Robust Factorization of Tensor Networks, With Applications in Quantum Chemistry

Edward Valeev
Department of Chemistry, Virginia Tech

SIAM PP22

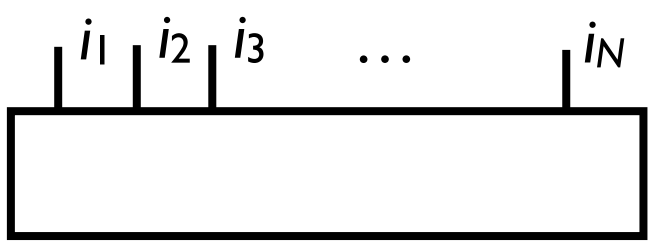
Minisymposium *"Parallel Algorithms for Tensor Computations and their Applications"*

Feb 26, 2022

Why Tensors? Consider N <u>Quantum</u> Particles …

state = $\psi(\boldsymbol{r}_1,\boldsymbol{r}_2,\ldots\boldsymbol{r}_N)$

expand in 1-particle basis: $\psi(\boldsymbol{r}_1,\boldsymbol{r}_2,\ldots\boldsymbol{r}_N) \approx \sum^M \psi_{i1i2..iN}\, \phi_{i1}(\boldsymbol{r}_1)\, \phi_{i2}(\boldsymbol{r}_2) \ldots \phi_{iN}(\boldsymbol{r}_N)$ **$M \sim$ (1..10) x N**

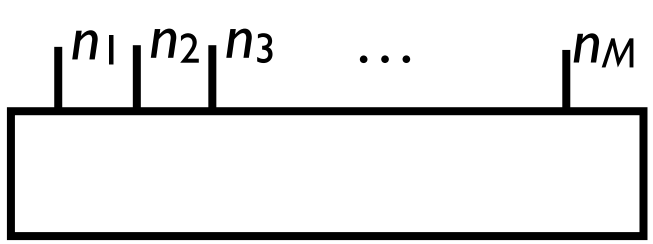


M^N elements

encodes probability of finding N particles in states $i_1..i_N$

(joint probability distribution)

or, in occupation number representation (convenient for indistinguishable particles)



K^M elements

encodes probabilities of specific occupancies of M states

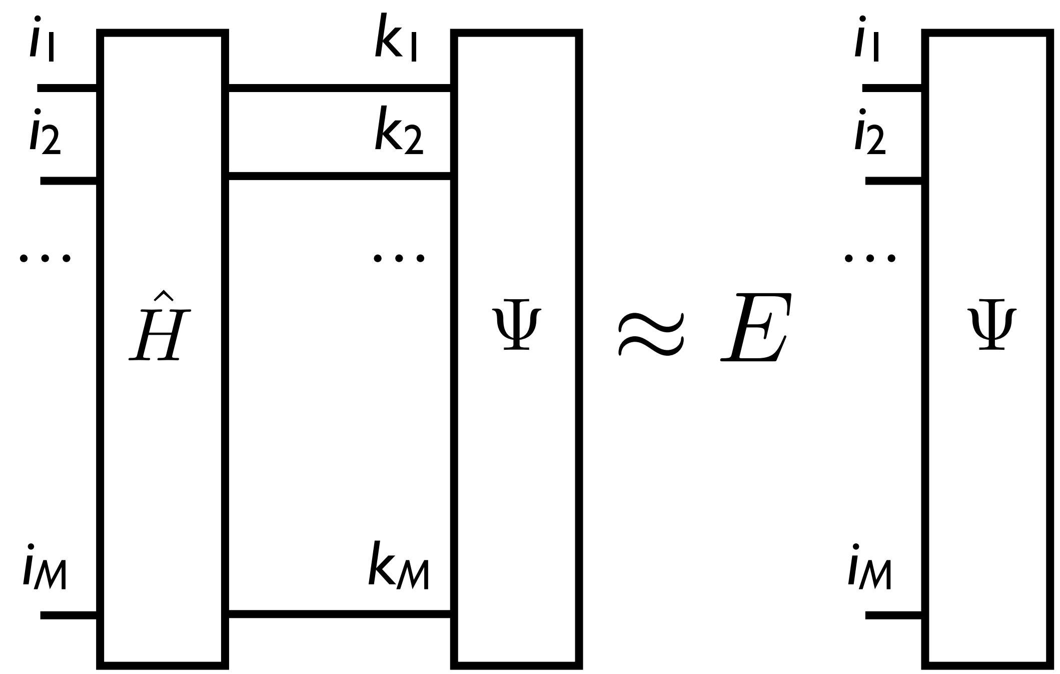
K = 2 or 4 for electrons, 2 for qubits, etc.

