

**Investigations of Pair
Fragments Screening and
Local Interaction Correlation
within the
Divide-Expand-Consolidate (DEC)
Local Correlation Method
Internship Project Report**

For an Internship at the
qLEAP Center for Theoretical Chemistry
at Aarhus University

Sven Kähler

Utrecht, August 2014

Chapter 1

Introduction

One of the challenges electronic structure theory currently faces is to extend the applicability of the highly accurate methods, that are nowadays well established for small molecular systems, towards macromolecular systems. The problem which these methods encounters when applied to larger systems is that the computational costs required to carry out calculations increases so steeply with system size that a wide range of systems is not accessible to highly accurate calculations. Therefore, modifications to these methods that lead to an improved scaling behavior are necessary with the ultimate goal to develop methods which scale linearly.

Linear scaling methods for Hartree-Fock (HF) and density functional theory (DFT) are well established. Due to the lack of systematic improvability and the lack of adequate dispersion interaction representation found in DFT[1, 2], wave function methods for electron correlation, like second order Møller-Plesset perturbation theory (MP2) or the hierarchy of Coupled Cluster methods (CC), are necessary but still under development. Since electron correlation is a predominantly local phenomenon, it has been proposed already early on to use a local correlation ansatz[3, 4].

Various methods have been developed. For MP2, a Laplace transformed approach has gained some attention[5]. For local CC methods, different schemes with different choices of local orbitals to replace the delocalized canonical molecular orbitals have been presented[6, 7, 8, 9, 10]. Before the development of the Divide-Expand-Consolidate (DEC)[11] scheme used here, the problems with the previous models have been to control for the error introduced by the local approximations and to represent virtual orbitals in a way that is physically sound[12], while allowing for the desired reduction in computational scaling.

In the DEC approach, which provides both linear scaling MP2[13] and CC methods[11, 14], local orbitals are generated by new localization algorithms

which have been specifically developed to yield virtual orbitals suitable for local electron correlation[15, 16, 17]. Based on these local orbitals, the molecular system can then be divided into subsystems, in the following referred to as atomic fragments (AFs), among which the localized orbitals are distributed. The correlation energy E_{corr} of the system can then be described as the sum of the correlation energy contributions of each AF and of pairs of AFs, referred to as pair fragments (PFs), as shown in equation 1.1, where the indices P and Q run over the set of AFs.

$$E_{corr} = \sum_P E_P + \sum_{P>Q} \Delta E_{PQ} \quad (1.1)$$

For this fragment-wise correlation energy approach to remain accurate, it would be insufficient to limit the set of virtual orbitals, to which excitations are included in the correlation energy calculation for each AF, to only those virtual orbitals that were assigned to the respective this AF. Instead, error control is introduced by expanding the virtual orbital spaces to also include virtual orbitals from surrounding fragments until the threshold for the desired level of accuracy is reached. In the following, this threshold will be referred to as fragment optimization threshold (FOT).

The localization and determination of orbital spaces in the DEC method will be described in more detail in chapter 2. For a truly linear scaling method, the pair fragment contributions in equation 1.1 must not contain all PFs, because this would result in quadratic scaling. Therefore the sum needs to be truncated. Since the PFs represent dispersion interactions they decay quickly with increasing distance between the two fragments. The simplest approach for the truncation is to choose some conservative distance threshold and only include PFs for which the distance between their AFs is below this threshold. In practise it has proven more efficient to generate estimates for the expected contribution of each pair fragment and only then include those fragments that are required to reach FOT precision in the overall calculation. In chapter 3 a systematic investigation of this screening procedure is given. Since the DEC procedure involves describing the overall correlation energy of an electronic system as a combination of interactions within and between subsystems, it appears natural to try to apply this approach to the calculation of interaction correlation energies. If successful, the resulting method could provide highly accurate interaction energies for macromolecular systems for a fraction of the costs of the existing methods. In chapter 4, results of preliminary investigations towards this goal are reported.

Chapter 2

The Divide-Expand-Consolidate (DEC) Scheme

2.1 Localization

The localization of molecular orbitals for the purpose of electron correlation calculations on macromolecular systems requires that occupied, as well as virtual orbitals, are highly local and that the optimization procedure is efficient enough to be feasible for such large systems. To effectively minimize the extend of orbital spaces required in the DEC procedure, the tails of the orbital functions should decay quickly.

While various methods for molecular orbital localization with different approaches on how to measure the locality of an orbital have been developed[18, 19, 20, 21], only few yield sufficiently compact, if at all local, results for virtual orbitals. Also the convergence of many localization algorithms is often poor due to the presence of negative Hessian eigenvalues. The reliability of any local orbital approach thus depends on combining a locality measure that is adequate for virtual orbitals and enforces compactness of orbitals with an algorithm that is sufficiently stable to avoid the shortcomings of previous attempts.

$$\mu_2^p = \langle p | (\hat{r} - \langle p | \hat{r} | p \rangle)^2 | p \rangle \quad (2.1)$$

$$\mu_4^p = \langle p | (\hat{r} - \langle p | \hat{r} | p \rangle)^4 | p \rangle \quad (2.2)$$

In the localization approach chosen to be used for the DEC method, the locality of an orbital p is measured via the second and fourth central moment,

given in equations 2.1 and 2.2, where small distribution moments indicate high locality. To further improve locality the localization function of an orbital that is to be minimized can also be defined as the square $(\mu_2^p)^2$ (or $(\mu_4^p)^2$) of the central moment so that an additional penalty is put on large moments. As measure for the collective locality of ensembles of molecular orbitals, the maximum of the locality measures of all orbitals has been chosen instead of the arithmetic mean, because arithmetic means would possibly tolerate outliers in sets of otherwise well localized orbitals. For the purpose of modelling local electron correlation it is preferable to ensure that the least local orbitals are still as local as possible, because those could otherwise cause considerable computational bottle-necks. The minimization of the locality function is then carried out using a quasi-Newton-based trust region approach that is both highly stable and efficient[22].

