European Summer School of Quantum Chemistry

2022 Torre Normanna Sicily

Lecture 3: Local Correlation Approaches

Frank Neese

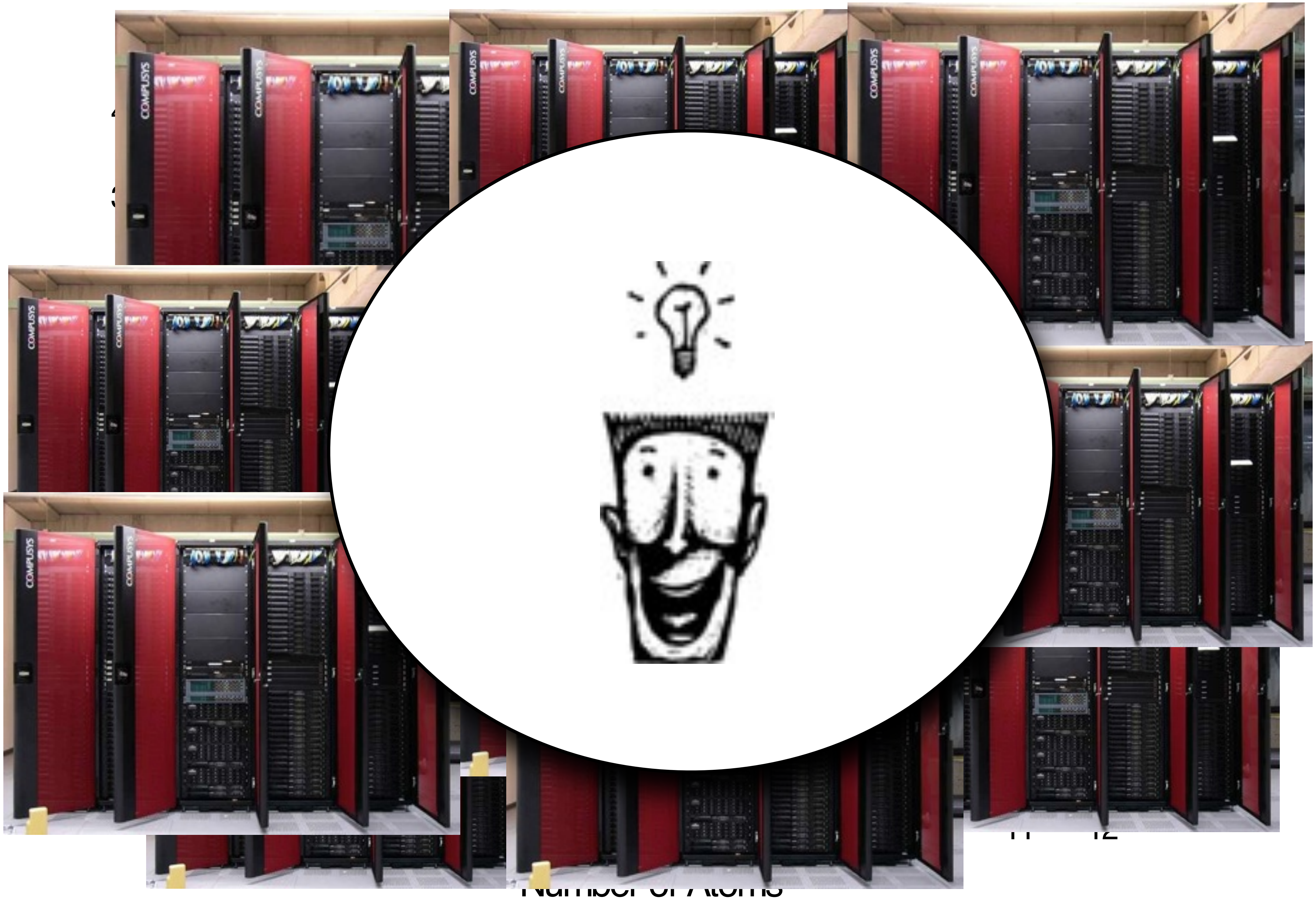
MAX-PLANCK-GESELLSCHAFT

MPI für Kohlenforschung

Kaiser-Wilhelm Platz 1

45470 Mülheim an der Ruhr

Problem with Wavefunction Methods



Early Local Correlation Approaches

The idea of local correlation is almost as old as correlation theory itself and has been suggested long before even Hartree-Fock calculation on medium sized molecules were feasible

O Sinanoglu *Adv. Chem. Phys.*, **1964**, 6,315
RK Nesbet, *Adv. Chem. Phys.*, **1965**, 9, 321

it took almost 20 years before it was taken up again. An important paper is from Cullen and Zerner (received no attention, perhaps because it was in a semi-empirical context)

JM Cullen, MC Zerner *J. Chem. Phys.*, **1982**, 77, 4088

Followed by the pioneering work of Pulay and Saebo (CISD, MP4)

P Pulay, *Chem. Phys. Lett.* **1983,** 100, 151.; S Saebo, P Pulay, *Chem. Phys. Lett.* **1985,** 113 13.
P Pulay, S Saebø, *Theor. Chim. Acta* **1986** 69, 357.; S Saebo, P Pulay, *J. Chem. Phys.* **1987,** 87 914.

And the early coupled cluster work (mostly CCD)

RJ Bartlett, GD Purvis, *Int. J. Quantum Chem*. 14, 561 **1978** WD Laidig, GD Purvis III, RJ Bartlett, *Int. J. Quantum Chem., Symp.* 16, 561 **1982**. WD Laidig, GD Purvis III RJ Bartlett, *Chem. Phys. Lett.* 97, 209 **1983**; WD Laidig, GD Purvis III RJ Bartlett, *J. Phys. Chem.* 89, 2161 **1985**; W Förner, J Ladik, P Otto, J Čížek, *Chem. Phys.* 97, 251 **1985** W Förner, *Chem. Phys.* 114, 21 **1987** M Takahashi J Paldus, *Phys. Rev. B* 31, 5121 **1985**

… given the hard- and software limitation at the time real applications were not feasible

Local Correlation: Importance of Accuracy Goals

The idea of local correlation methods is:

- ✓ preserve - *as much as possible* - the accuracy of wave function based correlation approaches.
- ✓ Reduce the unfavorable scaling with system size - *ideally to linear*

HOWEVER

- ➡ **One will only get wave function based ab initio quality, if the error that we introduce by exploiting the locality is not spoiling the intrinsic accuracy of the method!**

- ➡ Example: In a large molecule the correlation energy is **~10 Eh=6270** kcal/mol

- ➡ Chemical accuracy is ~1 kcal/mol

- ➡ **The target accuracy MUST be 99.9 to 99.99% of E_c to preserve the features of the methods**
- ➡ **Error cancellation in local approximations is NOT better than about 1 order of magnitude**

Principles of Local Correlation Theory

Pretty much all local correlation methods:

Approximation 1

$$E_{corr} = \sum_{Chunks\ K=1}^{N_K} \delta E_K \approx \sum_{Chunks\ K'=1}^{N'_K < N_K} \delta E_K^{(approx)}$$

Approximation 2

where „chunks“ =

Fragments, **Atoms**, **Atom Pairs**, **Orbitals**, **Orbital Pairs**, ..

Decomposition of the Exact Correlation Energy

Start from the Schrödinger equation $\hat{H}_{BO}\Psi = E\Psi$

Insert the full CI expansion

$$\hat{H}_{BO}\left(\Phi_{HF} + \sum_{ia} C_a^i \Phi_i^a + \left(\tfrac{1}{2!}\right)^2 \sum_{ijab} C_{ab}^{ij} \Phi_{ij}^{ab} + \ldots\right) = E\left(C_0 \Phi_{HF} + \sum_{ia} C_a^i \Phi_i^a + \left(\tfrac{1}{2!}\right)^2 \sum_{ijab} C_{ab}^{ij} \Phi_{ij}^{ab} + \ldots\right)$$

Multiply with the HF function from the left:

$$\underbrace{\left\langle \Phi_{HF} \middle| \hat{H}_{BO} \middle| \Phi_{HF} \right\rangle}_{E_{HF}} + \sum_{ia} C_i^a \underbrace{\left\langle \Phi_{HF} \middle| \hat{H}_{BO} \middle| \Phi_i^a \right\rangle}_{F_{ia}} \quad \text{=0 (Brillouin)} \quad + \tfrac{1}{4} \sum_{ijab} C_{ij}^{ab} \underbrace{\left\langle \Phi_{HF} \middle| \hat{H}_{BO} \middle| \Phi_{ij}^{ab} \right\rangle}_{\langle ij \| ab \rangle}$$

$$= E\left(\underbrace{\left\langle \Phi_{HF} \middle| \Phi_{HF} \right\rangle}_{1} + \sum_{ia} C_i^a \underbrace{\left\langle \Phi_{HF} \middle| \Phi_i^a \right\rangle}_{0} + \tfrac{1}{4} \sum_{ijab} C_{ij}^{ab} \underbrace{\left\langle \Phi_{HF} \middle| \Phi_{ij}^{ab} \right\rangle}_{0} \right)$$

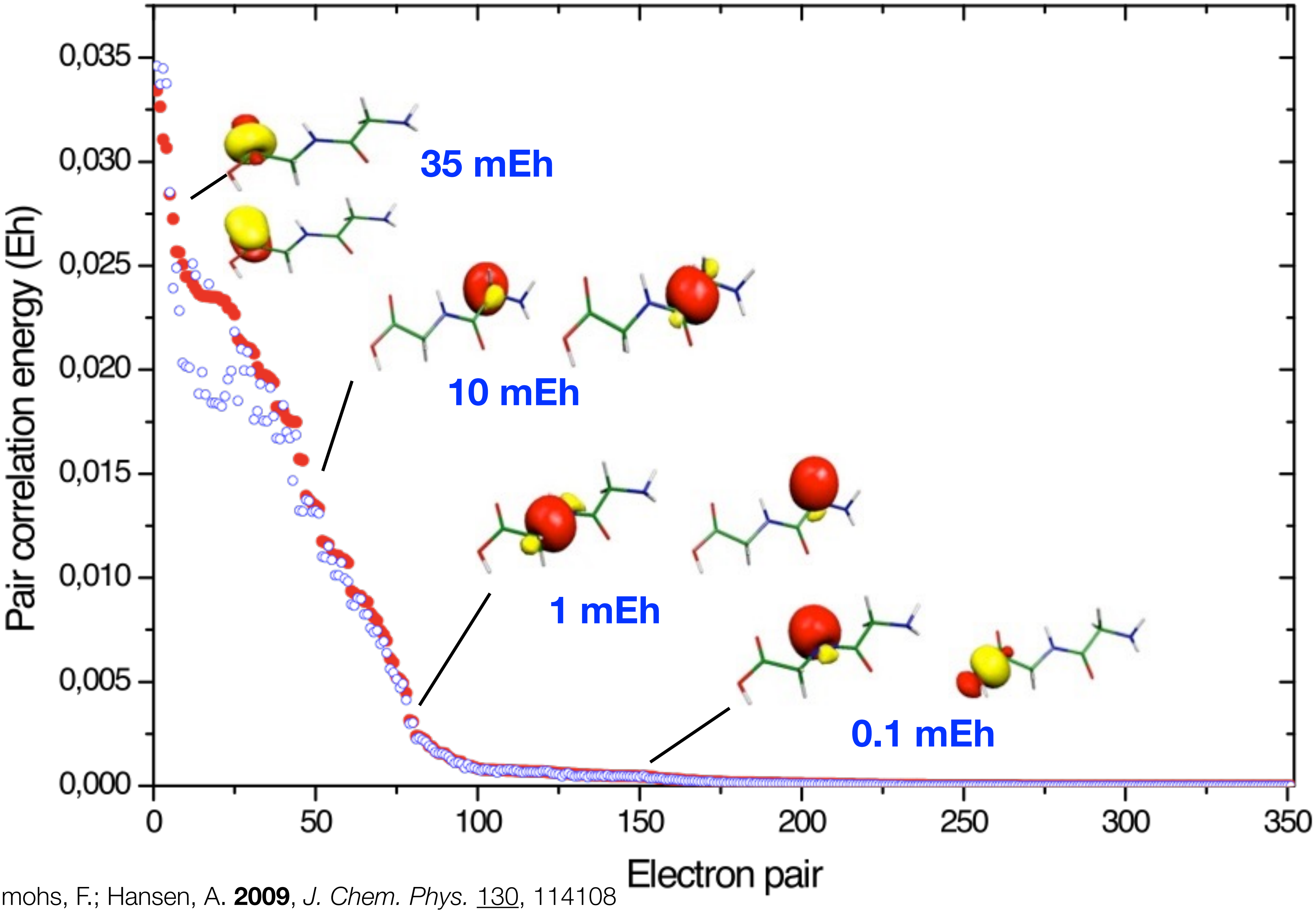
Thus:

$$\boxed{\tfrac{1}{4} \sum_{ijab} C_{ab}^{ij} \left\langle ij \| ab \right\rangle = \enclose{circle}{\tfrac{1}{2} \sum_{ij} \varepsilon_{ij}} = E_{corr}}$$

(Nesbet‘s theorem)

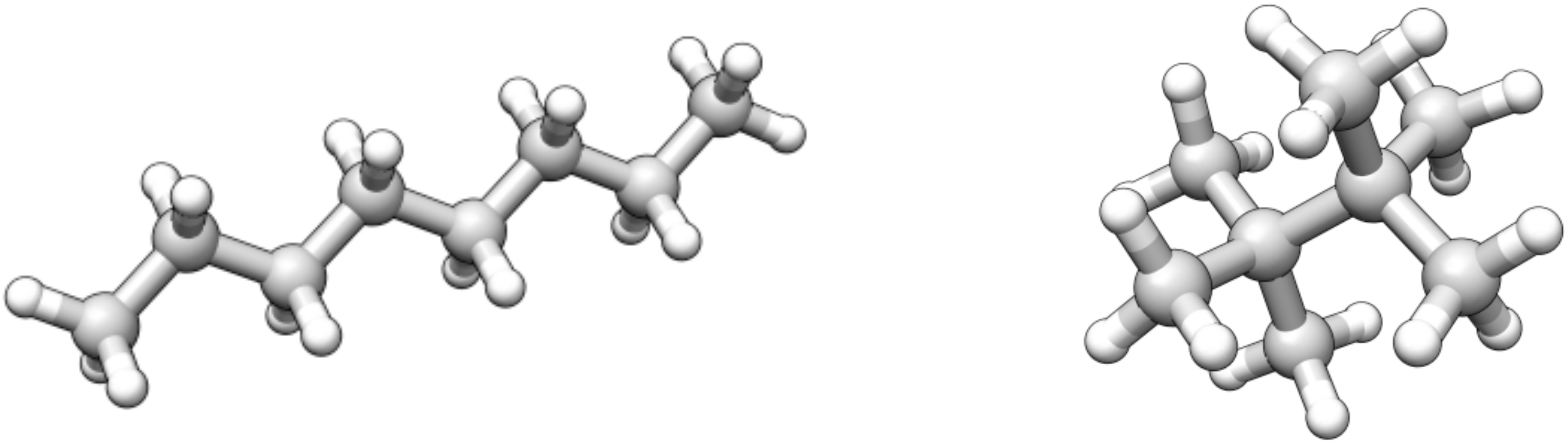
If we know the precise values of the double excitation coefficients we know the EXACT correlation energy! It is a sum of PAIR CORRELATION ENERGIES

Approximation 1: Locality of Pair Correlation Energies



FN, Wennmohs, F.; Hansen, A. **2009**, *J. Chem. Phys.* 130, 114108

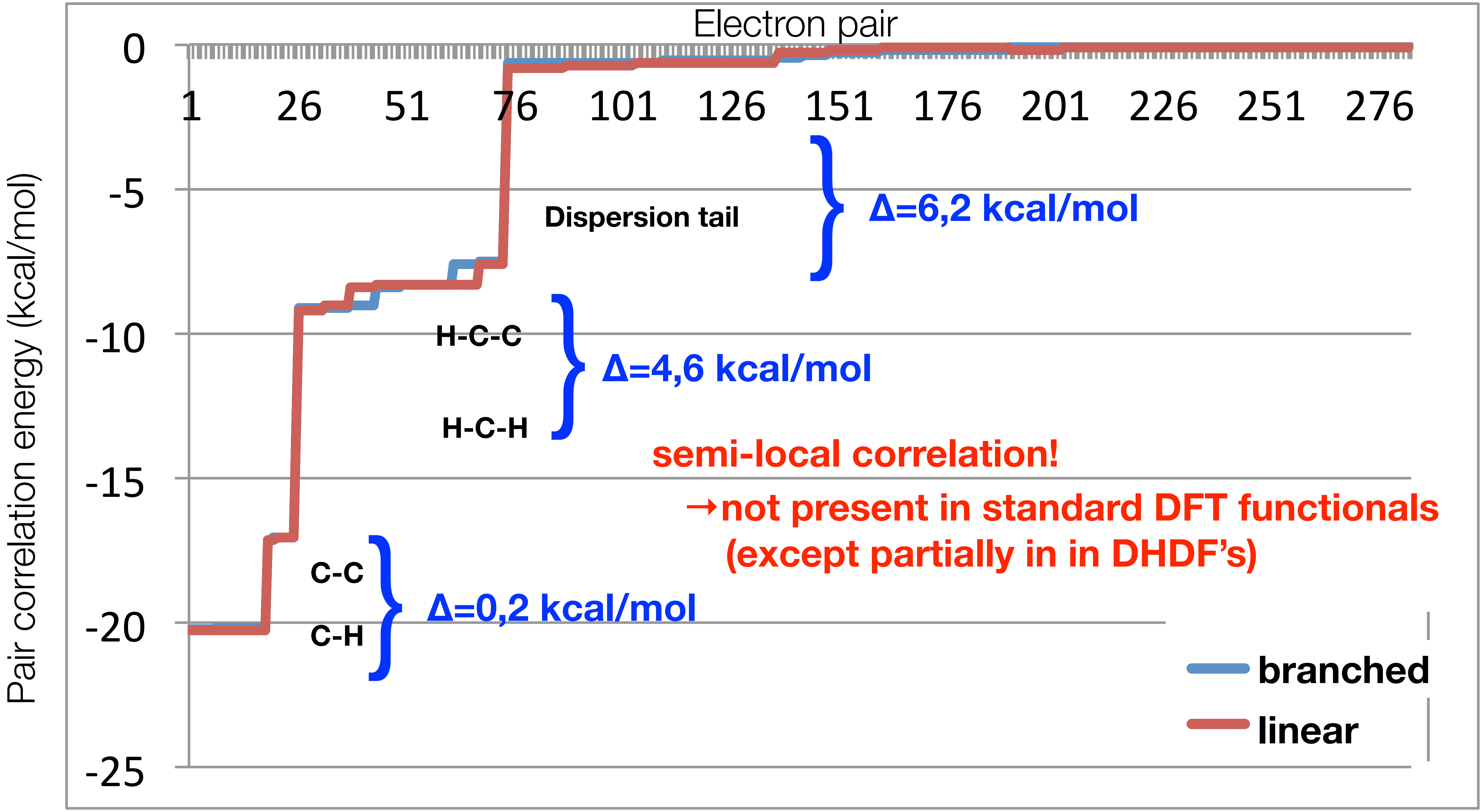
Chemically Speaking: How Local is the Correlation?



ΔE	=	+1.9±0.5	kcal/mol	Exp.
		-11.4	kcal/mol	HF
		-0,4	kcal/mol	CCSD
		-8.4	kcal/mol	B3LYP
		-9.9	kcal/mol	BLYP

Grimme S, *Angew. Chem.* **2006**, 118, 4571 –4575

Semi-Local Correlation Effects



Δ=6,2 kcal/mol

Δ=4,6 kcal/mol

Δ=0,2 kcal/mol

semi-local correlation!
→not present in standard DFT functionals
(except partially in in DHDF's)

We really need to capture both: semi-local and dispersion effects

Correlation is not quite as local as we'd like it to be!

FN, unpublished

The Leading Term of Electron Correlation

Coulomb **Exchange**

$$\varepsilon_{ij}^{MP2} = -\sum_{ab} \frac{4(ia\,|\,jb)(ia\,|\,jb) - 2(ia\,|\,jb)(ib\,|\,ja)}{\varepsilon_a + \varepsilon_b - \varepsilon_i - \varepsilon_j}$$

Assumptions:

- Occupied orbitals are localized
- Orbital energies in the denominator can be replaced by diagonal Fock matrix elements (**semi-canonical approximation**)
- Drop the exchange part for the long range behavior (*it falls off exponentially*)

$$\varepsilon_{ij}^{SC-MP2} \approx -4\sum_{ab} \frac{(ia\,|\,jb)^2}{\varepsilon_a + \varepsilon_b - F_{ii} - F_{jj}}$$

Analysis:

- For non-zero contributions, orbitals i and a and j and b must be „close“
- If the charge distributions $\rho_{ia}(\mathbf{r})=i(\mathbf{r})a(\mathbf{r})$ and $\rho_{jb}(\mathbf{r})=j(\mathbf{r})b(\mathbf{r})$ are well separated, we can make a **multipole expansion.**

Bipolar Expansion

Use the bipolar expansion in real spherical harmonics (local coordinate systems aligned!):

$$(ia\,|\,jb) \approx \sum_{l_a}\sum_{l_b}\sum_{m=l_<}^{l_>} d^{m}_{l_a l_b}\,\frac{Q^{ia}_{l_a m}\,Q^{jb}_{l_b m}}{R^{l_a+l_b+1}}$$

$$\left(d^{m}_{l_a l_b} = \frac{4\pi}{(2l_a+1)(2l_b+1)}\,\frac{(-1)^{l_a m}(l_a+l_b)!}{\sqrt{(l_a+m)(l_b+m)(l_b-m)(l_b-m)}}\right)$$

Distance between the center of the charge distributions

$$Q^{ia}_{lm} = \int r^{l} S_{lm}(\theta,\phi)\,\rho_{ia}(\mathbf{r})\,d\mathbf{r}$$

Multipole moments of the charge distributions

Since occupied and virtual orbitals are orthogonal, they have **no monopole**.

➡ the leading term is the **dipole-dipole-interaction**

$$(ia\,|\,jb) \propto R^{-3} \quad \Rightarrow \varepsilon^{SC-MP2}_{ij} \propto R^{-6}$$

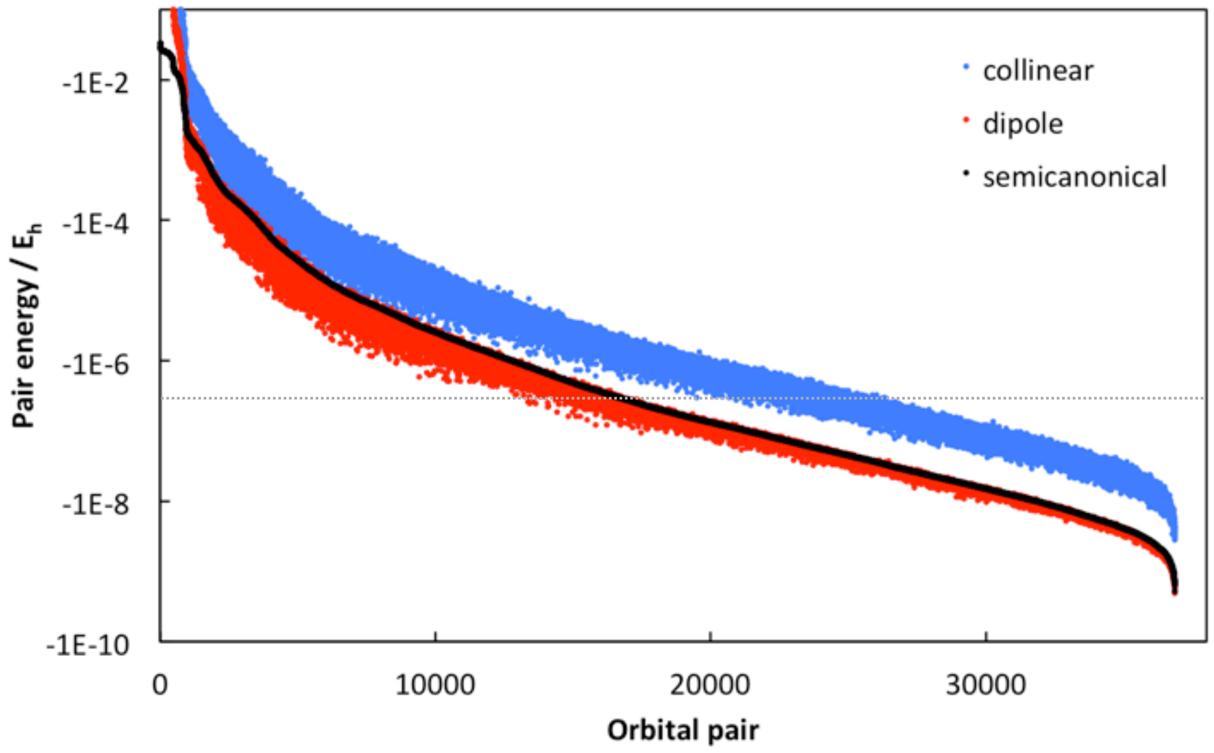
This is the **pure dispersion (induced dipole-induced dipole) interaction**.