factorial storage/operation complexities ("curse of dimensionality")

Properties of Quantum States Are Also Tensors

N-Body Schrödinger Equation = Tensor Eigenvalue Problem

$$\hat{H}\Psi = E\Psi \qquad \Longrightarrow$$



Example: N_2 molecule

N=14 $\qquad$ M=60 (cc-pVTZ basis)

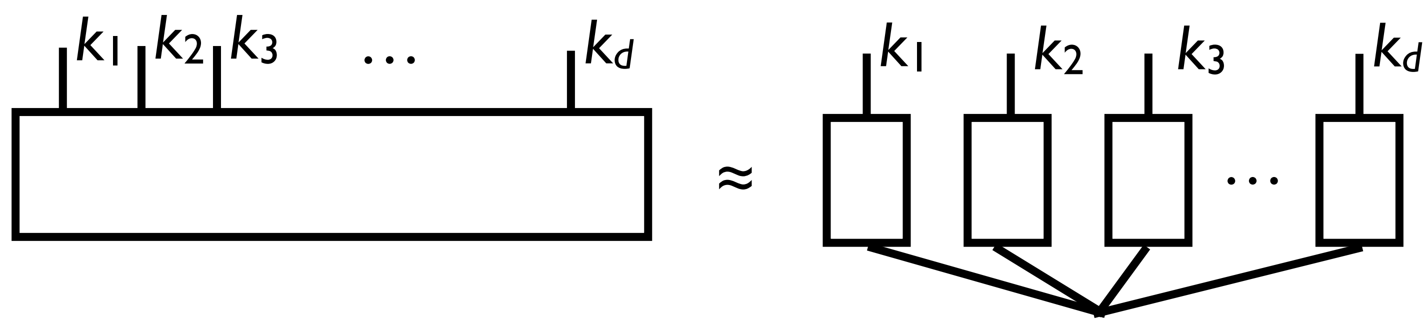
size(Ψ) = 1.5 x 10^{17} $\qquad$ but only <10^9 elements are significant!

element sparsity is useful, but it is essential to exploit more general *data sparsity* in H and Ψ !

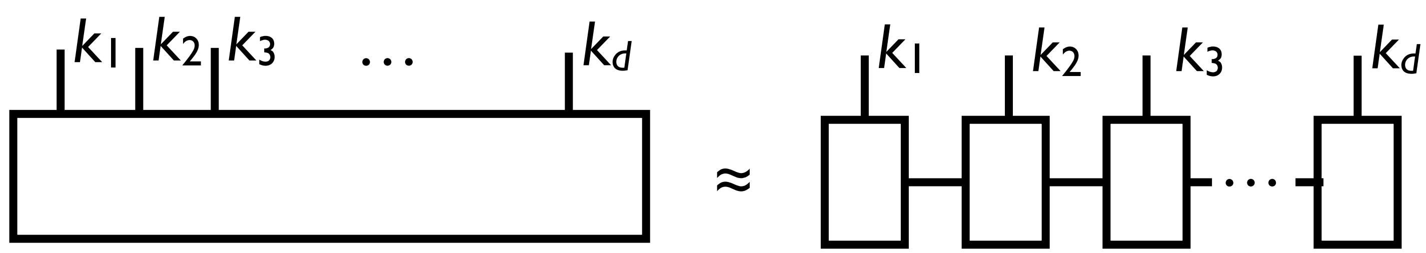
Tensor Factorization is Key to N-Body Quantum Simulation

