The subject of a density in coupled-cluster theory is approached from the viewpoint of response theory. Imagine one adds a perturbation of the form

$$H^{(\lambda)} = \lambda \sum_{pq} h_{pq}^{\lambda} E_p^q$$

to the Hamiltonian. Then it is always possible to cast the first derivative of the total energy in the form:

$$\frac{dE}{d\lambda} = \sum_{pq} D_{pq}^{(\text{response})} h_{pq}^{\lambda}$$

This is a nice result. The quantity $D_{pq}^{(\text{response})}$ is the so-called response density. In the case of CC theory where the energy is not obtained by variational optimization of an energy functional, the energy has to be replaced by a Lagrangian reading as follows:

$$L_{CC} = \langle \Phi_0 | \bar{H} | \Phi_0 \rangle + \sum_{\eta} \lambda_{\eta} \langle \Phi_{\eta} | \bar{H} | \Phi_0 \rangle + \sum_{ai} f_{ai} z_{ai}$$

Here $\langle \Phi_{\eta} |$ denotes any excited determinant (singly, doubly, triply, ….). There are two sets of Lagrange multipliers: the quantities z_{ai} that guarantee that the perturbed wavefunction fulfills the Hartree-Fock conditions by making the off-diagonal Fock matrix blocks zero and the quantities λ_{η} that guarantee that the coupled-cluster projection equations for the amplitudes are fulfilled. If both sets of conditions are fulfilled then the coupled-cluster Lagrangian simply evaluates to the coupled-cluster energy. The coupled-cluster Lagrangian can be made stationary with respect to the Lagrangian multipliers z_{ai} and λ_{η}. The response density is then defined through:

$$\frac{dL_{CC}}{d\lambda} = \sum_{pq} D_{pq}^{(\text{response})} h_{pq}^{\lambda}$$

The density D_{pq} appearing in this equation does not have the same properties as the density that would arise from an expectation value. For example, the response density can have eigenvalues lower than 0 or larger than 2. In practice, the response density is, however, the best "density" there is for coupled-cluster theory.

Unfortunately, the calculation of the coupled-cluster response density is quite involved because additional sets of equations need to be solved in order to determine the z_{ai} and λ_{η}. If only the equations for λ_{η} are solved one speaks of an "unrelaxed" coupled-cluster density. If both sets of equations are solved, one speaks of a "relaxed" coupled-cluster density. For most intents and purposes, the orbital relaxation effects incorporated into the relaxed density are small for a coupled-cluster density. This is so, because the coupled-cluster equations contain the exponential of the single excitation operator $e^{\hat{T}_1} = \exp(\sum_{ai} t_a^i E_i^a)$. This brings in most of the effects of orbital relaxation. In fact, replacing the $\hat{T}_1$ operator by the operator $\hat{\kappa} = \sum_{ai} \kappa_a^i (E_i^a - E_a^i)$ would provide all of the orbital relaxation thus leading to "orbital optimized coupled-cluster theory" (OOCC).

Not surprisingly, the equations that determine the coefficients λ_{η} (the Lambda equations) are as complicated as the coupled-cluster amplitude equations themselves. Hence, the calculation of the unrelaxed coupled-cluster density matrix is about twice as expensive as the calculation of the coupled-cluster energy (but not quite as with proper program organization terms can be reused and the Lambda equations are linear equations that converge somewhat better than the non-linear amplitude equations).

ORCA features the calculation of the unrelaxed coupled-cluster density on the basis of the Lambda equations for closed- and open-shell systems. If a fully relaxed coupled-cluster density is desired then ORCA still features the orbital-optimized coupled-cluster doubles method (OOCCD). This is not exactly equivalent to the fully relaxed CCSD density matrix because of the operator $\hat{\kappa}$ instead of $\hat{T}_1$. However, results are very close and orbital optimized coupled-cluster doubles is the method of choice if orbital relaxation effects are presumed to be large.

In terms of ORCA keywords, the coupled-cluster density is obtained through the following keywords:

```
#
# coupled-cluster density
#
%mdci   density    none
                   linearized
                   unrelaxed
                   orbopt
        end
```

which will work together with CCSD or QCISD (QCISD and CCSD are identical in the case of OOCCD because of the absence of single excitations). Note, that an unrelaxed density for CCSD(T) is NOT available.

Instead of using the density option "orbopt" in the mdci-block, OOCCD can also be invoked by using the keyword:

```
! OOCCD
```

´

3.10.4 Static versus Dynamic Correlation

Let us look at an "abuse" of the single reference correlation methods by studying (very superficially) a system which is not well described by a single HF determinant. This already occurs for the twisting of the double bond of C_2H_4. At a 90° twist angle the system behaves like a diradical and should be described by a multireference method (see section *Complete and Incomplete Active Space Self-Consistent Field (CASSCF and RAS/ORMAS)*)
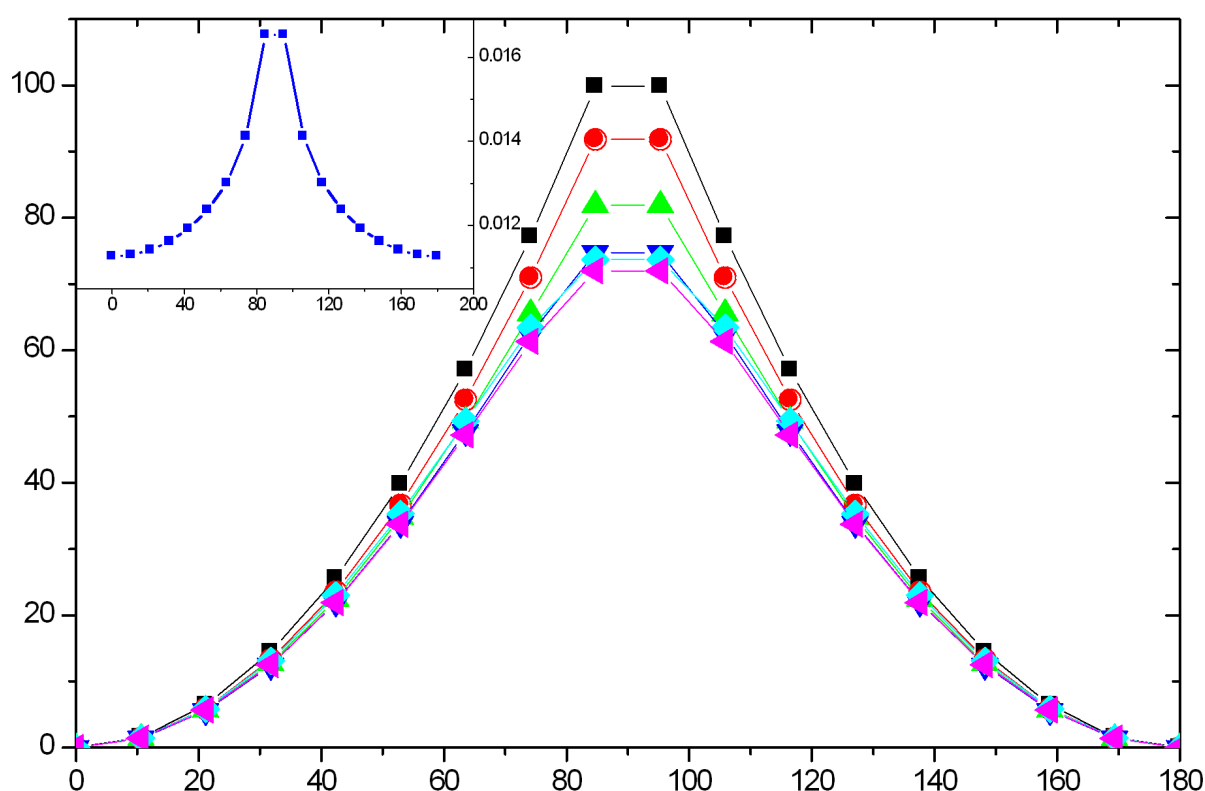

Fig. 3.4: A rigid scan along the twisting coordinate of C_2H_4. The inset shows the T_1 diagnostic for the CCSD calculation.

As can be seen in Fig. 3.4, there is a steep rise in energy as one approaches a 90° twist angle. The HF curve is actually discontinuous and has a cusp at 90°. This is immediately fixed by a simple CASSCF(2,2) calculation which gives a smooth potential energy surface. Dynamic correlation is treated on top of the CASSCF(2,2) method with the MRACPF approach as follows:

```
#
# twisting the double bond of C2H4
#
! SV(P) def2-TZVP/C SmallPrint NoPop MRACPF
%casscf nel         2
        norb        2
        mult        1
        nroots      1
```

```
        TrafoStep  RI
        end
%mrci   tsel 1e-10
        tpre 1e-10
        end
%method scanguess pmodel
        end
%paras  R= 1.3385
        Alpha=0,180,18
        end
* int 0 1
C   0   0   0   0      0    0
C   1   0   0 {R}      0    0
H   1   2   0 1.07   120    0
H   1   2   3 1.07   120  180
H   2   1   3 1.07   120 {Alpha}
H   2   1   3 1.07   120 {Alpha+180}
*
```

This is the reference calculation for this problem. One can see that the RHF curve is far from the MRACPF reference but the CASSCF calculation is very close. Thus, dynamic correlation is not important for this problem! It only appears to be important since the RHF determinant is such a poor choice. The MP2 correlation energy is insufficient in order to repair the RHF result. The CCSD method is better but still falls short of quantitative accuracy. Finally, the CCSD(T) curve is very close the MRACPF. This even holds for the total energy (inset of Fig. 3.5) which does not deviate by more than 2–3 mEh from each other. Thus, in this case one uses the powerful CCSD(T) method in an inappropriate way in order to describe a system that has multireference character. Nevertheless, the success of CCSD(T) shows how stable this method is even in tricky situations. The "alarm" bell for CCSD and CCSD(T) is the so-called "T_1-diagnostic"[6] that is also shown in Fig. 3.5. A rule of thumb says, that for a value of the diagnostic of larger than 0.02 the results are not to be trusted. In this calculation we have not quite reached this critical point although the T_1 diagnostic blows up around the $90°$ twist.

