Open-Source DLPNO-CCSD(T) in Psi4

By: Andy Jiang

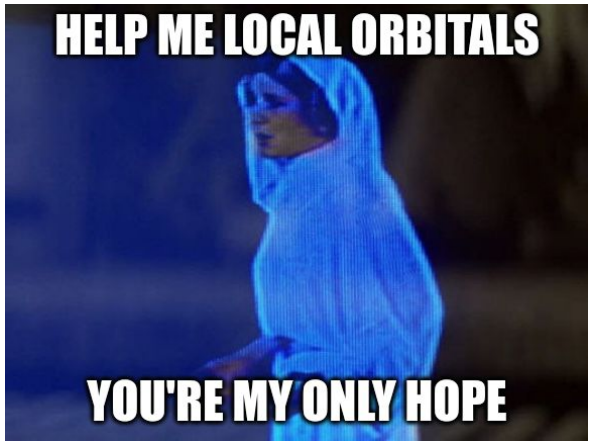
Coupled-Cluster Theory

- One of the most popular *ab-initio* quantum chemistry methods due to its **size extensivity** and **accuracy** (**< 1 kcal/mol deviation** from experiment with **CCSD(T)**)
- Unfortunately, the **cost** of coupled cluster is **high** because **canonical molecular orbitals** are inherently **delocalized** [***O(N^7)*** with **CCSD(T)**]
- Limits **system size** to around **30 atoms** on typical lab workstation

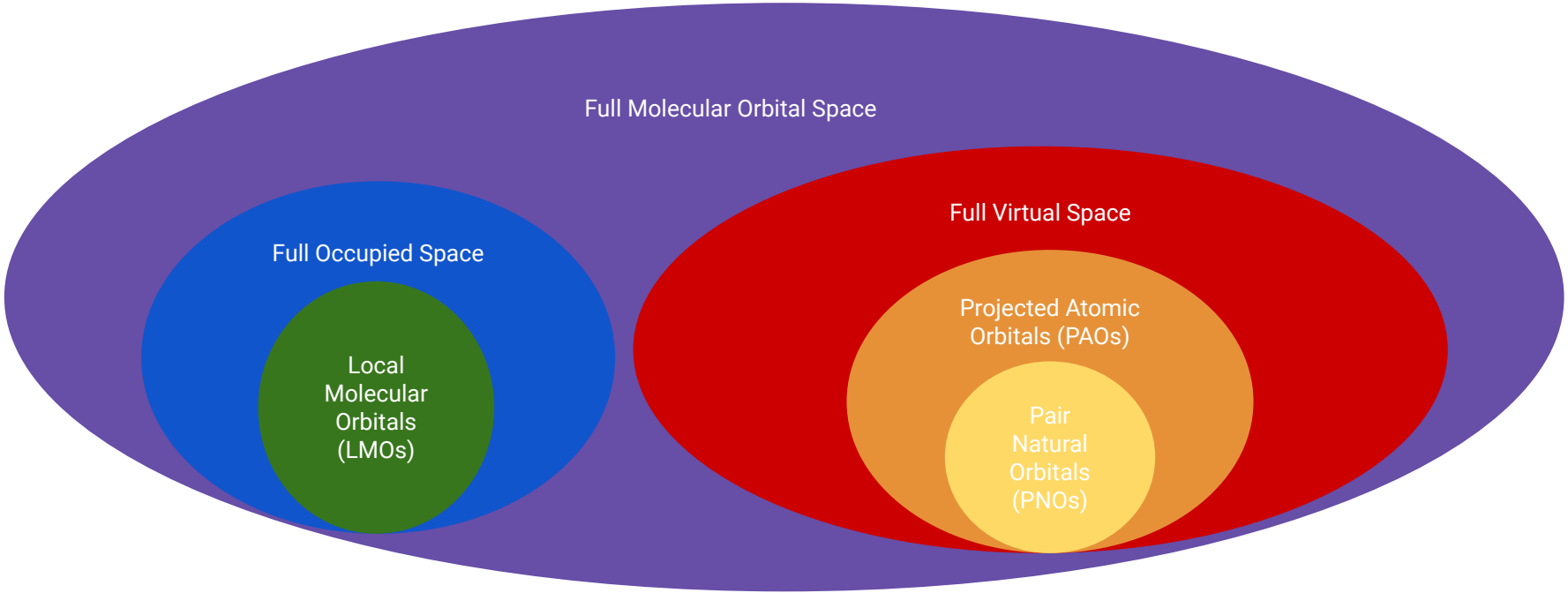
$$|\Psi\rangle = e^{T}|\Psi_0\rangle$$

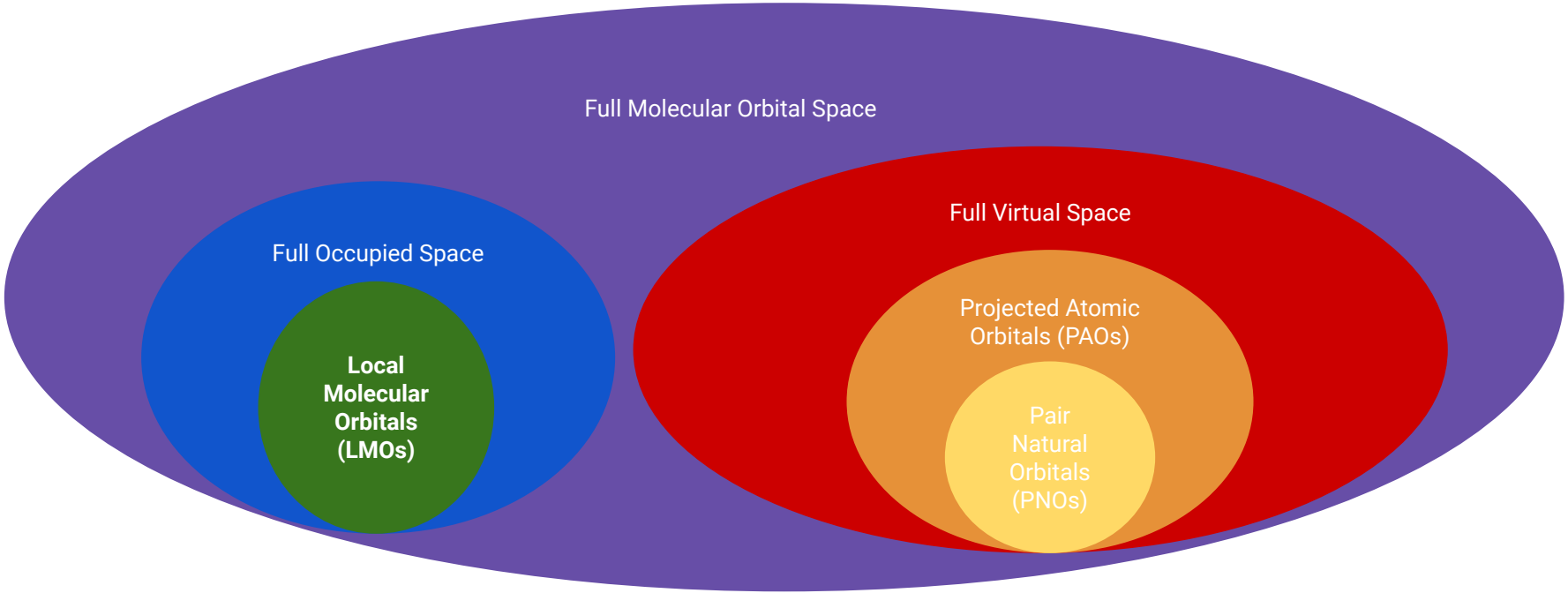
A New Hope: Local Coupled-Cluster

- Instead of using **canonical molecular orbitals**, we can use **local orbitals** as the basis to solve the **coupled-cluster equations**
- This approach **reduces** the **most expensive steps** of a **coupled-cluster** calculation to **linear scaling**
- **Formation** of the **local orbitals** remains **O(N^3)**
- Various methods in other **software** (**non open-source**):
  - DLPNO-CCSD(T) in ORCA
  - PNO-LCCSD(T) in MOLPRO
  - LNO-CCSD(T) in MRCC
- We want an **open-source** version!