tensor networks

CP

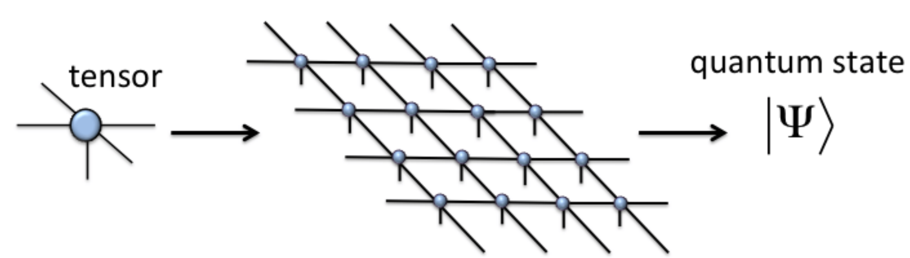


MPS/TT



$$A_{k_1 \ldots k_d} = \left(U^{(1)}\right)^{k_1}_{r_1} \left(U^{(2)}\right)^{k_2}_{r_1, r_2} \cdots \left(U^{(d-1)}\right)^{k_{d-1}}_{r_{d-2}, r_{d-1}} \left(U^{(d)}\right)^{k_d}_{r_{d-1}}$$

PEPS



\+ many more (MERA, etc.)

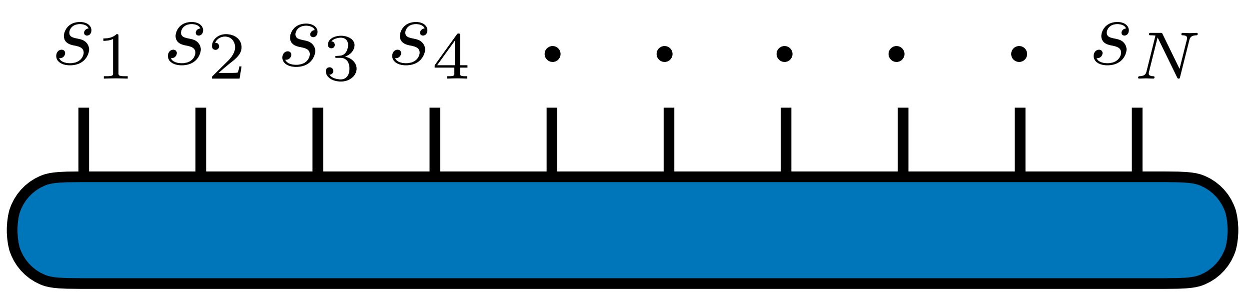
exact when factorization ranks → ∞

efficient only for special cases (e.g., DMRG for 1-d systems)

Tensor Networks via Cumulant/Perturbative Expansion

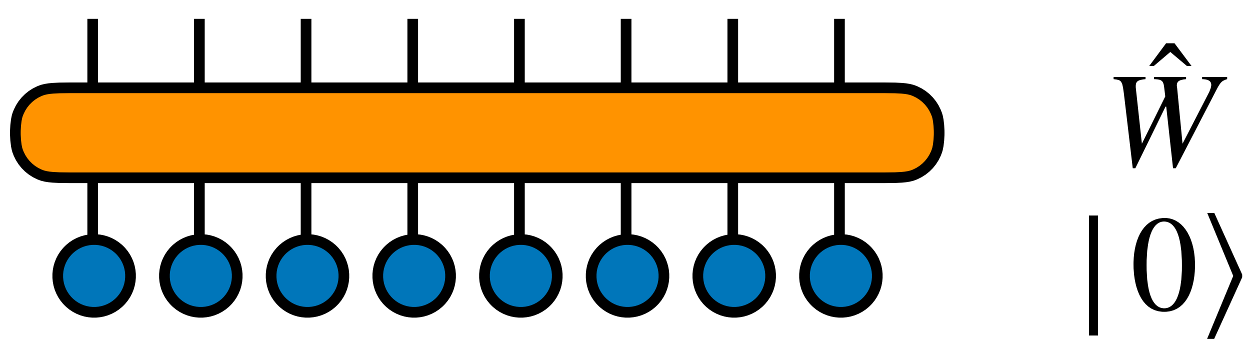
instead of encoding joint probability amplitudes

$$|\Psi\rangle \equiv$$



encode *differences* in probability amplitudes relative to simple (usually, uncorrelated) state