---

[6] It is defined as $\|T_1\| / N^{1/2}$ where T_1 are the singles amplitudes and N the number of correlated electrons. The original reference is [415]
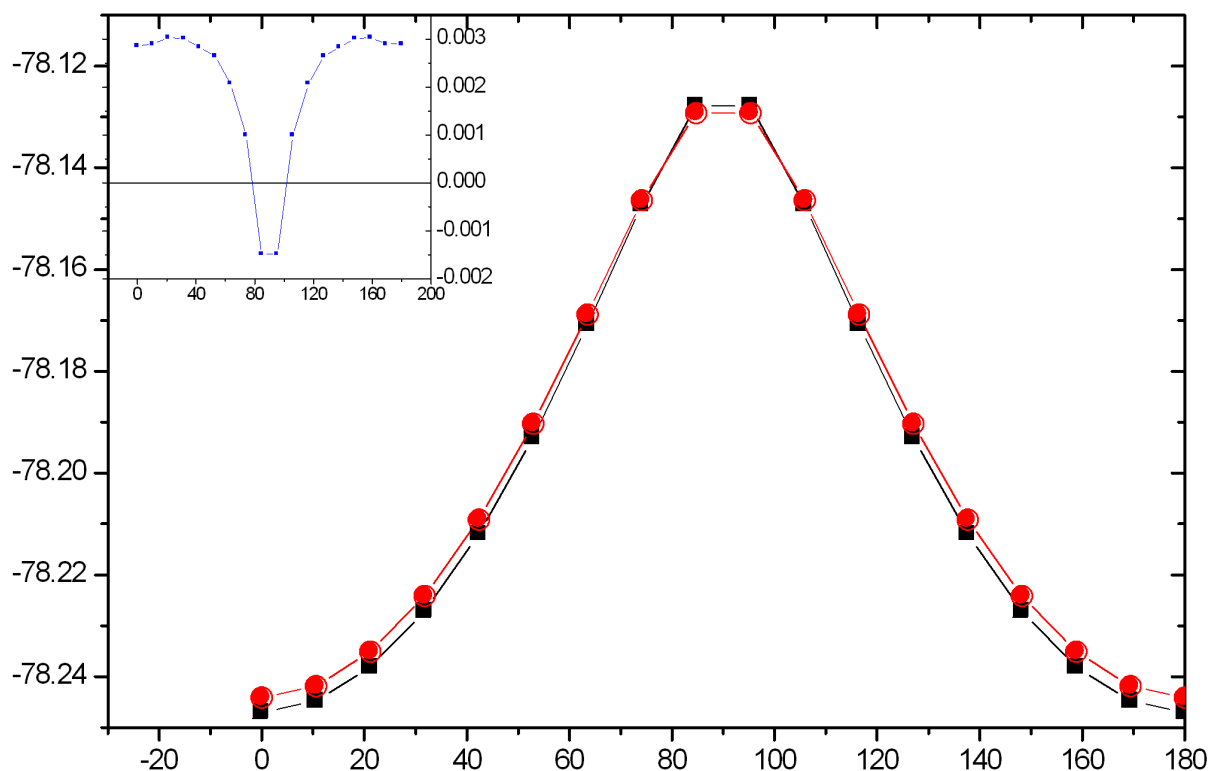
Fig. 3.5: Comparison of the CCSD(T) and MRACPF total energies of the C_2H_4 along the twisting coordinate. The inset shows the difference E(MRACPF)-E(CCSD(T)).

The computational cost (disregarding the triples) is such that the CCSD method is the most expensive followed by QCISD ($\sim$10% cheaper) and all methods (about 50% to a factor of two cheaper than CCSD). The most accurate method is generally CCSD(T). However, this is not so clear if the triples are omitted and in this regime the coupled pair methods (in particular CPF/1 and NCPF/1[7]) can compete with CCSD.

Let us look at the same type of situation from a slightly different perspective and dissociate the single bond of F_2. As is well known, the RHF approximation fails completely for this molecule and predicts it to be unbound. Again we use a much too small basis set for quantitative results but it is enough to illustrate the principle.

We first generate a "reference" PES with the MRACPF method:

```
! def2-SV def2-SVP/C MRACPF
%casscf nel    2
        norb   2
        nroots 1
        mult   1
        end
%mrci   tsel    1e-10
        tpre    1e-10
        end
%paras  R= 3.0,1.3,35
        end
* xyz 0 1
F 0 0 0
F 0 0 {R}
*
```

Note that we scan from outward to inward. This helps the program to find the correct potential energy surface since at large distances the σ and σ^* orbitals are close in energy and fall within the desired 2×2 window for the CASSCF

---

[7] The "N" methods have been suggested by [416] and are exclusive to ORCA. Please note that our NCPF/1 is different from the MCPF method in the literature [371]. The original CPF method — which we prefer — is from [368]; see also [417] for a nice review about the coupled pair approaches and the underlying philosophy.

calculation (see section *Complete and Incomplete Active Space Self-Consistent Field (CASSCF and RAS/ORMAS)*). Comparing the MRACPF and CASSCF curves it becomes evident that the dynamic correlation brought in by the MRACPF procedure is very important and changes the asymptote (loosely speaking the binding energy) by almost a factor of two (see Fig. 3.6). Around the minimum (roughly up to 2.0 Å) the CCSD(T) and MRACPF curves agree beautifully and are almost indistinguishable. Beyond this distance the CCSD(T) calculation begins to diverge and shows an unphysical behavior while the multireference method is able to describe the entire PES up to the dissociation limit. The CCSD curve is qualitatively OK but has pronounced quantitative shortcomings: it predicts a minimum that is much too short and a dissociation energy that is much too high. Thus, already for this rather "simple" molecule, the effect of the connected triple excitations is very important. Given this (rather unpleasant) situation, the behavior of the much simpler CEPA method is rather satisfying since it predicts a minimum and dissociation energy that is much closer to the reference MRACPF result than CCSD or CASSCF. It appears that in this particular case CEPA/1 and CEPA/2 predict the correct result.
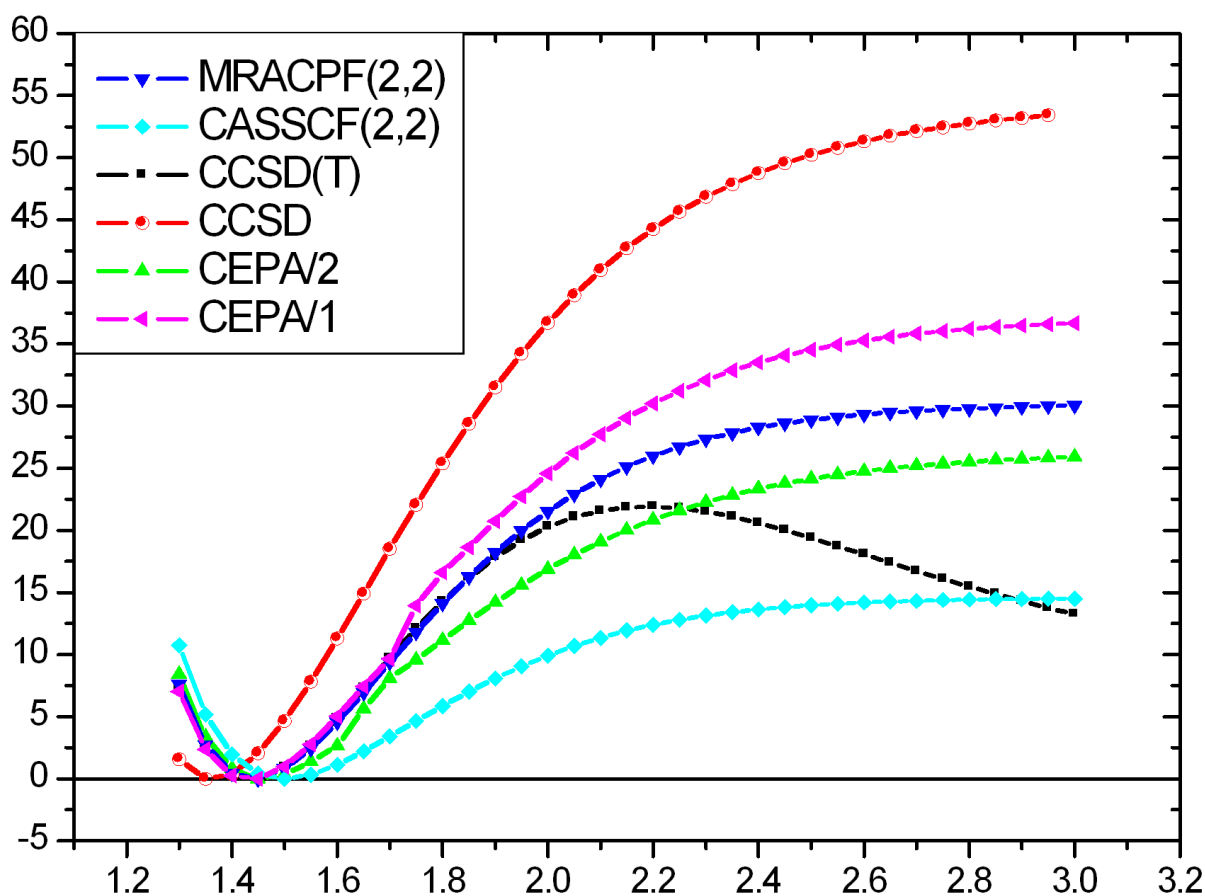

Fig. 3.6: Potential energy surface of the F_2 molecule calculated with some single-reference methods and compared to the MRACPF reference.