Molecular Orbital Hierarchy

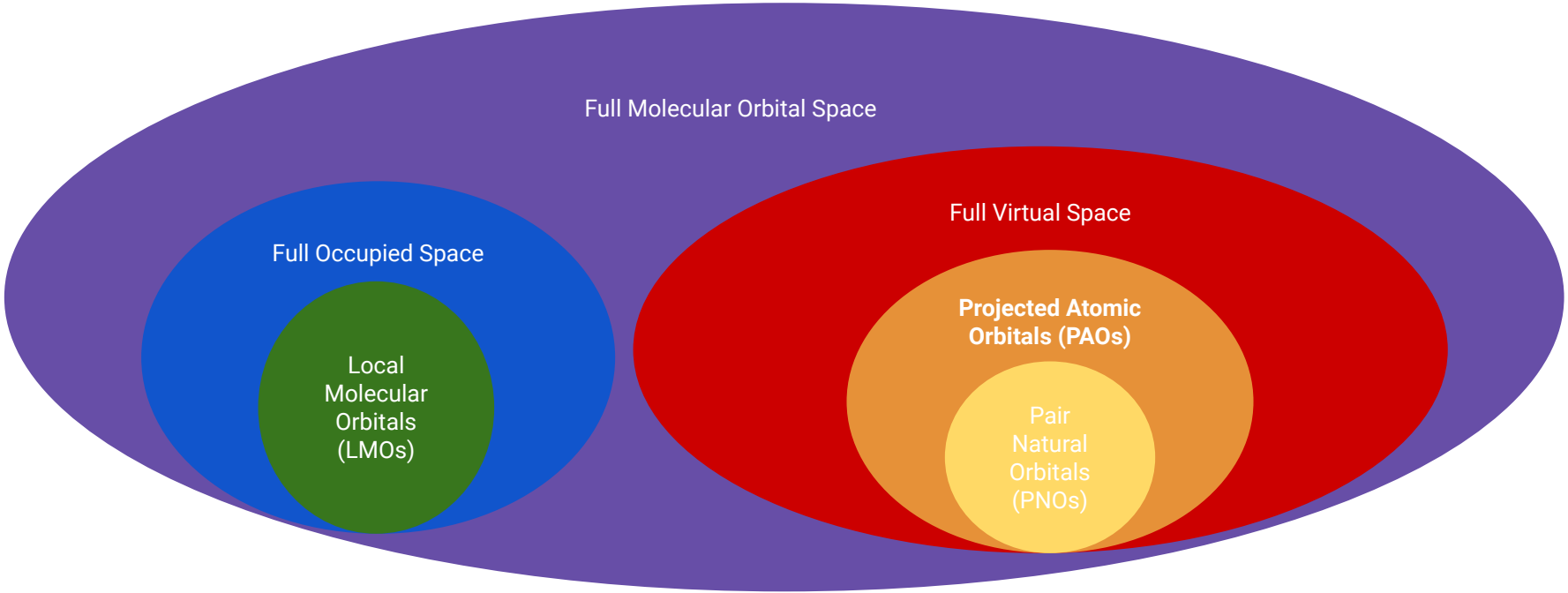



Full Molecular Orbital Space
Full Occupied Space
Local Molecular Orbitals (LMOs)
Full Virtual Space
Projected Atomic Orbitals (PAOs)
Pair Natural Orbitals (PNOs)

Local Molecular Orbitals (LMOs)

- **Localizing unitary transformations** of **occupied MO** coefficients
- The new **MO coefficients** have **limited spatial extent**
- Typically done with **Foster-Boys** or **Pipek-Mezey**

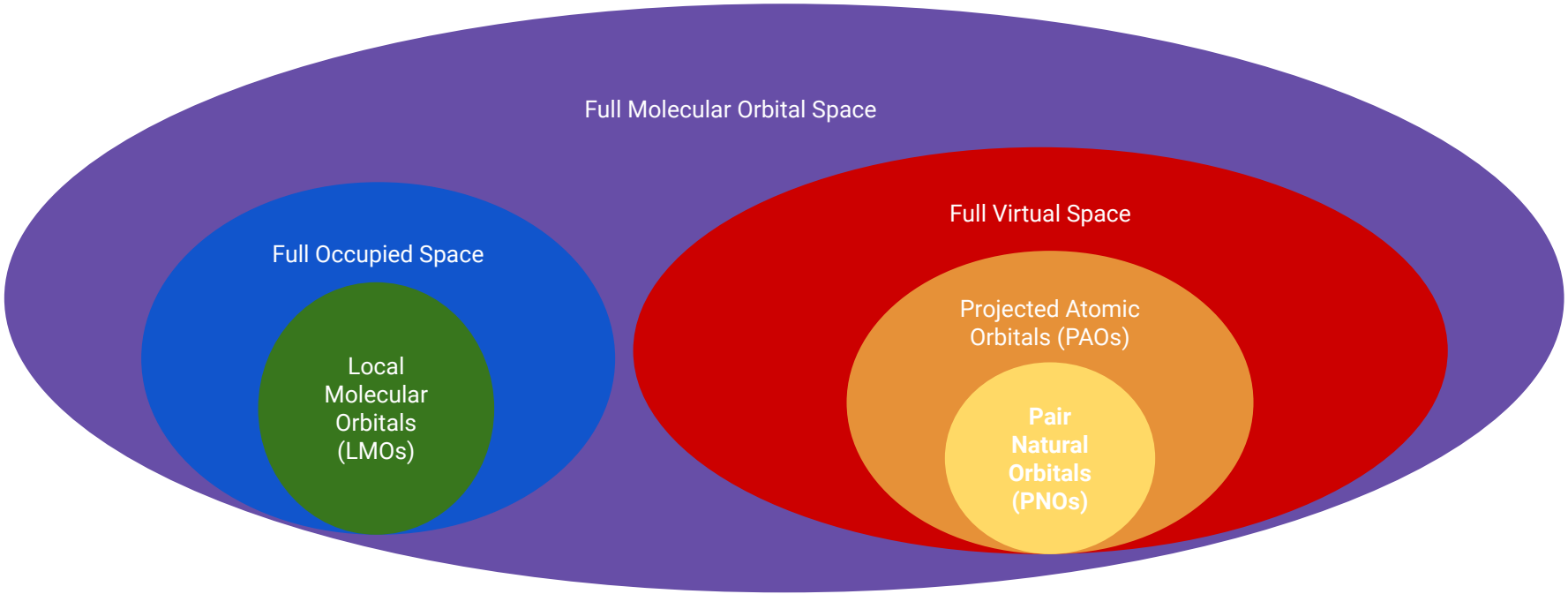
$$C^{L}_{\mu i} = C_{\mu j} U_{ji}$$


Full Molecular Orbital Space
Full Occupied Space
Local Molecular Orbitals (LMOs)
Full Virtual Space
Projected Atomic Orbitals (PAOs)
Pair Natural Orbitals (PNOs)