$$|\Psi\rangle \approx \hat{W}|0\rangle$$

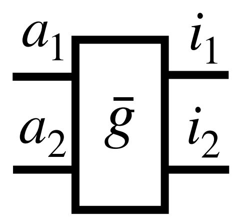


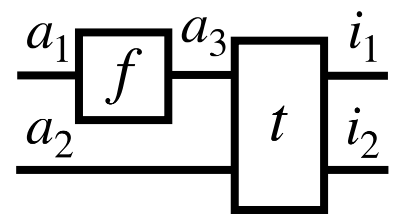
efficient if W limited to a sum of few-body terms (e.g, 2-body in CCSD)

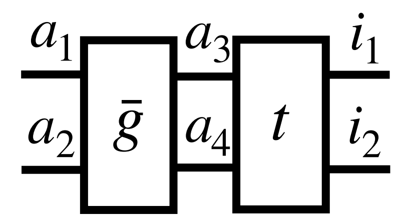
tensor ≈ sum of tensor networks

Example: CCD Equations

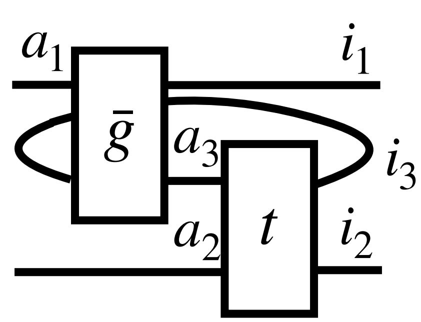
$$0 = \langle \Phi^{a_1a_2}_{i_1i_2} | \hat{H} \exp(\hat{T}_2) | 0 \rangle_c = \hat{A}^{a_1a_2}_{i_1i_2}\Big(\bar{g}^{i_1i_2}_{a_1a_2} \quad + \quad f^{a_3}_{a_1} t^{i_1i_2}_{a_3a_2} \quad + \quad \frac{1}{2}\bar{g}^{a_3a_4}_{a_1a_2} t^{i_1i_2}_{a_3a_4} \quad + \quad \cdots$$

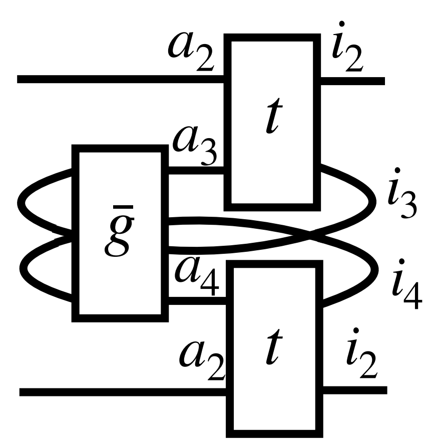






$$\bar{g}^{i_1a_3}_{a_1i_3} t^{i_3i_2}_{a_3a_2} \quad + \quad \frac{1}{2}\bar{g}^{a_3a_4}_{i_3i_4} t^{i_1i_3}_{a_1a_3} t^{i_4i_2}_{a_4a_2} \quad + \quad \cdots \Big)$$