3.10.5 Basis Sets for Correlated Calculations. The case of ANOs.

In HF and DFT calculations the generation and digestion of the two-electron repulsion integrals is usually the most expensive step of the entire calculation. Therefore, the most efficient approach is to use loosely contracted basis sets with as few primitives as possible — the Ahlrichs basis sets (SVP, TZVP, TZVPP, QZVP, def2-TZVPP, def2-QZVPP) are probably the best in this respect. Alternatively, the polarization-consistent basis sets pc-1 through pc-4 could be used, but they are only available for H-Ar. For large molecules such basis sets also lead to efficient prescreening and consequently efficient calculations.

This situation is different in highly correlated calculations such as CCSD and CCSD(T) where the effort scales steeply with the number of basis functions. In addition, the calculations are usually only feasible for a limited number of basis functions and are often run in the integral conventional mode, since high angular momentum basis functions are present and these are expensive to recompute all the time. Hence, a different strategy concerning the basis set

design seems logical. It would be good to use as few basis functions as possible but make them as accurate as possible. This is compatible with the philosophy of atomic natural orbital (ANO) basis sets. Such basis sets are generated from correlated atomic calculations and replicate the primitives of a given angular momentum for each basis function. Therefore, these basis sets are deeply contracted and expensive but the natural atomic orbitals form a beautiful basis for molecular calculations. In ORCA an accurate and systematic set of ANOs (ano-pVnZ, n = D, T, Q, 5 is incorporated). A related strategy underlies the design of the correlation-consistent basis sets (cc-pVnZ, n = D, T, Q, 5, 6,...) that are also generally contracted except for the outermost primitives of the "principal" orbitals and the polarization functions that are left uncontracted.

Let us study this subject in some detail using the H_2CO molecule at a standard geometry and compute the SCF and correlation energies with various basis sets. In judging the results one should view the total energy in conjunction with the number of basis functions and the total time elapsed. Looking at the data in the Table below, it is obvious that the by far lowest SCF energies for a given cardinal number (2 for double-zeta, 3 for triple zeta and 4 for quadruple-zeta) are provided by the ANO basis sets. Using specially optimized ANO integrals that are available since ORCA 2.7.0, the calculations are not even much more expensive than those with standard basis sets. Obviously, the correlation energies delivered by the ANO bases are also the best of all 12 basis sets tested. Hence, ANO basis sets are a very good choice for highly correlated calculations. The advantages are particularly large for the early members (DZ/TZ).

Table 3.38: Comparison of various basis sets for highly correlated calculations

Basis set	No. Basis Fcns	E(SCF)	E_C(CCSD(T))	E_{tot}(CCSD(T))	Total Time
cc-pVDZ	38	-113.876184	-0.34117952	-114.217364	2
cc-pVTZ	88	-113.911871	-0.42135475	-114.333226	40
cc-pVQZ	170	-113.920926	-0.44760332	-114.368529	695
def2-SVP	38	-113.778427	-0.34056109	-114.118988	2
def2-TZVPP	90	-113.917271	-0.41990287	-114.337174	46
def2-QZVPP	174	-113.922738	-0.44643753	-114.369175	730
pc-1	38	-113.840092	-0.33918253	-114.179274	2
pc-2	88	-113.914256	-0.41321906	-114.327475	43
pc-3	196	-113.922543	-0.44911659	-114.371660	1176
ano-pVDZ	38	-113.910571	-0.35822337	-114.268795	12
ano-pVTZ	88	-113.920389	-0.42772994	-114.348119	113
ano-pVQZ	170	-113.922788	-0.44995355	-114.372742	960
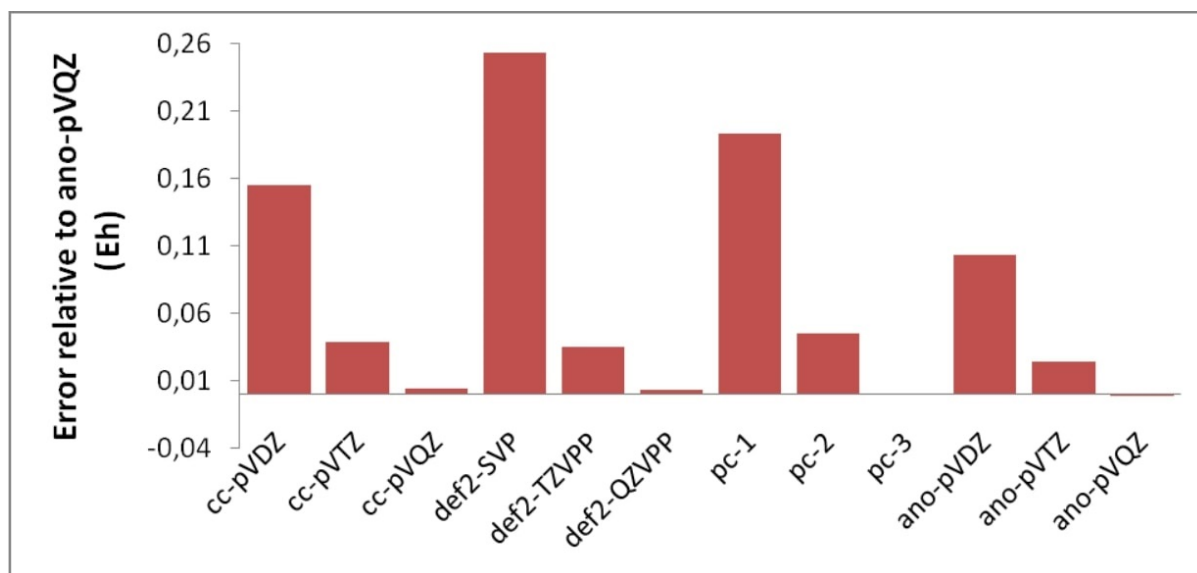
Fig. 3.7: Error in Eh for various basis sets for highly correlated calculations relative to the ano-pVQZ basis set.

Let us look at one more example in Table 3.39: the optimized structure of the N_2 molecule as a function of basis set using the MP2 method *(these calculations are a bit older from the time when the ano-pVnZ basis sets did not yet exist. Today, the ano-pVnZ would be preferred)* .

The highest quality basis set here is QZVP and it also gives the lowest total energy. However, this basis set contains up to g-functions and is very expensive. Not using g-functions and a set of f-functions (as in TZVPP) has a noticeable effect on the outcome of the calculations and leads to an overestimation of the bond distance of 0.2 pm — a small change but for benchmark calculations of this kind still significant. The error made by the TZVP basis set that lacks the second set of d-functions on the bond distance, binding energy and ionization potential is surprisingly small even though the deletion of the second d-set "costs" more than 20 mEh in the total energy as compared to TZV(2d,2p), and even more compared to the larger TZVPP.

A significant error on the order of $1 - 2$ pm in the calculated distances is produced by smaller DZP type basis sets, which underlines once more that such basis sets are really too small for correlated molecular calculations — the ANO-DZP basis sets are too strongly biased towards the atom, while the "usual" molecule targeted DZP basis sets like SVP have the d-set designed to cover polarization but not correlation (the correlating d-functions are steeper than the polarizing ones). The performance of the very economical SVP basis set should be considered as very good, and (a bit surprisingly) slightly better than cc-pVDZ despite that it gives a higher absolute energy.

Essentially the same picture is obtained by looking at the (uncorrected for ZPE) binding energy calculated at the MP2 level – the largest basis set, QZVP, gives the largest binding energy while the smaller basis sets underestimate it. The error of the DZP type basis sets is fairly large (≈ 2 eV) and therefore caution is advisable when using such bases.

Table 3.39: Comparison of various basis sets for correlated calculations.

Basis set	R_{eq} (pm)	E(2N-N$_2$) (eV)	IP(N/N$^+$) (eV)	E(MP2) (Eh)
SVP	112.2	9.67	14.45	**-109.1677**
cc-pVDZ	112.9	9.35	14.35	**-109.2672**
TZVP	111.5	10.41	14.37	**-109.3423**
TZV(2d,2p)	111.4	10.61	14.49	**-109.3683**
TZVPP	111.1	10.94	14.56	**-109.3973**
QZVP	110.9	11.52	14.60	**-109.4389**

3.10.6 The Coupled Cluster S-Diagnostic

There have been a number of diagnostic proposed for coupled cluster calculations that are designed to indicate "multi-reference" character. Unfortunately, the most widely used of them is the "T1-diagnostic". We have argued for a long time, that the T1-diagnostic has nothing to do with multi-reference character. Rather, it measures the extent of orbital relaxation in the dynamic correlation field. While this is a measure of how good or bad the Hartree-Fock approximation works for the given system, it is expressis verbis NOT related to the multi-configuration character that the CC wavefunction may or may not want to emulate. For a discussion of what we consider as multi-reference versus single-reference character, please consult the paper by Izsak et al.