2.2 Fragment Optimization

The initial atomic fragment orbital spaces are generated by assigning each molecular orbital to the atom on which it has the largest Löwdin charge. Orbitals assigned to hydrogen atoms by this procedure may be reassigned to the nearest non-hydrogen atom to give fewer and more evenly sized fragments. In the following the occupied/virtual orbitals assigned to an atomic site P will be denoted by \underline{P}/\bar{P} . Using this notation the E_P and ΔE_{PQ} terms in the correlation energy expression in equation 1.1 for the case of an MP2 calculation can be written as

$$E_P^{o,MP2} = \sum_{ij \in \underline{P}} \sum_{ab} t_{ij}^{ab} (2g_{iajb} - g_{ibja}) \quad (2.3)$$

and

$$\Delta E_{PQ}^{o,MP2} = \sum_{i \in \underline{P}, j \in \underline{Q}} \sum_{ab} t_{ij}^{ab} (2g_{iajb} - g_{ibja}) + \sum_{i \in \underline{Q}, j \in \underline{P}} \sum_{ab} t_{ij}^{ab} (2g_{iajb} - g_{ibja}), \quad (2.4)$$

where the indices a and b run over all virtual orbitals, t_{ij}^{ab} denotes the MP2 amplitudes and g_{iajb} and g_{ibja} are two-electron integrals. At this point no approximations with respect to a local treatment of electron correlation have been made. In order to introduce computational savings, the indices of the unoccupied orbitals can be restricted. Since both occupied and virtual orbitals are well localized, only orbitals that are spatially close to each other

lead to non-vanishing two-electron integrals. Therefore energy and amplitudes from a calculation restricted to the close spatial vicinity of the respective fragment will yield a reasonable approximation. When writing the set of occupied/virtual orbitals spacially close to P (which includes orbitals assigned to P) as $[P]/[\bar{P}]$, the E_P and ΔE_{PQ} terms become

$$E_P^{o,MP2} = \sum_{ij \in \underline{P}} \sum_{a \in [\bar{P}], b \in [\bar{P}]} t_{ij}^{ab} (2g_{iajb} - g_{ibja}) \quad (2.5)$$

and

$$\Delta E_{PQ}^{o,MP2} = \sum_{i \in \underline{P}, j \in \underline{Q}} \sum_{a \in [\bar{P}], b \in [\bar{Q}]} t_{ij}^{ab} (2g_{iajb} - g_{ibja}) + \sum_{i \in \underline{Q}, j \in \underline{P}} \sum_{a \in [\bar{Q}], b \in [\bar{P}]} t_{ij}^{ab} (2g_{iajb} - g_{ibja}). \quad (2.6)$$

To further increase the accuracy of the calculation up to FOT precision the orbital spaces are expanded step by step to include orbitals further and further away from the central atom of the fragment. After each step an energy calculation is carried out and compared to the previous iteration. The expansion is finished when the change in energy is below the FOT.

After the expansion is completed a reduction of the orbital spaces is carried out, where orbital are iteratively removed and the change in energy is checked after each step until no more orbitals can be removed without losing FOT precision. Multiple approaches for the specific expansion and reduction procedures have been and are being tested and parameters, for example the grouping in which orbitals are added or removed, are the topic of ongoing research.

2.3 DEC Correlation Energy

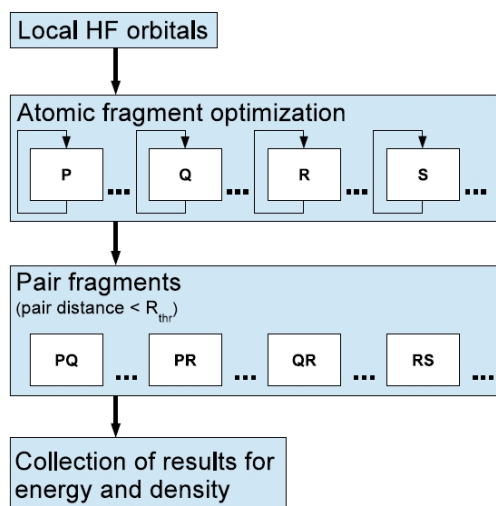


Figure 2.1: Overview over a DEC calculation (Adapted from [23]).

After all fragments are optimized, the resulting orbital spaces are used to evaluate the pair fragment energies. The overall energy of the system is then given by equation 1.1. While the equations used describe a MP2 calculation the DEC scheme can be applied analogously to a CC calculation. An overview over a DEC calculation is given in Figure 2.1.

The local, fragment-wise approach to electron correlation allows for well parallelized algorithms. All fragment optimizations and pair fragment calculations can be carried out independently. As a consequence, a massively parallel implementation of the DEC approach for MP2 and CCSD has been implemented which uses three levels of parallelization[23]:

1. Coarse-grained (independent fragment calculations)
2. Medium-grained (MPI-parallelization for each fragment)
3. Fine-grained (OpenMP-parallelization for each MPI-process)

Chapter 3

Pair Fragment Screening

3.1 Computational Savings From PF Screening

As explained in the introduction, the number of pair fragments scales quadratically with system size. Therefore pair fragments need to be discarded to yield a linear scaling algorithm which can be easily achieved, because the pair fragments describe dispersion type interactions that decay with the inter-fragment distance to the power of minus six. This means that a conservatively chosen distance cutoff already leads to a stable linear scaling algorithm.

However the use of the inter-fragment distance as a criterion is inefficient. As Figure 3.1 shows using the example of the peptide Valinomycin, in a typical macromolecular system pair fragment energies fall in a broad range for most inter-fragment distances. Any distance cutoff that is certain to include all relevant pairs adds many unnecessary ones at the same time. The ideal distance cutoff that includes all pair fragments exactly up to the most distant that meaningfully contributes to the calculation is not known prior to the calculation. Therefore the distance cutoff has to be chosen rather conservatively to ensure error control, where a cutoff value of 10 Å has been reported to form a viable compromise between safety and efficiency[24]. Since the pair fragment calculations are very numerous for the macromolecular systems the DEC scheme is intended for, they become the most time-consuming part of the total calculation.

The inclusion of unnecessary pair fragments can be avoided by the use of a prescreening procedure. The procedure calculates a rough estimate of the energy contribution of each pair using an MP2 calculation on a limited set of orbitals. Since the estimates do not depend on optimized orbital spaces, they can be carried out in parallel with the fragment optimization, so that their