Multipole Based Pair Prescreening

✓ Long Range Multipole Approximation

$$\varepsilon_{ij}^{DIP} = -\frac{4}{R_{ij}^{6}}\sum_{\tilde{\mu}_i\tilde{\nu}_j}\frac{\left\{\left\langle i\,|\,\mathbf{r}\,|\,\tilde{a}_i\right\rangle\left\langle j\,|\,\mathbf{r}\,|\,\tilde{b}_j\right\rangle-3\left(\left\langle i\,|\,\mathbf{r}\,|\,\tilde{a}_i\right\rangle\mathbf{R}_{ij}\right)\left(\left\langle j\,|\,\mathbf{r}\,|\,\tilde{b}_j\right\rangle\mathbf{R}_{ij}\right)\right\}^{2}}{\varepsilon_{\tilde{a}_i}+\varepsilon_{\tilde{b}_j}-F_{ii}-F_{jj}}$$

(1) .



- Drop pairs below a (conservative) selection threshold → linear scaling (ij)
- Correction for dropped pairs

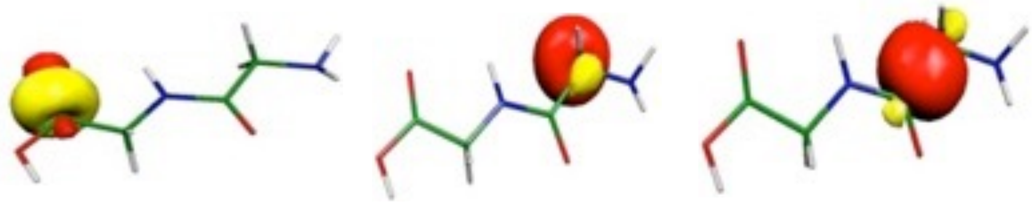
$$\Delta E_{\mathrm{Pre}} = \sum_{ij(dropped)} \varepsilon_{ij}^{DIP}$$

(this is actually very small relative to the residual error of the method)

Riplinger, C. FN *J. Chem. Phys*, **2013**, 138, 034106; Riplinger, C.; Pinski, P.; Becker, U.; Valeev, E. F.; Neese, F. *J. Chem. Phys.* **2016**, *144;* Hetzer, G.; Pulay, P.; Werner, HJ *Chem. Phys. Lett.*, **1998**, 290, 143

Approximation 2: Local Excitation Spaces

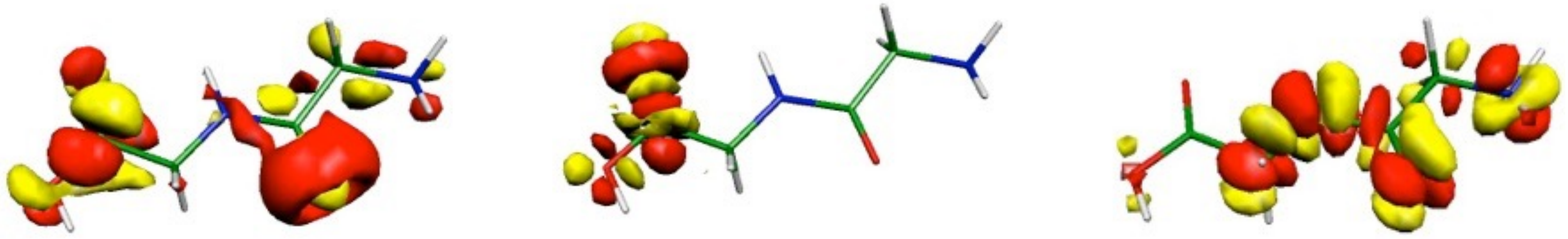
- ✓ The **occupied (internal) orbitals** localize nicely (mostly, that is)



  - ➡ *Significant MO coefficients extend over only a few atoms (1-5)*

- ✓ Pair correlation energies based on localized internal orbitals show locality with the expected R^{-6} decay

- ✓ The **virtual (external) orbitals** are problematic



  - ➡ *„Chaotic“, delocalized nature*
  - ➡ *‚Building higher and higher towers with smaller and smaller stones*
  - ➡ *Truncation schemes based on canonical MOs are unlikely to be highly successful*

Local Excitation Spaces

Let us go back to our analysis of the leading correlation term

$$\varepsilon_{ij}^{SC-MP2} \approx -4\sum_{ab} \frac{(ia\,|\,jb)^2}{\varepsilon_a + \varepsilon_b - F_{ii} - F_{jj}}$$

In order for this term to be significant

orbital a must be close to i AND orbital b must be close to j

in mathematical terms

The orbital pair ia and jb must have a significant *differential overlap*

Consequence: We can focus on local excitations and neglect long range charge transfer

However: A local representation of the virtual space is necessary

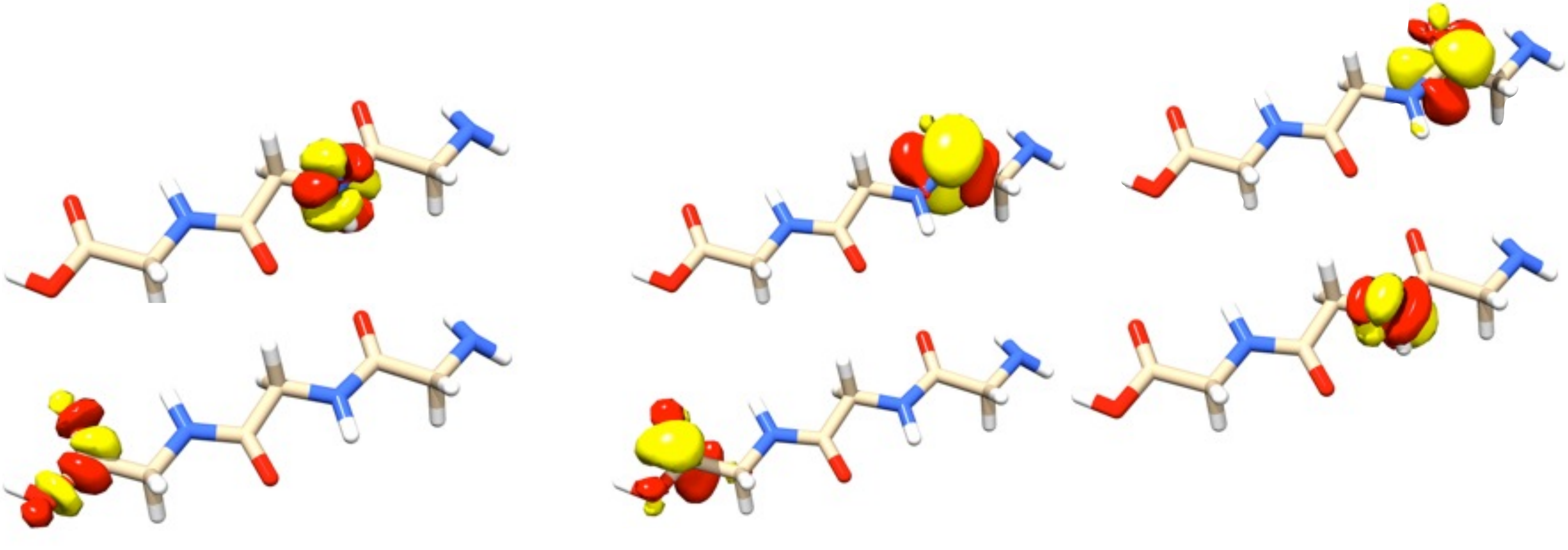
BUT: Standard localization schemes do not work well (but see Jörgensen et al)

- in particular for large basis sets the virtual orbitals do not localize well since the orthogonality constraint leads to highly oscillatory behavior
- Most researchers: Choose a non-orthogonal, local representation

Projected Atomic Orbitals

Projected atomic orbitals, PAOs, Pulay, P. *CPL*, **1983**, 100, 151

$$\left|\tilde{\mu}\right\rangle = \left(1 - \sum_i \left|i\right\rangle\left\langle i\right|\right)\left|\mu\right\rangle$$



- PAO’s are local close to the ‚parent‘ atom (but have significant tails)
- PAO’s span the virtual space and are orthogonal to the occupied space
- PAO’s are non-orthogonal and linearly dependent

Orthonormal Localized Virtual Orbitals

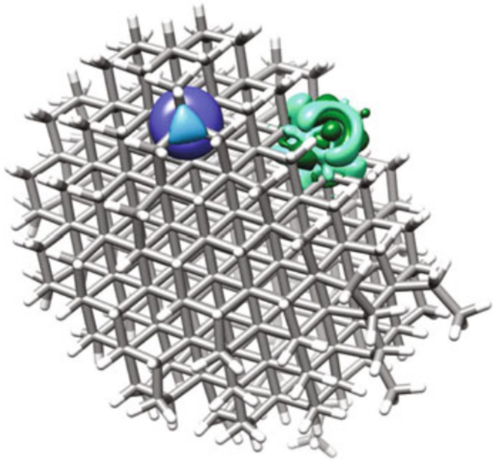
IM Hoyvik, K Kristensen, T Kjaergaard, P Jörgensen *Theo. Chem. Acc.*, **2014**, 133, 1417

- Careful comparison of PAOs and standard localized virtual MOs
- Suggested localization functional

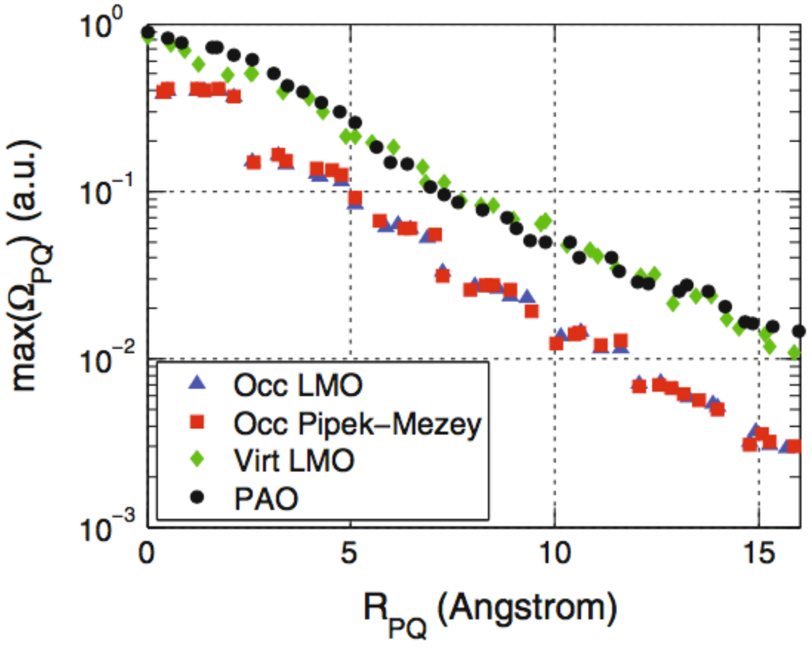
$$L = \sum_i \left| \left\langle \psi_i \,\middle|\, (\mathbf{r} - \mathbf{R}_i)^4 \,\middle|\, \psi_i \right\rangle \right|^2 \qquad \mathbf{R}_i = \left\langle \psi_i \,\middle|\, \mathbf{r} \,\middle|\, \psi_i \right\rangle$$

Fourth ‚central moment‘ (emphasizes the LMO tail region)

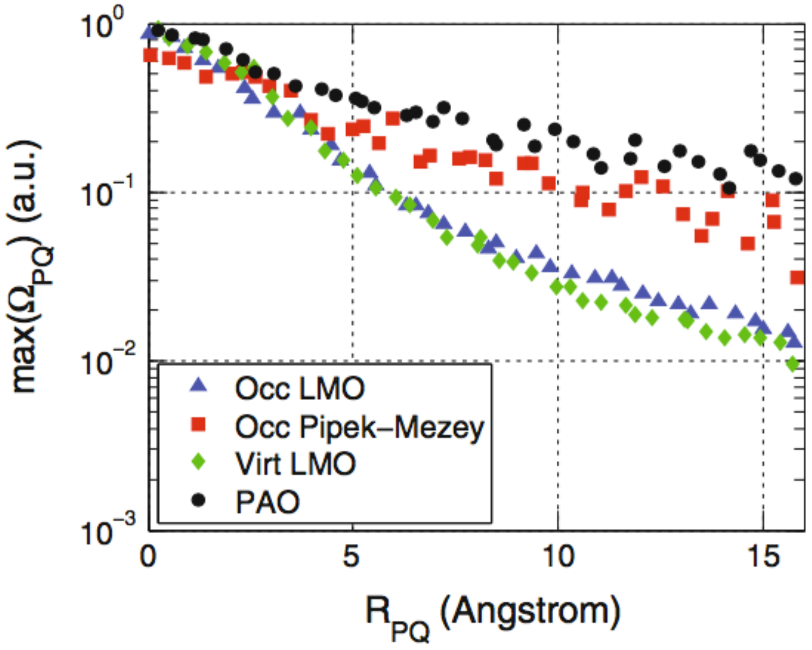
Center of gravity of MO i

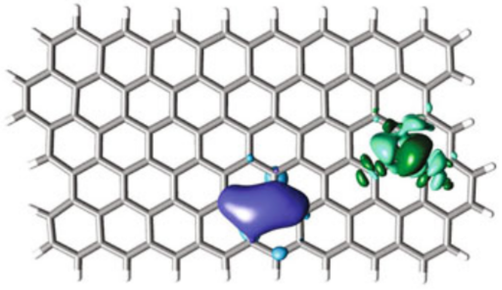


Diamond (localized)



Graphene (delocalized)





it seems possible to generate orthonormal virtual molecular orbitals that are as good or better than PAOs

Local Correlation Approaches

Methods to exploit the locality of electron correlation fall into two broad categories:

1) „Piecewise“ Local Approaches (Stoll, Piecuch, Kallay, Li, Jörgensen, Friedrich, ...)

- ✓ *Locality is used by dividing the molecule into subsystems (molecular fragments, orbital groups, ...).*
- ✓ *Small calculations are carried out on one, two, three ... subsystems at the time and*
- ✓ *Results are combined to estimate the total correlation energy*

2) „Direct“ Local Approaches (Pulay, Werner/Schütz, FN, ...)

- ✓ *Locality is used in the algorithm to avoid the computation of terms that are near zero or factors that are unity.*
- ✓ *Some kind of localized representation of the virtual space is required*

‚Piecewise‘ Local Correlation Schemes

The Cluster in Molecules (CIM) Approach

The Cluster in Molecules (CIM) approach was proposed by Li et al. and adopted by Piecuch et al. and Kallay et al.

Let us start from the (orbital invariant) coupled cluster energy expression

$$E_C = \sum_{ia} t_a^i f_{ia} + \frac{1}{4} \sum_{ijab} \left(t_{ab}^{ij} + t_a^i t_b^j\right) \left\langle ij \,||\, ab \right\rangle$$

and re-write it in terms of single-occupied orbital increments:

$$E_C = \sum_i \delta E_i \qquad \delta E_i = \sum_a t_a^i f_{ia} + \frac{1}{4} \sum_{jab} \left(t_{ab}^{ij} + t_a^i t_b^j\right) \left\langle ij \,||\, ab \right\rangle$$

let the occupied orbital be localized.

Replace: $\psi_i(\mathbf{r}) = \sum_\mu c_{\mu i} \mu(\mathbf{r}) \rightarrow \psi_i(\mathbf{r}) \approx \sum_{\mu \in \{i\}} c_{\mu i} \mu(\mathbf{r})$

{i} is the orbital domain of localized MO i and contains the AOs of all atoms to ensure a population of at least 1.98 when summed. Threshold τ Very small: 1-3 atoms!

S Li, J Ma, Y Jiang *J. Comp. Chem.*, **2002**, 23, 237; S Li, W Li, J Ma *Chin. J. Chem.*, **2003**, 21, 1422; S Li, J Shen, W Li, Y Jiang *J. Chem. Phys*, **2006**, 125, 074109

The Cluster in Molecules (CIM) Approach

Now use the off-diagonal Fock matrix elements $F_{ij} > \zeta_1$ to select orbitals j interacting with i

➡ **Environment [i]=[i, j_{1i}, j_{2i},…,j_{Ni}]** (Approximation 1)

Associated with the primary environment are the AOs that are the union of the AO domains of the orbitals in the environment

➡ **AO domain $[\mu]_i$**

Finally, the virtual space for the domain of orbital i is spanned by the PAO's belonging to the atoms that compose the AO domains

➡ **PAO domain $[\tilde{\mu}]_i$** (Approximation 2)

After orthogonalization, removal of linear dependencies and cutting small AO contributions, there is a set of orthonormal virtual orbitals that belong to the domain of i

(the actual algorithms are more involved than this, but the essence is just this)

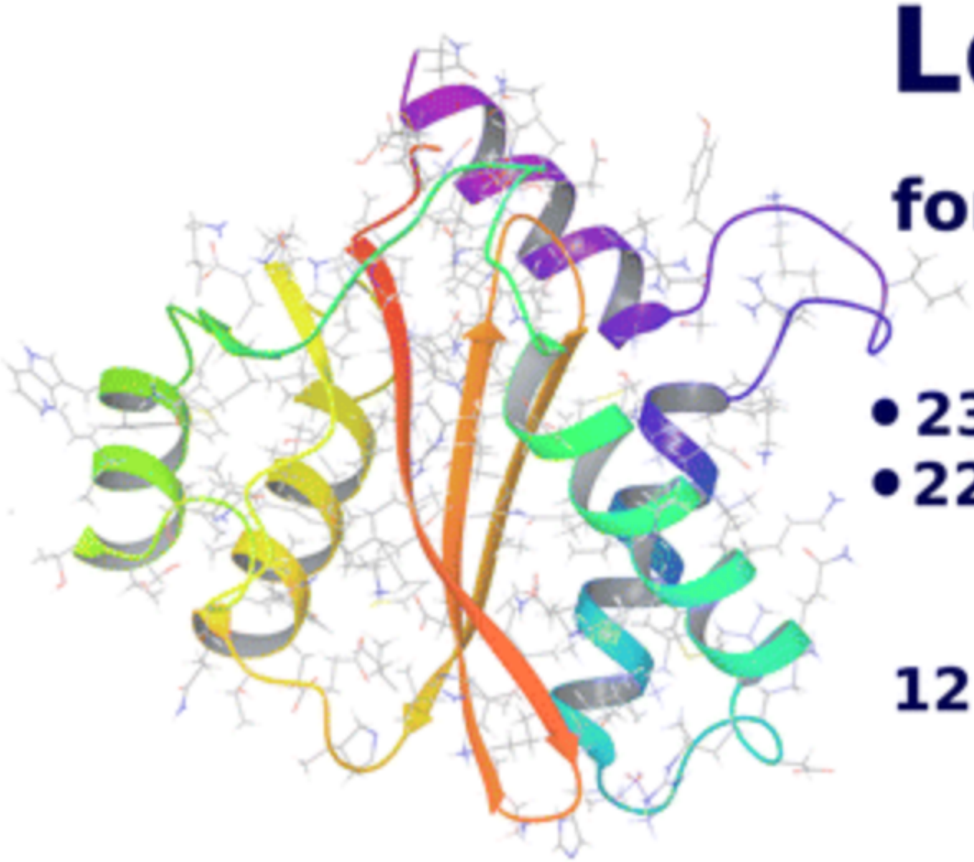
W Li, P Piecuch, JR Gour, S Li *j Chem. Phys*, **2009**, 131, 114109

An Integral-Direct Linear-Scaling Second-Order Møller–Plesset Approach

Péter R. Nagy[*], **Gyula Samu**, and **Mihály Kállay**[*]
MTA-BME Lendület Quantum Chemistry Research Group, Department of Physical Chemistry and Materials Science, Budapest University of Technology and Economics, P.O. Box 91, Budapest H-1521, Hungary

J. Chem. Theory Comput., Article ASAP
DOI: 10.1021/acs.jctc.6b00732

*E-mail: nagyrpeter@mail.bme.hu., *E-mail: kallay@mail.bme.hu.