ORCA has an alternative coupled cluster diagnostic, known as the "S-diagnostic" that was introduced by Perdersen et al. This diagnostic does indeed measure multi-reference character. It is available for canonical RHF and UHF CCSD and is invoked by

```
%mdci DoSDiagonstic true
end
```

Relevant Papers:

1. Liakos, D. G.; Neese, F. Interplay of Correlation and Relativistic Effects in Correlated Calculations on Transition-Metal Complexes: The Cu$_2$O$_2$$^{2+}$ Core Revisited. *J. Chem. Theory Comput.*, **2011**, 7, 1511–1523.

2. Izsák, Róbert; Ivanov, Aleksei V.; Blunt, Nick S.; Holzmann, Nicole; Neese, Frank. Measuring Electron Correlation: The Impact of Symmetry and Orbital Transformations. *Journal of Chemical Theory and Computation*, **2023**, 19 (10), 2703–2720. PMID: 37022051. arXiv:https://doi.org/10.1021/acs.jctc.3c00122, DOI: 10.1021/acs.jctc.3c00122.

3. Faulstich, Fabian M.; Kristiansen, Håkon E.; Csirik, Mihaly A.; Kvaal, Simen; Pedersen, Thomas Bondo; Laestadius, Andre. S-Diagnostic—An a Posteriori Error Assessment for Single-Reference Coupled-Cluster Methods. *The Journal of Physical Chemistry A*, **2023**, 127 (43), 9106–9120. PMID: 37874274. arXiv:https://doi.org/10.1021/acs.jpca.3c01575, DOI: 10.1021/acs.jpca.3c01575.

3.10.7 Explicitly Correlated Methods: F12-MP2 and F12-CCSD(T)

A physically perhaps somewhat more satisfying alternative to basis set extrapolation is the theory of explicit correlation. In this method terms are added to the wavefunction Ansatz that contain the interelectronic coordinates explicitly (hence the name "explicit correlation"). Initially these terms were linear in the interelectronic distances ("R12-methods"). However, it has later been found that better results can be obtained by using other functions, such as an exponential, of the interelectronic distance ("F12-methods"). These methods are known to yield near basis set limit results for correlation energies in conjunction with much smaller orbital basis sets.

In applying these methods several points are important:

- Special orbital basis sets are at least advantageous. The development of such basis sets is still in its infancy. For a restricted range of elements the basis sets cc-pVnZ-F12 are available (where n = D, T, Q) and are recommended. Note, that other than their names suggest, these are a fair bit larger than regular double, triple or quadruple-zeta basis sets

- In addition to an orbital basis set, a near-complete auxiliary basis set must be specified. This is the so-called "CABS" basis. For the three basis sets mentioned above these are called cc-pVnZ-F12-CABS. If you have elements that are not covered you are on your own to supply a CABS basis set. CABS basis sets can be read into ORCA in a way analogous to RI auxiliary basis sets (replace "AUX" by "CABS" in the input). There are automatic tools for building a CABS basis from an arbitrary orbital basis, e.g. AutoCABS[387]

- if the RI approximation is used in conjunction with F12, a third basis set is required - this can be the regular auxiliary "/C" basis, but we recommend to step one level up in the auxiliary basis set (e.g. use a cc-pVTZ/C fitting basis in conjunction with cc-pVDZ-F12)

- It is perfectly feasible to use RIJCOSX or RI-JK at the same time. In this case, you should provide a fourth basis set for the Coulomb fitting

- RHF and UHF are available, ROHF not. (Although, one can do a ROHF like calculation by using QROs)

- Gradients are not available

Doing explicitly correlated MP2 calculations is straightforward. For example look at the following calculation on the water molecule at a given geometry:

```
#
! F12-MP2 cc-pVDZ-F12 cc-pVDZ-F12-CABS  VeryTightSCF PModel

* xyz 0 1
    O    0.000000000000    0.000000000000    0.369372944000
    H    0.783975899000    0.000000000000   -0.184686472000
    H   -0.783975899000    0.000000000000   -0.184686472000
*
```

and similarly in conjunction with the RI approximation:

```
#
! F12-RI-MP2 cc-pVDZ-F12 cc-pVDZ-F12-CABS cc-pVTZ/C VeryTightSCF PModel

* xyz 0 1
    O    0.000000000000    0.000000000000    0.369372944000
    H    0.783975899000    0.000000000000   -0.184686472000
    H   -0.783975899000    0.000000000000   -0.184686472000
*
```

The output is relatively easy to interpret:

```
----------------
RI-MP2-F12 ENERGY
----------------


EMP2 correlation Energy            :       -0.241038994909
F12 correction                     :       -0.054735459470
```

(continues on next page)

```
                                   -----------------
MP2 basis set limit estimate    :     -0.295774454379

Hartree-Fock energy             :    -76.057963800414
(2)_S CABS correction to EHF    :     -0.003475342535
                                   -----------------
HF basis set limit estimate     :    -76.061439142949

MP2 total energy before F12     :    -76.299002795323
Total F12 correction            :     -0.058210802005
                                   -----------------
Final basis set limit MP2 estimate :  -76.357213597328
```

It consists of several parts. The first is the regular (RI-)MP2 correlation energy in the orbitals basis followed by the additive MP2 correction which are combined to provide an MP2 correlation energy basis set limit estimate. The second part consists of an estimate in the error in the underlying SCF energy. This is the "(2)_S CABS" correction. The combination of the SCF energy with this correction yields an estimate of the SCF basis set limit. The correction will typically undershoot somewhat, but the error is very smooth. Finally, the corrected correlation energy and the corrected SCF energy are added to yield the F12 total energy estimate at the basis set limit.

Let's look at some results and compare to extrapolation:

```
#
# Correlation energies of the water molecule: extrapolation versus F12
#
# cc-pVDZ MP2: -0.201380894
#      T     : -0.261263141
#      Q     : -0.282661311
#     T/Q    : -0.298276192
#     Q/5    : -0.300598282
#  F12-DZ    : -0.295775804
#  RI-F12-DZ : -0.295933560 (cc-pVDZ/C)
#              -0.295774489 (cc-pVTZ/C)
#  F12-TZ    : -0.299164006
#  RI-F12-TZ : -0.299163478 (cc-pVQZ/C)
#  F12-QZ    : -0.300130086
```

It is obvious that extrapolated and F12 correlation energies converge to the same number (in this case around 300 mEh). The best extrapolated result is still below the F12 result (this would primarily be meaningful in a variational calculation). However, first of all this was an expensive extrapolation and second, the small residual F12 error is very smooth and cancels in energy differences. In any case, already the F12-double-zeta (where "double zeta" is to be interpreted rather loosely) brings one into within 5 mEh of the basis set limit correlation energy and the F12-triple-zeta calculation to within 1 mEh, which is impressive.

The additional effort for the F12 calculation is rather high, since five types of additional two-electron integrals need to be calculated. Both integrals in CABS space and in the original orbital (OBS) space must be calculated and mixed Fock matrices are also required. Hence, one may wonder, whether a double-zeta F12 calculation actually saves any time over, say, a quadruple-zeta regular calculation. The actual answer to this question is: "NO". Given all possibilities of obtained approximate MP2 and SCF energies, we have investigated the question of how to obtain MP2 basis set limit energies most efficiently in some detail. The results show that in terms of timings, basis set extrapolation in combination with RI-JK is the method of choice for MP2.[388] However, energy differences are more reliable with F12-MP2. In combination with RI-JK or RIJCOSX F12-MP2 becomes also competitive in terms of computational efficiency.

This situation is different in the case of coupled-cluster methods, where F12 methods outperform extrapolation and are the method of choice.

For coupled-cluster theory, everything works in a very similar fashion:

```
# the keywords
! F12-CCSD(T)
```

```
# and
! CCSD(T)-F12
# are equivalent
```

A special feature of ORCA that can save large amounts of time, is to use the RI approximation only for the F12-part. The keyword here is:

```
! F12/RI-CCSD(T)
# or
! CCSD(T)-F12/RI
```

Everything else works as described for F12-MP2.

3.10.8 Automatic extrapolation to the basis set limit

> ℹ️ **Note**
>
> - This functionality is deprecated - it may still be usable but we will not actively maintain this part of code anymore. For basis set extrapolation please use the respective compound scripts.

As eluded to in the previous section, one of the biggest problems with correlation calculations is the slow convergence to the basis set limit. One possibility to overcome this problem is the use of explicitly correlated methods. The other possibility is to use basis set extrapolation techniques. Since this involves some fairly repetitive work, some procedures were hardwired into the ORCA program. So far, only energies are supported. For extrapolation, a systematic series of basis sets is required. This is, for example, provided by the cc-pVnZ, aug-cc-pVnZ or the corresponding ANO basis sets. Here n is the "cardinal number" that is 2 for the double-zeta basis sets, 3 for triple-zeta, etc.

The convergence of the HF energy to the basis set limit is assumed to be given by:

$$E_{\text{SCF}}^{(X)} = E_{\text{SCF}}^{(\infty)} + A \exp\left(-\alpha\sqrt{X}\right) \tag{3.123}$$

Here, $E_{\text{SCF}}^{(X)}$ is the SCF energy calculated with the basis set with cardinal number X, $E_{\text{SCF}}^{(\infty)}$ is the basis set limit SCF energy and A and α are constants. The approach taken in ORCA is to do a two-point extrapolation. This means that either A or α have to be known. Here, we take A as to be determined and α as a basis set specific constant.