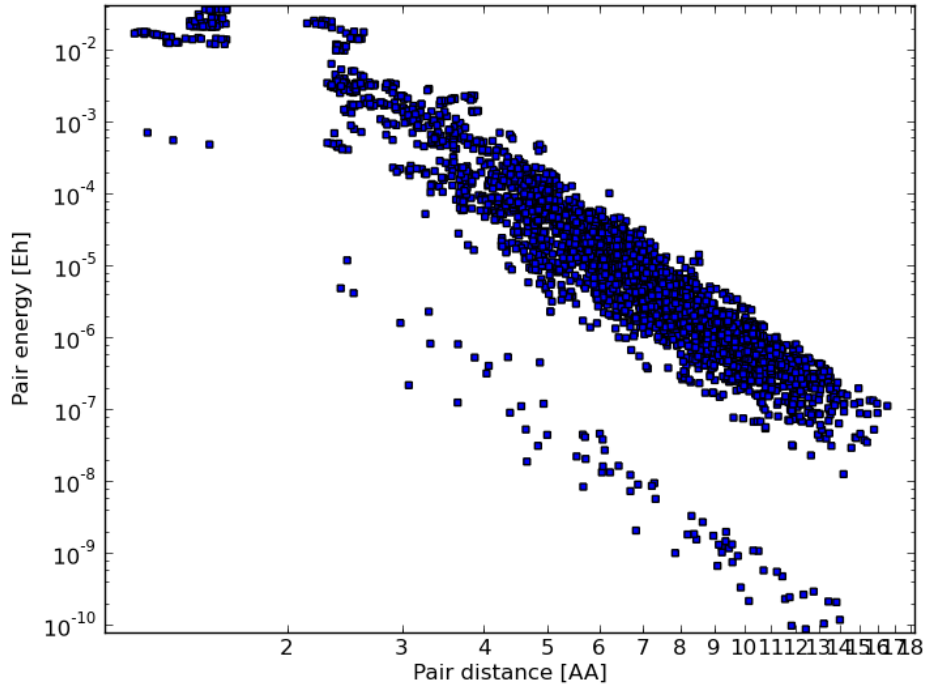


Figure 3.1: Distribution of Pair Fragment Energies as a Function of Inter-Fragment Distance

results are available before pair fragment calculations are started. Because of very limited orbital spaces used for the estimates, the calculations will be much faster than calculations in the large fragment spaces required for high precisions. For the same reason, they can obviously only provide an order of magnitude approximation.

Using these estimated pair fragment energies for each atomic fragment P , the set of substantially contributing pair fragments containing P is determined. For this purpose all fragments which include P are sorted by energy (in decreasing order) and a cutoff energy value, below which all contributions are discarded, is determined. The cutoff is given by the pair fragment energy of the last pair fragment PK needed to ensure that the sum of the missing contributions is below the FOT:

$$\frac{1}{2} \sum_{Q=K+1}^N |\Delta E_{PQ}^{est}| < FOT, \quad (3.1)$$

where N is the number of pair fragments that contain P . In the following results of an investigation of this pair fragment screening approach will be

reported which are concerned with the questions whether the estimates generated are sufficiently accurate, whether the errors produced by this pair screening procedure meet the accuracy targets and whether, and to what degree, the procedure as it is described above leads to computational savings. For this purpose conventional and DEC type MP2 calculations have been carried out for Palmitic Acid, a carbon acid, and for the peptide valinomycin. Palmitic Acid allows for a wide range of inter-fragment distances, while at the same time being computationally accessible for full conventional MP2 calculations, so that the accuracy of DEC results can be checked. The Valinomycin molecule is chosen as a typical moderate sized example for the type of systems that the DEC scheme is intended for and too large for standard MP2.

3.2 Results and Discussion

3.2.1 Accuracy of Pair Fragment Energy Estimates

In Figure 3.2 the energies of all possible pair fragments for the Palmitic Acid molecule are plotted against inter-pair distances. The pair energies have been constructed from the results of a standard MP2 calculation, which have afterwards been divided into fragments that are equivalent to the unrestricted ΔE_{PQ} terms given in equation 2.4. These pair fragment energies will be referred to as full pair energies in the following. The plot shows a linear decay of the pair fragment energies with distance. Since both axes are scaled logarithmically, the linear decay of the energies corresponds to the expected power law decay predicted for dispersion forces. Due to the regular linear shape of the Palimitic Acid molecule the range of energies encountered for a given inter-fragment distance is much narrower than the ones for Valinomycin seen in Figure 3.1.

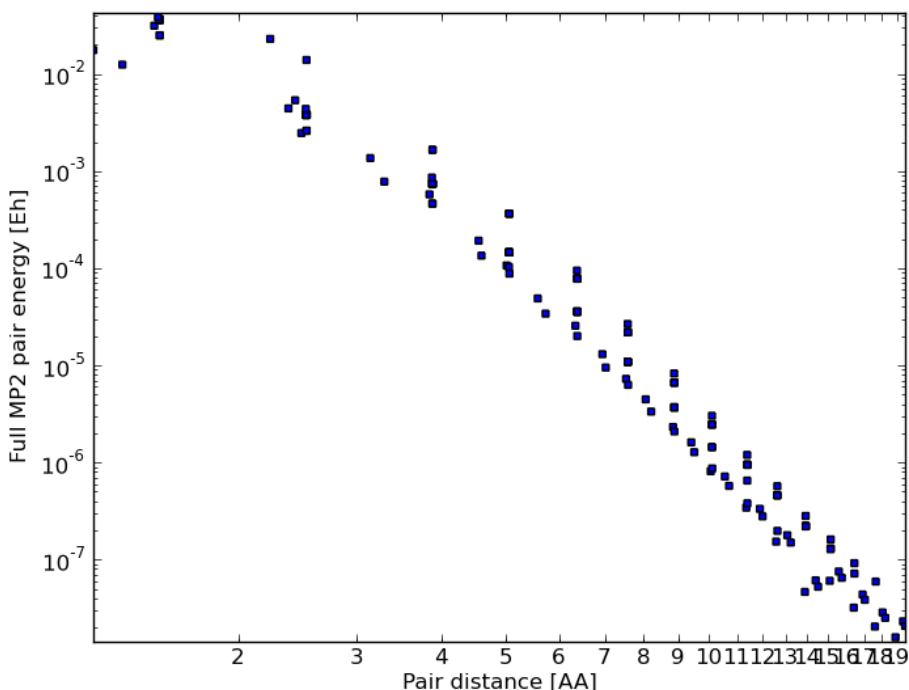


Figure 3.2: Distribution of Pair Fragment Energies as a Function of Inter-Fragment Distance for Palmitic Acid ($C_{16}H_{32}O_2$) constructed from conventional MP2 calculation

Plotting the estimated pair fragment energies against distance as a measure for the error of the estimation yields a distribution that is superficially so similar to the one shown in Figure 3.2, that it is not shown here. Instead the difference between the estimates and the full pair energies is plotted in Figure 3.3. The plot shows that the energy differences decay with a similar slope than the full pair energies shown in Figure 3.2, but with the pair energy differences for inter-fragment distances less than 4 \AA shifted to lower energy difference values.

For further analysis of the difference between estimated and full pair energies, the relative error, i.e. the ratio between the difference and the full pair energy, is plotted in Figure 3.4. The plot shows that below inter-fragment distances of 4 \AA the relative error is below 5 percent. Between distances of 4 and 5 \AA the relative error steeply increases to roughly 40 percent and remains on average at that level for distances larger than 5 \AA , with relative errors ranging between 25 and 50 percent.