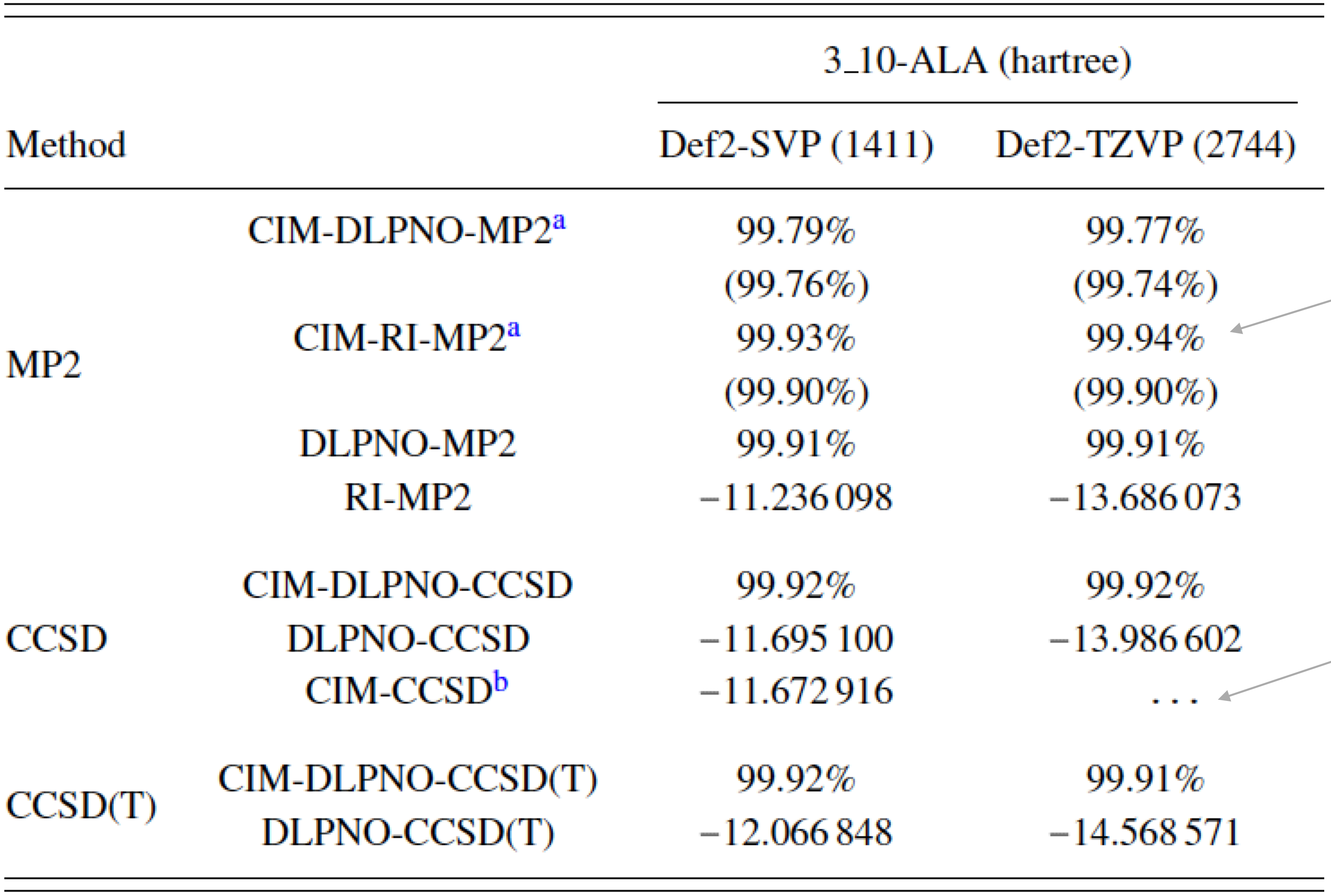
Local MP2

for HIV-1 integrase

- 2380 atoms
- 22621 atomic orbitals

12 h on a workstation

Problems with Piecewise Correlation Methods



		3_10-ALA (hartree)	
Method		Def2-SVP (1411)	Def2-TZVP (2744)
MP2	CIM-DLPNO-MP2[a]	99.79% (99.76%)	99.77% (99.74%)
	CIM-RI-MP2[a]	99.93% (99.90%)	99.94% (99.90%)
	DLPNO-MP2	99.91%	99.91%
	RI-MP2	−11.236 098	−13.686 073
CCSD	CIM-DLPNO-CCSD	99.92%	99.92%
	DLPNO-CCSD	−11.695 100	−13.986 602
	CIM-CCSD[b]	−11.672 916	. . .
CCSD(T)	CIM-DLPNO-CCSD(T)	99.92%	99.91%
	DLPNO-CCSD(T)	−12.066 848	−14.568 571

Good

Problem: the largest subsystem calculation is too large to be doable with a regular CCSD(T) program

Guo, Y.; Becker, U.; FN *JCP*, **2018**, 148 , 124117

Strengths and Weaknesses of Piecewise Schemes

Advantages

- ✓ Relatively **easy to implement** and intuitively appealing
- ✓ Extensive **reuse of canonical code** or even driving existing canonical programs
- ✓ **Readily extended to properties**
- ✓ **Parallelizes** with extremely high efficiency

Disadvantages

- ✓ **Redundancy**: Need for overlapping fragments leads to redundant integral calculations and amplitude optimizations.
- ✓ **Feasibility**: If the largest subsystem gets too large to be done by a regular correlation program, the method fails (Not a big problem for MP2, but very big problem for CCSD(T) or even higher)
- ✓ **Generality**: Fragmentation may not be straightforward in electronically complicated or heavily delocalized situations (does not apply to CIM and DEC)
- ✓ **Practicality**: Very few chemical applications; no extended benchmarks

Direct Local Correlation Schemes

Local MP2 Theory

Two complications relative to canonical MP2:

- At first sight the use of **non-orthogonal orbitals** appears to be „nightmare“ of added complexity. However, the PAO’s remain orthogonal to the occupied space and there are never more than two PAOs in any excited determinant
- In the local representation the **Fock matrix is no longer diagonal** and hence the usual Möller Plesset expansion does not apply.

Pulay and Saebo suggested to use the **Hylleraas functional** instead

$$E^{(2)} = \min(2\left\langle \Psi^{(1)} \mid H \mid \Psi^{(0)} \right\rangle + \left\langle \Psi^{(1)} \mid H_0 - E_0 \mid \Psi^{(1)} \right\rangle)$$

it readily leads to an **orbital invariant formulation of MP2**:

$$\Psi = \Psi^{(0)} + \Psi^{(1)} = \Psi_{HF} + \tfrac{1}{4}\sum_{ijab} C_{ab}^{ij} \Phi_{ij}^{ab}$$

$$H = H_0 + V \qquad H_0 = F, \quad V = H - H_0$$

P. Pulay, and S. Saebø, *Theor. Chim. Acta* 69, 357 (**1986**).

Local MP2 Theory

$$\left\langle \Psi^{(1)} \mid H \mid \Psi^{(0)} \right\rangle = \tfrac{1}{4}\sum_{ijab} C^{ij}_{ab} \left\langle ij \,||\, ab \right\rangle$$

$$\left\langle \Psi^{(1)} \mid H_0 - E_0 \mid \Psi^{(1)} \right\rangle = \tfrac{1}{8}\sum_{ijabc} C^{ij}_{ab} F_{cb} C^{ij}_{ac} - \tfrac{1}{8}\sum_{ijkab} C^{ij}_{ab} F_{jk} C^{ik}_{ab}$$

Minimization w.r.t. the coefficients C leads to the linear equation system:

$$R^{ij}_{ab} = \left\langle ij \,||\, ab \right\rangle + \sum_{c} (C^{ij}_{ac} F_{cb} + F_{ac} C^{ij}_{cb}) - \sum_{k} (C^{ik}_{ab} F_{kj} + F_{ik} C^{kj}_{ab}) = 0$$

Which immediately leads back to canonical MP2 if the Fock operator is diagonal.

If now the virtual orbitals are replaced by non-orthogonal PAO's, only a slight complication arises. The first order wave function is:

$$\Psi^{(1)} = \tfrac{1}{4}\sum_{ij\tilde{\mu}\tilde{\nu}} C^{ij}_{\tilde{\mu}\tilde{\nu}} \Phi^{\tilde{\mu}\tilde{\nu}}_{ij}$$

Local MP2 Theory

The residual becomes:

$$R^{ij}_{\tilde{\mu}\tilde{\nu}} = \left\langle ij \,||\, \tilde{\mu}\tilde{\nu} \right\rangle + \sum_{c} \left(F_{\tilde{\mu}\tilde{\kappa}} C^{ij}_{\tilde{\kappa}\tilde{\tau}} \tilde{S}_{\tilde{\tau}\tilde{\nu}} + \tilde{S}_{\tilde{\mu}\tilde{\kappa}} C^{ij}_{\tilde{\kappa}\tilde{\tau}} F_{\tilde{\tau}\tilde{\nu}} \right) - \sum_{k} \left(F_{ik} \tilde{S}_{\tilde{\mu}\tilde{\kappa}} C^{kj}_{\tilde{\kappa}\tilde{\tau}} \tilde{S}_{\tilde{\tau}\tilde{\nu}} + F_{kj} \tilde{S}_{\tilde{\mu}\tilde{\kappa}} C^{ik}_{\tilde{\kappa}\tilde{\tau}} \tilde{S}_{\tilde{\tau}\tilde{\nu}} \right) = 0$$

or: $$\mathbf{R}^{ij} = \mathbf{K}^{ij} + \left(\mathbf{F}\mathbf{C}^{ij}\tilde{\mathbf{S}} + \tilde{\mathbf{S}}\mathbf{C}^{ij}\mathbf{F} \right) - \sum_{k} \left(f_{ik} \tilde{\mathbf{S}}\mathbf{C}^{kj}\tilde{\mathbf{S}} + f_{kj} \tilde{\mathbf{S}}\mathbf{C}^{ik}\tilde{\mathbf{S}} \right) = 0$$

With the PAO overlap matrix $\tilde{S}_{\tilde{\mu}\tilde{\nu}} = \left\langle \tilde{\mu} \,|\, \tilde{\nu} \right\rangle$

So far: **no approximation!**

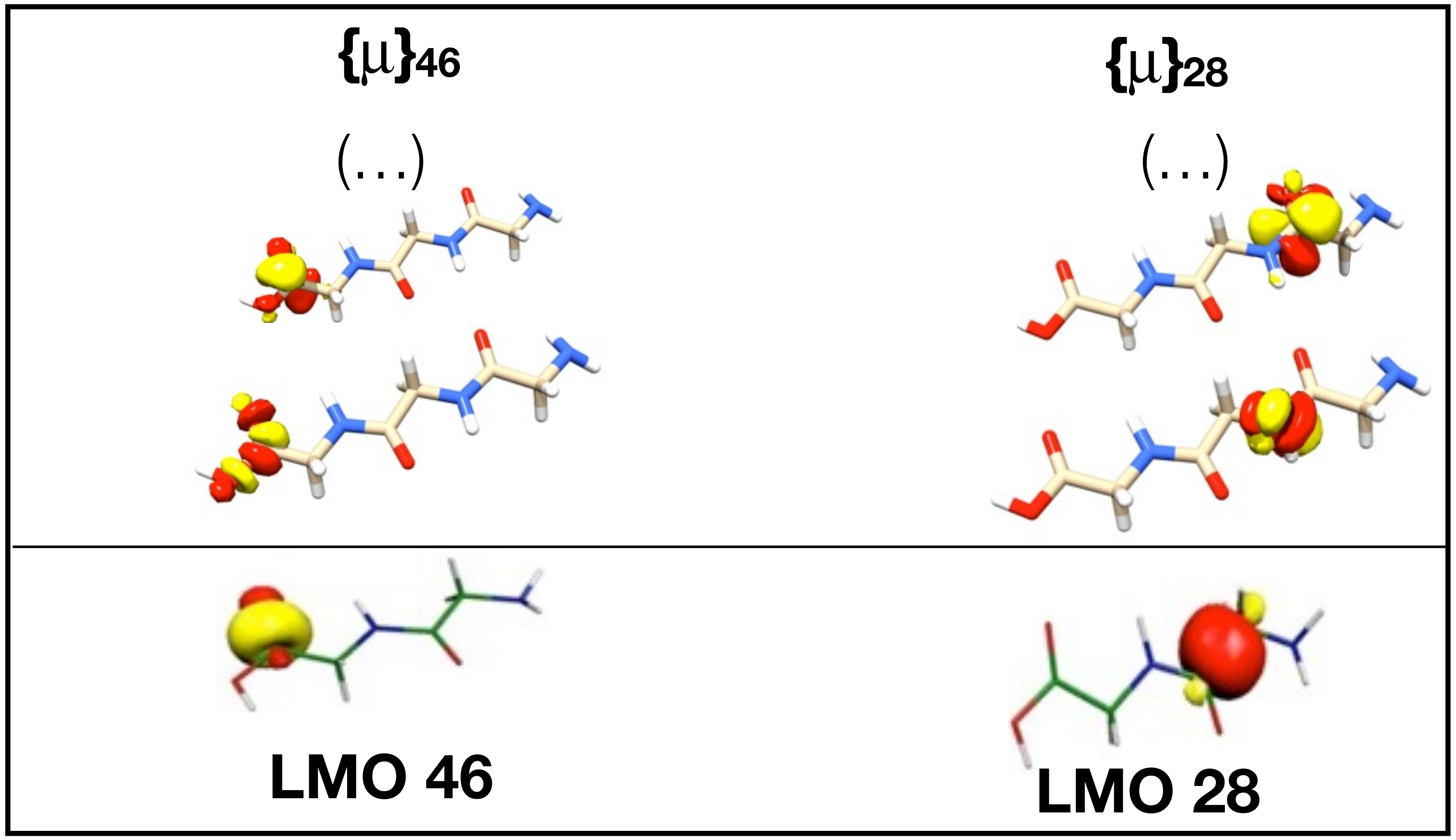
These equations represent a more complicated and ill-conditioned way to do MP2!

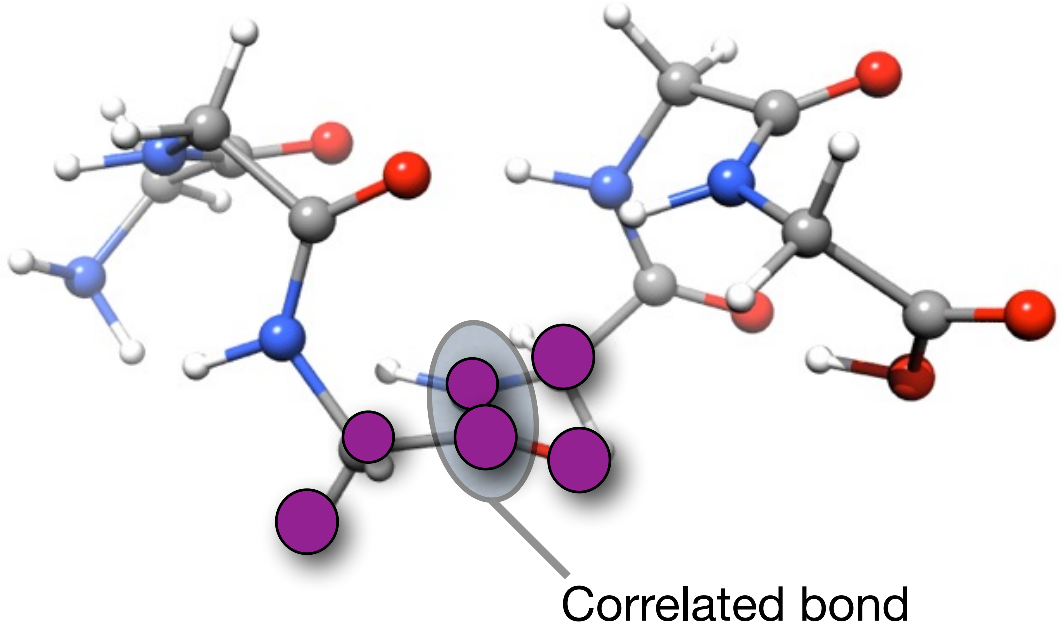
- Owing to the linear dependencies in the PAO set, the equations are singular.
- Removing the singularities by diagonalizing of the PAO overlap matrix and dropping the eigenvectors corresponding to zero eigenvalues leads back to canonical MP2

➡ **Introduce correlation domains for each electron pair ij**

PAO’s, Domains and Pair Domains

- A **domain** $\{\mu\}_i$ is a set of PAOs chosen for a given internal LMO according to some prescription





J. W. Boughton and P. Pulay, *J. Comp. Chem.* 14, 736 **?1993?**.

- A **pair domain** $\{\mu\}_{ii}$ is the union of the individual orbital domains

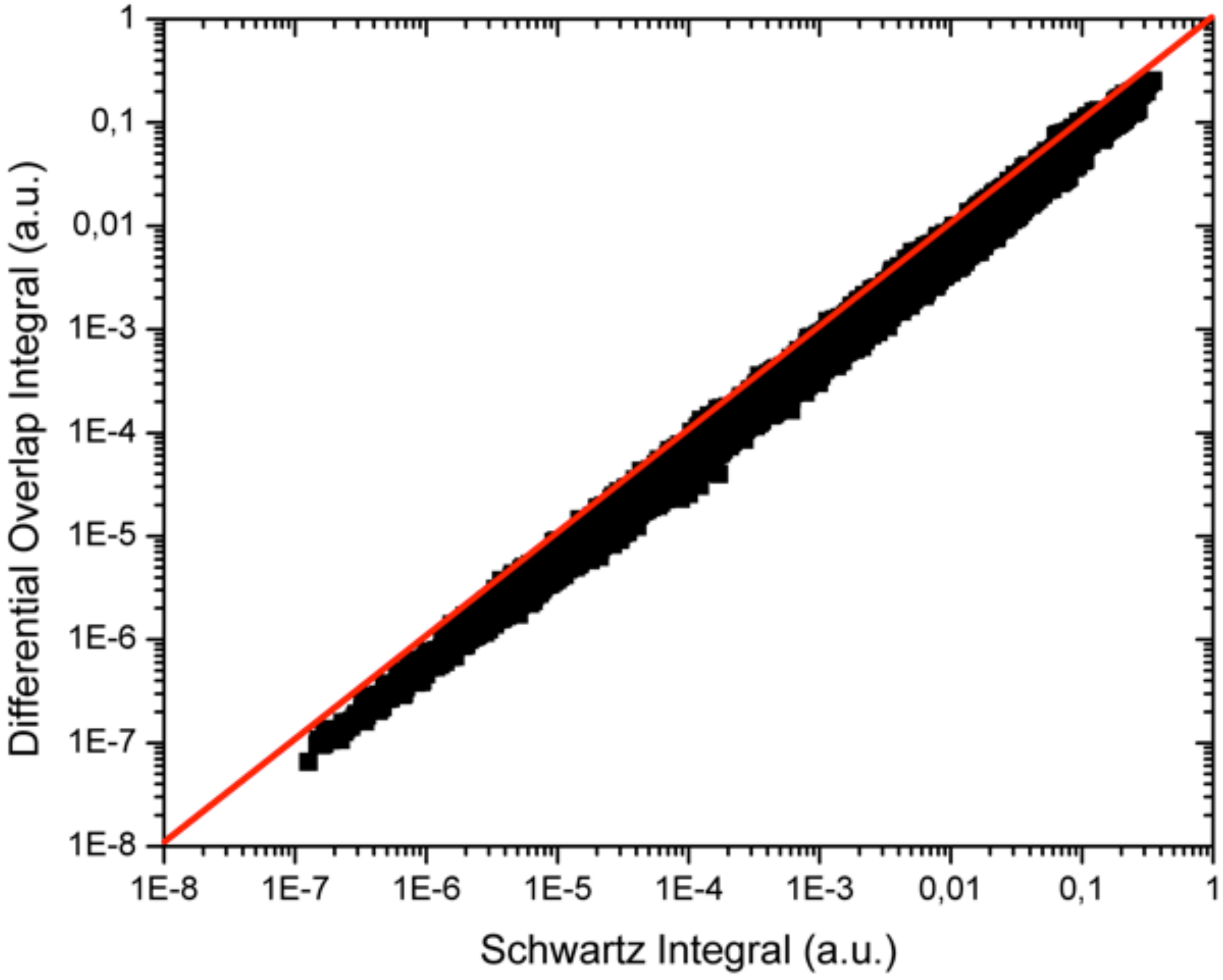
$$\{\mu\}_{ij} = \{\mu\}_i \cup \{\mu\}_j$$

- Fine, but for sufficient accuracy (99.9%) domains become impractically large
- Need „compaction“ of the space

Domain Construction

Our proposal: ***Differential Overlap Integral***

$$DOI_{ik} = \sqrt{\int |f_i(\mathbf{r})|^2 \, |g_k(\mathbf{r})|^2 \, d\mathbf{r}}$$



- ✓ PAOs that have a DOI above threshold with the LMO in question are added to the domain (then atom completed)

- Actual spatial extent of the virtual space is taken into account
- Excellent approximation to the the Schwartz screening integrals
- Easy to compute efficiently in linear scaling for any set of functions

Pinski, P.; Riplinger, C.; Valeev, E. F.; Neese, F. *J. Chem. Phys.* **2015**, *143*.

Properties of the PAO’s in Domains

$$|\tilde{\mu}\rangle = \left(\hat{1} - \sum_{i} |i\rangle\langle i|\right)|\mu\rangle$$

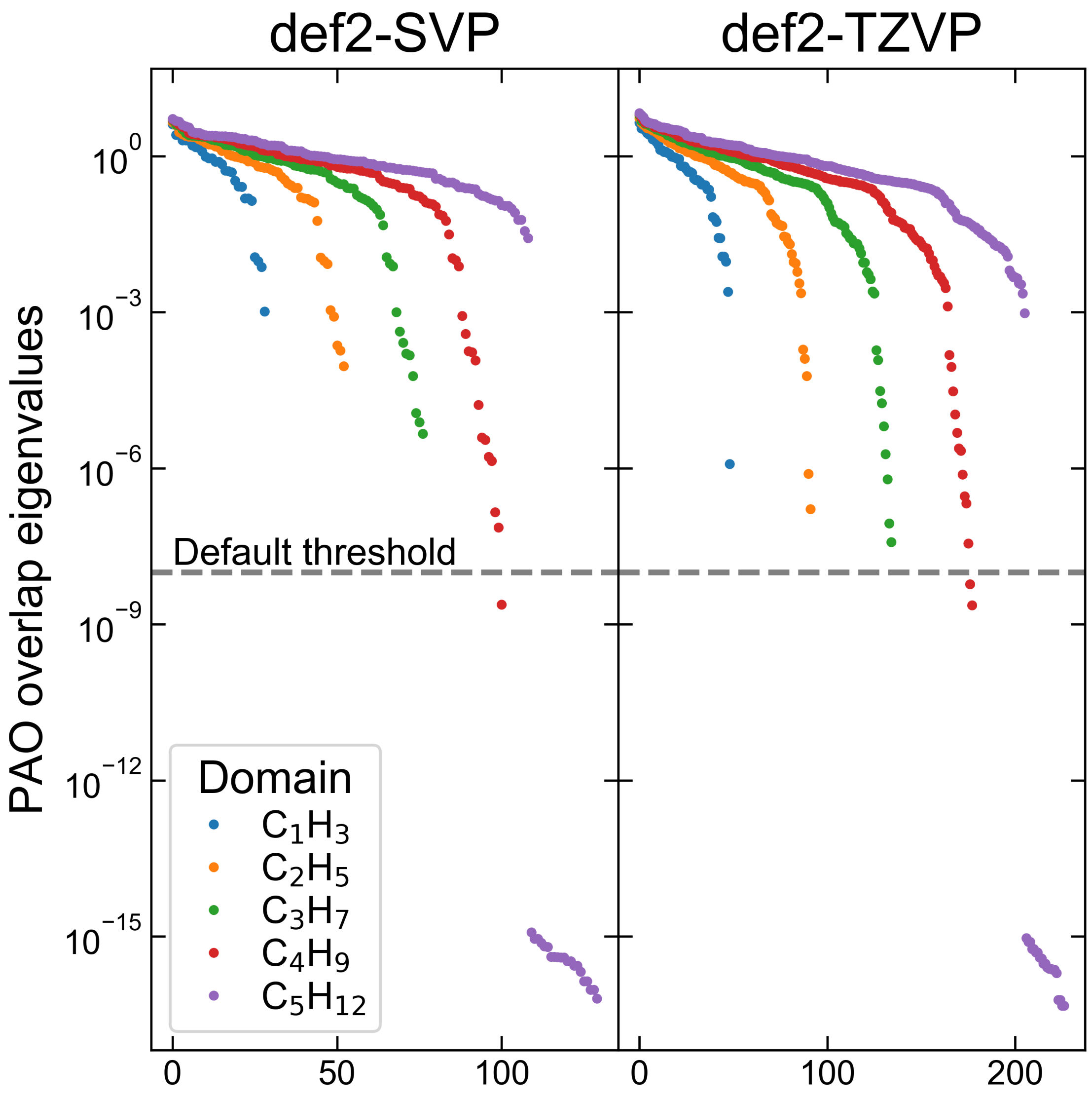
- Not normalized
- Guaranteed to be linearly dependent (spans the virtual space only and has full AO dimension)
- Somewhat local

$$\left\{\tilde{\mu}_{i}\right\}$$

- Now normalized but still not orthogonal and linearly dependent

- Diagonalize PAO overlap in a given domain and discard eigenvectors to eigenvalues <10^{-8} (smallest possible to retain numerical stability)

- Diagonalize the Fock operator in the non-redundant space to get to normlized, non-redundant, quasi-canonical PAO’s



Stoychev, G. L.; Auer, A. A.; Gauss, J.; Neese, F. *J. Chem. Phys.* **2021**, *154*, 27.