Projected Atomic Orbitals (PAOs)

- A **basis** for the **virtual space** after **removing** the **occupied MO** space from the entire **AO space**

$$C^{L}_{\mu\nu} = I_{\mu\nu} - C^{L}_{\mu i} C^{L}_{\sigma i} S_{\sigma\nu}$$

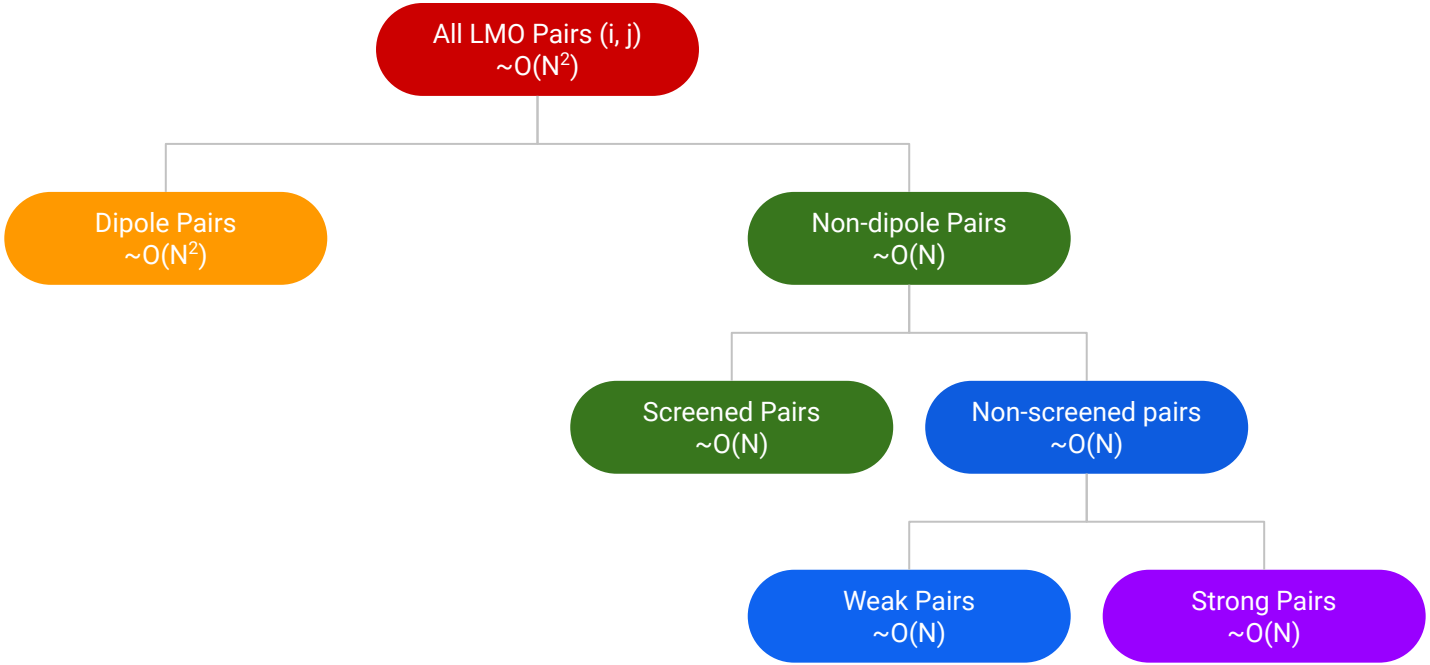

Full Molecular Orbital Space
Full Occupied Space
Local Molecular Orbitals (LMOs)
Full Virtual Space
Projected Atomic Orbitals (PAOs)
Pair Natural Orbitals (PNOs)

Pair Natural Orbitals (PNOs)

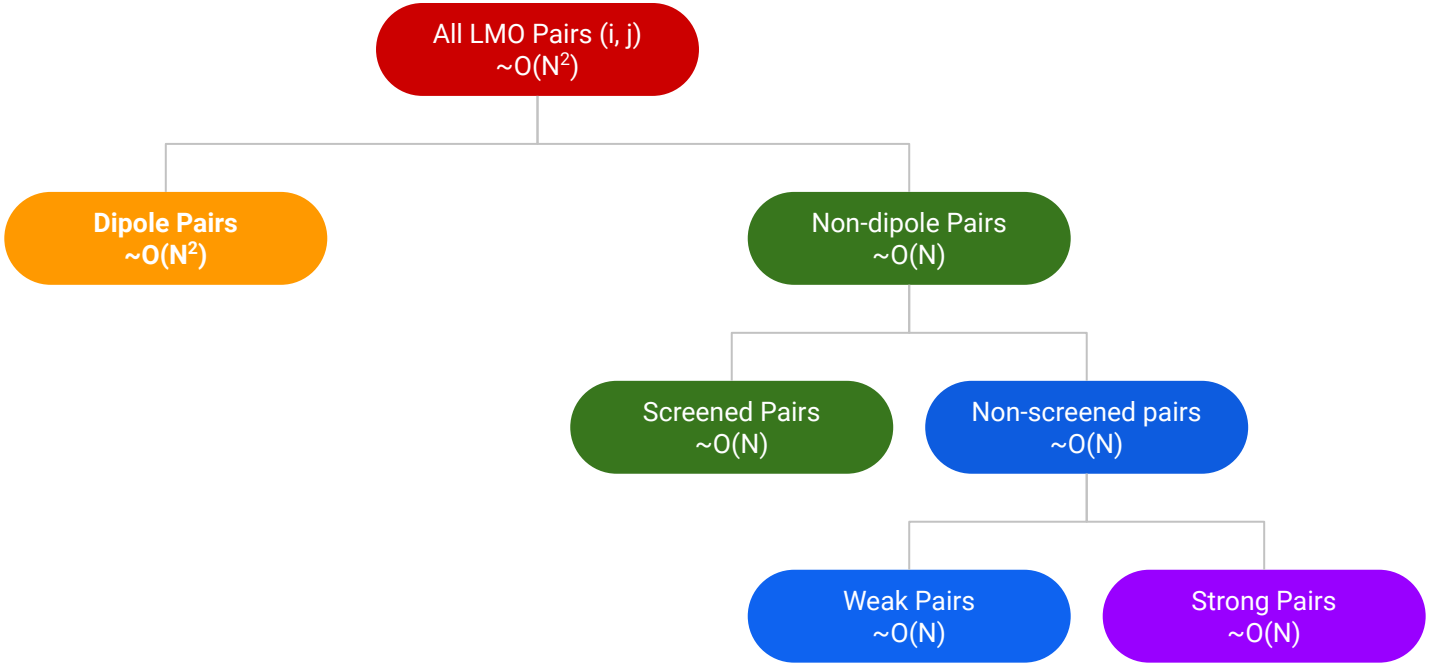
- **Eigenvectors** of the **pair density** of an **LMO pair** in **PAO basis**
- Truncated by **occupation number** (eigenvalues) to **reduce size** of **virtual space**
- **Truncation** determined by **T_CUT_PNO** (default: $\mathbf{10^{-8}}$)

$$D_{ij}^{\mu\nu} = \widetilde{T}_{ij}^{\mu\sigma} T_{ij}^{\nu\sigma} + \widetilde{T}_{ij}^{\sigma\mu} T_{ij}^{\sigma\nu}$$

$$D_{ij} X^{PNO,ij} = X^{PNO,ij} n^{occ,ij}$$



$$E_{DLPNO\text{-}CCSD} = E_{LCCSD} + \Delta E_{WEAK\_PAIRS} + \Delta E_{SCREENED\_PAIRS} + \Delta E_{DIPOLE\_PAIRS} + \Delta E_{PNO}$$



$$E_{DLPNO\text{-}CCSD} = E_{LCCSD} + \Delta E_{WEAK\_PAIRS} + \Delta E_{SCREENED\_PAIRS} + \mathbf{\Delta E_{DIPOLE\_PAIRS}} + \Delta E_{PNO}$$