The increase in relative error for pairs in which the fragments are more than 4 \AA appears as a shift in energy difference between estimate and full calculation. This can be explained by the radius up to which orbitals are included in the

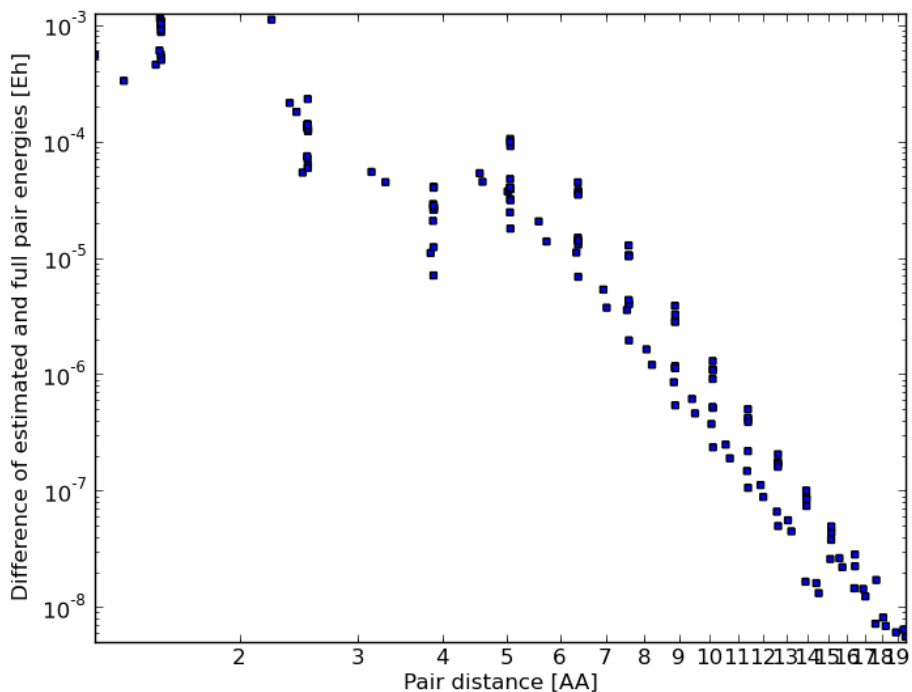


Figure 3.3: Difference between Estimated Pair Fragment Energies and Pair Fragment Energies Constructed From Full MP2 Calculation for Palmitic Acid

virtual spaces of the estimates, which is by default set to 2 Å. This means that beyond a distance of 4 Å orbitals spaces are no longer overlapping and poorer results for estimates are to be expected in those cases. This has been confirmed by repeating the calculations with modified radii for the estimates leading to the same increase in error for pairs with non-overlapping spaces compared to pairs with overlapping spaces.

For the validity of the pair screening process, this means that for the pair fragments with the highest energy contributions for which shorter fragment separations occur the estimates are of comparatively high quality. For further separated fragments, the quality of the estimate falls to a reliable order of magnitude approximation.

3.2.2 Comparison of DEC with Energy Screening and Full Calculation

In light of the result that the quality of estimated pair fragment energies decreases significantly when the orbital spaces of the involved fragments no longer overlap, the question arises whether the energies of DEC optimized

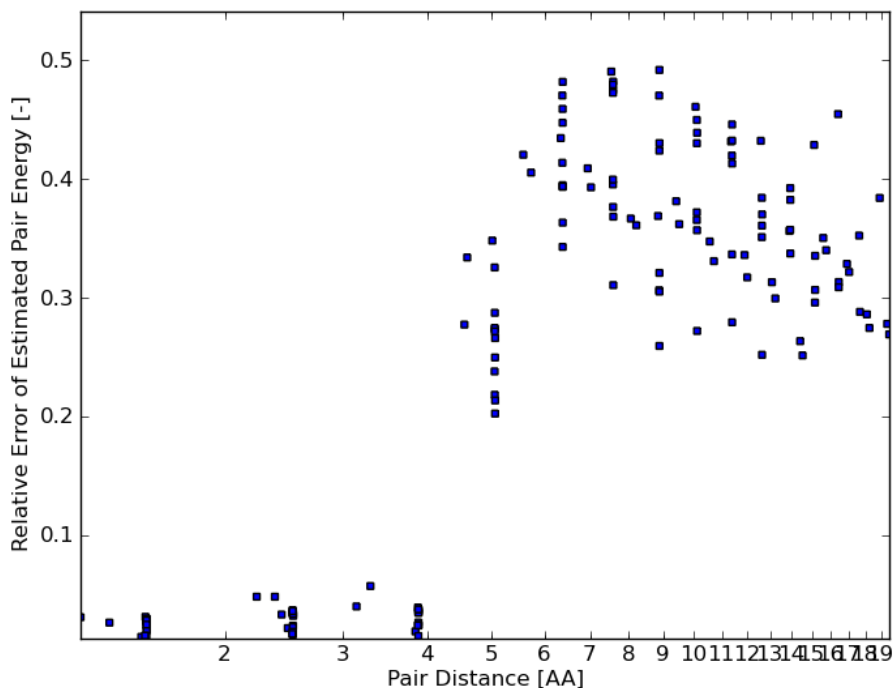


Figure 3.4: Relative Error of Estimated Pair Fragment Energies Compared to Pair Fragment Energies Constructed From Full MP2 Calculation for Palmitic Acid

fragments are subject to a similar limitation for non-overlapping optimized orbital spaces.

In Figure 3.5 the relative error of the DEC pair energies compared to the full pair energies is plotted against the full pair energy for a DEC calculation on Palmitic Acid with the FOT 10^{-5} . In addition the degree of orbital space overlap is indicated. The blue squares are used when the unextended orbital spaces of the fragments overlap, the yellow circles when the optimized orbital spaces overlap and the crosses when no overlap occurs. Furthermore the blue vertical lines indicate ideal cutoff values for the pair fragment screening, determined by applying the criterion described previously in equation 3.1 to full MP2 pair energies instead of estimated ones.