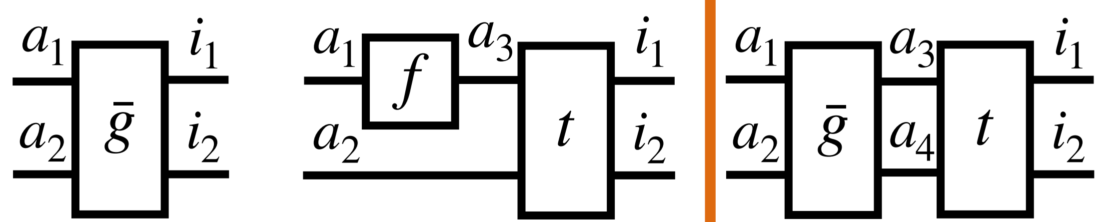


storage = $O(N^4)$
cost = $O(N^6)$

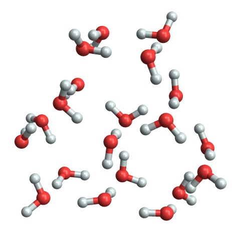
Example: CCD Equations

dominant tensor contraction

$$0 = \langle \Phi^{a_1a_2}_{i_1i_2} | \hat{H} \exp(\hat{T}_2) | 0 \rangle_c = \hat{A}^{a_1a_2}_{i_1i_2} \Big(\bar{g}^{i_1i_2}_{a_1a_2} \quad + \quad f^{a_3}_{a_1} t^{i_1i_2}_{a_3a_2} \quad + \quad \boxed{\frac{1}{2} \bar{g}^{a_3a_4}_{a_1a_2} t^{i_1i_2}_{a_3a_4} \quad +} \quad \cdots$$



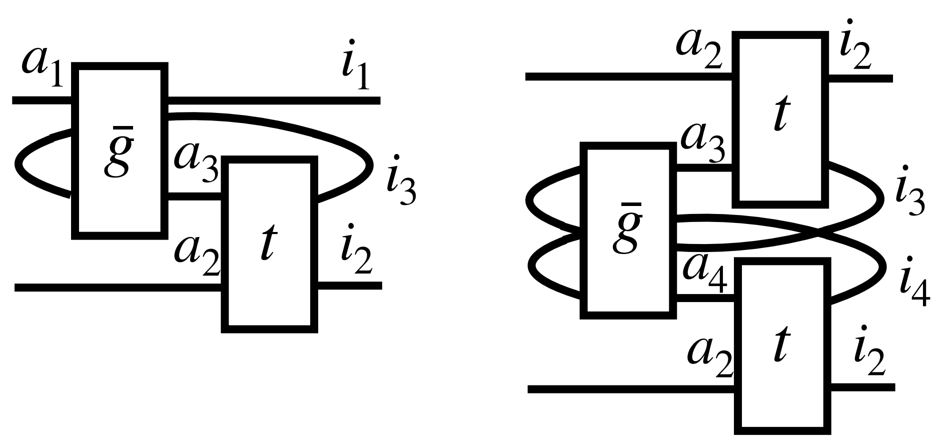
example:



extent(i) = N/2 = 80
extent(a) = M - N/2 = 860

size(t) = 37.9 GB
size(g) = 4.4 TB
cost ~ 3.5 x 10^{15} FLOPs

$$\bar{g}^{i_1a_3}_{a_1i_3} t^{i_3i_2}_{a_3a_2} \quad + \quad \frac{1}{2} \bar{g}^{a_3a_4}_{i_3i_4} t^{i_1i_3}_{a_1a_3} t^{i_4i_2}_{a_4a_2} \quad + \quad \cdots \quad \Big)$$



storage = $O(N^4)$
cost = $O(N^6)$

Gets Arbitrarily Complex: CCSDT

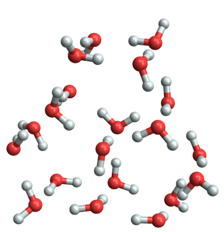
$$\begin{aligned}
\langle 0|\hat{A}_1\bar{H}\,|0\rangle = (0 + -1 \times A^{a_1}_{i_1} g^{i_1a_2}_{i_2a_1} t^{i_2}_{a_2} + -1 \times A^{a_1}_{i_1} f^{i_1}_{i_2} t^{i_2}_{a_1} + A^{a_1}_{i_1} f^{a_2}_{a_1} t^{i_1}_{a_2} + -\frac{1}{2} \times A^{a_1}_{i_1} g^{i_1a_2}_{i_2i_3} t^{i_2i_3}_{a_1a_2} + \\
-\frac{1}{2} \times A^{a_1}_{i_1} g^{a_2a_3}_{i_2a_1} t^{i_1i_2}_{a_2a_3} + A^{a_1}_{i_1} f^{i_1}_{a_1} + A^{a_1}_{i_1} g^{a_2a_3}_{i_2i_3} t^{i_2}_{a_1} t^{i_3}_{a_2} t^{i_1}_{a_3} + A^{a_1}_{i_1} f^{a_2}_{i_2} t^{i_1i_2}_{a_1a_2} + \frac{1}{4} \times A^{a_1}_{i_1} g^{a_2a_3}_{i_2i_3} t^{i_1i_2i_3}_{a_1a_2a_3} + \\
-1 \times A^{a_1}_{i_1} g^{a_2a_3}_{i_2a_1} t^{i_1}_{a_2} t^{i_2}_{a_3} + -1 \times A^{a_1}_{i_1} g^{i_1a_2}_{i_2i_3} t^{i_2}_{a_1} t^{i_3}_{a_2} + -1 \times A^{a_1}_{i_1} f^{a_2}_{i_2} t^{i_2}_{a_1} t^{i_1}_{a_2} + -\frac{1}{2} \times A^{a_1}_{i_1} g^{a_2a_3}_{i_2i_3} t^{i_2}_{a_1} t^{i_1i_3}_{a_2a_3} + -\frac{1}{2} \times A^{a_1}_{i_1} g^{a_2a_3}_{i_2i_3} t^{i_1}_{a_2} t^{i_2i_3}_{a_1a_3} + \\
A^{a_1}_{i_1} g^{a_2a_3}_{i_2i_3} t^{i_2}_{a_2} t^{i_1i_3}_{a_1a_3})
\end{aligned}$$