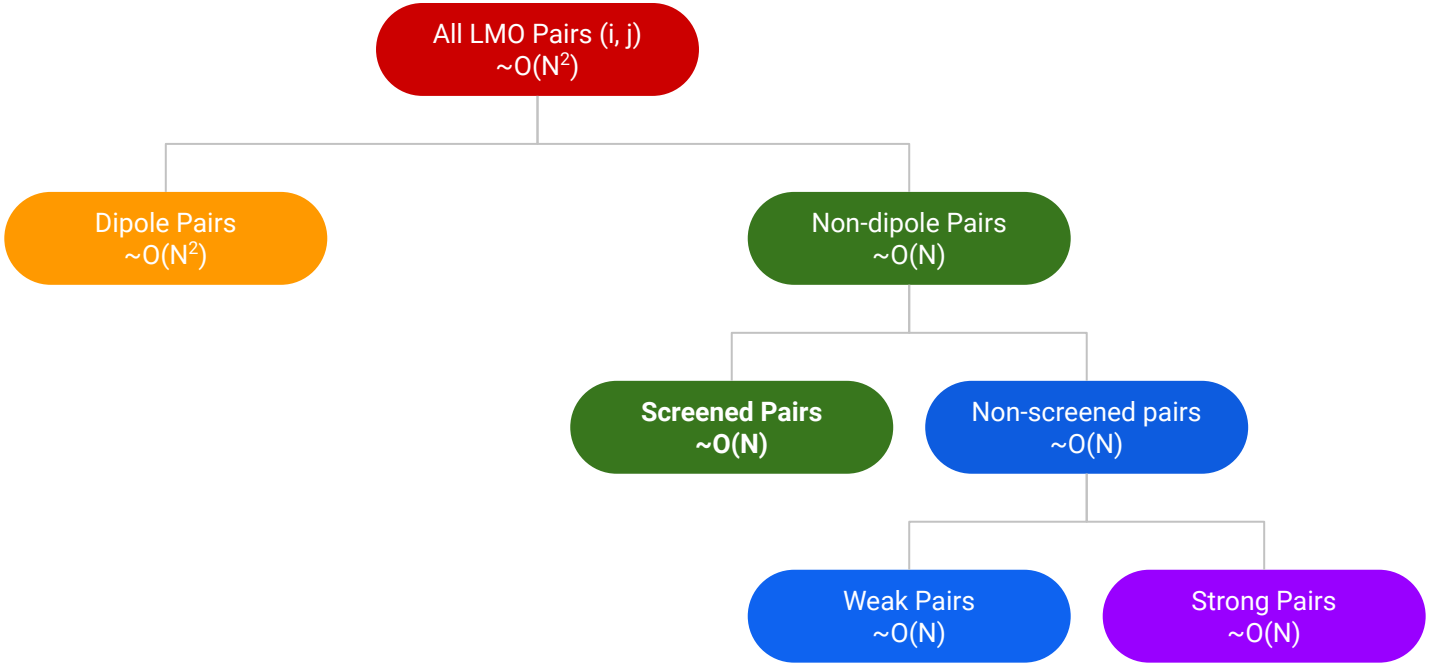
Dipole Energy Correction

- The energy of **LMO pairs** with **limited spatial overlap** (**T_CUT_DO_IJ** = 10^{-5}), are first computed through a dipole estimate[1]

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

- If the energy falls below (**T_CUT_PRE** = 10^{-7}), the pair is considered a **dipole pair** and its energy contribution is added to $\mathbf{\Delta E_{DIPOLE\_PAIRS}}$, and **not further considered**

[1]P. Pinski, C. Riplinger, E. F. Valeev, F. Neese; *J. Chem. Phys.* 143 (3), 034108 (2015).

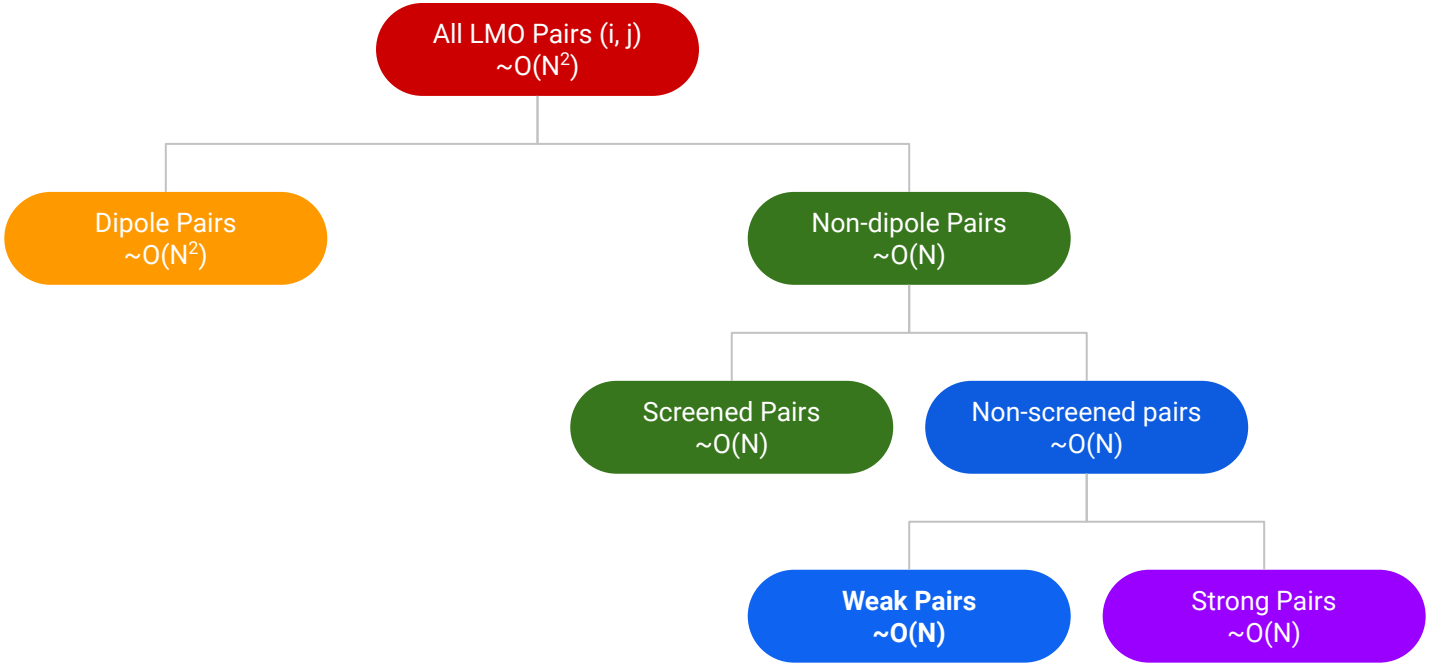


$$E_{DLPNO\text{-}CCSD} = E_{LCCSD} + \Delta E_{WEAK\_PAIRS} + \Delta E_{SCREENED\_PAIRS} + \Delta E_{DIPOLE\_PAIRS} + \Delta E_{PNO}$$