The relative errors found between DEC and full pair energies are overall much lower than the ones found in the comparison of estimated and full pair energies, with the maximum error less than 5 percent. For higher energy pair fragments the relative errors are at the order of 0.1 percent, thus still significantly lower. For pair fragment energies below 10^{-5} Eh the relative errors begin to increase and become larger for smaller pair energies. The overlap

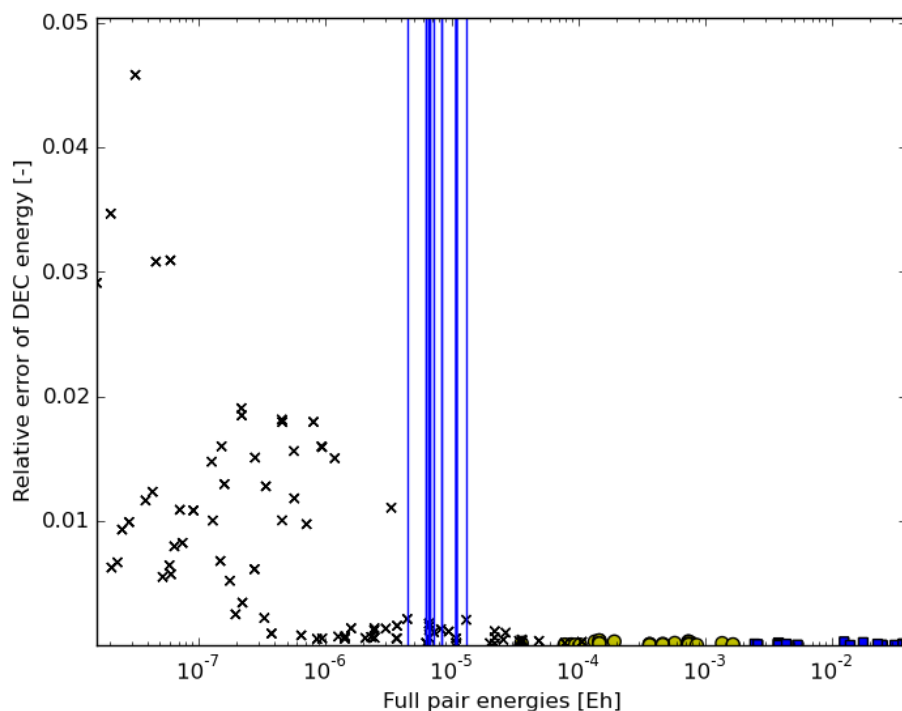


Figure 3.5: Relative Error of Estimated Pair Fragment Energies Compared to DEC Pair Fragment Energies with FOT 10E-5 for Palmitic Acid with Fragment Space Overlaps and Ideal Cutoffs

indicators show that orbital spaces no longer overlap for pair fragments with energies lower than 10^{-4} Eh. The cutoff values are all grouped around 10^{-5} Eh.

These observations show that all DEC pair fragment energies which are used in the calculation of the total energy accurately approximate the full MP2 values regardless of whether orbital spaces do or do not overlap. As seen before for the estimated pair fragment energies the errors are largest for the least relevant contributions and because of the optimized orbital spaces used it is ensured that all significant contributions are included regardless of whether they originate from overlapping or non-overlapping spaces. Further investigations for different FOT values have shown that the cutoff values calculated are close to the FOT and all pair fragment energies larger than the FOT are of high accuracy.

3.2.3 Computational Savings

It remains to be discussed whether the pair fragment screening leads to substantial savings in computational effort which are worth the added computational costs for the calculation of pair energy estimates. For this purpose DEC calculations on Valinomycin have been carried out, one using the pair screening described above and one calculating all possible pair fragment energies. Additionally the computational effort required for calculating the estimated pair energies has been determined. Since the pair estimates are calculated in parallel with the fragment optimizations, the required time has been obtained as the difference in time to solution between the optimization with and without calculation of pair energy estimates.

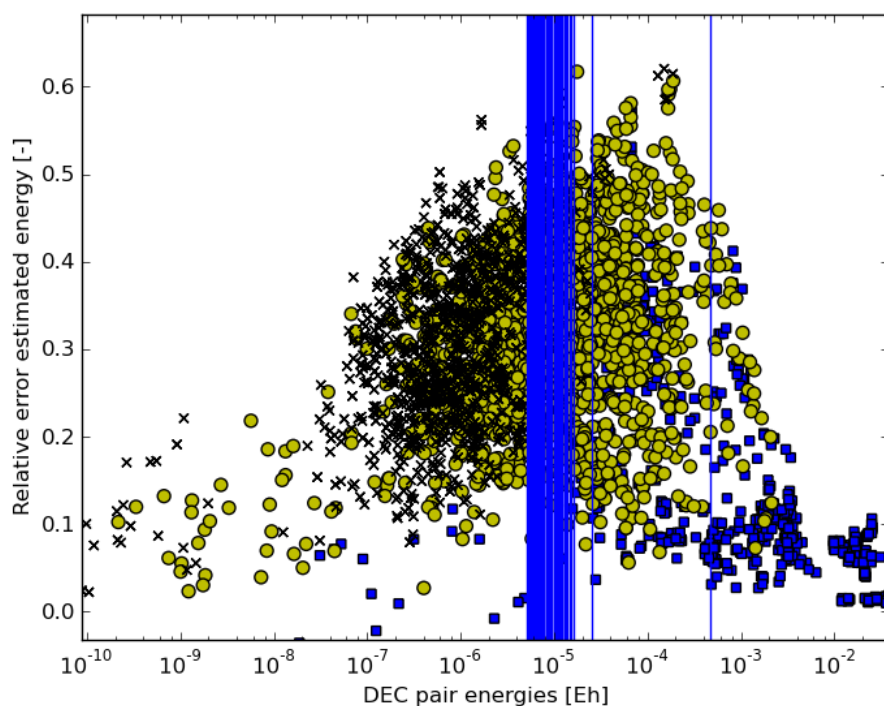


Figure 3.6: Distribution of Pair Fragment Energies as a Function of Inter-Fragment Distance

In Figure 3.6 relative errors of estimated pair fragment energies compared to DEC pair energies are shown together with cutoff energies based on the pair estimates. The distribution of relative errors approximately resembles a gaussian distribution centered around 10^{-5} Eh. The cutoff values are spread around 10^{-5} Eh as well and thus roughly split the distribution of pair fragments in half.

As the plot already indicates a significant number of pair fragments do not need to be calculated accurately in this specific model system. A detailed analysis shows that of the total 3003 pair fragments only 979 remain to be calculated when using the energy screening criterion whereas 2250 remain when a 10 Å distance cutoff is used. The timing analysis shows that the calculation of the pair estimates has required 180 seconds more for the fragment optimization/estimate procedure than was required for a pure optimization procedure. The computational savings from the pair screening can only be extrapolated from the number of pairs and timings of calculations that either included all pairs (8 h 18 m 19 s) or only those remaining after energy screening (1 h 52 m 26 s). In conclusion, the use of estimated pair energies saves between two and three hours of calculating unnecessary pair fragment energies while costing itself only three minutes for calculating the estimates. The savings and timings discussed here are obviously highly specific to the molecular system chosen. In particular the exact costs of calculating the pair estimates depends heavily on details in the scheduling of the fragment optimization. Because of the large decrease in computational costs shown for a model system that well represents the intended use of the DEC scheme, this nevertheless demonstrates the advantage of the pair fragment energy screening approach over the distance cutoff criterion.