$$\begin{aligned}
\langle 0|\hat{A}_2\bar{H}\,|0\rangle = (\frac{1}{4} \times A^{a_1a_2}_{i_1i_2} g^{a_3a_4}_{i_3i_4} t^{i_3}_{a_1} t^{i_4}_{a_2} t^{i_1}_{a_3} t^{i_2}_{a_4} + \frac{1}{8} \times A^{a_1a_2}_{i_1i_2} g^{a_3a_4}_{a_1a_2} t^{i_1i_2}_{a_3a_4} + \frac{1}{4} \times A^{a_1a_2}_{i_1i_2} f^{a_3}_{i_3} t^{i_1i_2i_3}_{a_1a_2a_3} + \frac{1}{4} \times A^{a_1a_2}_{i_1i_2} g^{a_3a_4}_{i_3a_1} t^{i_1i_2i_3}_{a_2a_3a_4} + \frac{1}{4} \times A^{a_1a_2}_{i_1i_2} g^{i_1a_3}_{i_3i_4} t^{i_2i_3i_4}_{a_1a_2a_3} + \\
-\frac{1}{2} \times A^{a_1a_2}_{i_1i_2} f^{a_3}_{a_1} t^{i_1i_2}_{a_2a_3} + \frac{1}{2} \times A^{a_1a_2}_{i_1i_2} f^{i_1}_{i_3} t^{i_2i_3}_{a_1a_2} + \frac{1}{8} \times A^{a_1a_2}_{i_1i_2} g^{i_1i_2}_{i_3i_4} t^{i_3i_4}_{a_1a_2} + -1 \times A^{a_1a_2}_{i_1i_2} g^{i_1a_3}_{i_3a_1} t^{i_2i_3}_{a_2a_3} + \frac{1}{4} \times A^{a_1a_2}_{i_1i_2} g^{i_1i_2}_{a_1a_2} + \\
\frac{1}{2} \times A^{a_1a_2}_{i_1i_2} g^{i_1a_3}_{a_1a_2} t^{i_2}_{a_3} + \frac{1}{2} \times A^{a_1a_2}_{i_1i_2} g^{i_1i_2}_{i_3a_1} t^{i_3}_{a_2} + \frac{1}{2} \times A^{a_1a_2}_{i_1i_2} f^{a_3}_{i_3} t^{i_1}_{a_3} t^{i_2i_3}_{a_1a_2} + \frac{1}{2} \times A^{a_1a_2}_{i_1i_2} f^{a_3}_{i_3} t^{i_3}_{a_1} t^{i_1i_2}_{a_2a_3} + \frac{1}{4} \times A^{a_1a_2}_{i_1i_2} g^{i_1a_3}_{i_3i_4} t^{i_2}_{a_3} t^{i_3i_4}_{a_1a_2} + \\
-\frac{1}{2} \times A^{a_1a_2}_{i_1i_2} g^{i_1a_3}_{i_3i_4} t^{i_3}_{a_3} t^{i_2i_4}_{a_1a_2} + -1 \times A^{a_1a_2}_{i_1i_2} g^{i_1a_3}_{i_3i_4} t^{i_3}_{a_1} t^{i_2i_4}_{a_2a_3} + -1 \times A^{a_1a_2}_{i_1i_2} g^{a_3a_4}_{i_3a_1} t^{i_1}_{a_3} t^{i_2i_3}_{a_2a_4} + -\frac{1}{2} \times A^{a_1a_2}_{i_1i_2} g^{a_3a_4}_{i_3a_1} t^{i_3}_{a_3} t^{i_1i_2}_{a_2a_4} + \frac{1}{4} \times A^{a_1a_2}_{i_1i_2} g^{a_3a_4}_{i_3a_1} t^{i_3}_{a_2} t^{i_1i_2}_{a_3a_4} + \\