Screened Pair Energy Correction

- Computed using **semi-canonical LMP2 amplitudes** in the **PAO basis**[1]
- **Cheaper** than full iterative LMP2
- If the energy estimate is less than (**T_CUT_PAIRS_MP2** = 10^{-6}), the **energy contribution** from the **pair** is **summed** onto $\Delta E_{\text{SCREENED\_PAIRS'}}$ and is **not further considered**

$$E_{ij}^{\text{SC-LMP2}} = \frac{(i\mu_{ij}|j\nu_{ij})[2(i\mu_{ij}|j\nu_{ij}) - (i\nu_{ij}|j\mu_{ij})]}{F_{ii} + F_{jj} - \epsilon_{\mu_{ij}} - \epsilon_{\nu_{ij}}}$$



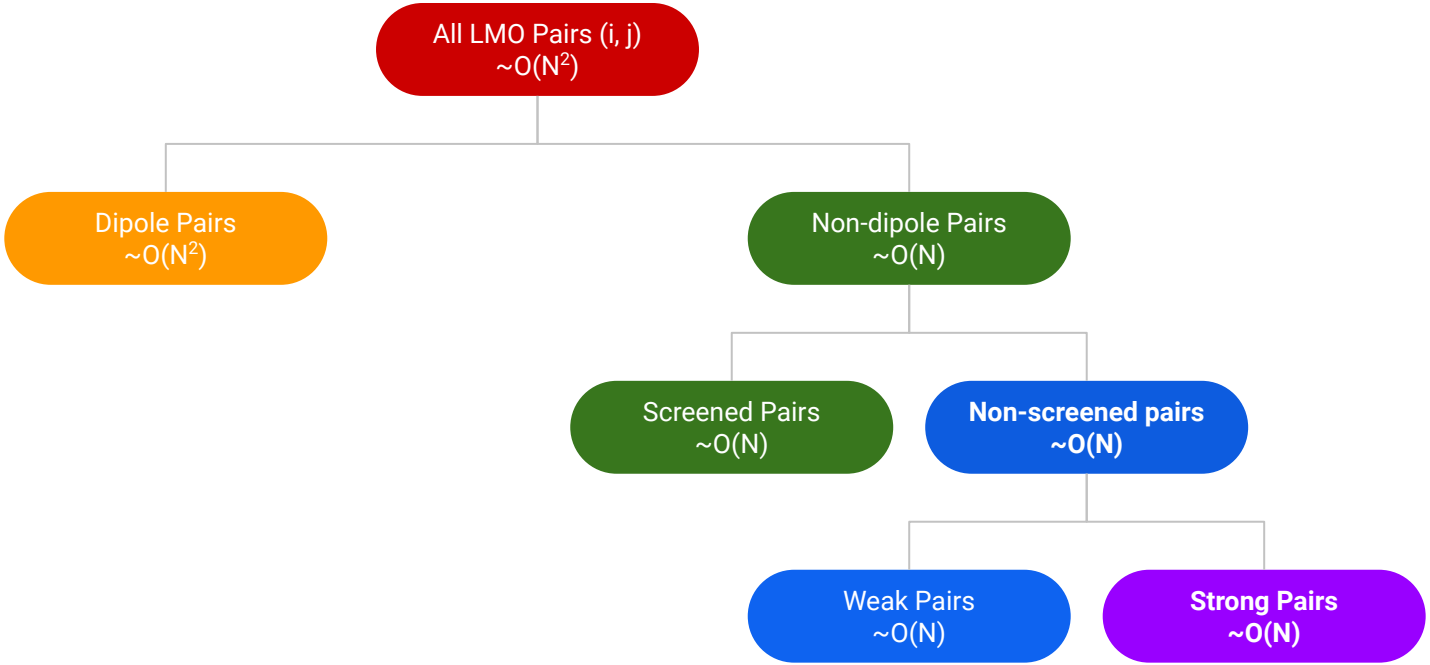
$$E_{DLPNO\text{-}CCSD} = E_{LCCSD} + \Delta E_{WEAK\_PAIRS} + \Delta E_{SCREENED\_PAIRS} + \Delta E_{DIPOLE\_PAIRS} + \Delta E_{PNO}$$

Weak Pair Energy Correction

- Next, an **iterative LMP2 energy** estimate is obtained for **each surviving pair** by solving the **LMP2 equations** in the **PAO basis**[1]
- If the energy estimate is less than (**T_CUT_PAIRS** = 10^{-5}), the **energy contribution** from the **pair** is **summed** onto $\Delta E_{\text{WEAK\_PAIRS}}$, and **kept for (T)**

$$R_{\tilde{\mu}\tilde{\nu}}^{ij} = K_{\tilde{\mu}\tilde{\nu}}^{ij} + \left(\varepsilon_{\tilde{\mu}} + \varepsilon_{\tilde{\nu}} - F_{ii} - F_{jj}\right) T_{\tilde{\mu}\tilde{\nu}}^{ij} - \sum_{k \neq i, \tilde{\kappa}\tilde{\tau}} F_{ik} S_{\tilde{\mu}\tilde{\kappa}} T_{\tilde{\kappa}\tilde{\tau}}^{kj} S_{\tilde{\tau}\tilde{\nu}} - \sum_{k \neq j, \tilde{\kappa}\tilde{\tau}} F_{kj} S_{\tilde{\mu}\tilde{\kappa}} T_{\tilde{\kappa}\tilde{\tau}}^{ik} S_{\tilde{\tau}\tilde{\nu}}$$

[1]P. Pinski, C. Riplinger, E. F. Valeev, F. Neese; *J. Chem. Phys.* 143 (3), 034108 (2015).