Chapter 4

Interaction Energies

4.1 Interaction Energies From Conventional And From Local Calculations

The DEC scheme describes the total correlation energy of a molecular system as a combination of correlation within fragments and the interaction between fragments. As a natural consequence the question arises whether this approach could also be used to efficiently calculate interaction correlation energies for large molecular systems.

The interaction energy between two systems ΔE with given positions of the systems relative to each other is defined as the difference between the energy of the combined system E_{AB} and the energy of the systems at a distance where the systems do not interact. In practice the energy for non-interacting systems is obtained by separate calculation on each of the systems yielding E_A and E_B . The interaction energy is thus calculated as:

$$\Delta E = E_{AB} - E_A - E_B. \quad (4.1)$$

Even though this definition is very simple and intuitive, complications arise with the basis set approximations used for practical calculations. The naive approach when using finite basis sets would be to use one basis set with functions located on nuclei in system A when calculating the energy of system A, functions on B for the energy of B and the combination of both basis sets for the combined system. This however results in an imbalance in the quality of basis sets used. In the combined system the additional basis functions on one subsystem improve the description of the wave function of the other subsystem and vice versa, while this improvement does not occur for the calculations on the isolated systems. As a result, unphysical stabilization of the combined system occurs that leads to an error for interaction the

energy (too attractive interaction energies). This error is denoted the basis set superposition error (BSSE).

One approach to resolve the basis set imbalance is to include the basis set of subsystem B in the calculation of subsystem A (and vice versa). These basis functions would be located at the same relative positions as in the calculation for the combined system, just without the nuclei. This approach is called the counterpoise correction (CPC)[25]. The corrected interaction energy is given by

$$\Delta E^{CP} = E_{AB}^{\{AB\}} - E_A^{\{AB\}} - E_B^{\{AB\}}, \quad (4.2)$$

where $\{AB\}$ is used to emphasize that the basis used contains both the functions of both system A and B.

An alternative, potentially very efficient approach for calculating the correlation part of interaction energies arises from the DEC scheme. Here the interaction correlation energy is calculated as the sum of those pair fragment energy contributions where the two fragments are located in different subsystems, leading to the following expression for the interaction correlation energy:

$$\Delta E_{DEC} = \sum_{P \in A, Q \in B} \Delta E_{PQ}. \quad (4.3)$$

In the following this approach is investigated by comparing DEC interaction correlation energies to counterpoise-corrected interaction correlation energies for the neon dimer and the water dimer with a range of basis sets. For these calculations the pair fragment energies are obtained by carrying out full MP2, CCSD and CCSD(T) calculations using local orbitals and afterwards reconstructing the pair fragment contributions. This approach is equivalent to a DEC calculation with unrestricted virtual orbital spaces. It has been chosen for these preliminary investigations for the simplicity of implementation and eliminations of errors that the truncations of virtual spaces in DEC calculations introduce. The molecular geometries used are taken from an earlier publication[26, 27] on interaction energies.

4.2 Results and Discussion

4.2.1 Comparison of Counterpoise-Corrected and DEC Interaction Correlation Energies

Calculations of both DEC, counterpoise-corrected and uncorrected interaction correlation energies have been carried out on the neon dimer at the MP2,

Table 4.1: MP2 Interaction Correlation Energies for Neon Dimer

Basis	InteractE	InteractE CPC	InteractE DEC
DZ	-2.84929000e-05	-1.24003000e-05	-2.16874535e-05
TZ	-1.06024500e-04	-4.54909000e-05	-7.53574075e-05
QZ	-1.40647800e-04	-8.60588000e-05	-1.49145520e-04
5Z	-1.39284000e-04	-1.16213600e-04	-1.55332607e-04
6Z	-1.50015400e-04	-1.36658000e-04	-1.76240000e-04

Table 4.2: CCSD Interaction Correlation Energies for Neon Dimer

Basis	InteractE	InteractE CPC	InteractE DEC
DZ	-2.61270000e-05	-1.34532000e-05	-2.31156652e-05
TZ	-9.08456000e-05	-4.59278000e-05	-7.91620165e-05
QZ	-1.29650300e-04	-8.90665000e-05	-1.32620017e-04
5Z	-1.44494000e-04	-1.23766600e-04	-1.70794487e-04
6Z	-1.60164700e-04	-1.48158300e-04	-1.98278789e-04

Table 4.3: CCSD(T) Interaction Correlation Energies for Neon Dimer

Basis	InteractE	InteractE CPC	InteractE DEC
DZ	-2.73537000e-05	-1.39647000e-05	-2.33684607e-05
TZ	-1.00310600e-04	-4.83648000e-05	-8.12112010e-05
QZ	-1.45758700e-04	-9.69199000e-05	-1.38226449e-04
5Z	-1.61783900e-04	-1.37763700e-04	-1.79728168e-04
6Z	-1.80901500e-04	-1.66958100e-04	-2.10081138e-04

CCSD, and CCSD(T) levels and with basis sets ranging from double to hex-tuple zeta. The results are shown in Table 4.1 for the MP2 calculations, in 4.2 for CCSD and in 4.3 for CCSD(T).

On all levels of theory both the CPC corrected and the DEC interaction correlation energy can be seen to slowly converge with increasing basis set size, but not to the same value. The DEC values are for larger basis set between 30 to 40 percent larger than the CPC corrected values and thereby also significantly larger than the uncorrected values. The exact interaction correlation energy is usually larger than the uncorrected but smaller than the counterpoise-corrected value. Since the relative difference between the DEC and the CPC values appears to be stable with basis set size the error seems to be systematic.

4.2.2 Connection between Imperfect Localization and Quality of DEC Interaction Correlation Energies

One possible explanation for the incorrect DEC interaction energies is the imperfect localization of molecular orbitals. Even localized orbitals on subsystem A still overlap to a small degree with subsystem B. This leads to unphysical energy contributions from small charge contributions to the other subsystem that could be sizeable enough to cause the errors found for the DEC interaction energies.

Another explanation for the errors of the DEC interaction energies is that the pair fragment energies do not necessarily contain all contributions relevant to the calculation of interaction energies. From investigating the influence of the localization error conclusions can be drawn to what degree the localization error and to what degree contributions missing in the DEC interaction energy are to blame for the error.