Solving Local MP2 Theory

Using the domains, the residual becomes

$$\mathbf{R}^{ij} = \mathbf{K}^{ij} + (\mathbf{F}^{(ij)}\mathbf{C}^{ij}\tilde{\mathbf{S}}^{(ij,ij)} + \tilde{\mathbf{S}}^{(ij,ij)}\mathbf{C}^{ij}\mathbf{F}^{(ij)}) - \sum_k (f_{ik}\tilde{\mathbf{S}}^{(ij,jk)}\mathbf{C}^{kj}\tilde{\mathbf{S}}^{(kj,ij)} + f_{kj}\tilde{\mathbf{S}}^{(ij,ik)}\mathbf{C}^{ik}\tilde{\mathbf{S}}^{(ik,ij)}) = 0$$

Where all matrices are now local:

$$\left.\begin{aligned} R^{ij}_{\tilde{\mu}\tilde{\nu}} &\longrightarrow R^{ij}_{\tilde{\mu}\in\{ij\},\tilde{\nu}\in\{ij\}} \\ K^{ij}_{\tilde{\mu}\tilde{\nu}} &\longrightarrow K^{ij}_{\tilde{\mu}\in\{ij\},\tilde{\nu}\in\{ij\}} \\ F^{(ij)}_{\tilde{\mu}\tilde{\nu}} &\longrightarrow F_{\tilde{\mu}\in\{ij\},\tilde{\nu}\in\{ij\}} \\ \tilde{S}^{(ij,kl)}_{\tilde{\mu}\tilde{\nu}} &\longrightarrow \tilde{S}_{\tilde{\mu}\in\{ij\},\tilde{\nu}\in\{kl\}} \end{aligned}\right\}$$

✓ Sub-matrices can either be stored or constructed on the fly

Two more steps are necessary to reach linear scaling

1. **Discard weakly interacting electron pairs:**

   In almost all local treatments done by *distance criteria* (R_{ij}=distance between orbital centroids)

2. **Discard small terms in the sum over k** by analyzing f_{ik},f_{jk}

   Threshold F_{Cut}~10^{-5} Eh

PAO based Local Correlation Treatments

JOURNAL OF CHEMICAL PHYSICS VOLUME 111, NUMBER 13 1 OCTOBER 1999

Low-order scaling local electron correlation methods. I. Linear scaling local MP2

Martin Schütz, Georg Hetzer, and Hans-Joachim Werner[a)]

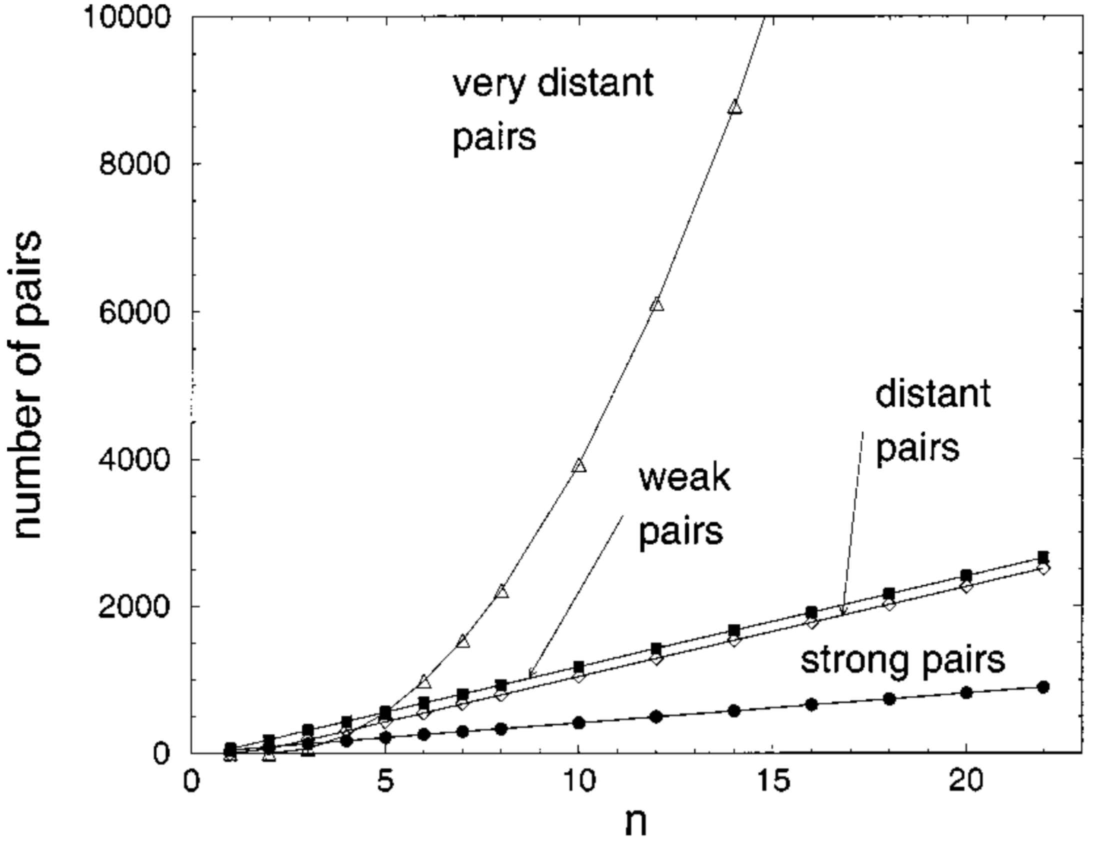


FIG. 1. The number of strong, weak, distant and very distant pairs as a function of the size n of a polyglycine peptide chain $[\mathrm{gly}]_n$. The number of strong, weak, and distant pairs all scale linearly with the molecular size, whereas the number of very distant pairs scales quadratically with n.

Hierarchical treatment of electron pairs, multipole approximations, careful thresholding lead to efficient, linear scaling algorithms

Problems with PAO based treatments

NJ Russ, TD Crawford J. Chem. Phys., **2004**, 121, 691

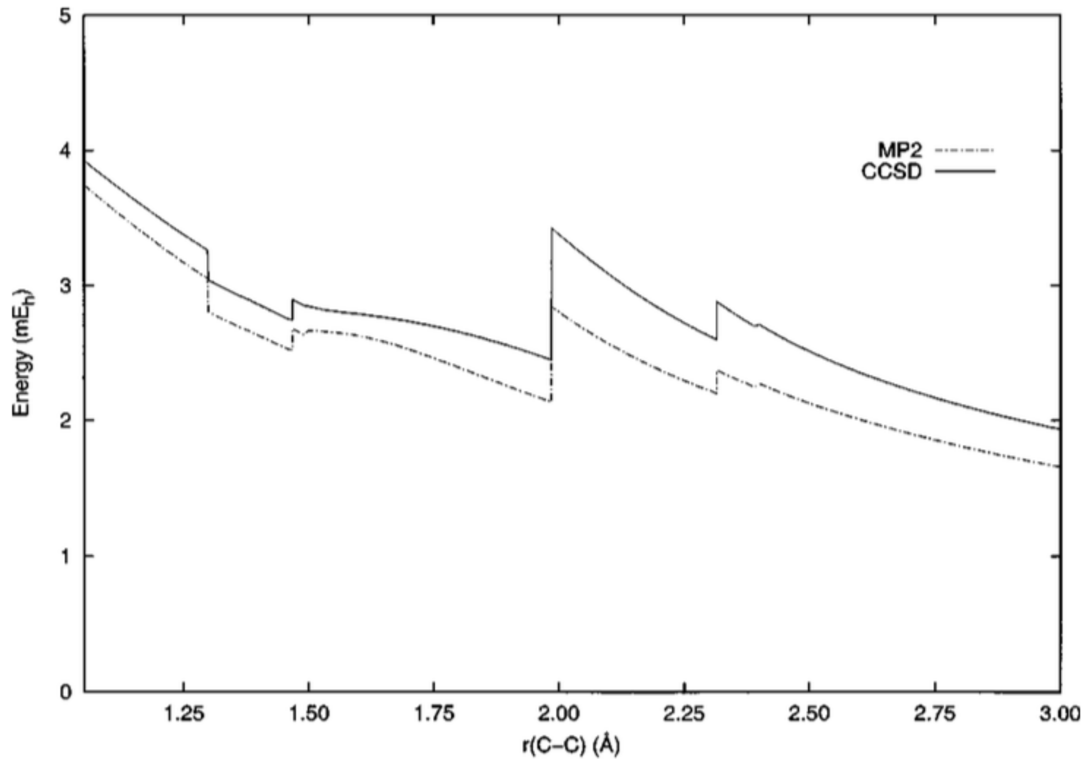


FIG. 7. LMP2 and LCCSD localization errors (in mE_h) for singlet ketene dissociation, where the four discontinuities discussed in the text are clearly visible.

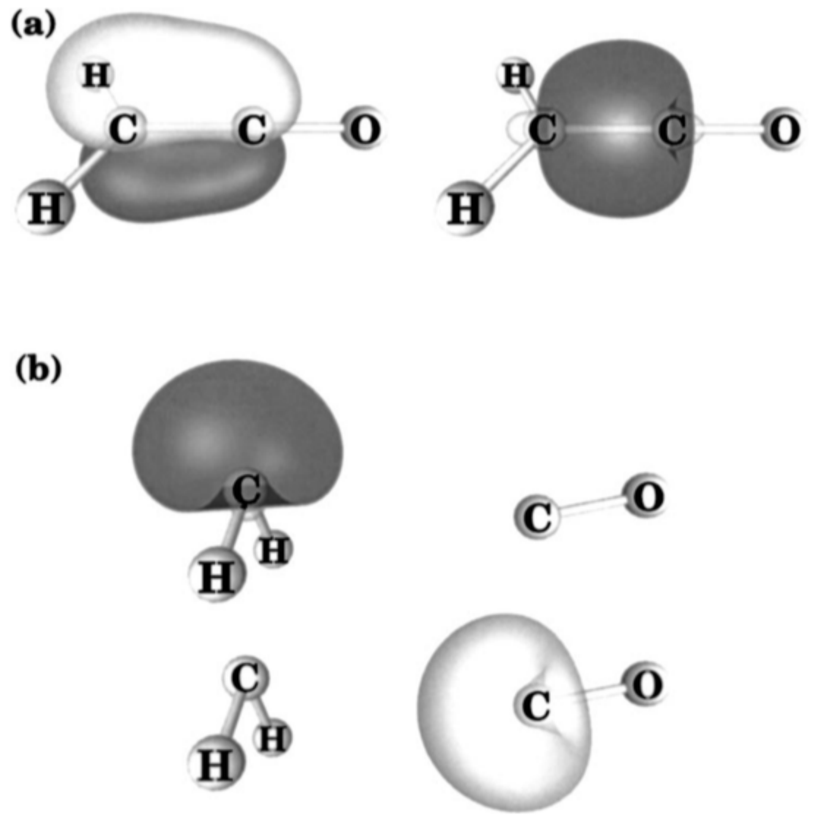
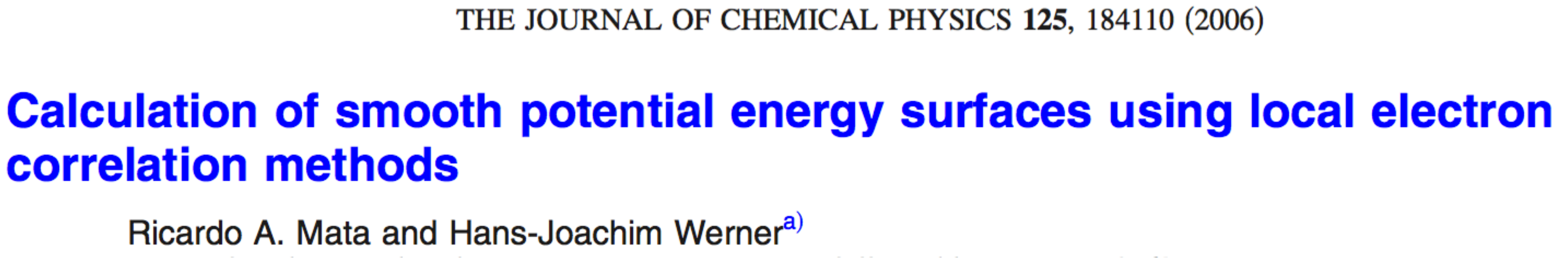


FIG. 8. Contour plots of the relevant Pipek–Mezey localized orbitals for singlet ketene: (a) The π and σ bonding orbitals near the equilibrium geometry and (b) the corresponding lone-pair dissociated MOs of singlet methylene and carbon monoxide.

- ➡ Discontinuous potential energy surfaces due to small and changing domains along the PES
- ➡ Reply:



THE JOURNAL OF CHEMICAL PHYSICS **125**, 184110 (2006)

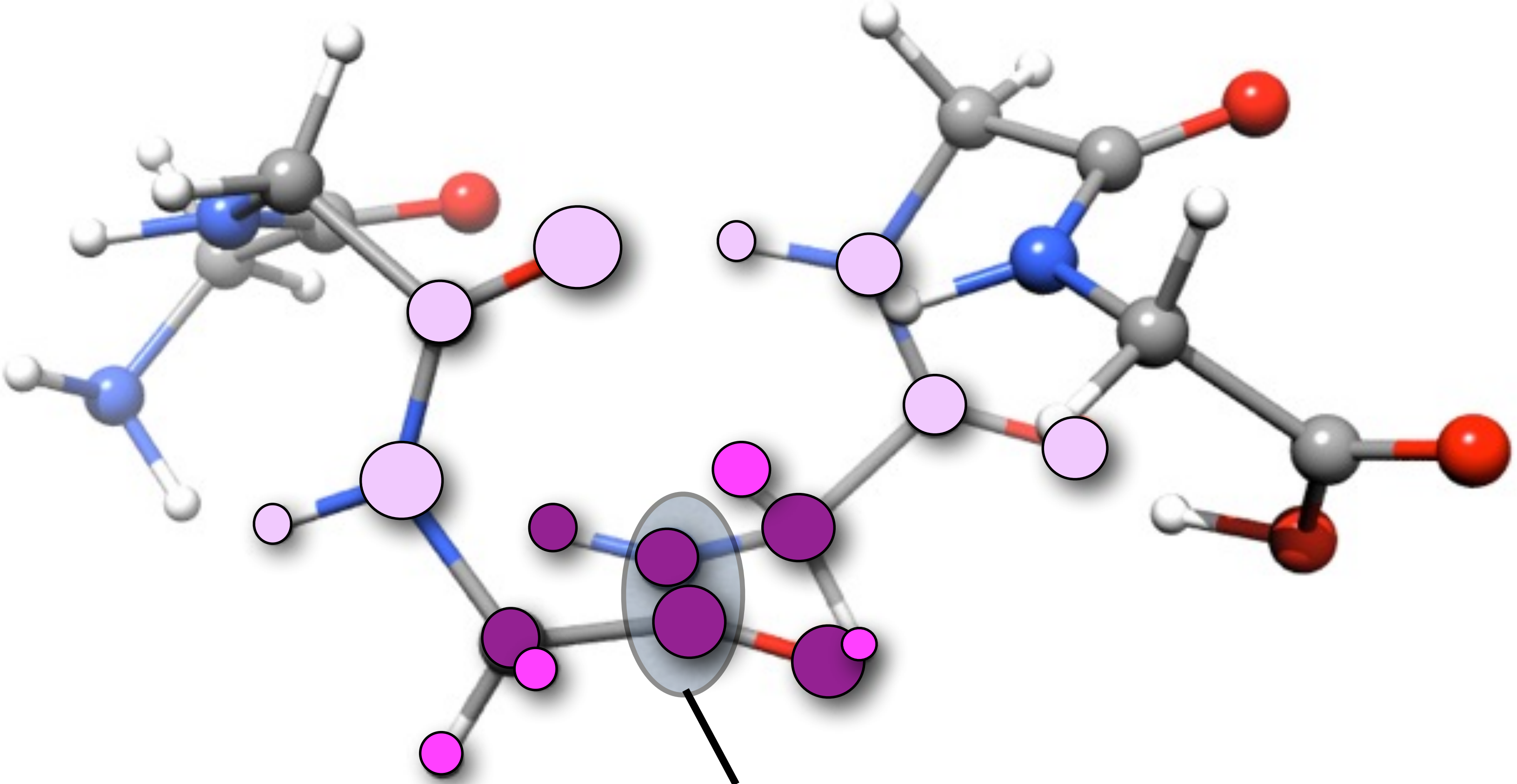
Calculation of smooth potential energy surfaces using local electron correlation methods

Ricardo A. Mata and Hans-Joachim Werner[a)]

In either PAO based or CIM based procedures the correlation energy recovered depends critically on the PAO domains.

… how large do they have to be in order to lead to an accurate result?

How Large do Domains have to be?



Correlated bond

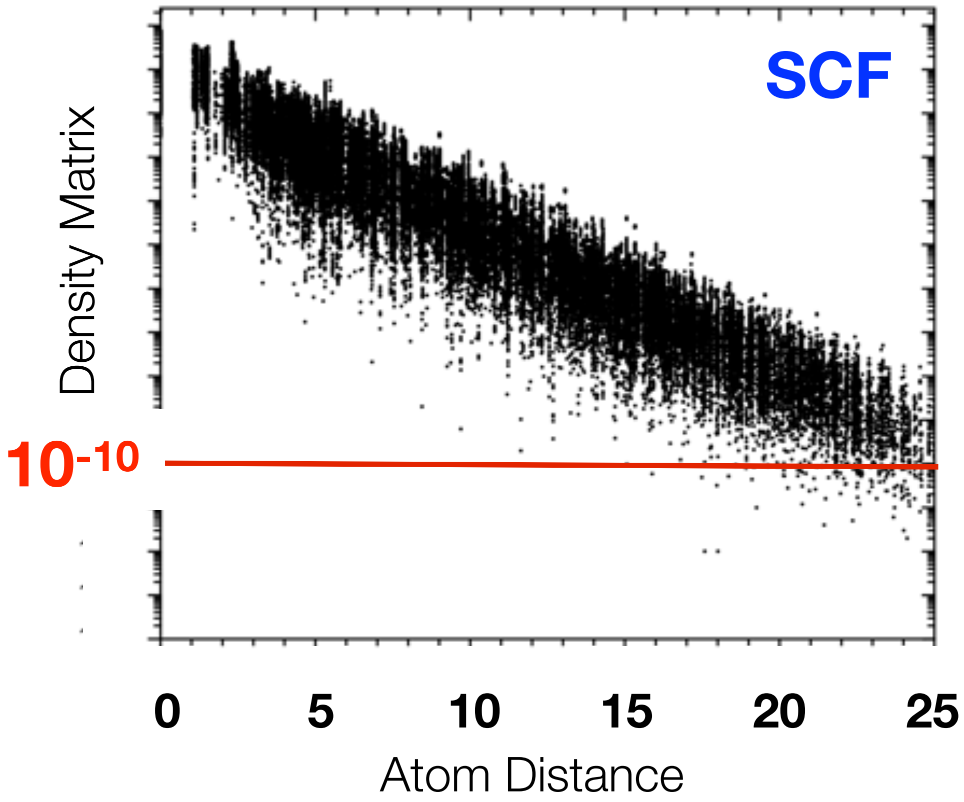
$T_{DO}=0.1$	98.4% E_{corr}	$N^{av}_{PAO}=115$
$T_{DO}=0.01$	99.7% E_{corr}	$N^{av}_{PAO}=588$
$T_{DO}=0.001$	99.9% E_{corr}	$N^{av}_{PAO}=935$

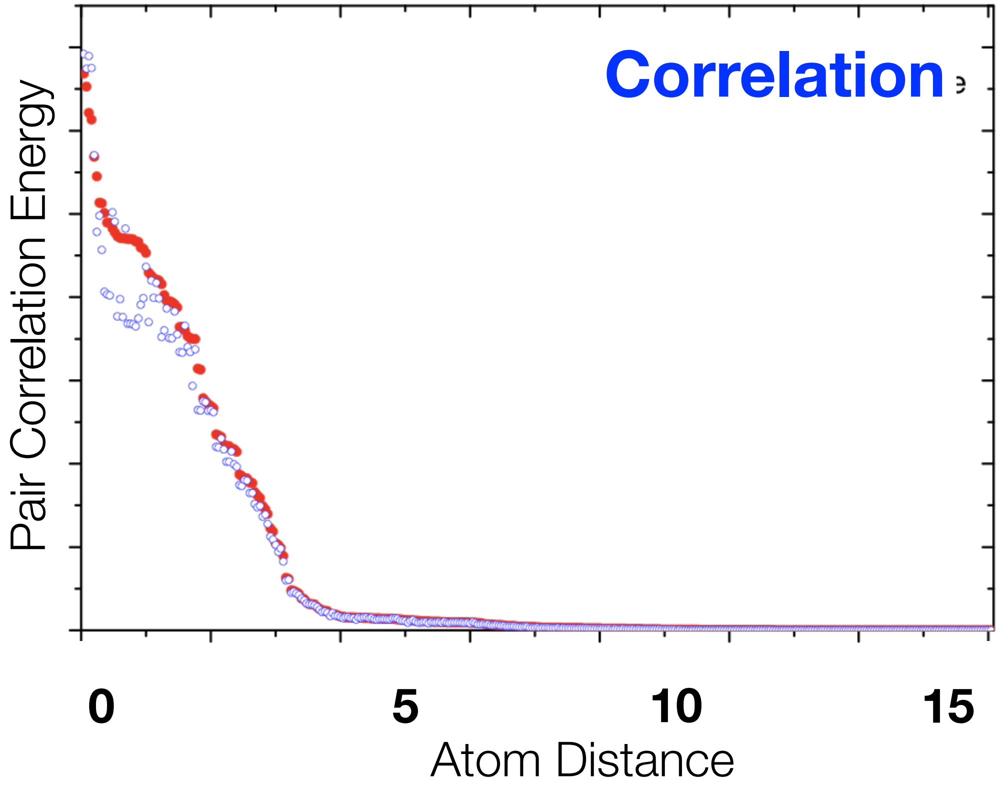
- At the domain size one reaches target accuracy the average number of PAOs per domain is too large for the calculation to be efficient or even doable
- There are important correlation effects that are not *that* local

Virtual Space Compaction: Pair Natural orbitals

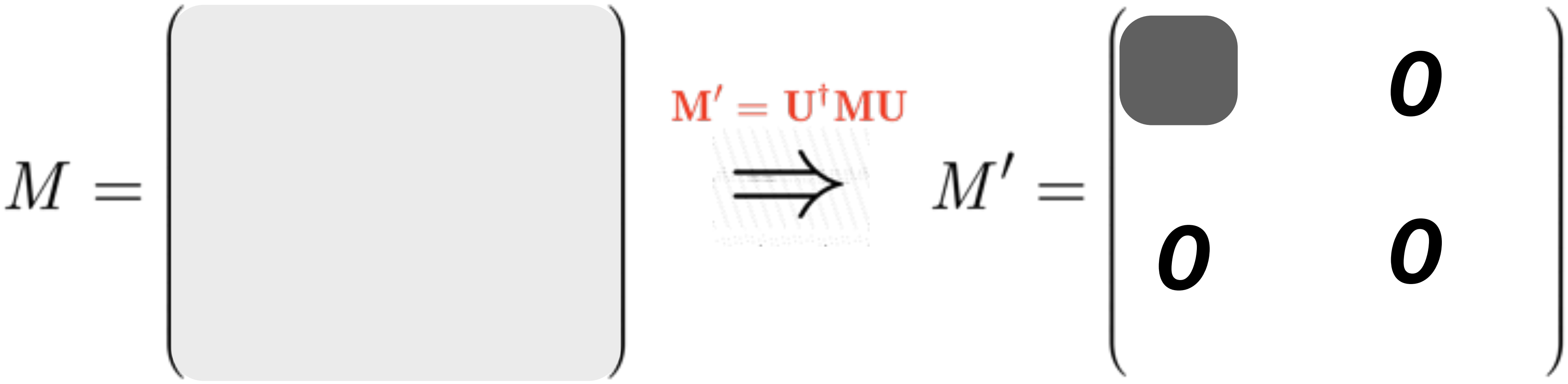
Saving Time in Electronic Structure Calculations

Exploit Sparsity!