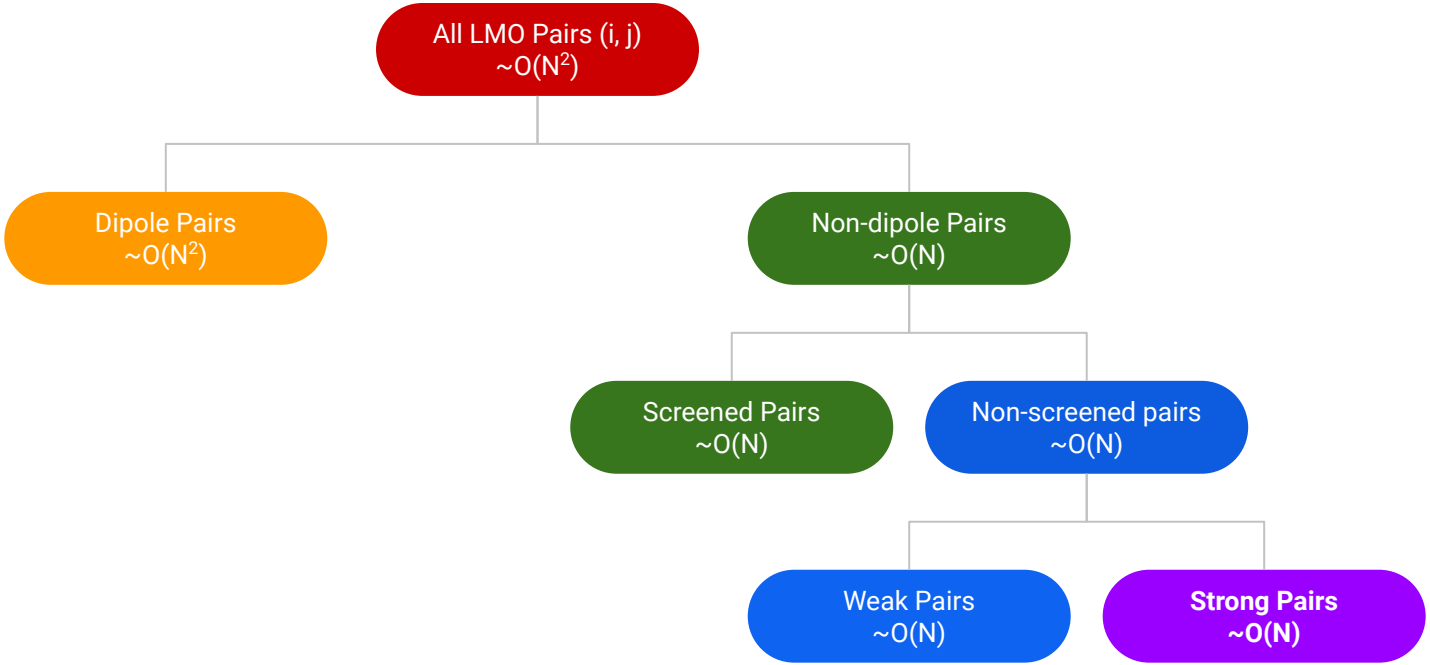


$$E_{DLPNO\text{-}CCSD} = E_{LCCSD} + \Delta E_{WEAK\_PAIRS} + \Delta E_{SCREENED\_PAIRS} + \Delta E_{DIPOLE\_PAIRS} + \Delta E_{PNO}$$

PNO Truncation Error

- After **PNOs** are formed from **LMP2 PAO amplitudes**, the **difference between the LMP2 energy** for the **strong pairs** in the **PAO** vs **PNO** basis is defined as the **PNO truncation error**

$$\Delta E_{\text{PNO}} = \sum_{ij \in \text{strong pairs}} E_{ij}^{\text{PAO-LMP2}} - E_{ij}^{\text{PNO-LMP2}}$$



$$E_{DLPNO\text{-}CCSD} = E_{LCCSD} + \Delta E_{WEAK\_PAIRS} + \Delta E_{SCREENED\_PAIRS} + \Delta E_{DIPOLE\_PAIRS} + \Delta E_{PNO}$$

LCCSD Iteration Energy

- **Solve the CCSD residual equations** for each **strong pair** in **PNO basis**

$$T_i^{a_{ii}} -= \frac{R_i^{a_{ii}}}{D_i^{a_{ii}}} \qquad T_{ij}^{a_{ij}b_{ij}} -= \frac{R_{ij}^{a_{ij}b_{ij}}}{D_{ij}^{a_{ij}b_{ij}}}$$

$$E_{\mathrm{LCCSD}} = \sum_{ij \in \text{strong pairs}} (T_{ij}^{a_{ij}b_{ij}} + S_{a_{ii}}^{a_{ij}} T_i^{a_{ii}} T_j^{b_{jj}} S_{b_{jj}}^{b_{ij}}) L_{ij}^{a_{ij}b_{ij}}$$

DLPNO-(T) Energy Contributions

$$E_{DLPNO\text{-}CCSD(T)} = E_{DLPNO\text{-}CCSD} + E_{LCCSD(T0)} + \Delta E_{(T)} + \Delta E_{SCREENED\_TRIPLETS}$$

DLPNO-(T) Energy Contributions

$$E_{DLPNO\text{-}CCSD(T)} = E_{DLPNO\text{-}CCSD} + E_{LCCSD(T0)} + \Delta E_{(T)} + \Delta E_{\text{[illegible]}}$$

DLPNO-(T) Energy Contributions

$$E_{\text{DLPNO-CCSD(T)}} = E_{\text{DLPNO-CCSD}} + E_{\text{LCCSD(T0)}} + \Delta E_{\text{(T)}} + \Delta E_{\text{SCREENED\_TRIPLETS}}$$

Screened Triplets Correction

- Compute the **semi-canonical (T0)** energy using a lower **TNO tolerance** (PNO analogs for triplets) (**T_CUT_TNO_PRE** = 10^{-7}), for all **relevant triplets**
- A **relevant triplet ijk** contains at least **two strong pairs** in (**ij, jk, ik**), the other may be a **weak pair**
- If the **energy** falls below (**T_CUT_TNO_WEAK** = 10^{-7}), the triplet is **not further considered**, with its **energy contribution** added to $\Delta E_{\text{SCREENED\_TRIPLETS}}$

$$E_{(T)} = \sum_{i \le j \le k} \frac{t_{ijk}^{abc}}{1 + (\delta_{ij} + \delta_{jk} + \delta_{ik}) + 2\delta_{ij}\delta_{jk}\delta_{ik}} \cdot (8V_{ijk}^{abc} - 4V_{ijk}^{bac} - 4V_{ijk}^{acb} - 4V_{ijk}^{cab} + 2V_{ijk}^{bca} + 2V_{ijk}^{cab})$$

DLPNO-(T) Energy Contributions

$$E_{DLPNO\text{-}CCSD(T)} = E_{DLPNO\text{-}CCSD} + E_{LCCSD(T0)} + \Delta E_{(T)} + \Delta E_{SCREENED\_TRIPLETS}$$

Semicanonical (T0) Energy

- Compute **semi-canonical (T0) energies** for **surviving triplets**, at tighter tolerance (**T_CUT_TNO** = 10^{-9})

$$W_{ijk}^{a_{ijk}b_{ijk}c_{ijk}} = P_L[(ia_{ijk}|b_{ijk}d_{ijk})S_{d_{kj}d_{ijk}}^{kj,ijk} t_{kj}^{d_{kj}c_{kj}} S_{c_{kj}c_{ijk}}^{kj,ijk} - S_{a_{il}a_{ijk}}^{il,ijk} T_{il}^{a_{il}b_{il}} S_{b_{il}d_{ijk}}^{il,ijk} (jl_{ijk}|kc_{ijk})]$$

$$V_{ijk}^{a_{ijk}b_{ijk}c_{ijk}} = W_{ijk}^{a_{ijk}b_{ijk}c_{ijk}} + P_S[S_{a_{ii}a_{ijk}}^{ii,ijk} t_i^{a_{ii}} (jb_{ijk}|kc_{ijk})]$$

$$T_{ijk}^{a_{ijk}b_{ijk}c_{ijk}} = -\frac{W_{ijk}^{a_{ijk}b_{ijk}c_{ijk}}}{\epsilon_{a_{ijk}} + \epsilon_{b_{ijk}} + \epsilon_{c_{ijk}} - f_{ii} - f_{jj} - f_{kk}}$$