As a measure for the charge contribution to the other subsystem the summed up Löwdin charges of orbitals from one system on the other system are calculated. For a water dimer the Löwdin charges are calculated at equilibrium distance and doubled equilibrium distance separation and for different localization methods (fourth moment, second moment and Pipek-Mezey).

In Table 4.2.2 the uncorrected, CPC and DEC interaction correlation energies for the water dimer are compared for different the different localization methods, for equilibrium and double equilibrium distance and for double zeta, augmented double zeta, quadruple zeta and augmented quadruple zeta basis sets. For the same set of calculations for each subsystem the Löwdin charges on the other subsystem are shown in Table 4.2.2 as well as the combined total Löwdin charge on other subsystems and the DEC interaction correlation energy.

For the interaction correlation energies in Table 4.2.2 the same slow convergence of the energies with basis set size is visible that was already observed for the neon dimer calculations. The localization method used appears to have very little influence on the resulting interaction energies. For doubled equilibrium distance calculations the uncorrected and CPC energies are positive for both uncorrected and CPC calculations, indicating that the correlation part of the interaction energy acts repulsive in these cases, while the DEC energies are of similar absolute value but negative. For equilibrium distances the DEC energy shows the same error as for the neon dimer and gives close to twice the CPC energy for the QZ and augQZ and worse for DZ and augDZ. The Löwdin charges in Table 4.2.2 increase with basis set size and depend more strongly on the localization method than the energies. The charges are much lower for the calculation carried out at doubled equilibrium distance. In

Table 4.4: Water Dimer Interaction Correlation Energies

Calculation	InteractE	InteractE CPC	InteractE DEC
DZ pfm	-2.54104050e-03	-4.09998800e-04	-2.36311566e-03
DZ psm	-2.54104050e-03	-4.09998800e-04	-2.34136908e-03
DZ pml	-2.54104050e-03	-4.09998800e-04	-2.31884328e-03
augDZ pfm	-2.24086730e-03	-1.27427240e-03	-3.33553812e-03
augDZ psm	-2.24086730e-03	-1.27427240e-03	-3.31774115e-03
augDZ pml	-2.24086730e-03	-1.27427240e-03	-3.28388460e-03
DZ pfm dist2	5.89340000e-05	6.05417000e-05	-1.23834987e-05
DZ psm dist2	5.89340000e-05	6.05417000e-05	-1.23833574e-05
DZ pml dist2	5.89341000e-05	6.05418000e-05	-1.24203598e-05
augDZ pfm dist2	3.62735000e-05	5.49908000e-05	-2.84544811e-05
augDZ psm dist2	3.62735000e-05	5.49908000e-05	-2.84546346e-05
augDZ pml dist2	3.62735000e-05	5.49908000e-05	-2.84950623e-05
QZ pfm	-2.54152160e-03	-1.81736960e-03	-3.74890275e-03
QZ psm	-2.54152150e-03	-1.81736950e-03	-3.72620365e-03
QZ pml	-2.54152150e-03	-1.81736930e-03	-3.71425541e-03
augQZ pfm	-2.36122370e-03	-2.03111940e-03	-3.94093734e-03
augQZ psm	-2.36122390e-03	-2.03111960e-03	-3.91984295e-03
augQZ pml	-2.36122360e-03	-2.03111990e-03	-3.88785212e-03
QZ pfm dist2	2.89423000e-05	4.38978000e-05	-2.55276445e-05
QZ psm dist2	2.89421000e-05	4.38973000e-05	-2.55261127e-05
QZ pml dist2	2.89419000e-05	4.38974000e-05	-2.55618086e-05
augQZ pfm dist2	3.55496000e-05	4.29065000e-05	-3.09175457e-05
augQZ psm dist2	3.55495000e-05	4.29063000e-05	-3.09180661e-05
augQZ pml dist2	3.55495000e-05	4.29067000e-05	-3.11194087e-05

general increased Löwdin charges coincide with more attractive DEC interaction correlation energies but the relation is clearly not linear. The relative size of the error of the DEC interaction energy compared to the CPC energy decreases with basis set size and is larger of double distance while the opposite is true for the Löwdin charges.

The most plausible explanation is that the changes in both the Löwdin charges and the DEC interaction energies are both caused by changes in the basis set size. The data discussed suggests that the localization error is not the main contribution to the error of the DEC interaction correlation energy. Instead the investigation of missing contributions to the DEC interaction energy should be further investigated.

Table 4.5: Löwdin Charges on other subsystem

Basis	AF1 LOWDIN	AF2 LOWDIN	LOWDIN TOTAL	InteractE DEC
DZ pfm	5.03872000e-03	5.25533400e-02	5.75920600e-02	-2.36311566e-03
DZ psm	4.67458000e-03	5.21891400e-02	5.68637200e-02	-2.34136908e-03
DZ pml	4.16508000e-03	5.16797200e-02	5.58448000e-02	-2.31884328e-03
augDZ pfm	2.66629800e-02	9.52887400e-02	1.21951720e-01	-3.33553812e-03
augDZ psm	2.63951800e-02	9.50209000e-02	1.21416080e-01	-3.31774115e-03
augDZ pml	2.58517000e-02	9.44774200e-02	1.20329120e-01	-3.28388460e-03
DZ pfm dist2	2.00000000e-07	1.19200000e-05	1.21200000e-05	-1.23834987e-05
DZ psm dist2	1.80000000e-07	1.19200000e-05	1.21000000e-05	-1.23833574e-05
DZ pml dist2	5.20000000e-07	1.22200000e-05	1.27400000e-05	-1.24203598e-05
augDZ pfm dist2	3.14400000e-05	3.17920000e-04	3.49360000e-04	-2.84544811e-05
augDZ psm dist2	3.14600000e-05	3.17880000e-04	3.49340000e-04	-2.84546346e-05
augDZ pml dist2	3.15400000e-05	3.17980000e-04	3.49520000e-04	-2.84950623e-05
QZ pfm	1.05916600e-02	9.39267000e-02	1.04518360e-01	-3.74890275e-03
QZ psm	1.02448400e-02	9.35799000e-02	1.03824740e-01	-3.72620365e-03
QZ pml	9.68528000e-03	9.30203800e-02	1.02705660e-01	-3.71425541e-03
augQZ pfm	5.63507000e-02	2.28765560e-01	2.85116260e-01	-3.94093734e-03
augQZ psm	5.61310000e-02	2.28545880e-01	2.84676880e-01	-3.91984295e-03
augQZ pml	5.57448800e-02	2.28159780e-01	2.83904660e-01	-3.88785212e-03
QZ pfm dist2	2.20000000e-06	1.03340000e-04	1.05540000e-04	-2.55276445e-05
QZ psm dist2	2.16000000e-06	1.03260000e-04	1.05420000e-04	-2.55261127e-05
QZ pml dist2	2.10000000e-06	1.03180000e-04	1.05280000e-04	-2.55618086e-05
augQZ pfm dist2	5.24200000e-05	3.05000000e-04	3.57420000e-04	-3.09175457e-05
augQZ psm dist2	5.23800000e-05	3.05000000e-04	3.57380000e-04	-3.09180661e-05
augQZ pml dist2	5.36200000e-05	3.06220000e-04	3.59840000e-04	-3.11194087e-05