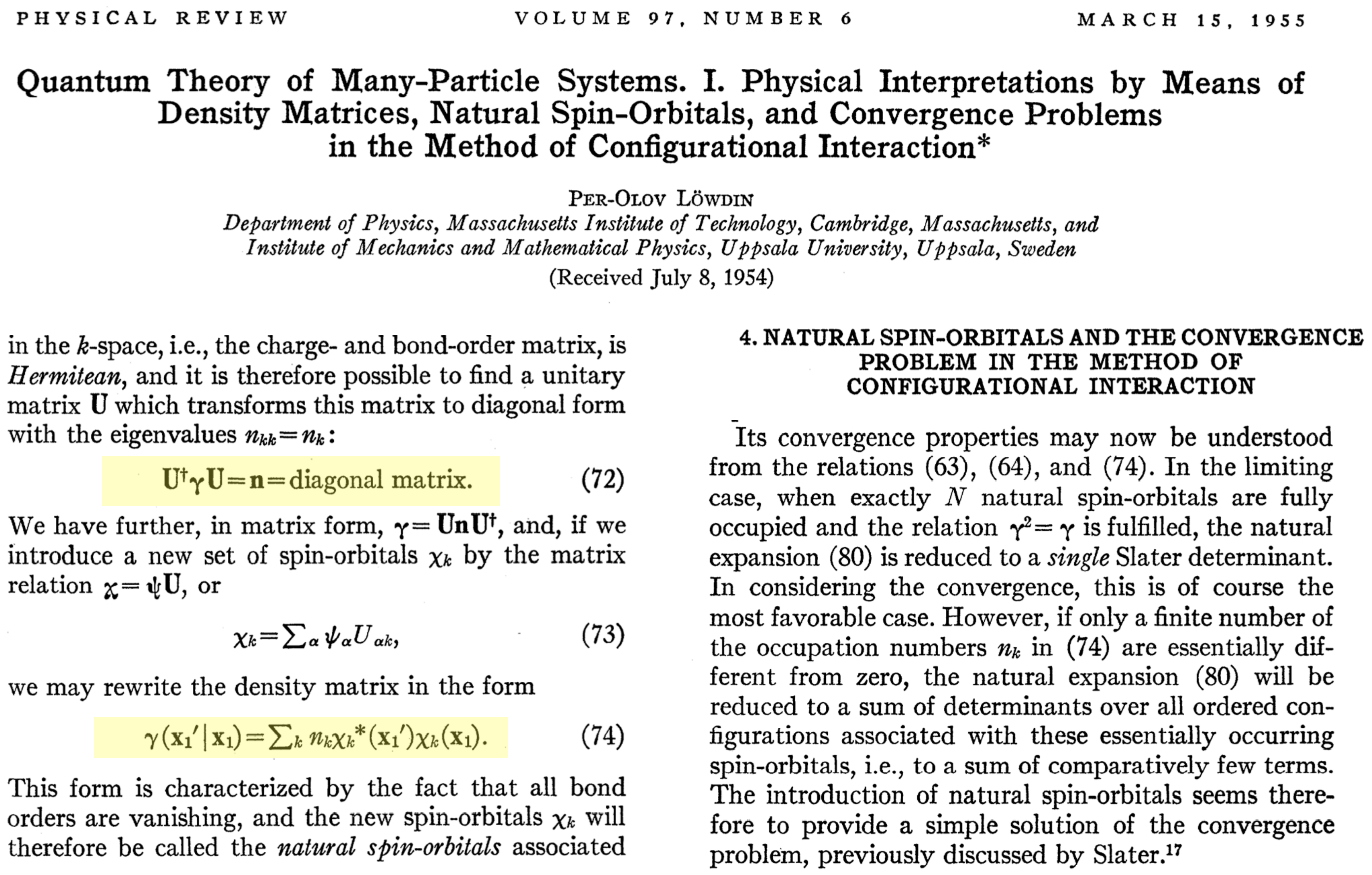




Compress Data!



Natural Orbitals



PHYSICAL REVIEW VOLUME 97, NUMBER 6 MARCH 15, 1955

Quantum Theory of Many-Particle Systems. I. Physical Interpretations by Means of Density Matrices, Natural Spin-Orbitals, and Convergence Problems in the Method of Configurational Interaction*

PER-OLOV LÖWDIN
Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts, and Institute of Mechanics and Mathematical Physics, Uppsala University, Uppsala, Sweden

(Received July 8, 1954)

in the k-space, i.e., the charge- and bond-order matrix, is *Hermitean*, and it is therefore possible to find a unitary matrix $\mathbf{U}$ which transforms this matrix to diagonal form with the eigenvalues $n_{kk}=n_k$:

$$\mathbf{U}^\dagger\gamma\mathbf{U}=\mathbf{n}=\text{diagonal matrix.} \qquad (72)$$

We have further, in matrix form, $\gamma=\mathbf{UnU}^\dagger$, and, if we introduce a new set of spin-orbitals χ_k by the matrix relation $\boldsymbol{\chi}=\boldsymbol{\psi}\mathbf{U}$, or

$$\chi_k=\sum\nolimits_\alpha \psi_\alpha U_{\alpha k}, \qquad (73)$$

we may rewrite the density matrix in the form

$$\gamma(\mathbf{x}_1'|\mathbf{x}_1)=\sum\nolimits_k n_k\chi_k^*(\mathbf{x}_1')\chi_k(\mathbf{x}_1). \qquad (74)$$

This form is characterized by the fact that all bond orders are vanishing, and the new spin-orbitals χ_k will therefore be called the *natural spin-orbitals* associated

4. NATURAL SPIN-ORBITALS AND THE CONVERGENCE PROBLEM IN THE METHOD OF CONFIGURATIONAL INTERACTION

Its convergence properties may now be understood from the relations (63), (64), and (74). In the limiting case, when exactly N natural spin-orbitals are fully occupied and the relation $\gamma^2=\gamma$ is fulfilled, the natural expansion (80) is reduced to a *single* Slater determinant. In considering the convergence, this is of course the most favorable case. However, if only a finite number of the occupation numbers n_k in (74) are essentially different from zero, the natural expansion (80) will be reduced to a sum of determinants over all ordered configurations associated with these essentially occurring spin-orbitals, i.e., to a sum of comparatively few terms. The introduction of natural spin-orbitals seems therefore to provide a simple solution of the convergence problem, previously discussed by Slater.[17]

Note added in proof.—It is desirable to have also a more exact mathematical measure for the rapidity of convergence of the two configurational interaction series (66) and (80). We note that, according to (60) and (63), the charge order $\gamma(k)$ gives the probability for the ordinary spin-orbital ψ_k to occur in the expansion of the total wave function Ψ. If only M of the numbers $\gamma(k)$, $k=1, 2, 3, \cdots$, are essentially different from zero, then the number of essential terms in (66) is given by the corresponding number of possible configurations: $M!/N!\,(M-N)!$. In using this procedure, however, it is necessary to evaluate the individual quantities $\gamma(k)$ and to distinguish between essential and unessential charge orders.

A still simpler measure of convergence may be constructed by observing that the charge orders always lie between 0 and 1 and that, in the limiting cases $\gamma(k)=0$ and $\gamma(k)=1$, the corresponding spin-orbital ψ_k occurs in none or in all of the terms in (66), respectively, without contributing to the slowing down of the convergence of the series. The eventual slowness of the convergence of (66) depends instead on the possibility for an electron to be distributed over two or more spin-orbitals, giving charge orders of an intermediate order of magnitude, $0<\gamma(k)<1$. The rapidity of convergence of (66) may therefore be measured by the smallness of the quantity

$$\vartheta = (1/N)\,\Sigma_k\{1-\gamma(k)\}\gamma(k) = 1-(1/N)\,\Sigma_k\{\gamma(k)\}^2,$$

which fulfills the inequality of $0\leqq\vartheta<1$. In considering different basic sets $\psi_1, \psi_2, \psi_3, \cdots$ for the description of the same total wave function Ψ, it is clear that the natural spin-orbitals χ_k are characterized by having the *smallest* ϑ *value possible.* According to (72), we have $\boldsymbol{\gamma}=\mathbf{U}\mathbf{n}\mathbf{U}^\dagger$ and $\boldsymbol{\gamma}^2=\mathbf{U}\mathbf{n}^2\mathbf{U}^\dagger$, leading to $\mathrm{Tr}(\boldsymbol{\gamma}^2)=\mathrm{Tr}(\mathbf{n}^2)$ and

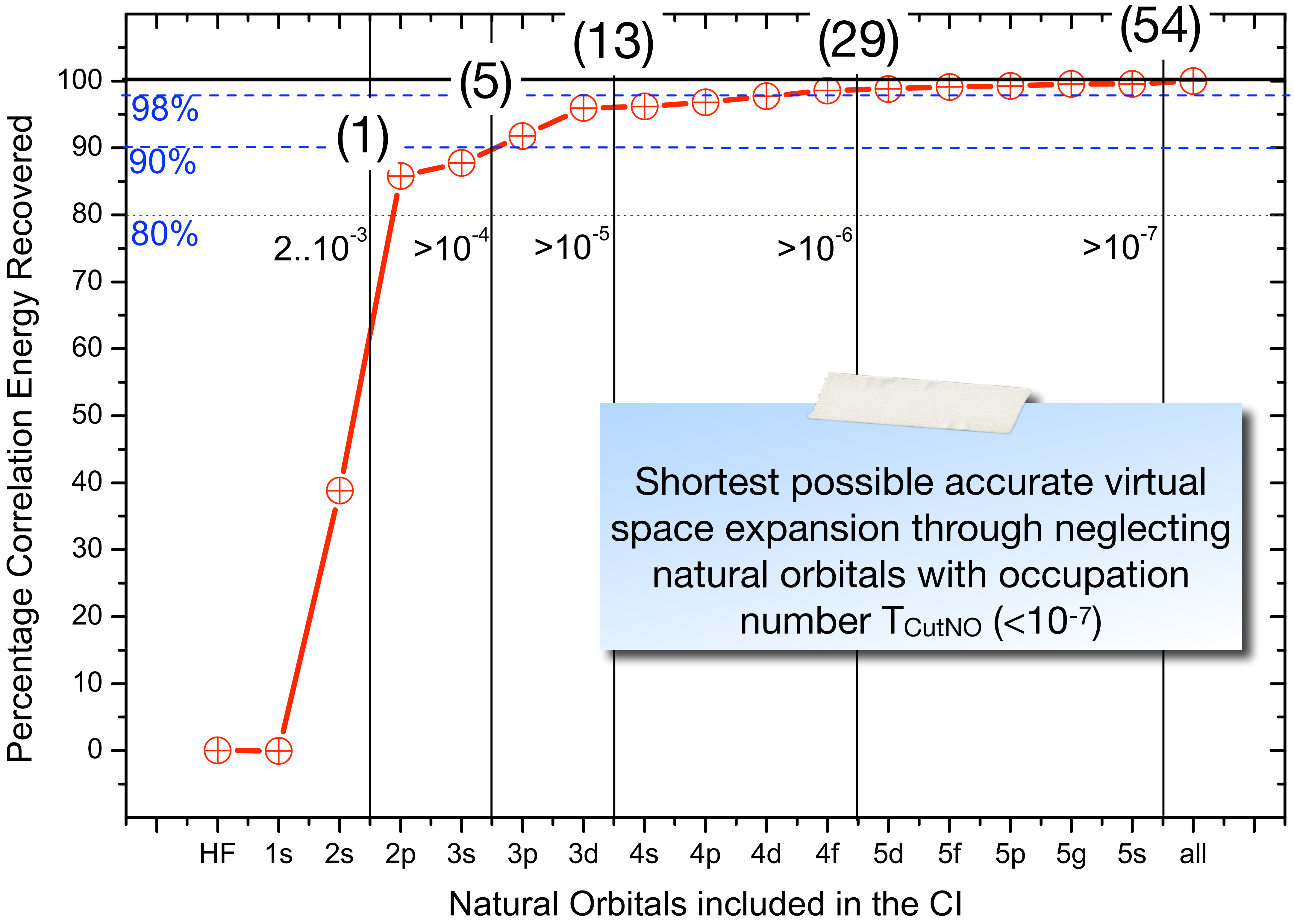
$$\Sigma_k\gamma_k^2 = \Sigma_k n_k^2 - \sum_{l\neq k}|\gamma_{kl}|^2 \leq \Sigma_k n_k^2,$$

with the final result

$$1-(1/N)\,\Sigma_k n_k^2 \leq 1-(1/N)\,\Sigma_k \gamma_k^2,$$

which proves our theorem. This means that the natural spin-orbitals are distinguished not only by having vanishing bond orders but also by giving the smallest number of essential charge orders possible. By investigating the quantity ϑ, one can therefore easily estimate how much improvement one can expect in the convergence of a given configurational interaction series by introducing the natural spin-orbitals.

The Natural Expansion of He



Shortest possible accurate virtual space expansion through neglecting natural orbitals with occupation number T_{CutNO} ($<10^{-7}$)

THE JOURNAL OF CHEMICAL PHYSICS VOLUME 45, NUMBER 5 1 SEPTEMBER 1966

Pseudonatural Orbitals as a Basis for the Superposition of Configurations. I. He_2^+

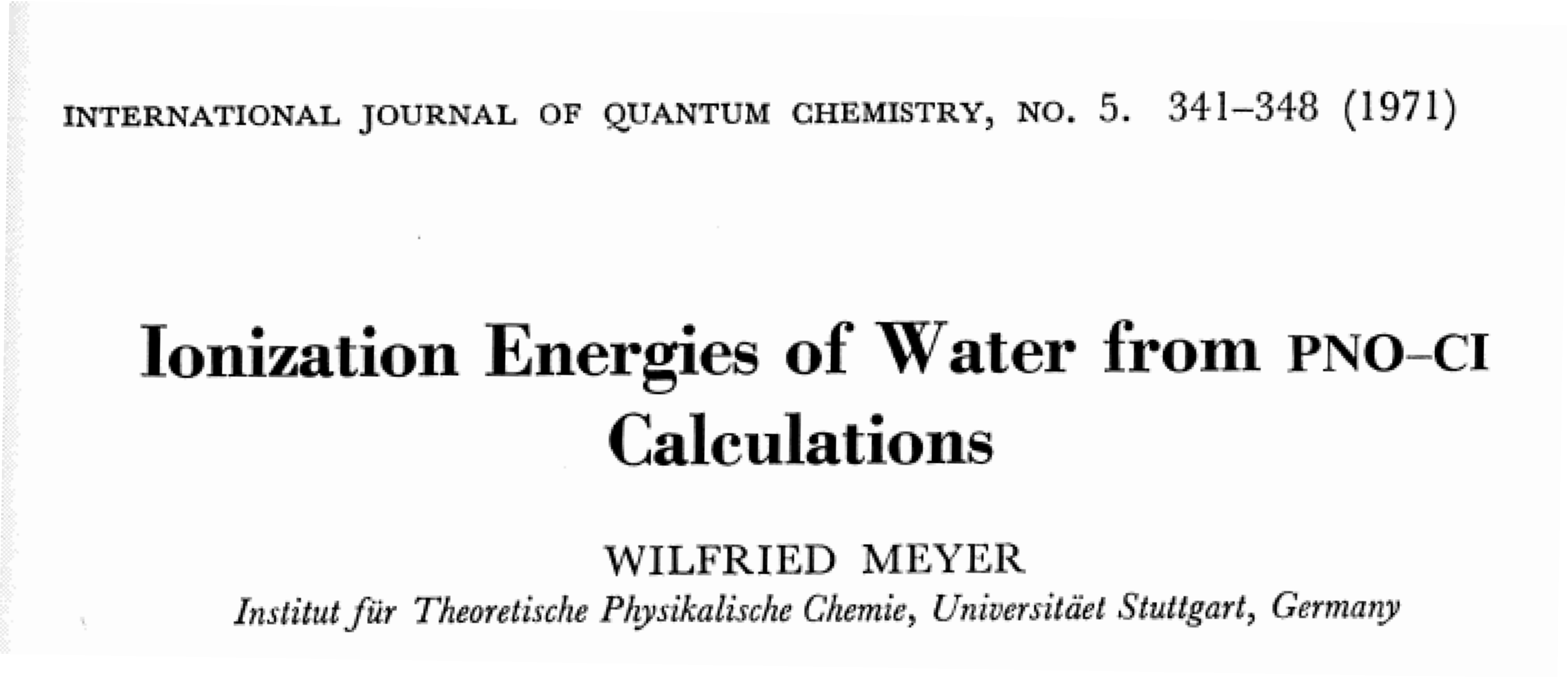
C. EDMISTON*
University of Wyoming, Laramie, Wyoming
AND
M. KRAUSS
National Bureau of Standards, Washington, D. C.

(Received 1 December 1965)

The use of pseudonatural orbitals (PNO) is proposed to improve the rate of convergence in the superposition of configurations (SOC). Natural orbitals are determined for selected electron pairs in the Hartree–Fock field of the $n-2$ electron core and are then used as the basis for the total SOC calculation. Since these natural orbitals are not natural for the n-electron system they are considered false or pseudonatural orbitals when used in the n-electron problem.

The PNO basis has been applied to He_2^+ and H_3 to test the convergence. Complete results are reported here only for He_2^+. The PNO's are quite successful in speeding up the convergence of the SOC and rendering the calculation of correlation energy quite practical in general. Gaussian-type orbitals (GTO) are used throughout and were not a serious impediment to obtaining quantitative accuracy. In fact the large number of unoccupied Hartree–Fock orbitals consequent upon the use of a GTO basis permit a straightforward determination of the PNO orbitals.

Early Applications of Pair Natural Orbitals

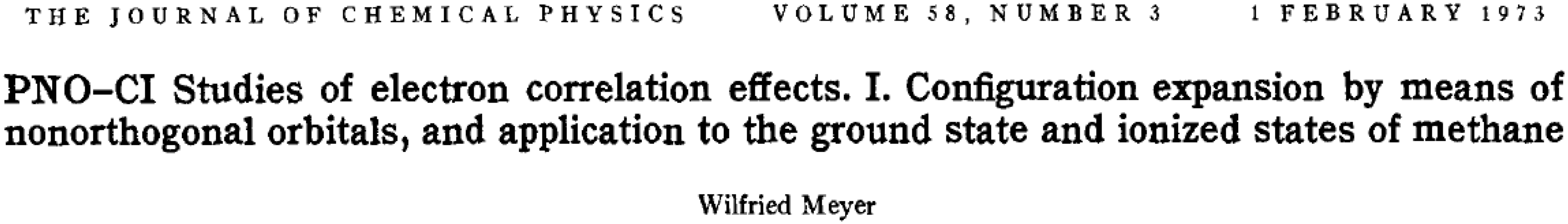


INTERNATIONAL JOURNAL OF QUANTUM CHEMISTRY, NO. 5. 341–348 (1971)

Ionization Energies of Water from PNO–CI Calculations

WILFRIED MEYER

Institut für Theoretische Physikalische Chemie, Universitäet Stuttgart, Germany



THE JOURNAL OF CHEMICAL PHYSICS VOLUME 58, NUMBER 3 1 FEBRUARY 1973

PNO–CI Studies of electron correlation effects. I. Configuration expansion by means of nonorthogonal orbitals, and application to the ground state and ionized states of methane

Wilfried Meyer

Pair Natural Orbitals and the Virtual Space

- ✓ Exact density $$D_{ab} = \sum_{P\equiv ij} D_{ab}^{(P)}$$

a,b= canonical virtual orbitals

$P \equiv (ij)$ Pair label

- ✓ Diagonalize each pair density individually

$$\mathbf{D}^{(P)}\mathbf{d}^{(P)} = \mathbf{n}^{(P)}\mathbf{d}^{(P)}$$

- ✓ PNO expansion $$\left|\tilde{a}_P\right\rangle = \sum_{a} d_{a\tilde{a}_P}^{(P)} \left|a\right\rangle$$

(exact without truncation)

  - PNOs are orthogonal to all occupied orbitals
  - PNOs of a given pair are orthonormal
  - PNOs of different pairs are not orthogonal

$$\left\langle \tilde{a}_P \mid \tilde{b}_Q \right\rangle \equiv S_{\tilde{a}_P,\tilde{b}_Q}^{(P,Q)} = \sum_{ab} d_{a\tilde{a}_P}^{P} d_{b\tilde{b}_Q}^{Q} \underbrace{\left\langle a \mid b \right\rangle}_{\delta_{ab}} = (\mathbf{d}^{P})^{T}\mathbf{d}^{Q}$$

  - PNOs based on local occupied orbitals are also local

Pair Natural Orbitals and the Virtual Space

✓ Truncation according to $n_a^P < T_{\mathrm{CutPNO}}$

- Asymptotically constant number of PNOs per pair
- Minimal error for a given expansion length

✓ Estimate of PNO error: $\Delta E_{\mathrm{PNO}} = \sum_P \varepsilon_P^{(\mathrm{full})} - \varepsilon_P^{(\mathrm{selected})}$

✓ After truncation, try to expand the PNOs of one pair in terms of another pair PNOs:

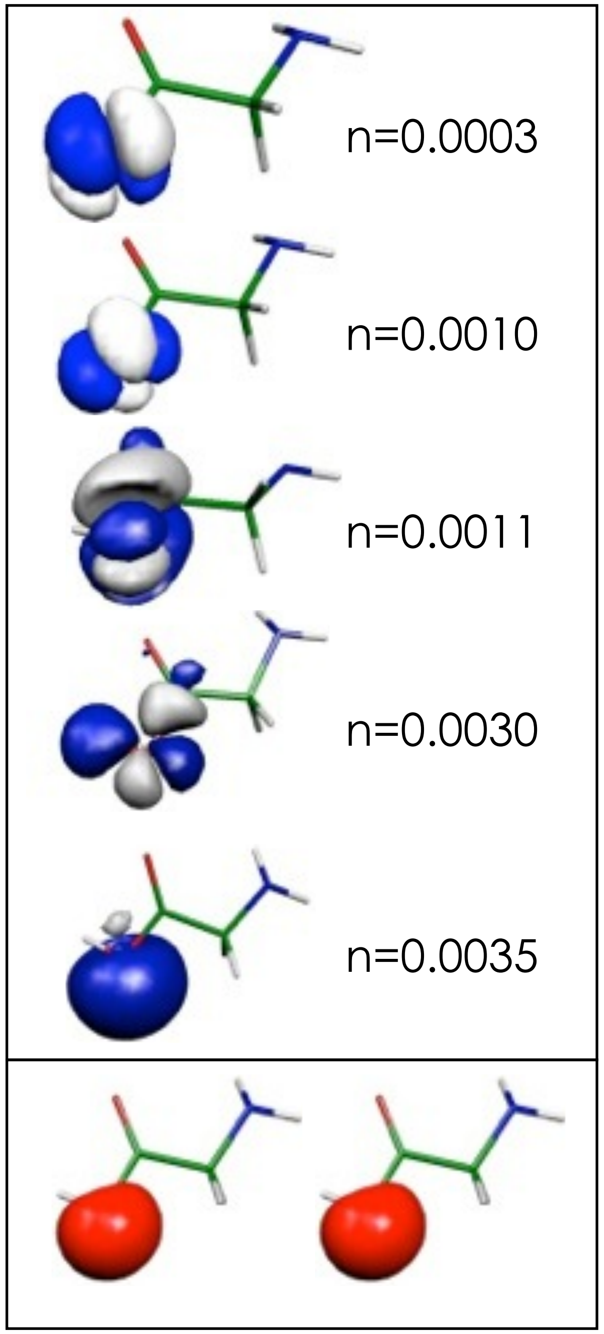
$$\left|a_P\right\rangle \approx \sum_{b_Q} c_{b_Q a_P} \left|b_Q\right\rangle$$

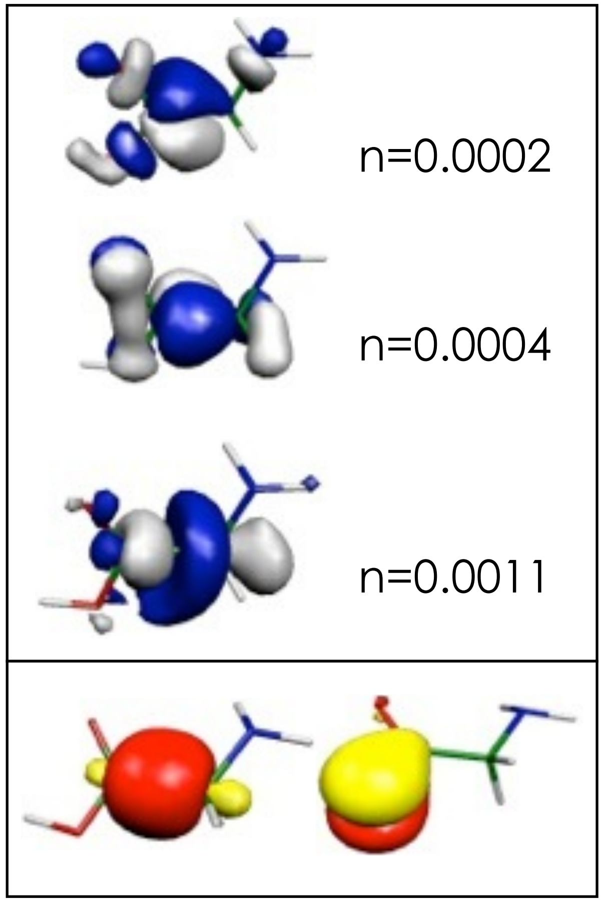
$$c_{b_Q a_P} = \left\langle b_Q \mid a_P \right\rangle = S^{(Q,P)}_{b_Q, a_P}$$

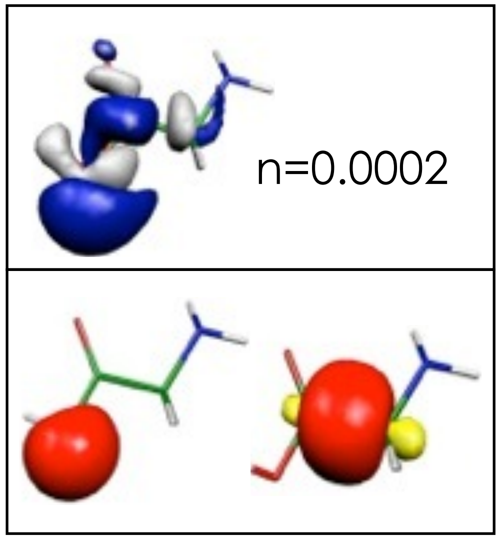
- Expansion is approximate for truncated PNO expansions! (it can be **very** bad)
- The overlap is a **projector**

Pair Natural Orbitals (PNOs)

- Small number of significant PNOs per electron pair
- Vanishing (0-5) PNOs for weak pairs
- Located in the same region of space as the internal pair but as delocalized as necessary
- Orthonormal within one pair, non-orthogonal between pairs







FN; Wennmohs, F.; Hansen, A. *J. Chem. Phys.* **2009**, 130, 114108

Obstacles

✓ At threshold 0, each pair dimension = full VMO	▸ Nightmare!
✓ Need to know the exact density to get PNOs	▸ Absurd!
✓ Many more PNOs than VMOs	▸ Integral generation is frightening

➡ PNO method that expand PNOs in virtual MOs are possible (and maybe sometimes desirable! e.g. first generation LPNO methods), but it only becomes efficient and linear scaling with further approximations

Domain Based **L**ocal **P**air **N**atural **O**rbital Methods

✓ Logical approximation: Expand the PNOs in terms of local virtual orbitals, e.g. PAO's taken from large pair-specific domains

$$\left|a_{ij}\right\rangle = \sum_{\tilde{\mu}_{ij}} d^{(ij)}_{\mu_{ij}a_{ij}} \left|\tilde{\mu}_{ij}\right\rangle$$

✓ The pair density is approximate and comes from (semi-canonical) local MP2:

$$\mathbf{D}^{(ij)} = \mathbf{T}^{(ij)}\mathbf{T}^{(ij)+} + \mathbf{T}^{(ij)+}\mathbf{T}^{(ij)}$$

$$T^{ij}_{\tilde{\mu}_{ij}\tilde{\nu}_{ij}} = -\frac{\left(i\tilde{\mu}_{ij} \mid j\tilde{\nu}_{ij}\right)}{F_{ii} - F_{jj} - \varepsilon_{\tilde{\mu}_{ij}} - \varepsilon_{\tilde{\nu}_{ij}}}$$

➡ The generated PNOs are rather approximate, but experience shows that only minor improvements are possible by making more elaborate choices; the domain approximation is is more problematic.