The correlation energy is supposed to converge as:

$$E_{\text{corr}}^{(\infty)} = \frac{X^\beta E_{\text{corr}}^{(X)} - Y^\beta E_{\text{corr}}^{(Y)}}{X^\beta - Y^\beta} \tag{3.124}$$

The theoretical value for β is 3.0. However, it was found by Truhlar and confirmed by us, that for 2/3 extrapolations $\beta = 2.4$ performs considerably better.

For a number of basis sets, we have determined the optimum values for α and β[48]:

	α_{23}	β_{23}	α_{34}	β_{34}
cc-pVnZ	4.42	2.46	5.46	3.05
pc-n	7.02	2.01	9.78	4.09
def2	10.39	2.40	7.88	2.97
ano-pVnZ	5.41	2.43	4.48	2.97
saug-ano-pVnZ	5.48	2.21	4.18	2.83
aug-ano-pVnZ	5.12	2.41		

Since the β values for 2/3 are close to 2.4, we always take this value. Likewise, all 3/4 and higher extrapolations are done with $\beta = 3$. However, the optimized values for α are taken throughout.

Using the keyword `! Extrapolate(X/Y,basis)`, where `X` and `Y` are the corresponding successive cardinal numbers and `basis` is the type of basis set requested (= `cc`, `aug-cc`, `cc-core`, `ano`, `saug-ano`, `aug-ano`, `def2`) ORCA will calculate the SCF and optionally the MP2 or MDCI energies with two basis sets and separately extrapolate.

The keyword works also in the following way: `! Extrapolate(n,basis)` where `n` is the is the number of energies to be used. In this way the program will start from a double-zeta basis and perform calculations with `n` cardinal numbers and then extrapolate the different pairs of basis sets. Thus for example the keyword `! Extrapolate(3, CC)` will perform calculations with cc-pVDZ, cc-pVTZ and cc-pVQZ basis sets and then estimate the extrapolation results of both cc-pVDZ/cc-pVTZ and cc-pVTZ/cc-pVQZ combinations.

Let us take the example of the H2O molecule at the B3LYP/TZVP optimized geometry. The reference values have been determined from a HF calculation with the decontracted aug-cc-pV6Z basis set and the correlation energy was obtained from the cc-pV5Z/cc-pV6Z extrapolation. This gives:

```
E(SCF,CBS)          = -76.066958 Eh
EC(CCSD(T),CBS)     =  -0.30866 Eh
Etot(CCSD(T),CBS)   = -76.37561 Eh
```

Now we can see what extrapolation can bring in:

```
!CCSD(T) Extrapolate(2/3) TightSCF Conv Bohrs
* int 0 1
O 0 0 0  0 0 0
H 1 0 0  1.81975 0 0
H 1 2 0  1.81975 105.237 0
*
```

NOTE:

- The RI-JK and RIJCOSX approximations work well together with this option and RI-MP2 is also possible. Auxiliary basis sets are automatically chosen and can not be changed.

- All other basis set choices, externally defined bases etc. will be ignored — the automatic procedure only works with the default basis sets!

- The basis sets with the "core" postfix contain core correlation functions. By default it is assumed that this means that the core electrons are also to be correlated and the frozen core approximation is turned off. However, this can be overridden in the method block by choosing, e.g. `%method frozencore fc_electrons end!`

- So far, the extrapolation is only implemented for single points and not for gradients. Hence, geometry optimizations cannot be done in this way.

- The extrapolation method should only be used with very tight SCF convergence criteria. For open shell methods, additional caution is advised.

This gives:

```
Alpha(2/3)     :  4.420 (SCF Extrapolation)
Beta(2/3)      :  2.460 (correlation extrapolation)

SCF energy with basis cc-pVDZ:                      -76.026430944
SCF energy with basis cc-pVTZ:                      -76.056728252
Extrapolated CBS SCF energy (2/3) :                 -76.066581429 (-0.
↪009853177)

MDCI energy with basis cc-pVDZ:                      -0.214591061
MDCI energy with basis cc-pVTZ:                      -0.275383015
Extrapolated CBS correlation energy (2/3) :          -0.310905962 (-0.
↪035522947)
```

(continues on next page)

```
Estimated CBS total energy (2/3) :                      -76.377487391
```

Thus, the error in the total energy is indeed strongly reduced. Let us look at the more rigorous 3/4 extrapolation:

```
Alpha(3/4)     :  5.460 (SCF Extrapolation)
Beta(3/4)      :  3.050 (correlation extrapolation)

SCF energy with basis cc-pVTZ:                          -76.056728252
SCF energy with basis cc-pVQZ:                          -76.064381269
Extrapolated CBS SCF energy (3/4) :                     -76.066687152 (-0.
↪002305884)

MDCI energy with basis cc-pVTZ:                          -0.275383016
MDCI energy with basis cc-pVQZ:                          -0.295324345
Extrapolated CBS correlation energy (3/4) :             -0.309520368 (-0.
↪014196023)

Estimated CBS total energy (3/4) :                      -76.376207520
```

In our experience, the ANO basis sets extrapolate similarly to the cc-basis sets. Hence, repeating the entire calculation with `Extrapolate(3,ANO)` gives:

```
Estimated CBS total energy (2/3) :                      -76.377652792
Estimated CBS total energy (3/4) :                      -76.376983433
```

Which is within 1 mEh of the estimated CCSD(T) basis set limit energy in the case of the 3/4 extrapolation and within 2 mEh for the 2/3 extrapolation.

For larger molecules, the bottleneck of the calculation will be the CCSD(T) calculation with the larger basis set. In order to avoid this expensive (or prohibitive) calculation, it is possible to estimate the CCSD(T) energy at the basis set limit as:

$$E_{\text{corr}}^{(\text{CCSD(T)};Y)} \approx E_{\text{corr}}^{(\text{CCSD(T)};X)} + E_{\text{corr}}^{(\text{MP2};\infty)} - E_{\text{corr}}^{(\text{MP2};X)} \tag{3.125}$$

This assumes that the basis set dependence of MP2 and CCSD(T) is similar. One can then extrapolate as before. Alternatively, the standard way — as extensively exercised by Hobza and co-workers — is to simply use:

$$E_{\text{total}}^{(\text{CCSD(T)};\text{CBS})} \approx E_{\text{SCF}}^{(Y)} + E_{\text{corr}}^{(\text{CCSD(T)};X)} + E_{\text{corr}}^{(\text{MP2};\infty)} - E_{\text{corr}}^{(\text{MP2};X)} \tag{3.126}$$

The appropriate keyword is:

```
! ExtrapolateEP2(2/3,ANO,MP2) TightSCF Conv Bohrs
* int 0 1
O 0 0 0  0 0 0
H 1 0 0  1.81975 0 0
H 1 2 0  1.81975 105.237 0
*
```

This creates the following output:

```
Alpha     :  5.410 (SCF Extrapolation)
Beta      :  2.430 (correlation extrapolation)

SCF energy with basis ano-pVDZ:                         -76.059178452
SCF energy with basis ano-pVTZ:                         -76.064774379
Extrapolated CBS SCF energy :                           -76.065995735 (-0.
↪001221356)

MP2 energy with basis ano-pVDZ:                          -0.219202871
MP2 energy with basis ano-pVTZ:                          -0.267058634
```

```
Extrapolated CBS correlation energy :                    -0.295568604 (-0.
↪028509970)

CCSD(T) correlation energy with basis ano-pVDZ:          -0.229478341
CCSD(T) - MP2 energy with basis ano-pVDZ:                -0.010275470

Estimated CBS total energy :                             -76.371839809
```

The estimated correlation energy is not really bad — within 3 mEh from the basis set limit.

Using the `ExtrapolateEP2(n/m,bas,[method, method-details])` keyword one can use a generalization of the above method where instead of MP2 any available correlation method can be used as described in Ref. [389]. `method` is optional and can be either MP2 or DLPNO-CCSD(T), the latter being the default. In case the method is DLPNO-CCSD(T) in the `method-details` option one can ask for LoosePNO, NormalPNO or TightPNO.

$$E_{\text{corr}}^{(\text{CCSD(T)};CBS)} \approx E_{\text{corr}}^{(\text{CCSD(T)};X)} + E_{\text{corr}}^{(\text{M};CBS)}(X, X+1) - E_{\text{corr}}^{(\text{M};X)} \tag{3.127}$$

Here M represents any correlation method one would like to use. For the previous water molecule the input of a calculation that uses DLPNO-CCSD(T) (which is the default now) instead of MP2 would look like:

```
! ExtrapolateEP2(2/3,cc,DLPNO-CCSD(T)) TightSCF Conv Bohrs
* int 0 1
O 0 0 0  0 0 0
H 1 0 0  1.81975 0 0
H 1 2 0  1.81975 105.237 0
*
```

and it would produce the following output:

```
Alpha     :   4.420 (SCF Extrapolation)
Beta      :   2.460 (correlation extrapolation)

SCF energy with basis cc-pVDZ:                           -76.026430944
SCF energy with basis cc-pVTZ:                           -76.056728252
Extrapolated CBS SCF energy :                            -76.066581429 (-0.
↪009853177)

MDCI energy with basis cc-pVDZ:                          -0.214429497
MDCI energy with basis cc-pVTZ:                          -0.275299699
Extrapolated CBS correlation energy :                    -0.310868368 (-0.
↪035568670)

CCSD(T) correlation energy with basis cc-pVDZ:           -0.214548320
CCSD(T) - MDCI energy with basis cc-pVDZ:                -0.000118824

Estimated CBS total energy :                             -76.377568621
```

which is less than 2 mEh from the basis set limit. Finally it was shown [389] that instead of extrapolating the cheap method, M, using cardinal numbers X and $X+1$ it is better to use cardinal numbers $X+1$ and $X+2$.