Bibliography

- [1] E.R. Johnson, R.A. Wolkow, and G.A. DiLabio. Application of 25 density functionals to dispersion-bound homomolecular dimers. *Chemical Physics Letters*, 394:334–338, 2004.
- [2] Sergio Felipe Sousa, Pedro Alexandrino Fernandes, and Maria Joao Ramos. General performance of density functionals. *Journal of Physical Chemistry A*, 111:10439–10452, 2007.
- [3] Peter Pulay. Localizability of dynamic electron correlation. *Chemical Physics Letters*, 100:151–154, 1983.
- [4] S. Saebo and P. Pulay. Local treatment of electron correlation. *Annual Review of Physical Chemistry*, 44:213–236, 1993.
- [5] Bernd Doser, Daniel S. Lambrecht, Jörg Kussmann, and Christian Ochsenfeld. Linear-scaling atomic orbital-based second-order møller-plesset perturbation theory by rigorous integral screening criteria. *The Journal of Chemical Physics*, 130:064107, 2009.
- [6] C. Hampel and H.-J. Werner. Local treatment of electron correlation in coupled cluster theory. *Journal of Chemical Physics*, 104:6286–6297, 1996.
- [7] G.E. Scuseria and P.Y. Ayala. Linear scaling coupled cluster and perturbation series in the atomic orbital basis. *Journal of Chemical Physics*, 111:8330–8343, 1999.
- [8] N. Flocke and R.J. Bartlett. A natural linear scaling coupled-cluster method. *Journal of Chemical Physics*, 121:10935–10944, 2004.
- [9] Wei Li, Piotr Piecuch, Jeffrey R. Gour, and Shuhua Li. Local correlation calculations using standard and renormalized coupled-cluster approaches. *Journal of Chemical Physics*, 131:114109, 2009.

- [10] Masato Kobayashi and Hiromi Nakai. Extension of linear-scaling divide-and-conquer-based correlation method to coupled cluster theory with singles and doubles excitations. *Journal of Chemical Physics*, 129:044103, 2008.
- [11] Marcin Ziólkowski, Branislav Jansík, Thomas Kjægaard, and Poul Jørgensen. Linear scaling coupled cluster method with correlation energy based error control. *The Journal of Chemical Physics*, 133:014107, 2010.
- [12] Christine Krause and Hans-Joachim Werner. Comparison of explicitly correlated local coupled-cluster methods with various choices of virtual orbitals. *Phys. Chem. Chem. Phys.*, 14:7591–7604, 2012.
- [13] Kasper Kristensen, Ida-Marie Høyvik, Branislav Jansík, Poul Jørgensen, Thomas Kjægaard, Simen Reine, and Jacek Jakowski. Mp2 energy and density for large molecular systems with internal error control using the divide-expand-consolidate scheme. *Phys. Chem. Chem. Phys.*, 14:15706–15714, 2012.
- [14] Kasper Kristensen, Marcin Ziólkowski, Branislav Jansík, Thomas Kjægaard, and Poul Jørgensen. A locality analysis of the divide-expand-consolidate coupled cluster amplitude equations. *Journal of Chemical Theory and Computation*, 7:1677–1694, 2011.
- [15] M. Ziólkowski, B. Jansík, P. Jørgensen, and J. Olsen. Maximum locality in occupied and virtual orbital spaces using a least-change strategy. *Journal of Chemical Physics*, 131:124112, 2009.
- [16] Ida-Marie Høyvik, Branislav Jansík, and Poul Jørgensen. Trust region minimization of orbital localization function. *Journal of Chemical Theory and Computation*, 8:3137–3146, 2012.
- [17] Ida-Marie Høyvik, Branislav Jansík, and Poul Jørgensen. Orbital localization using fourth central moment minimization. *The Journal of Chemical Physics*, 137:224114, 2012.
- [18] S.F. Boys. Construction of some molecular orbitals to be approximately invariant for changes from one molecule to another. *Reviews of Modern Physics*, 32:296–299, 1960.
- [19] C. Edmiston and K. Ruedenberg. Localized atomic and molecular orbitals. *Reviews of Modern Physics*, 35:457–464, 1963.

- [20] J. Pipek and P.G. Mezey. A fast intrinsic localization procedure applicable for abinitio and semiempirical linear combination of atomic orbital wave-functions. *Journal of Chemical Physics*, 90:4916–4926, 1989.
- [21] J.E. Subotnik, A.D. Dutoi, and M. Head-Gordon. Fast localized orthonormal virtual orbitals which depend smoothly on nuclear coordinates. *Journal of Chemical Physics*, 123:114108, 2005.
- [22] Ida-Marie Høyvik, Branislav Jansík, Kasper Kristensen, and Poul Jørgensen. Local hartree-fock orbitals using a three-level optimization strategy for the energy. *Journal of Computational Chemistry*, 34:1311–1320, 2013.
- [23] Kasper Kristensen, Thomas Kjægaard, Ida-Marie Høyvik, Patrick Etenhuber, Poul Jørgensen, Branislav Jansík, Simen Reine, and Jacek Jakowski. The divide-expand-consolidate mp2 scheme goes massively parallel. *Molecular Physics*, 111:1196–1210, 2013.
- [24] Ida-Marie Høyvik, Kasper Kristensen, Branislav Jansík, and Poul Jørgensen. The divide-expand-consolidate family of coupled cluster methods: Numerical illustrations using second order møller-plesset perturbation theory. *The Journal of Chemical Physics*, 136:014105, 2012.
- [25] Frans B. van Duijnveldt, Jeanne G.C.M. van Duijnveldt-van de Rijdt, and Joop H. van Lenthe. State of the art in counterpoise theory. *Chemical Reviews*, 94:1873–1885, 1994.
- [26] Ronald A. Aziz and M.J. Slaman. The ne-ne interaction potential revisited. *Chemical Physics*, 130:187–194, 1989.
- [27] Asger Halkier, Wim Klopper, Trygve Helgaker, Poul Jørgensen, and Peter R. Taylor. Basis set convergence of the interaction energy of hydrogen-bonded complexes. *Journal of Chemical Physics*, 111:9157–9167, 1999.