Local Correlation with PNOs: PNO-MP2

Rcall: In Nonorthogonal virtuals (exact equation!):

$$\mathbf{R}^{ij} = \mathbf{K}^{ij} + (\mathbf{F}\mathbf{C}^{ij}\tilde{\mathbf{S}} + \tilde{\mathbf{S}}\mathbf{C}^{ij}\mathbf{F}) - \sum_k (f_{ik}\tilde{\mathbf{S}}\mathbf{C}^{kj}\tilde{\mathbf{S}} + f_{kj}\tilde{\mathbf{S}}\mathbf{C}^{ik}\tilde{\mathbf{S}}) = 0 \qquad \tilde{S}_{\tilde{a}\tilde{b}} = \left\langle \tilde{a} \,|\, \tilde{b} \right\rangle$$

- Throwing out negligible electron pairs (ij) does not change the equations
- Introducing pair-specific truncated virtual spaces does change the notation:

$$\underset{\mathrm{N_{ij}xN_{ij}}}{\mathbf{R}^{ij}} = \underset{\mathrm{N_{ij}xN_{ij}}}{\mathbf{K}^{ij}} + (\underset{\mathrm{N_{ij}xN_{ij}}}{\mathbf{F}^{(ij)}}\underset{\mathrm{N_{ij}xN_{ij}}}{\mathbf{C}^{ij}} \underbrace{\tilde{\mathbf{S}}^{(ij,ij)}}_{\mathbf{1}\,in\,PNOs} + \underbrace{\tilde{\mathbf{S}}^{(ij,ij)}}_{\mathbf{1}\,in\,PNOs} \underset{\mathrm{N_{ij}xN_{ij}}}{\mathbf{C}^{ij}}\underset{\mathrm{N_{ij}xN_{ij}}}{\mathbf{F}^{(ij)}}) - \sum_k (f_{ik}\underset{\mathrm{N_{ij}xN_{kj}}}{\tilde{\mathbf{S}}^{(ij,jk)}}\overset{\mathrm{N_{kj}xN_{kj}}}{\mathbf{C}^{kj}}\underset{\mathrm{N_{kj}xN_{ij}}}{\tilde{\mathbf{S}}^{(kj,ij)}} + f_{kj}\underset{\mathrm{N_{ij}xN_{ik}}}{\tilde{\mathbf{S}}^{(ij,ik)}}\overset{\mathrm{N_{ik}xN_{ik}}}{\mathbf{C}^{ik}}\underset{\mathrm{N_{ik}xN_{ij}}}{\tilde{\mathbf{S}}^{(ik,ij)}}) =$$

$$\mathrm{S}_{\mathrm{a,b}}^{(\mathrm{P,Q})} = \left\langle \mathrm{a_P} \,|\, \mathrm{b_Q} \right\rangle$$

Thus there is an emerging **„cooking recipe“**:

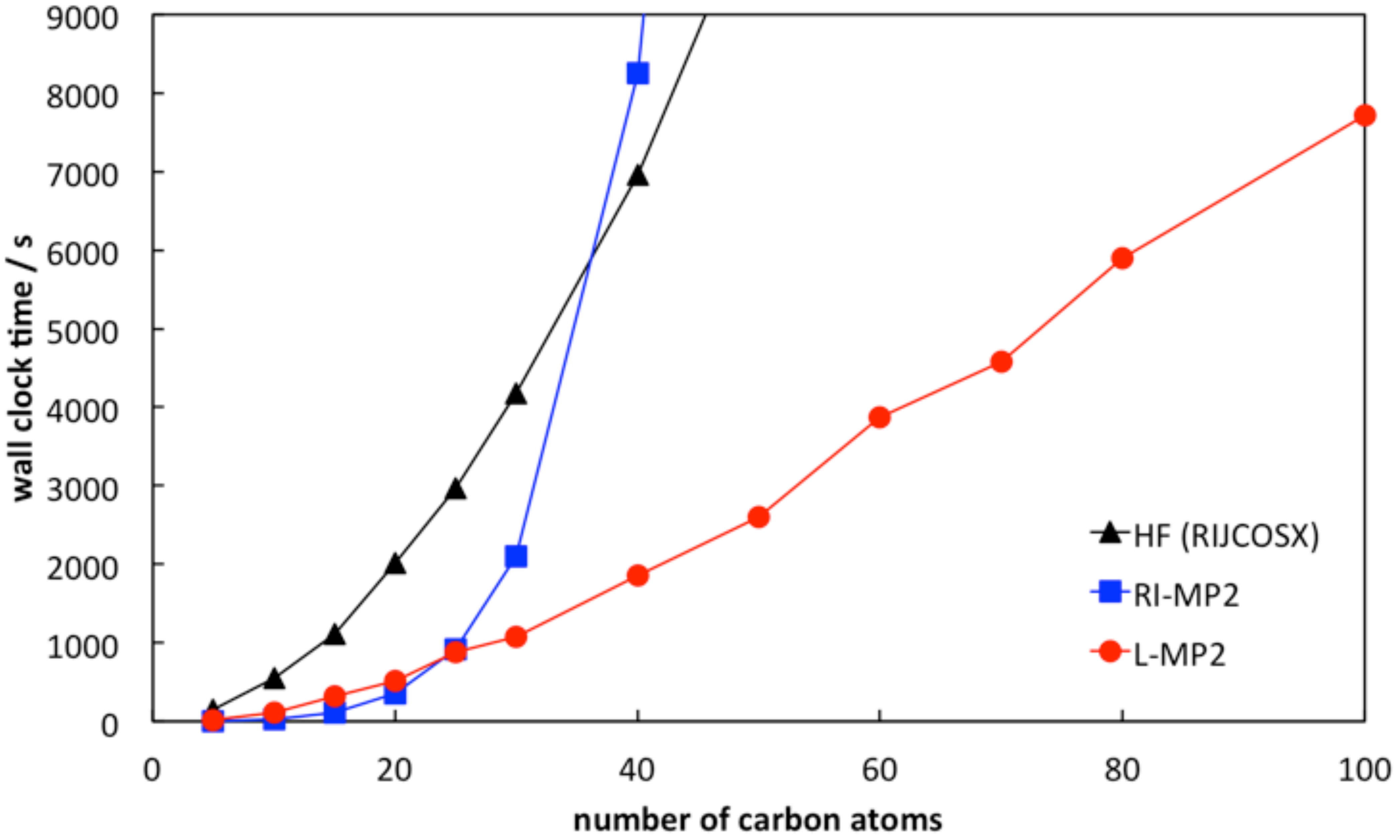
- Whenever two pairs are „connected“ there will be mismatched dimensions on the matrices to be multiplied: insert the pair-pair overlap in these terms.
- In reality it is not quite that simple …

P. Pulay, and S. Saebø, *Theor. Chim. Acta* 69, 357 (**1986**).

Steps in DLPNO-MP2

- ✓ Perform **HF** calculation
- ✓ **Localize** occupied MOs – separately for core and valence
- ✓ Construct normalized redundant **PAOs**
- ✓ Select PAO **domains** based on DOI (T_{CutDO})
- ✓ **Screen** *ij*-pairs based on a dipole approximation of the pair energy
- ✓ For every pair domain, construct **quasi-canonical non-redundant PAOs**
- ✓ Calculate the **semi-canonical amplitudes**
- ✓ **Diagonalize** the pair density
- ✓ **Keep PNO**s with occupation > T_{CutPNO}
- ✓ **Solve** MP2 residual equations in PNO basis

DLPNO-MP2: Efficiency and Scaling



- **ALWAYS faster than even accelerated Hartree-Fock**
- **Early crossover with the canonical RI-MP2 method!**

P Pinski, C Riplinger, E Valeev, FN, *J Chem Phys.* **2015,** 143, 034108

(Domain Based Local)
Pair Natural Orbital Coupled Cluster Theory

The DLPNO-CCSD Approach

The DLPNO approach to CCSD follows as a natural extension to MP2 (although historically, DLPNO-CCSD was much earlier than DLPNO-MP2)

The **cluster operator** is written in the PNO basis:

$$T_1 + T_2 = \sum_{ia} t^{i}_{\tilde{a}_i} a^{+}_{\tilde{a}_i} a_i + \frac{1}{4} \sum_{ijab} t^{ij}_{\tilde{a}_{ij}\tilde{b}_{ij}} a^{+}_{\tilde{b}_{ij}} a^{+}_{\tilde{a}_{ij}} a_j a_i$$

- *PNO's for doubles are made from DLPNO-MP2 and cut with* T_{CutPNO}
- *PNO's for singles are identical to the PNOs of the diagonal pairs and cut with* $0.01 T_{CutPNO}$

The **PNOs** are expanded in large PAO domains (using $\mathrm{T_{CutDO}}$) for domain construction and $\mathrm{T_{CutMKN}}$ for aux-domain construction)

$$\left| \tilde{a}_{ij} \right\rangle = \sum_{\tilde{\mu} \in \{ij\}} d^{ij}_{\tilde{\mu}\tilde{a}} \left| \tilde{\mu} \right\rangle$$

Natural Triple Excitations

Our suggestion: **Natural triples orbitals** (TNO's)

✓ Three-pair density: $\mathbf{D}^{ijk} = \frac{1}{3}(\mathbf{D}^{ij} + \mathbf{D}^{ik} + \mathbf{D}^{jk})$

(The operator $\hat{D}^{ijk} = \sum_{\bar{a}_{ij}} \left|\bar{a}_{ij}\right\rangle\left\langle\bar{a}_{ij}\right| + \sum_{\bar{a}_{ik}} \left|\bar{a}_{ik}\right\rangle\left\langle\bar{a}_{ik}\right| + \sum_{\bar{a}_{jk}} \left|\bar{a}_{jk}\right\rangle\left\langle\bar{a}_{jk}\right|$ *projects onto the joint PNO space of the three pairs)*

✓ Formation of the three pair density in the **PAO basis** is linear scaling:

✓ Eigenfunctions: $\mathbf{D}^{ijk}\mathbf{x}^{ijk} = n^{ijk}\mathbf{x}^{ijk}$ (cut-off below a given n^{ijk}(min) just as for PNOs)

✓ Recanonicalize: $\mathbf{x}^{ijk+}\mathbf{F}\mathbf{x}^{ijk}$

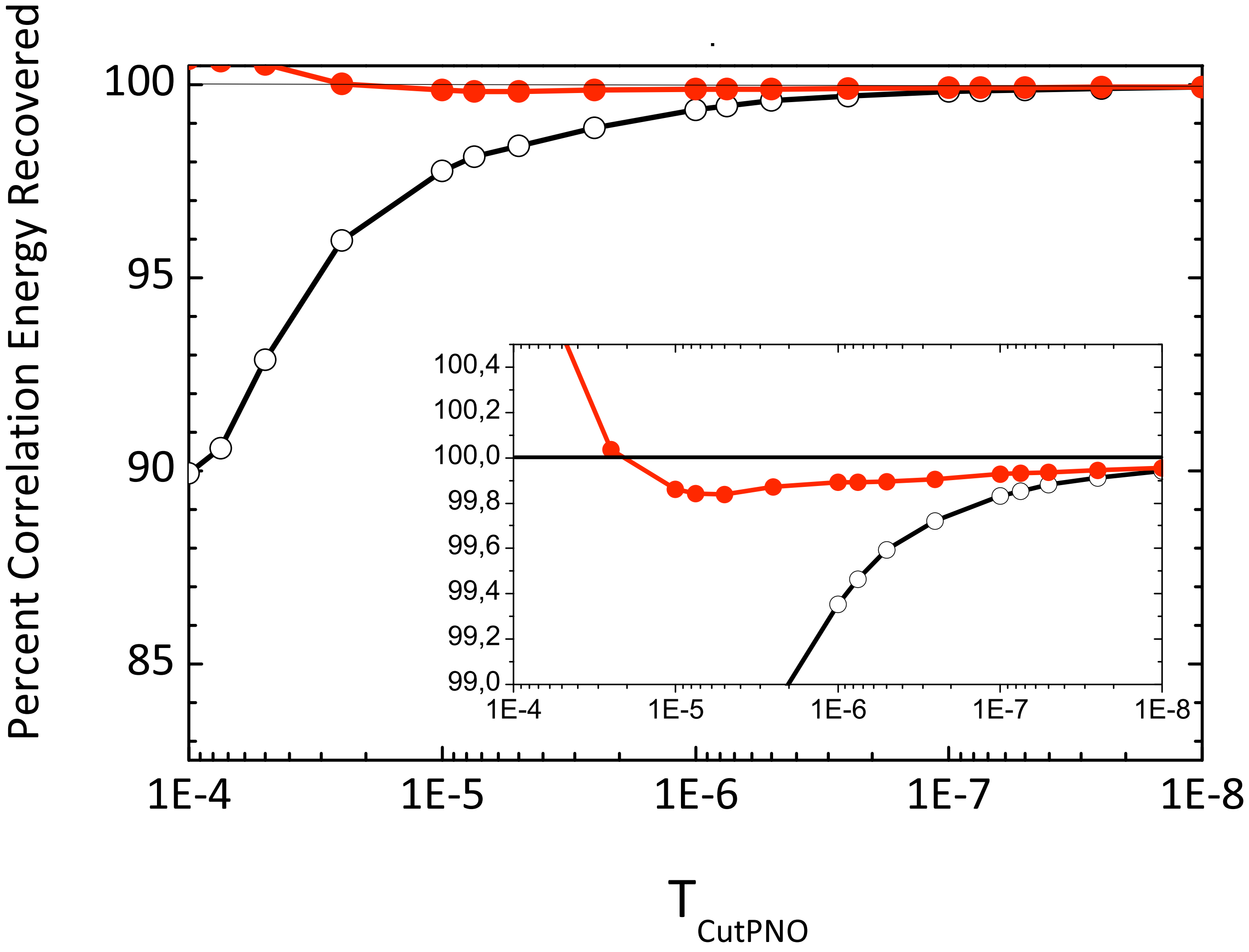
✓ Amplitudes are projected into the TNO basis: $T^{ij;TNO}_{\tilde{a}_{ijk},\tilde{b}_{ijk}} = S^{ijk,ij}_{\tilde{a}_{ijk},\tilde{c}_{ij}} T^{ij;PNO}_{\tilde{c}_{ij},\tilde{d}_{ij}} S^{ijk,ij}_{\tilde{b}_{ijk},\tilde{d}_{ij}}$

➡ Integrals over TNOs must be generated for each triple (bookkeeping complicated but linear scaling) *avoiding projection*

➡ *Linear scaling implementation achieved (Dr. Christoph Riplinger)*

Riplinger, Sandhöfer, Hansen, FN, JCP, **2013**,139(13):134101

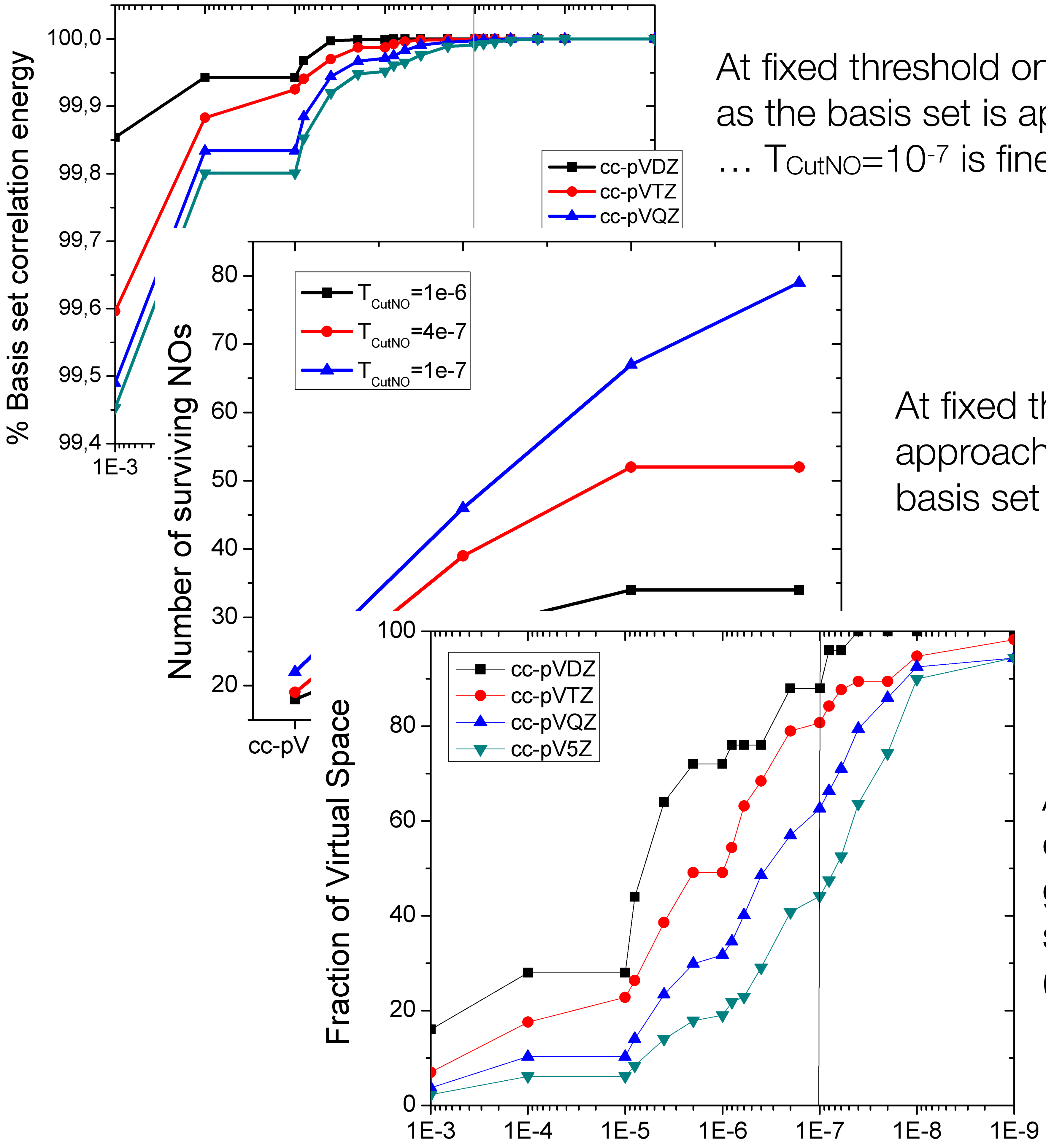
Convergence of DLPNO-CCSD(T)



- ✓ Typically 99.8-99.9% of the canonical correlation energy
- ✓ Energetics of the canonical counterpart methods is reproduced to a few tenth of kcal/mol. Maximum achievable accuracy ~0.1 kcal/mol of the canonical result.
- ✓ The methods are robust and completely black box in character

Riplinger, C. FN *J. Chem. Phys*, **2013**, 138, 034106; FN; A. Hansen, D.G. Liakos,, *J. Chem. Phys.*, **2009** 131, 064103

Li-Dimer @ 1.5 Angström



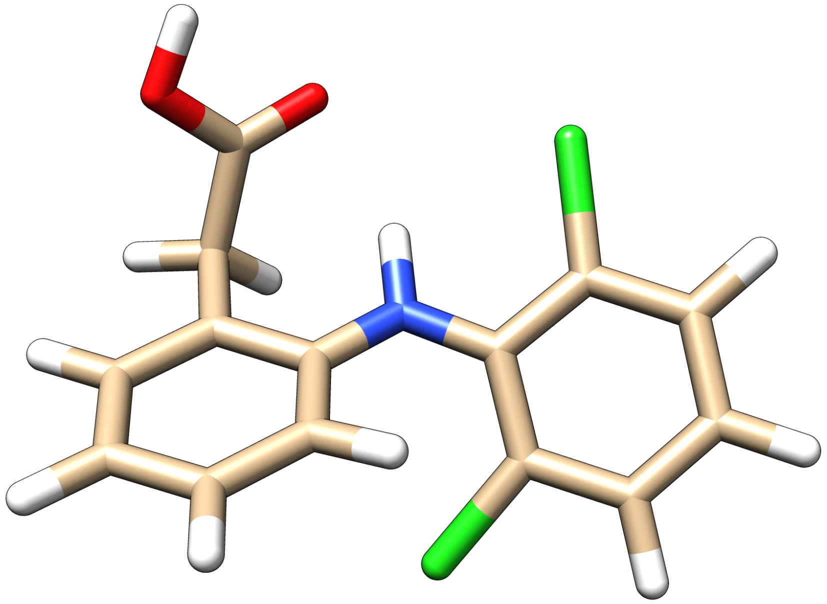
At fixed threshold one recovers less E_C as the basis set is approaching completeness … $T_{CutNO}=10^{-7}$ is fine even for large bases

At fixed threshold the number of NOs *is* approaching **a constant (!)** as the the basis set saturates

At fixed threshold the fraction of the virtual space treated gets smaller as the basis set saturates
(cost reduction ~ (fraction)^4)

Real Life Basis Set Behavior of DLPNO-CCSD(T)

In real life convergence of the PNO expansion is *more* favorable than for, say He, since weakly interacting electron pairs saturate more quickly with basis set!