$$E_{\text{corr}}^{(\text{CCSD(T)};CBS)} \approx E_{\text{corr}}^{(\text{CCSD(T)};X)} + E_{\text{corr}}^{(\text{M};CBS)}(X+1, X+2) - E_{\text{corr}}^{(\text{M};X)} \tag{3.128}$$

This can be done using the `ExtrapolateEP3(bas,[method,method-details])` keyword:

```
! ExtrapolateEP3(CC) TightSCF Conv Bohrs
```

and the corresponding output would be:

```
Alpha     :   5.460 (SCF Extrapolation)
Beta      :   3.050 (correlation extrapolation)
```

3.10. Coupled Cluster and CI Theories (MDCI)

```
SCF energy with basis cc-pVDZ:                          -76.026430944
SCF energy with basis cc-pVTZ:                          -76.056728252
SCF energy with basis cc-pVQZ:                          -76.064381269
Extrapolated CBS SCF energy :                           -76.066687152 (-0.
↪002305884)

MDCI energy with basis cc-pVDZ:                          -0.214429497
MDCI energy with basis cc-pVTZ:                          -0.275299699
MDCI energy with basis cc-pVQZ:                          -0.295229871
Extrapolated CBS correlation energy :                   -0.309417951 (-0.
↪014188080)

CCSD(T) correlation energy with basis cc-pVDZ:          -0.214548319
CCSD(T) - MDCI energy with basis cc-pVDZ:               -0.000118822

Estimated CBS total energy :                            -76.376223926
```

For the ExtrapolateEP2, and ExtrapolateEP3 keywords the default cheap method is the DLPNO-CCSD(T) with the NormalPNO thresholds. There also available options with MP2, and DLPNO-CCSD(T) with LoosePNO and TightPNO settings.

3.10.9 Cluster in molecules (CIM)

Cluster in molecules (CIM) approach is a linear scaling local correlation method developed by Li and the coworkers in 2002.[390] It was further improved by Li, Piecuch, Kállay and other groups recently.[391, 392, 393, 394, 395] The CIM is inspired by the early local correlation method developed by Förner and coworkers.[396] The total correlation energy of a closed-shell molecule can be considered as a summation of correlation energies of each occupied LMOs.

$$E_{\text{corr}} = \sum_i^{occ} E_i = \sum_i^{occ} \frac{1}{4} \sum_{j,ab} \langle ij||ab\rangle T_{ab}^{ij} \tag{3.129}$$

For each occupied LMO, it only correlates with its nearby occupied LMOs and virtual MOs. To reproduce the correlation energy of each occupied LMO, only a subset of occupied and virtual LMOs are needed in the correlation calculation. Instead of doing the correlation calculation of the whole molecule, the correlation energies of all LMOs can be obtained within various subsystems.

The CIM approach implemented in ORCA is following an algorithm proposed by Guo and coworkers with a few improvements.[394, 395]

1. To avoid the real space cutoff, the differential overlap integral (DOI) is used instead of distance threshold. There is only one parameter 'CIMTHRESH' in CIM approach, controlling the construction of CIM subsystems. If the DOI between LMO *i* and LMO *j* is larger than CIMTHRESH, LMO *j* will be included into the MO domain of *i*. By including all nearby LMO of *i*, one can construct a subsystem for MO *i*. The default value of CIMTHRESH is 0.001. If accurate results are needed, a tighter CIMTHRESH must be used.

2. Since ORCA 4.1, the neglected correlations between LMO *i* and LMOs outside the MO domain of *i* are considered as well. These weak correlations are approximately evaluated by dipole moment integrals. With this correction, the CIM results of 3 dimensional proteins are significantly improved. About 99.8% of the correlation energies are recovered.

The CIM can invoke different single reference correlation methods for the subsystem calculations. In ORCA the CIM-RI-MP2, CIM-CCSD(T), CIM-DLPNO-MP2 and CIM-DLPNO-CCSD(T) methods are available. The CIM-RI-MP2 and CIM-DLPNO-CCSD(T) have been proved to be very efficient and accurate methods to compute correlation energies of very big molecules, containing a few thousand atoms.[395]

The usage of CIM in ORCA is simple. For CIM-RI-MP2,

```
#
# CIM-RI-MP2 calculation
```

```
#
! RI-MP2 cc-pVDZ cc-pVDZ/C CIM
%CIM
  CIMTHRESH 0.0005  # Default value is 0.001
end

* xyzfile 0 1 CIM.xyz
```

For CIM-DLPNO-CCSD(T),

```
#
# CIM-DLPNO-CCSD calculation
#
! DLPNO-CCSD(T) cc-pVDZ  cc-pVDZ/C CIM
* xyzfile 0 1 CIM.xyz
```

The parallel efficiency of CIM has been significantly improved.[395] Except for a few domain construction sub-steps, the CIM algorithm can achieve very high parallel efficiency. Since ORCA 4.1, the parallel version does not support Windows platform anymore due to the parallelization strategy. The generalization of CIM from closed-shell to open-shell (multi-reference) will also be implemented in the near future.

3.10.10 Local correlation (DLPNO)

ORCA features the extremely powerful "domain-based local pair natural orbital" approximation (DLPNO),[397] which us an extension of the "local pair natural orbital" ansatz (LPNO).[383, 384, 385, 398] These methods are designed to provide results as close as possible to the canonical coupled-cluster results while gaining orders of magnitude of efficiency through a series of well-controlled approximations. In fact, typically about 99.8% to 99.9% of the canonical correlation energy is recovered in such calculations. Even higher accuracy is achievable but will ultimately come at much higher computational cost. The default cut-offs are set such that the vast majority of chemically meaningful energy differences are reproduced to better than 1 kcal/mol relative to the canonical results with the same basis set.

With ORCA 6.0, the LPNO variants are no longer supported, as they involve some higher order scaling steps (up to N^5) while DLPNO is linear scaling and of similar accuracy. The comparison between LPNO-CCSD and DLPNO-CCSD is shown in Fig. 3.8. It is obvious that DLPNO-CCSD is (almost) never slower than LPNO-CCSD. However, its true advantages do become most apparent for molecules with more than approximately 60 atoms. The triples correction, that was added with our second paper from 2013, shows a perfect linear scaling, as is shown in part (a) of Fig. 3.8. For large systems it adds about 10%–20% to the DLPNO-CCSD computation time, hence its addition is possible for all systems for which the latter can still be obtained. Since 2016, the entire DLPNO-CCSD(T) algorithm is linear scaling. The improvements of the linear-scaling algorithm, compared to DLPNO2013-CCSD(T), start to become significant at system sizes of about 300 atoms, as becomes evident in part (b) of Fig. 3.8.

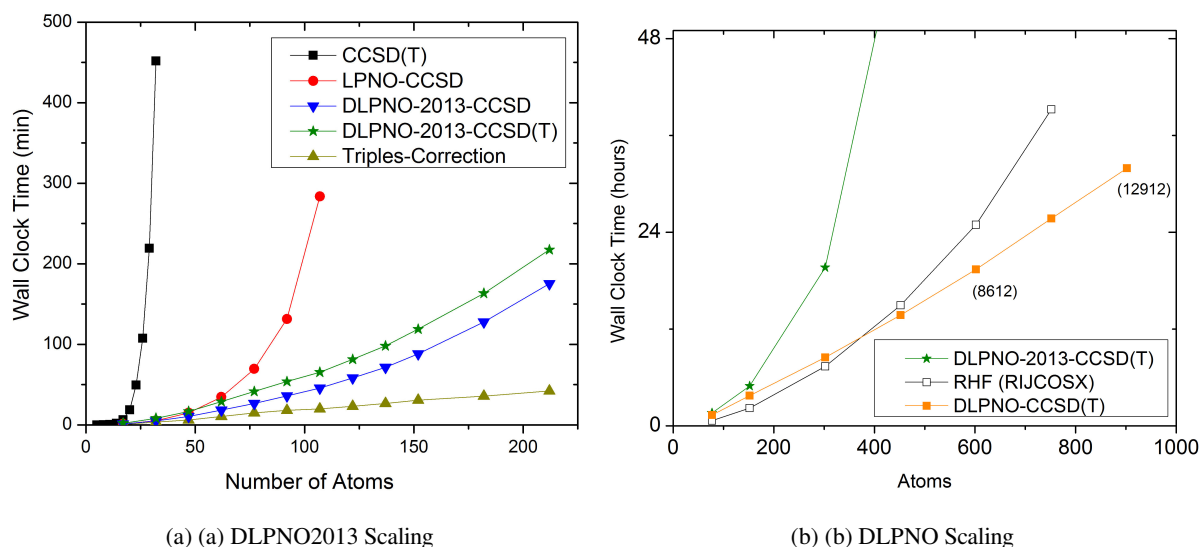(a) (a) DLPNO2013 Scaling

(b) (b) DLPNO Scaling

Fig. 3.8: a) Scaling behavior of the canonical CCSD, LPNO-CCSD and DLPNO2013-CCSD(T) methods. It is obvious that only DLPNO2013-CCSD and DLPNO2013-CCSD(T) can be applied to large molecules. The advantages of DLPNO2013-CCSD over LPNO-CCSD do not show before the system has reached a size of about 60 atoms. b) Scaling behavior of DLPNO2013-CCSD(T), DLPNO-CCSD(T) and RHF using RIJCOSX. It is obvious that only DLPNO-CCSD(T) can be applied to truly large molecules, is faster than the DLPNO2013 version, and even has a crossover with RHF at about 400 atoms.

