. Pulay, Comparison of the Boys and Pipek-Mezey localizations in the local correlation approach and automatic virtual basis selection, J. Comput. Chem. 14 (1993) 736–740.

[18] A. ElAzhary, G. Rauhut, P. Pulay, H.-J. Werner, Analytical energy gradients for local second-order Møller-Plesset perturbation theory, J. Chem. Phys. 108 (1998) 5185–5193.

[19] G. Rauhut, H.-J. Werner, Analytical energy gradients for local coupled-cluster methods, Phys. Chem. Chem. Phys. 3 (2001) 4853–4862.

[20] M. Schütz, H.-J. Werner, R. Lindh, F. R. Manby, Analytical energy gradients for local second-order Møller-Plesset perturbation theory using density fitting approximations, J. Chem. Phys. 121 (2004) 737–750.

[21] G. Rauhut, A. E. Azhary, F. Eckert, U. Schumann, H.-J. Werner, Impact of local approximations on MP2 vibrational frequencies, Spectrochim. Acta A 55 (1999) 647–658.

[22] G. Rauhut, H.-J. Werner, The vibrational spectra of furoxan and dichlorofuroxan: a comparative theoretical study using density functional theory and local electron correlation methods, Phys. Chem. Chem. Phys. 5 (2003) 2001–2008.

[23] T. Henar, G. Rauhut, H.-J. Werner, Impact of local and density fitting approximations on harmonic vibrational frequencies, J. Phys. Chem. A (2006) 2060–2064.

[24] T. Korona, K. Pflüger, H.-J. Werner, The effect of local approximations in coupled-cluster wavefunctions on dipole moments and static dipole polarisabilities, Phys. Chem. Chem. Phys. 6 (2004) 2059–2065.

[25] N. J. Russ, T. D. Crawford, Local correlation in coupled cluster calculations of molecular response properties, Chem. Phys. Lett. 400 (2004) 104–111.

[26] J. Gauss, H.-J. Werner, NMR chemical shift calculations within local correlation methods: the GIAO-LMP2 approach, Phys. Chem. Chem. Phys. 2 (2000) 2083–2090.

[27] N. J. Russ, T. D. Crawford, Potential energy surface discontinuities in local correlation methods, J. Chem. Phys. 121 (2) (2004) 691–696.

[28] R. Mata, H.-J. Werner, Application of local electron correlation methods to chemical reactions, to be published.

[29] P. Y. Ayala, G. E. Scuseria, Linear scaling second-order Møller-Plesset theory in the atomic orbital basis for large molecular systems, J. Chem. Phys. 110 (1999) 3660–3671.

[30] G. E. Scuseria, P. Y. Ayala, Linear scaling coupled cluster and perturbation theories in the atomic orbital basis, J. Chem. Phys. 111 (1999) 8330–8343.

[31] H.-J. Werner, F. R. Manby, P. Knowles, Fast linear scaling second-order Møller-Plesset perturbation theory (MP2) using local and density fitting approximations, J. Chem. Phys. 118 (2003) 8149–8160.

[32] M. Schütz, F. R. Manby, Linear scaling local coupled cluster theory with density fitting. I : 4-external integrals, Phys. Chem. Chem. Phys. 5 (2003) 3349–3358.

[33] R. Polly, H.-J. Werner, F. R. Manby, P. J. Knowles, Fast Hartree-Fock theory using local density fitting approximations, Mol. Phys. 102 (2004) 2311–2321.

[34] H.-J. Werner, M. Schütz, Density fitting approximations in local coupled cluster (lccsd(t)) calculations, to be published.

[35] H.-J. Werner, P. J. Knowles, R. Lindh, M. Schütz, P. Celani, T. Korona, F. R. Manby, G. Rauhut, R. D. Amos, A. Bernhardsson, A. Berning, D. L. Cooper, M. J. O. Deegan, A. J. Dobbyn, F. Eckert, C. Hampel, G. Hetzer, A. W. Lloyd, S. J. McNicholas, W. Meyer, M. E. Mura, A. Nicklass, P. Palmieri, R. Pitzer, U. Schumann, H. Stoll, A. J. Stone, R. Tarroni, T. Thorsteinsson, Molpro, version 2002.10, a package of ab initio programs, see http://www.molpro.net (2005).

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

[37] F. Claeyssens, J. N. Harvey, F. R. Manby, R. A. Mata, A. J. Mulholland, K. E. Ranaghan, M. Schütz, S. Thiel, W. Thiel, H.-J. Werner, Converged quantum chemistry for enzymatic catalysis using local correlation methods, to be published.

[38] K. Pflüger, H.-J. Werner, A. Schäfer, M. Schütz, The effect of local approximations on reaction energies, to be published.

[39] J. T.H. Dunning, K. A. Peterson, A. Wilson, Gaussian basis sets for use in correlated molecular calculations. X. The atoms aluminum through argon revisited, J. Chem. Phys. 114 (2001) 9244–9253.

[40] W. Meyer, L. Frommhold, Ab initio calculation of the dipole moment of He-Ar and the collision-induced absorption spectra, Phys. Rev. A 33 (1986) 3807–3814.

[41] S. Saebø, W. Tong, P. Pulay, Efficient elimination of basis set superposition errors by the local correlation method: Accurate ab initio studies of the water dimer, J. Chem. Phys. 98 (1993) 2170–2175.

[42] M. Schütz, G. Rauhut, H.-J. Werner, Local treatment of electron correlation in molecular clusters: Structures and stabilities of $(H_2O)_n$, n = 2-4, J. Phys. Chem. A 102 (1998) 5997–6003.

[43] I. Barin, Thermochemical data of pure substances, Verlag Chemie, Weinheim, 1989.

[44] D. R. Lide, Handbook of chemistry and physics, 76th edition, CRC Press, Boca Raton, FL (USA), 1995.

[45] A. Schäfer, H. Horn, R. Ahlrichs, Fully optimized contracted gaussian basis sets for atoms Li to Kr, J. Chem. Phys. 97 (1992) 2571–2577.

Fig. 1. Werner and Pflüger