Diclophenac

(30 atoms)

	N_{Bas}	N_{PNO}^{av}	**Fraction**	**Time(s)**	
cc-pVDZ	329	19	5.7	723	
cc-pVTZ	723	27	3.7	3782	(5.2x)
cc-pVQZ	1383	32	2.3	13676	(3.6x)
cc-pV5Z	3243	36	1.1	46675	(3.4x)
cc-pV6Z	3669	39	1.0	144141	(3.0x)

=1.6 days/16 cores

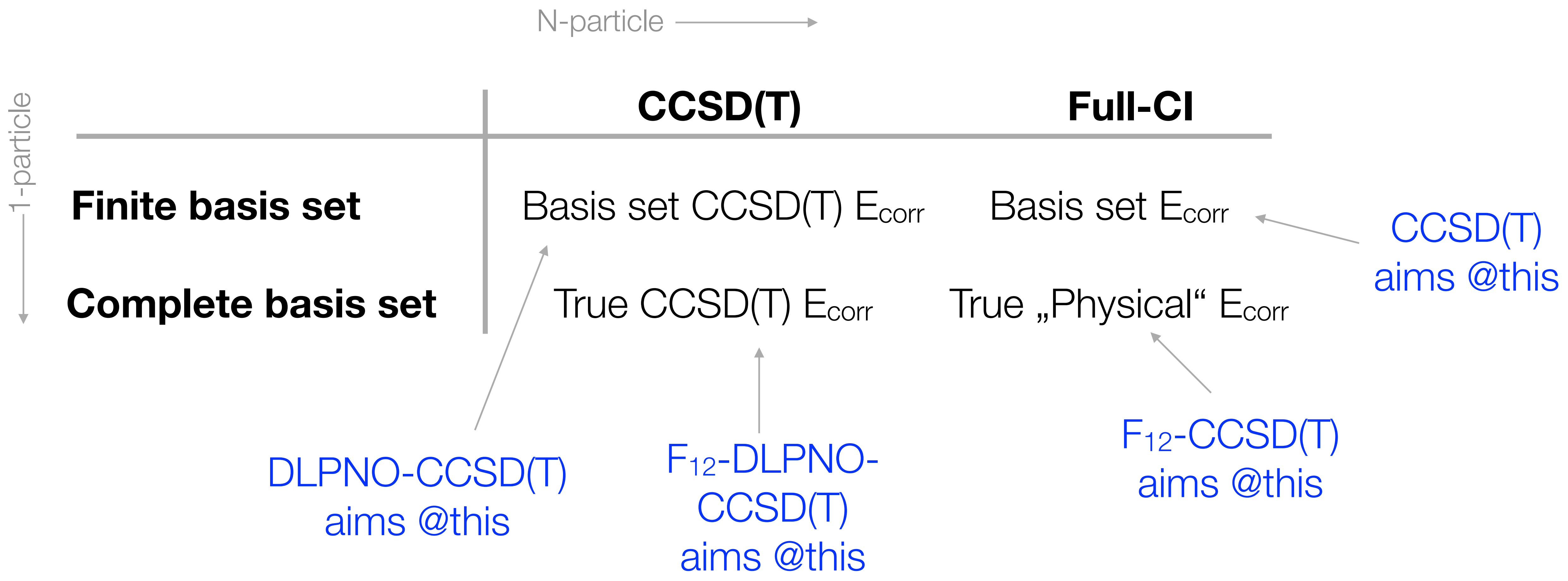
↓

Savings ~10^8!

Increase of computer time with cardinal number: DLPNO-CCSD(T) **~factor 3-4**

Canonical CCSD(T) **~factor 10-100**

PNO Truncation, Basis Sets, Correlation Energy



	CCSD(T)	**Full-CI**
Finite basis set	Basis set CCSD(T) E_{corr}	Basis set E_{corr}
Complete basis set	True CCSD(T) E_{corr}	True „Physical“ E_{corr}

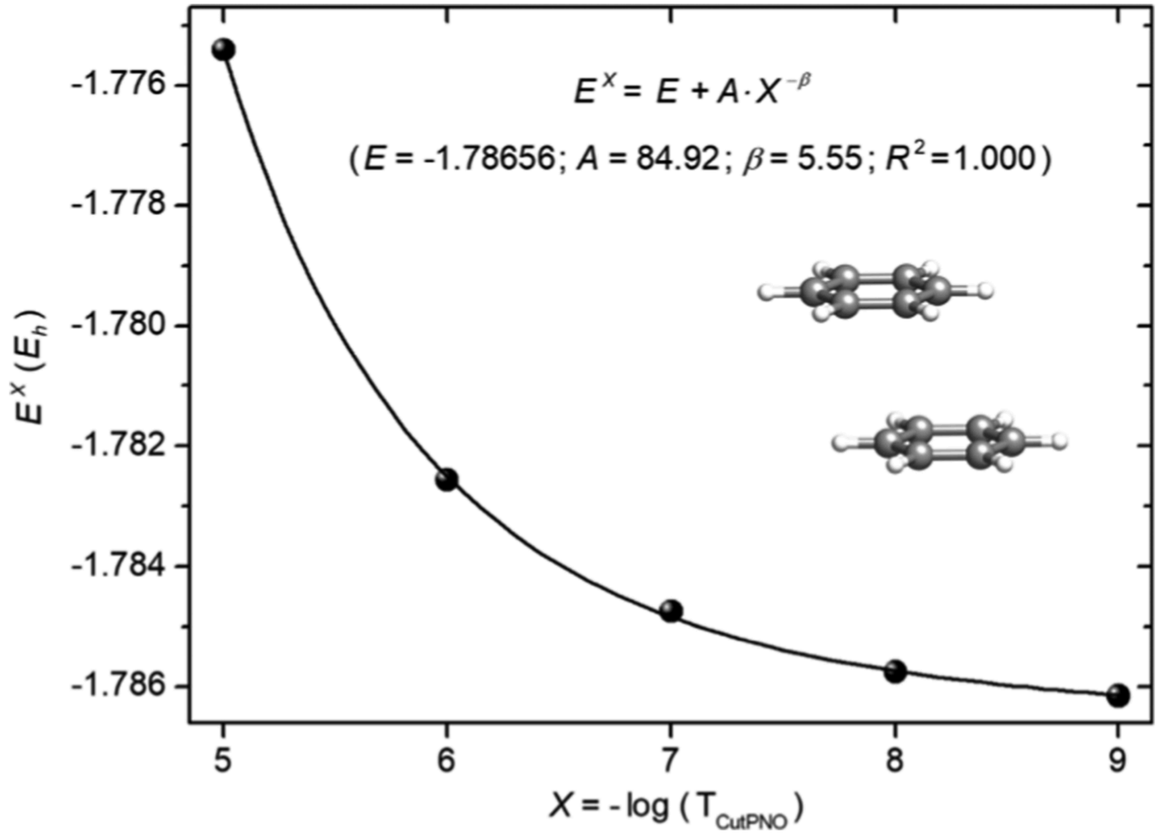
- ➡ **One should only judge DLPNO-CCSD(T) relative to how well it approximates CCSD(T) energy in the same basis set!**
- ➡ **Extrapolation to the 1-particle (basis set) limit is something separate**
- ➡ **Extrapolation to the N-particle (full CI) limit is something separate**

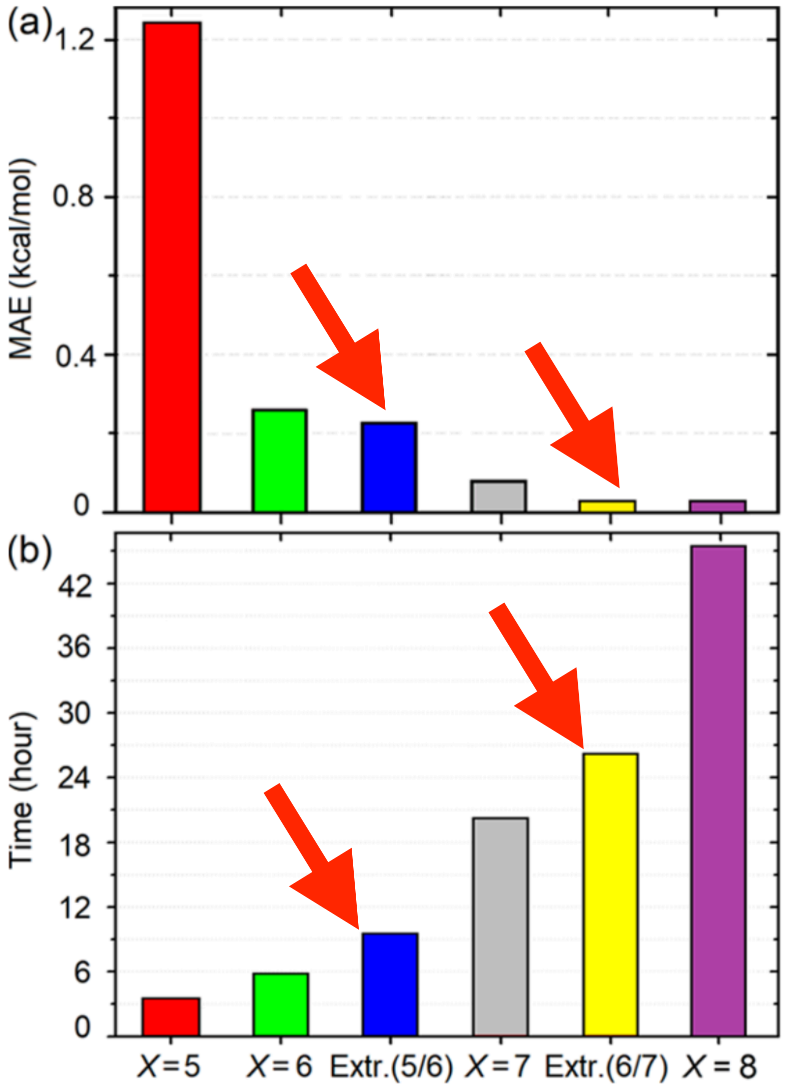
PNO Extrapolation

One can extrapolate the PNO energy towards 0 threshold:

$$E(T=0) \approx E(T=10^{-X}) + F\left(E(T=10^{-Y}) - E(T=10^{-X})\right)$$

$$T = PNO\ truncation\ threshold \qquad F = 1.5 \qquad Y > X$$





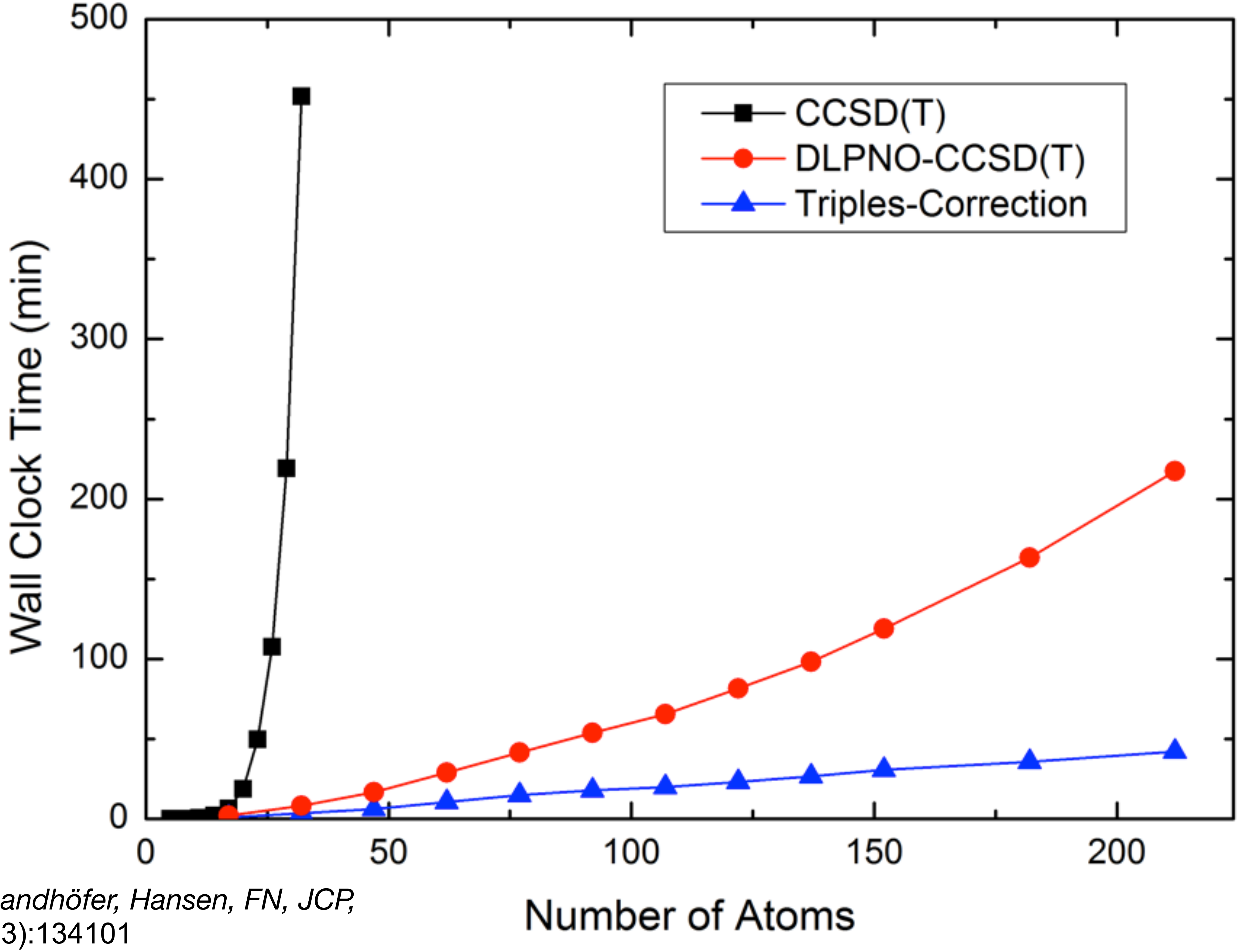
S66 set
TightPNO/
CBS

Note:

$$E^{(\infty)} = \frac{e^{\alpha\sqrt{Y}}E^{(Y)} - e^{\alpha\sqrt{X}}E^{(X)}}{e^{\alpha\sqrt{Y}} - e^{\alpha\sqrt{X}}} \longrightarrow F = \frac{e^{\alpha\sqrt{Y}}}{e^{\alpha\sqrt{Y}} - e^{\alpha\sqrt{X}}}$$

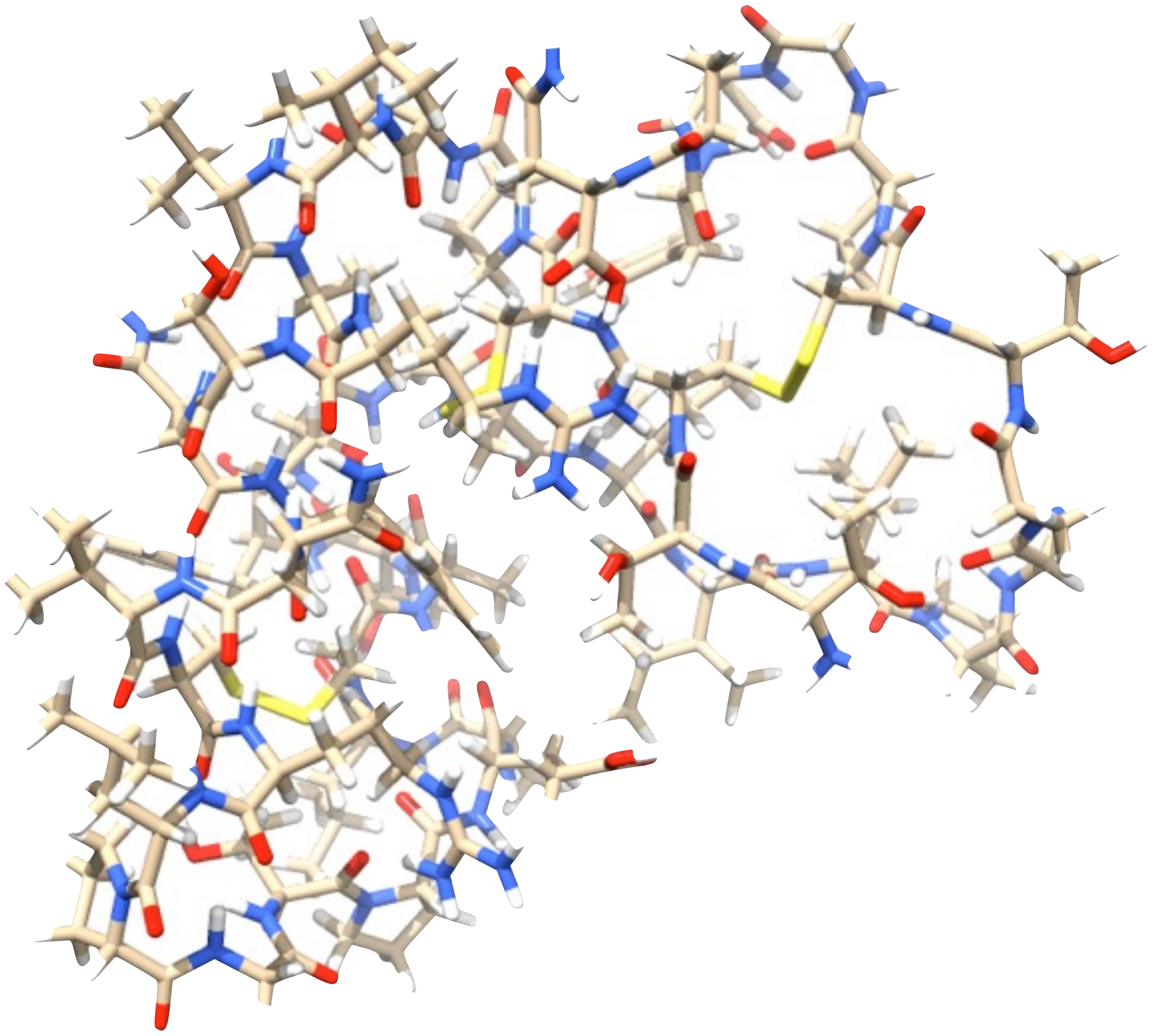
Altun, A.; Neese, F.; Bistoni, G. *J. Chem. Theo. Comp.* **2020**, *16*, 6142-6149.

Scaling of DLPNO-CCSD(T)



Riplinger, Sandhöfer, Hansen, FN, JCP,
2013,139(13):134101

CCSD(T) Calculations on Entire Proteins



Crambin

644 atoms
def2-SV(P)/6187 basis functions

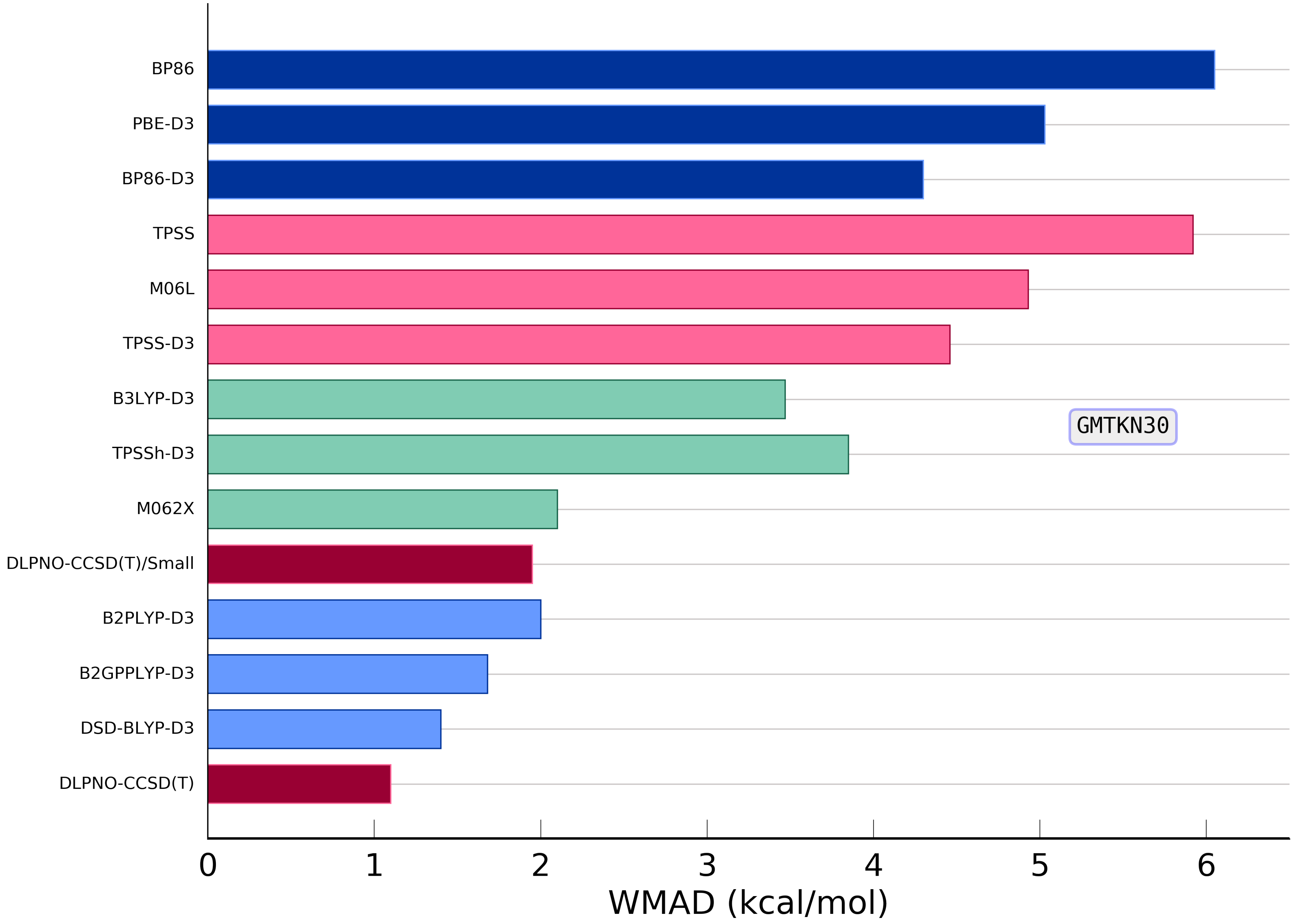
Canonical computation time
~5 Million Years