Using the DLPNO-CCSD(T) approach it was possible for the first time (in 2013) to perform a CCSD(T) level calculation on an entire protein (Crambin with more than 650 atoms, Fig. 3.9). While the calculation using a double-zeta basis took about 4 weeks on one CPU with DLPNO2013-CCSD(T), it takes only about 4 days to complete with DLPNO-CCSD(T). With DLPNO-CCSD(T) even the triple-zeta basis calculation can be completed within reasonable time, taking 2 weeks on 4 CPUs.
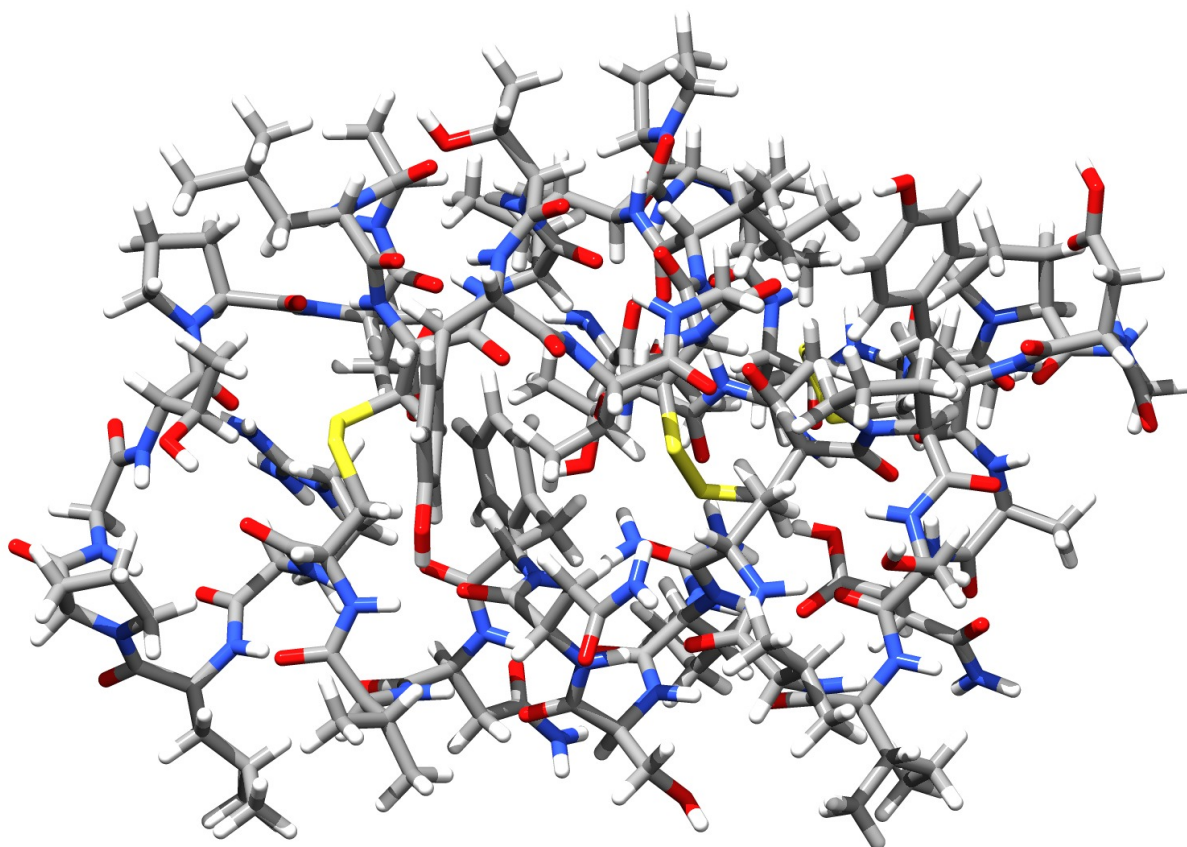
Fig. 3.9: Structure of the Crambin protein - the first protein to be treated with a CCSD(T) level ab initio method

It is important to understand that the LPNO and DLPNO implementations are intimately tied to the RI approximation. Hence, in these calculations one *must* specify a fitting basis set. The same rules as for RI-MP2 apply: the auxiliary basis sets optimized for MP2 are just fine for PNO calculations. You can specify several aux bases for the same job and the program will sort it out correctly.

The theory of the LPNO and DLPNO methods has been thoroughly described in the literature in a number of original research papers:

- F. Neese, A. Hansen, D. G. Liakos: Efficient and accurate local approximations to the coupled-cluster singles and doubles method using a truncated pair natural orbital basis.[383]

- F. Neese, A. Hansen, F. Wennmohs, S. Grimme: Accurate Theoretical Chemistry with Coupled Electron Pair Models.[385]

- F. Neese, F. Wennmohs, A. Hansen: Efficient and accurate local approximations to coupled electron pair approaches. An attempt to revive the pair-natural orbital method.[398]

- D. G. Liakos, A. Hansen, F. Neese: Weak molecular interactions studied with parallel implementations of the local pair natural orbital coupled pair and coupled-cluster methods.[399]

- A. Hansen, D. G. Liakos, F. Neese: Efficient and accurate local single reference correlation methods for high-spin open-shell molecules using pair natural orbitals.[384]

- C. Riplinger, F. Neese: An efficient and near linear scaling pair natural orbital based local coupled-cluster method.[397]

- C. Riplinger, B. Sandhoefer, A. Hansen, F. Neese: Natural triple excitations in local coupled-cluster calculations with pair natural orbitals.[400]

- C. Riplinger, P. Pinski, U. Becker, E. F. Valeev, F. Neese: Sparse maps - A systematic infrastructure for reduced-scaling electronic structure methods. II. Linear scaling domain based pair natural orbital coupled cluster theory.[401]

- D. Datta, S. Kossmann, F. Neese: Analytic energy derivatives for the calculation of the first-order molecular properties using the domain-based local pair-natural orbital coupled-cluster theory[402]

- M. Saitow, U. Becker, C. Riplinger, E. F. Valeev, F. Neese: A new linear scaling, efficient and accurate, open-shell domain based pair natural orbital coupled cluster singles and doubles theory.[403]

- Y. Guo, C. Riplinger, U. Becker, D.G. Liakos, Y. Minenkov, L. Cavallo, and F. Neese: An improved linear scaling perturbative triples correction for the domain based local pair-natural orbital based singles and doubles coupled cluster method (DLPNO-CCSD(T)).[404]

Hence, it is sufficient to only point to a few significant design principles and features of these methods:

1. The correlation energy of any molecule can be written as a sum over the correlation energies of pairs of electrons, labelled by the internal indices (ij) of pairs of orbitals that are occupied in the reference determinant. If the orbitals (i) and (j) are localized, the pair correlation energy ϵ_{ij} falls off very quickly with distance, quite typically by about an order of magnitude per chemical bond that is separating orbitals (i) and (j). Hence, by using a suitable cut-off for a reasonable pair correlation energy estimate many electron pairs can be removed from the high-level treatment and only a linear scaling number of electron pairs will make a significant contribution to the correlation energy.

2. The natural estimate for the pair correlation energy comes from second order many body perturbation theory (MP2). However, canonical MP2 is scaling with the fifth power of the molecular size and hence, is not really attractive from a theoretical nor computational point of view. Owing to the small pre-factor RI-MP2 goes a long way to provide reasonably cheap estimates for pair correlation energies. However, if one uses localized internal orbitals, then the MP2 energy expression must be cast in form of linear equations. On the other hand, if one uses canonical virtual orbitals together with localized internal orbitals and neglects the coupling terms coming from purely internal Fock matrix elements F(i,k) and F(k,j) then one ends up with a fair approximation that is termed "semi-canonical MP2" in ORCA. It serves as a guess in the older LPNO method. For DLPNO this method is also not attractive.

3. In DLPNO, the guess is more sophisticated. Here the virtual space is spanned by projected atomic orbitals (PAOs) that are assigned to domains of atoms that are associated with each local internal orbital (i) and with the union of two such domains (i) and (j) for the electron pair (ij). If one applies the semi-local approximation, one obtains an excellent approximation to the semi-canonical MP2 energy. This is called the "semi-local" approximation and it scales linearly with respect to computational effort. If one iterates these equations to self-consistency to eliminate the coupling terms F(i,k) and F(k,j) then one obtains the full local MP2 method (LMP2 or L-MP2). By making the domains large enough the results approach the canonical MP2 energy to arbitrary accuracy while still being linear scaling with respect to computational resources. This method is the default for the DLPNO method.