DLPNO-(T) Energy Contributions

$$E_{DLPNO\text{-}CCSD(T)} = E_{DLPNO\text{-}CCSD} + E_{LCCSD(T0)} + \Delta E_{(T)} + \Delta E_{SCREENED\_TRIPLETS}$$

Iterative (T) Energy Correction

- Compute **(T0) energies** and **amplitudes** at **weaker tolerances** (**T_CUT_TNO_STRONG** = 10^{-8}, **T_CUT_TNO_WEAK** = 10^{-7})
- **Strong triplets** are triplets which account for **90% of the (T0) energy**, and **weak triplets** are the remaining triplets
- **Iteratively update triples amplitudes**

$$R_{ijk}^{a_{ijk}b_{ijk}c_{ijk}} = W_{ijk}^{a_{ijk}b_{ijk}c_{ijk}} - T_{ijk}^{a_{ijk}b_{ijk}c_{ijk}}(\epsilon_{a_{ijk}} + \epsilon_{b_{ijk}} + \epsilon_{c_{ijk}} - f_{ii} - f_{jj} - f_{kk})$$
$$-\sum_{l\neq i} f_{il} T_{ljk}^{a_{ljk}b_{ljk}c_{ljk}} S_{a_{ljk}b_{ljk}c_{ljk}}^{a_{ijk}b_{ijk}c_{ijk}} - \sum_{l\neq j} f_{jl} T_{ilk}^{a_{ilk}b_{ilk}c_{ilk}} S_{a_{ilk}b_{ilk}c_{ilk}}^{a_{ijk}b_{ijk}c_{ijk}} - \sum_{l\neq k} f_{kl} T_{ijl}^{a_{ijl}b_{ijl}c_{ijl}} S_{a_{ijl}b_{ajl}c_{ajl}}^{a_{ijk}b_{ijk}c_{ijk}}$$

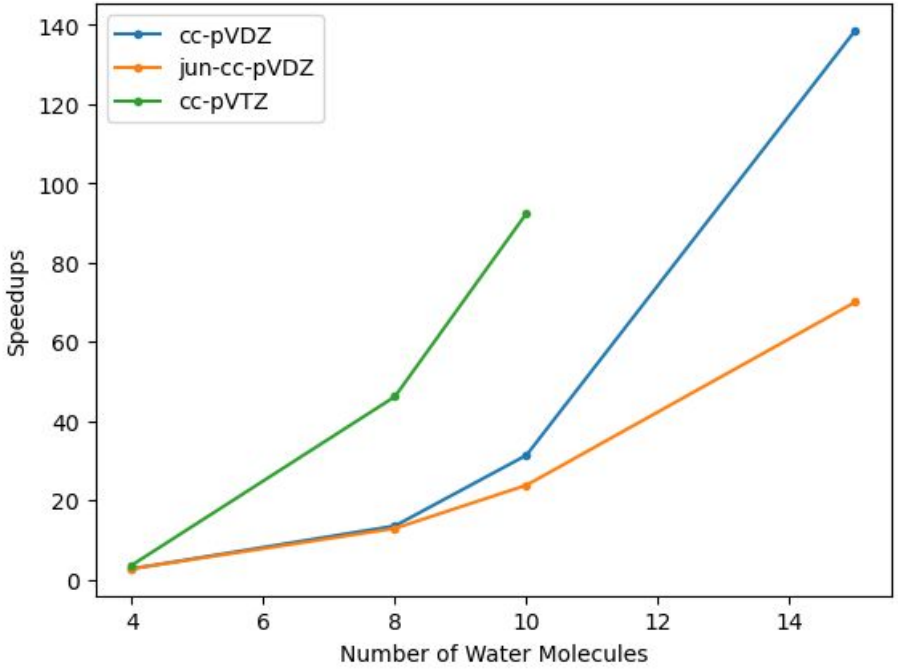
$$\Delta E_{(T)} = \sum_{ijk\in\text{strong triplets}} E_{(T)}^{ijk} - E_{(T0)}^{ijk}(\text{strong}) + \sum_{ijk\in\text{weak triplets}} E_{(T)}^{ijk} - E_{(T0)}^{ijk}(\text{weak})$$

S22 Interaction Energy Errors

Basis Set	H Bonded	Disp Bound	Mixed	Overall
cc-pVDZ	0.12	0.23	0.10	0.15
jun-cc-pVDZ	0.10	0.23	0.05	0.13
cc-pVTZ	0.11	0.22	0.11	0.15

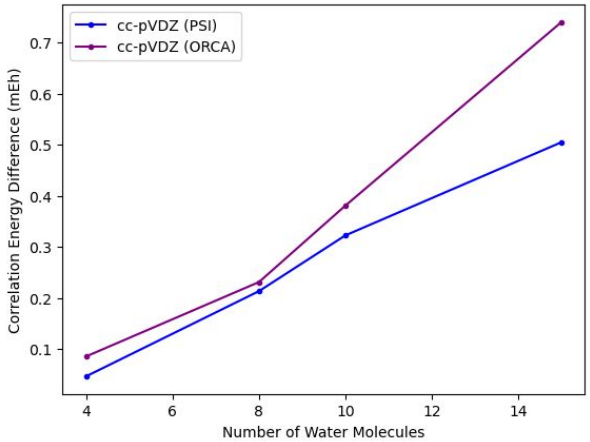
Table I: Interaction Energy Errors (kcal/mol) relative to DF-CCSD(T) reference, CP Corrected

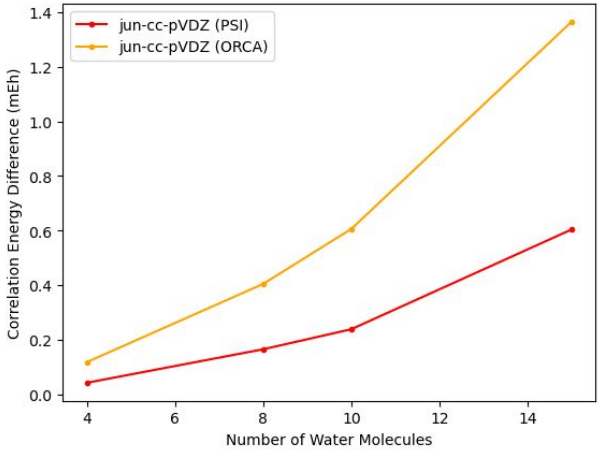
Speedups compared to CCSD(T)

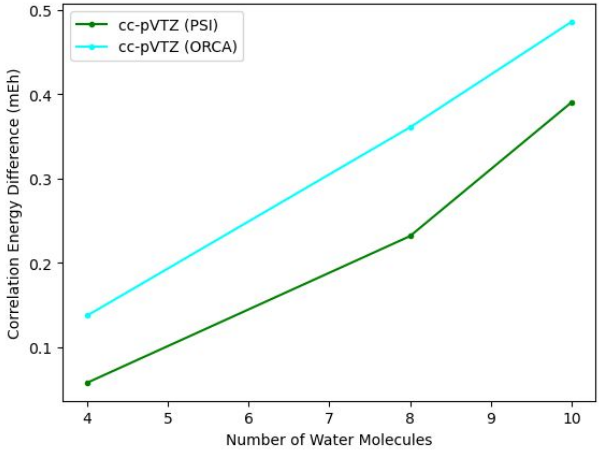


DLPNO-CCSD(T) compared to **FNO-DF-CCSD(T)** [Previous state of the art in PSI]