DLPNO-CCSD(T)
~3 weeks/1 Core

Riplinger, Sandhöfer, Hansen, FN, JCP, **2013**,139(13):134101

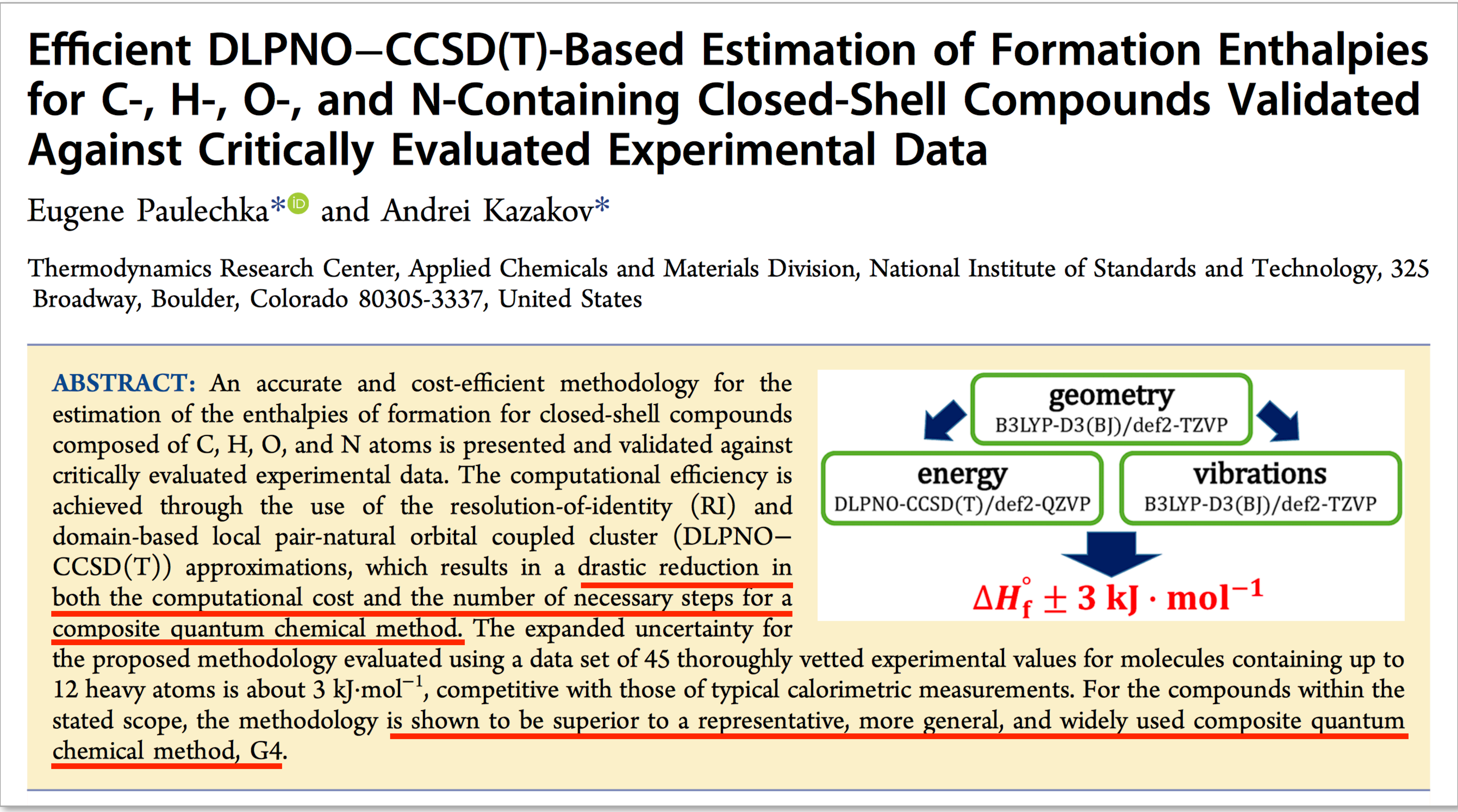
http://www.physicstoday.org/daily_edition/physics_update/coupled_cluster_theory_tackles_a_protein

Benchmark Results for DLPNO-CCSD(T)



L. Goerigk and S. Grimme, J. Chem. Theory Comput. **2011**, 7, 291-309

High Accuracy Thermochemistry



Efficient DLPNO–CCSD(T)-Based Estimation of Formation Enthalpies for C-, H-, O-, and N-Containing Closed-Shell Compounds Validated Against Critically Evaluated Experimental Data

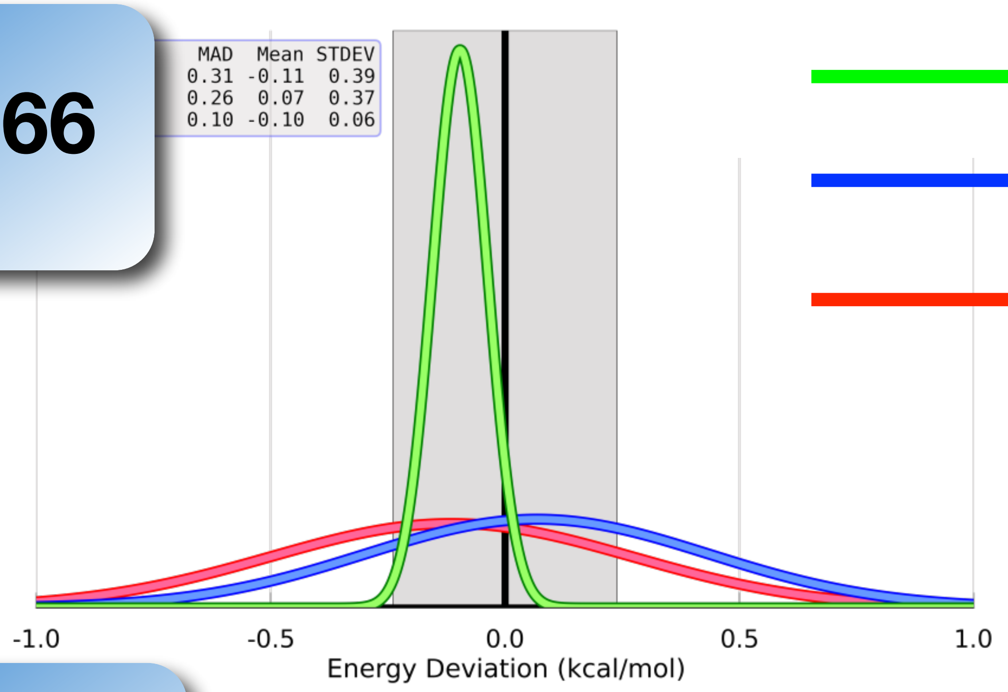
Eugene Paulechka* and Andrei Kazakov*

Thermodynamics Research Center, Applied Chemicals and Materials Division, National Institute of Standards and Technology, 325 Broadway, Boulder, Colorado 80305-3337, United States

ABSTRACT: An accurate and cost-efficient methodology for the estimation of the enthalpies of formation for closed-shell compounds composed of C, H, O, and N atoms is presented and validated against critically evaluated experimental data. The computational efficiency is achieved through the use of the resolution-of-identity (RI) and domain-based local pair-natural orbital coupled cluster (DLPNO–CCSD(T)) approximations, which results in a drastic reduction in both the computational cost and the number of necessary steps for a composite quantum chemical method. The expanded uncertainty for the proposed methodology evaluated using a data set of 45 thoroughly vetted experimental values for molecules containing up to 12 heavy atoms is about 3 kJ·mol^{-1}, competitive with those of typical calorimetric measurements. For the compounds within the stated scope, the methodology is shown to be superior to a representative, more general, and widely used composite quantum chemical method, G4.

How accurate can you get?

S66



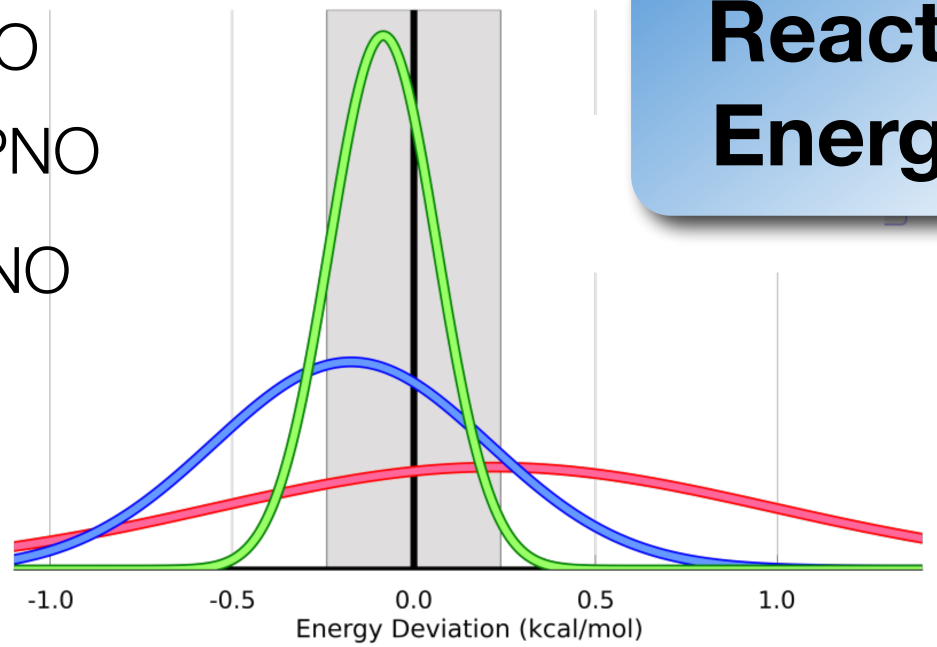
MAD	Mean	STDEV
0.31	-0.11	0.39
0.26	0.07	0.37
0.10	-0.10	0.06

🟩 TightPNO

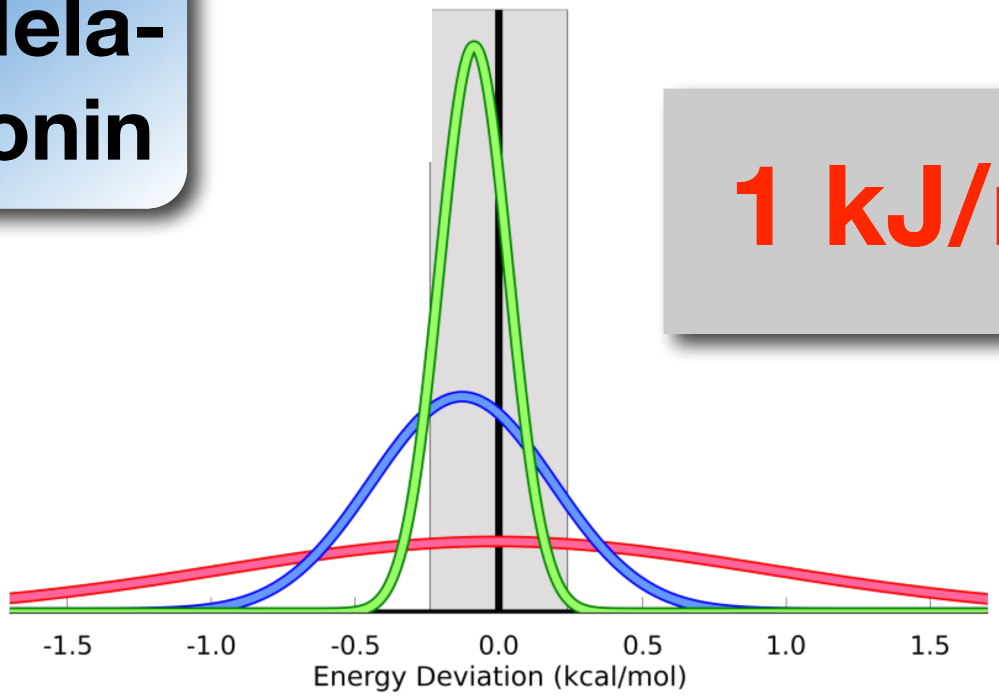
🟦 NormalPNO

🟥 LoosePNO

Reaction Energies

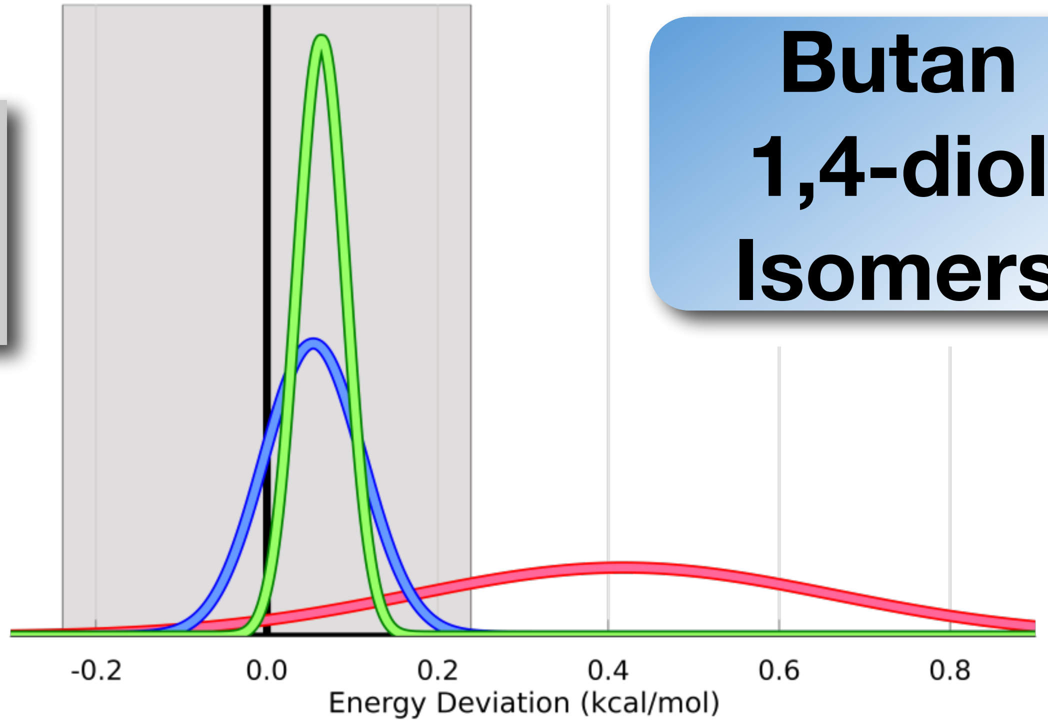


Mela-tonin



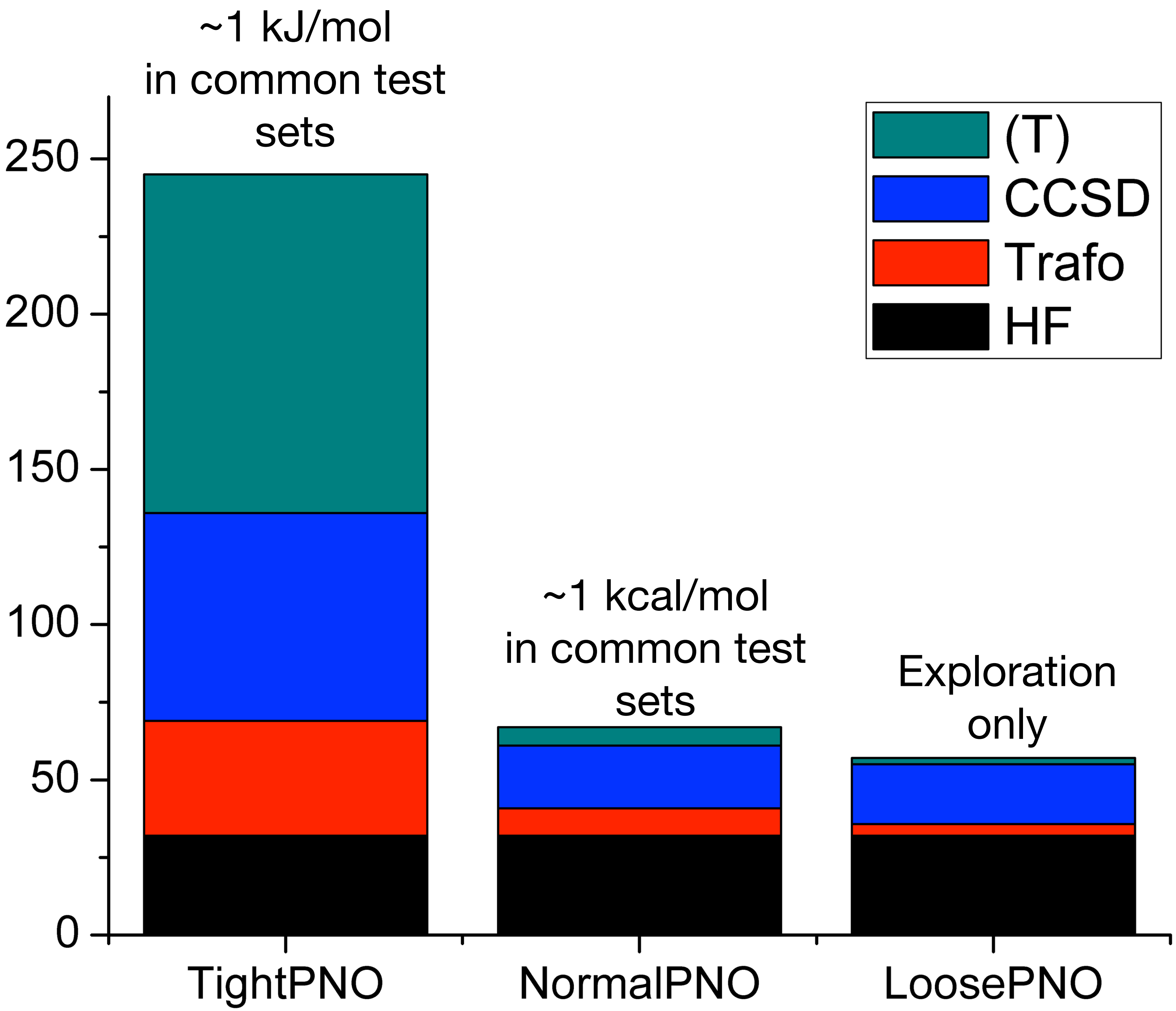
1 kJ/mol

Butan 1,4-diol Isomers



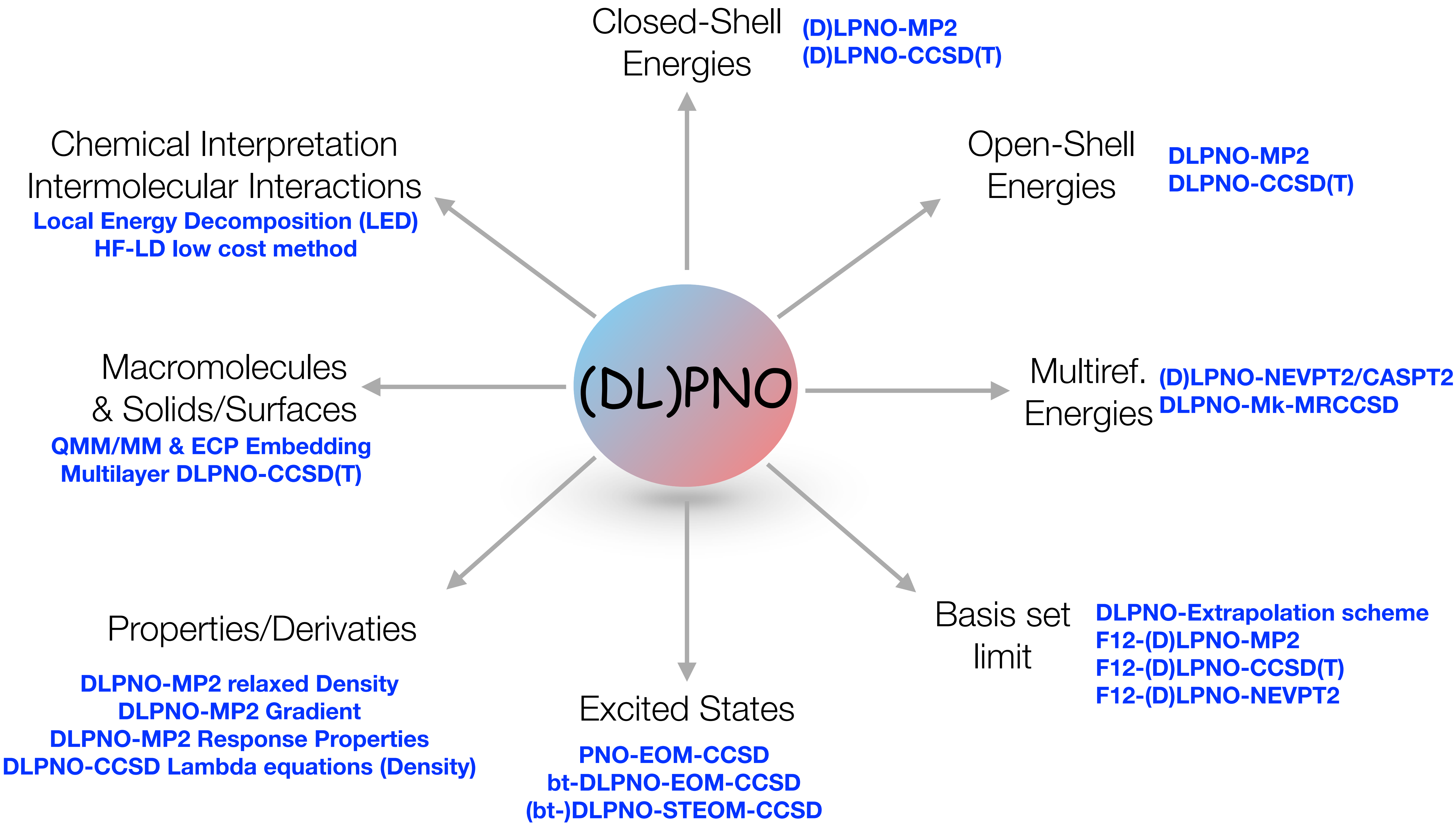
D.G Liakos, M. Sparta, J.M.L. Martin and F. Neese *JCTC, asap*

Cost *vs* Accuracy



D.G Liakos, M. Sparta, J.M.L. Martin and F. Neese *JCTC, **2015**,* 11, 1525

Beyond Closed Shell Energies



(DL)PNO

Closed-Shell Energies

(D)LPNO-MP2
(D)LPNO-CCSD(T)

Open-Shell Energies

DLPNO-MP2
DLPNO-CCSD(T)

Multiref. Energies

(D)LPNO-NEVPT2/CASPT2
DLPNO-Mk-MRCCSD

Basis set limit

DLPNO-Extrapolation scheme
F12-(D)LPNO-MP2
F12-(D)LPNO-CCSD(T)
F12-(D)LPNO-NEVPT2

Excited States

PNO-EOM-CCSD
bt-DLPNO-EOM-CCSD
(bt-)DLPNO-STEOM-CCSD

Properties/Derivaties

DLPNO-MP2 relaxed Density
DLPNO-MP2 Gradient
DLPNO-MP2 Response Properties
DLPNO-CCSD Lambda equations (Density)

Macromolecules & Solids/Surfaces

QMM/MM & ECP Embedding
Multilayer DLPNO-CCSD(T)

Chemical Interpretation
Intermolecular Interactions

Local Energy Decomposition (LED)
HF-LD low cost method

PNO based coupled cluster methods: Summary

(1) **(relative) Simplicity.** Only one critical cut-off (T_{CutPNO}); local approximations only ‚boost‘ efficiency. T_{CutPNO} can be use to control the absolute desired accuracy

(2) **No real-space cut-offs** and **no fragmentation** necessary

(3) **No redundant integral generation** or amplitude optimizations

(4) **No *reliance* on sparsity** (e.g. not linear scaling ‚by construction‘)

(5) Optimal correlation spaces: a) small for weak pairs, b) as delocalized as necessary

(6) **Excellent behavior with basis set size**

(7) Only local method with **proven accuracy** (better than 1 kcal/mol) and **proven efficiency** (approaching SCF/DFT times) for real life applications.

(8) **Very weak or no dependence on the localization method.** Well localized internal space not even required

(9) **Very smooth error**; no kinks and jumps in PESs

(10) **Black box character**

(11) Meanwhile developed for open shells, excited states, properties, F12, multireference, (gradients),…

Does Local Correlation „Solve all Problems“?

NO

1. Local correlation- by design - will fail in almost exactly the same way as canonical CCSD(T) does. Hence, it is accurate in the **regime of applicability of CC theory**, not beyond. However, **multi-reference local methods are as successful.**

2. There is a **residual error** that is **size intensive**. It is typically **<1 kcal/mol** (*relative to canonical CCSD(T)*) However, that is **still not zero** (*relative to basis set limit full CI*) and to ‚hammer‘ the total energy down to even higher accuracy is difficult without drastically slowing things down. (Hence, CCSDT and beyond may not be attractive)

3. One ‚only‘ gets an accurate **electronic energy**. For **zero-point**, **thermal** and **entropy** corrections as well as **solvation** or geometry errors one is stuck with the same errors as before - this is actually now the limiting factor for applications!