4. Basically, the high-spin open-shell version of the DLPNO uses the same strategy as the closed-shell variant to efficiently generate the open-shell PNOs in a consistent manner to the closed-shell formalism. Through the development of the UHF-LPNO-CCSD method, we have realized that use of the unrestricted MP2 (UMP2) approach to define the open-shell PNOs introduces a few complexities; (1) the PNOs for β spin orbitals cannot be defined for α-α electron pairs and vice versa, (2) the diagonal PNOs for singly occupied orbitals cannot be properly defined, and (3) the PNO space does not become identical to that in the closed-shell LPNO framework in the closed-shell limit. However, to program all the unrestricted CCSD terms in the LPNO basis, those PNOs are certainly necessary. Therefore, in the UHF-LPNO-CCSD implementation, several terms, which, in many cases, give rather minor contributions in the correlation energy are omitted. Due to these facts, the UHF-LPNO-CCSD does not give identical results to that of RHF-LPNO-CCSD in the closed-shell limit. In addition, screening of the weak pairs on the basis of the semi-canonical UMP2 pair-energy results in somewhat unbalanced treatment of the closed- and open-shell states in some cases, leading to rather larger errors in the reaction energies. To overcome those issues, in the high-spin open-shell DLPNO-CCSD method, the PNOs are generated in the framework of semi-canonical NEVPT2 which smoothly converges to the RHF-MP2 counterpart in the closed-shell limit. The open-shell DLPNO-CCSD, which is made available from ORCA 4.0, includes a full set of the open-shell CCSD equations and is designed as a natural extension to the RHF-DLPNO-CCSD.

5. Screening of the electron pairs according to a truncation parameter (in ORCA it is called T_{CutPairs}) is not sufficient to obtain a highly performing local electronic structure method. The original work of Pulay suggested to limit excitations out of the internal orbitals (i) and (j) to the domain associated with the pair (ij). While this works well and has been implemented to perfection by Werner, Schütz and co-workers over the years,[405,

406, 407, 408] the pre-factor of such calculations is high, since the domains have to be chosen large in order to recover 99.9% or more of the canonical correlation energy.

6. The ORCA developers have therefore turned to an approach that has been used with a high degree of success in the early 1970s by Meyer, Kutzelnigg, Staemmler and their co-workers, namely the method of "pair natural orbitals" (PNOs).[409, 410, 411, 412]

   As shown by Loewdin in his seminal paper from 1955, natural orbitals (the eigenfunctions of the one-particle density matrix) provide the fastest possible convergence of the correlated wavefunction with respect to the number of one-particle functions included in the virtual space. It has been amply established that approximate natural orbitals are almost as succesful as the true natural orbitals (which would require the knowledge of the exact wavefunction) in this respect. While the success of approximate correlation treatments of many particle systems that use approximate natural orbitals of the whole systems are somewhat limited, this is not the case for pair natural orbitals. The latter have first been suggested as a basis for correlation calculations by England and co-workers and, at the time, were given the name "pseudonatural orbitals", a term that was used by Meyer throughout his pioneering work.

7. The PNOs are approximate natural orbitals of a given electron pair. In order to generate them one requires a one particle pair density matrix D_{ij} the eigenfunctions of which are the PNOs themselves while the corresponding eigenvalues are the PNO occupation numbers. While there are many creative possibilities that can lead to slightly different PNO sets, a quite useful and natural approximation is to generate such a density from the MP2 amplitudes as an expectation value (the "unrelaxed" MP2 density. One then uses a second threshold (in ORCAT_{CutPNO}) that controls the PNOs to be included in the calculation. PNOs with an occupation number $< T_{\text{CutPNO}}$ are neglected.

8. PNOs of a given electron pair form an orthonormal set. However, PNOs belonging to different electron pairs are not orthonormal and hence they overlap. This non-orthogonality leads to surprisingly few complications because the PNOs stay orthogonal to all occupied orbitals. In practice, the equations for PNO-based correlation calculations are hardly more complex than the canonical equations.

9. The nice feature of these pair densities is that they become small when the pair interaction becomes small. Hence weak pairs are correlated by very few PNOs. Therefore, the PNO expansion of the wavefunction is extremely compact and there only is a linear scaling number of significant excitation amplitudes that need to be considered.

10. A great feature of the PNOs is that they are "self-adapting" to the correlation situation that they are supposed to describe. Hence, they are as delocalized as required by the physics and there is no ad-hoc assumption about their location in space. However, it is clear that the PNOs are located in the same region of space as the internal orbitals that they correlate because otherwise they would not contribute to the correlation energy.

11. In the iterative local MP2, a set of PNOs is calculated for the MP2 calculation from the semicanonical amplitudes first using the cutoff `TCutPNO` $\times$`LMP2ScaleTCutPNO`. The size of the resulting PNO space should be comparable with DLPNO-MP2. After the iterations have converged, a (smaller) set of PNOs for the CCSD double excitation amplitudes is generated from the iterated local MP2 amplitudes in the (larger) PNO basis. The PNOs for the single excitation amplitudes are always calculated using the semicanonical MP2 amplitudes, as they require a much more conservative PNO truncation threshold.

12. Capitalizing on this feature one can define generous domains and expand the PNOs in terms of the PAOs and auxiliary fit functions (for the RI approximations) that are contained in these domains. In ORCA this is controlled by the third significant truncation parameter T_{CutMKN}. This is the basis of the DLPNO method. In the older LPNO method, the PNOs are expanded in terms of the canonical virtual orbitals and T_{CutMKN} is only used for a local fit to the PNOs. In the linear-scaling DLPNO implementation the domains expanding the PNOs in terms of the PAOs are controlled via T_{CutDO}.

13. PNO expansions have the amazing advantage that the PNOs converge to a well-defined set as the basis set is approaching completeness. Hence, the increase in the number of PNOs per electron pair is very small upon enlargement of the orbital basis. In other words, correlation calculations with large basis sets are not that much more expensive than correlation calculations with small basis sets. Thus, the advantage of PNO over canonical calculations increases with the size of the basis set. This is a great feature as large and flexible basis sets are a requirement for meaningful correlation calculations.

14. In summary, DLPNO and LPNO calculations are controlled by only three cut-off parameters with well-defined meanings: a) T_{CutPairs}, the cut-off for the electron pairs to be included in the coupled-pair or coupled-cluster iterations, b) T_{CutPNO} which controls how many PNOs are retained for a given electron pair and c) T_{CutMKN} that

---

3.10. Coupled Cluster and CI Theories (MDCI)   **321**

controls how large the domains are that the PNOs expand over. For the linear-scaling DLPNO calculations the domain sizes are controlled via T_{CutDO}.

15. It is clear that owing to the truncations a certain amount of error is introduced in the results. However, having the LMP2 results available, one can compensate for the errors coming from T_{CutPairs} and T_{CutPNO}. This is done in ORCA and the correction is added to the final correlation energy, thus bringing it very close (to mEh accuracy or better) to the canonical result. T_{CutMKN} is best dealt with by making it conservative (at $1e-3$ to $1e-4$ the domains are about 20–30 atoms large, which is sufficient for an accurate treatment).

16. Note that the LPNO and DLPNO methods do not introduce any real space cut-offs. We refrain from doing so and insist in our method development by making all truncations based on wavefunction or energy parameters. We feel that this is the most unbiased approach and it involves no element of "chemical intuition" or "prejudice".

17. In the DLPNO method a highly efficient screening mechanism is operative. In this method one first obtains a (quadratically scaling) multipole estimate for the pair correlation energy that is extremely fast to compute (a few seconds, even for entire proteins). Only if this estimate is large enough, a given electron pair is even considered for a LMP2 treatment. Quite typically pairs with energy contributions of $1e-6$ Eh and smaller are very well described by the dipole-dipole estimate. Specifically, we drop pairs with estimated energies $< 0.01 \times T_{\text{CutPairs}}$ and add their multipole energy sum to the final correlation energy. These corrections tend to be extremely small, even for large molecules and are insignificant for the energy. However, importantly, the multipole estimate ensures linear scaling in the MP2 treatment. The pairs that then do not survive the pair-prescreening are called "weak pairs" in the ORCA or DLPNO sense. They still play a role in the calculation of the triple excitation correction.

18. The calculation of triple excitation contributions is more involved and one does not have a perturbative estimate available since the (T) contribution is perturbative itself. While the (T) contribution is much smaller than the CCSD correlation energy, the errors introduced by the various local and PNO approximations can be significant. We found that one has to include triples with at least one pair being a "weak" LMP2 pair (with its LMP2 amplitudes) in order to arrive at sufficiently accurate results.

Basic Usage

The use of the DLPNO methods is simple and requires little special attention from the user:

```
# Local Pair Natural Orbital Test
! cc-pVTZ cc-pVTZ/C DLPNO-CCSD TightSCF
%maxcore 2000

# these are the default values - they need not to be touched!
%mdci  TCutPNO   3.33e-7   # cutoff for PNO occupation numbers. This
                                is the main truncation parameter
       TCutPairs 1e-4  # cut-off for estimated pair correlation energies.
                           This exploits the locality in the internal space
       TCutMKN   1e-3  # this is a technical parameter here that controls the
↪domain
                           size for the local fit to the PNOs. It is conservative.
       end

* xyz 0 1
... (coordinates)
*
```

The LPNO methods used to run in strict analogy to their canonical variants. In fact, one should not view the LPNO methods as new model chemistry - they are designed to reproduce the canonical results, including BSSE. This is different from the domain based local correlation methods that do constitute a new model chemistry with properties that are different from the original methods.

In some situations, it may be appropriate to adapt the accuracy of the calculation. Sensible defaults have been determined from extensive benchmark calculations and are accessible via LoosePNO, NormalPNO and TightPNO keywords in the simple input line.[413]