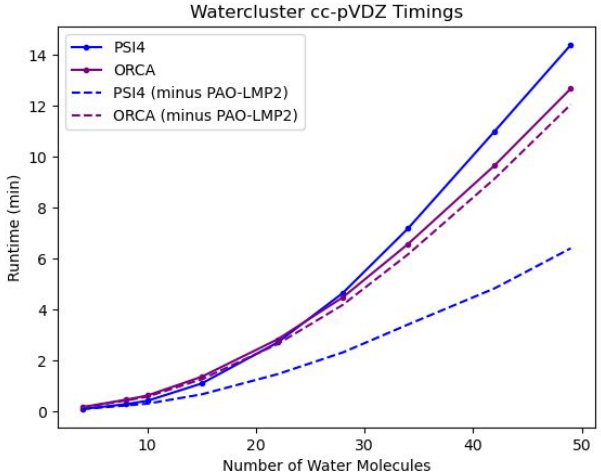
PSI vs ORCA Accuracy (Water Clusters)

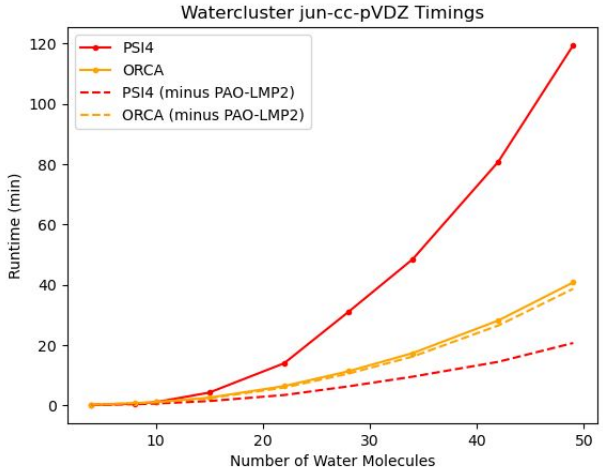


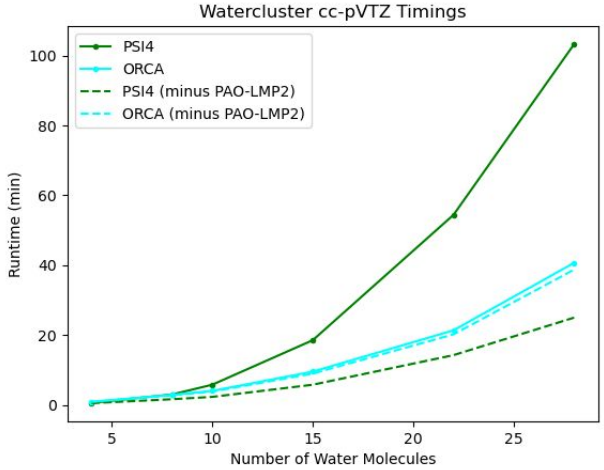


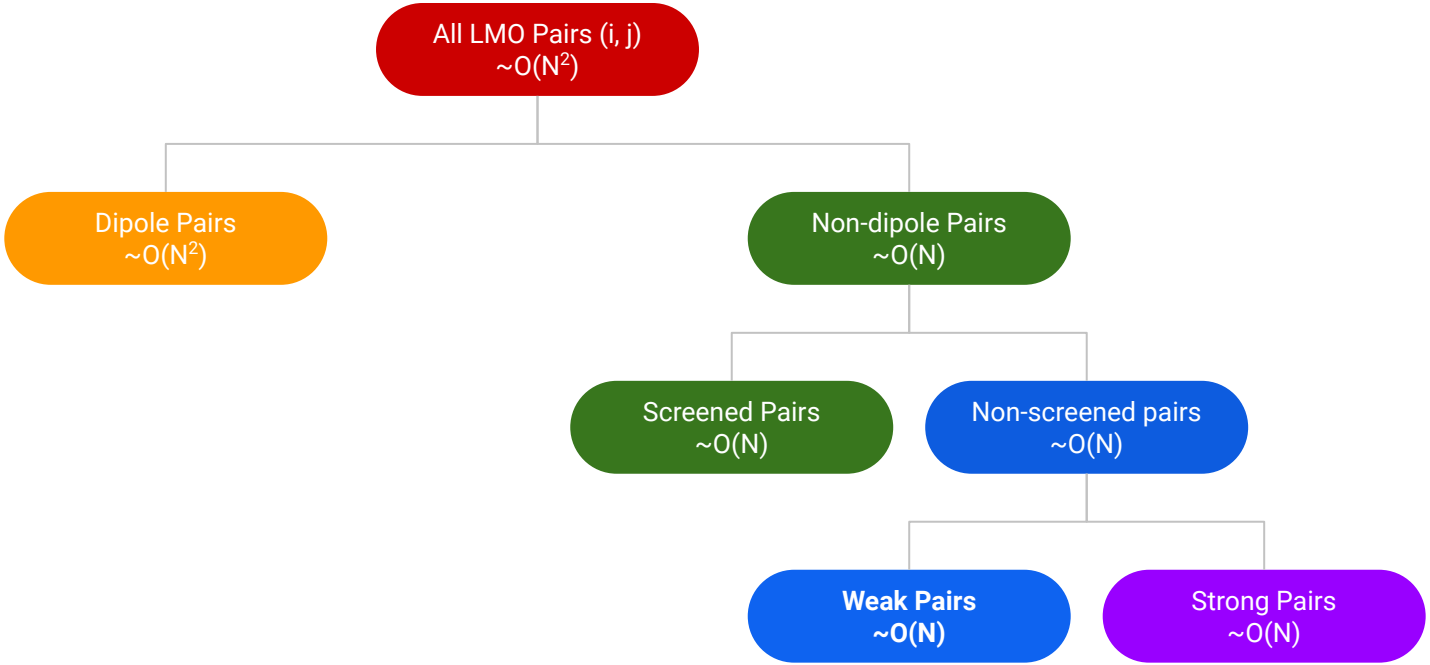


PSI vs ORCA Timings (Water Clusters)









$$E_{DLPNO\text{-}CCSD} = E_{LCCSD} + \Delta E_{WEAK\_PAIRS} + \Delta E_{SCREENED\_PAIRS} + \Delta E_{DIPOLE\_PAIRS} + \Delta E_{PNO}$$

Future Work

- **Optimize PAO-LMP2 prescreening** to make code **more competitive** with ORCA [**Jan 2024**]
- **Merge DLPNO-CCSD(T)** code into **PSI** [**Mar 2024**]
- **Apply DLPNO-CCSD(T)** to **large molecules** (i.e. **protein-drug systems**)
- **Incorporate** other **rank reduction** methods, such as **tensor hypercontraction**, to further **optimize** algorithm

Acknowledgements

- **Schaefer Group, CCQC**
  - Dr. Henry F. Schaefer
  - Dr. Justin Turney
- **Sherrill Group**
  - Dr. David Sherrill
  - Dr. Lori Burns
  - Dr. David Poole

The End

- Questions, comments, concerns?
- For further questions
  - **Email:** Andy.Jiang@uga.edu