\frac{1}{4} \times A^{a_1a_2}_{i_1i_2} g^{a_3a_4}_{a_1a_2} t^{i_1}_{a_3} t^{i_2}_{a_4} + A^{a_1a_2}_{i_1i_2} g^{i_1a_3}_{i_3a_1} t^{i_3}_{a_2} t^{i_2}_{a_3} + \frac{1}{4} \times A^{a_1a_2}_{i_1i_2} g^{i_1i_2}_{i_3i_4} t^{i_3}_{a_1} t^{i_4}_{a_2} + \frac{1}{2} \times A^{a_1a_2}_{i_1i_2} g^{a_3a_4}_{i_3i_4} t^{i_1i_3}_{a_1a_3} t^{i_2i_4}_{a_2a_4} + -\frac{1}{4} \times A^{a_1a_2}_{i_1i_2} g^{a_3a_4}_{i_3i_4} t^{i_3i_4}_{a_1a_3} t^{i_1i_2}_{a_2a_4} + \\
\frac{1}{16} \times A^{a_1a_2}_{i_1i_2} g^{a_3a_4}_{i_3i_4} t^{i_3i_4}_{a_1a_2} t^{i_1i_2}_{a_3a_4} + -\frac{1}{4} \times A^{a_1a_2}_{i_1i_2} g^{a_3a_4}_{i_3i_4} t^{i_1i_3}_{a_1a_2} t^{i_2i_4}_{a_3a_4} + \frac{1}{4} \times A^{a_1a_2}_{i_1i_2} g^{a_3a_4}_{i_3i_4} t^{i_3}_{a_3} t^{i_1i_2i_4}_{a_1a_2a_4} + \frac{1}{4} \times A^{a_1a_2}_{i_1i_2} g^{a_3a_4}_{i_3i_4} t^{i_1}_{a_3} t^{i_2i_3i_4}_{a_1a_2a_4} + \frac{1}{4} \times A^{a_1a_2}_{i_1i_2} g^{a_3a_4}_{i_3i_4} t^{i_3}_{a_1} t^{i_1i_2i_4}_{a_2a_3a_4} + \\
\frac{1}{8} \times A^{a_1a_2}_{i_1i_2} g^{a_3a_4}_{i_3i_4} t^{i_1}_{a_3} t^{i_2}_{a_4} t^{i_3i_4}_{a_1a_2} + \frac{1}{2} \times A^{a_1a_2}_{i_1i_2} g^{a_3a_4}_{i_3i_4} t^{i_3}_{a_3} t^{i_1}_{a_4} t^{i_2i_4}_{a_1a_2} + \frac{1}{8} \times A^{a_1a_2}_{i_1i_2} g^{a_3a_4}_{i_3i_4} t^{i_3}_{a_1} t^{i_4}_{a_2} t^{i_1i_2}_{a_3a_4} + -1 \times A^{a_1a_2}_{i_1i_2} g^{a_3a_4}_{i_3i_4} t^{i_3}_{a_1} t^{i_1}_{a_3} t^{i_2i_4}_{a_2a_4} + -\frac{1}{2} \times A^{a_1a_2}_{i_1i_2} g^{a_3a_4}_{i_3i_4} t^{i_3}_{a_1} t^{i_4}_{a_3} t^{i_1i_2}_{a_2a_4} + \\
\frac{1}{2} \times A^{a_1a_2}_{i_1i_2} g^{a_3a_4}_{i_3a_1} t^{i_3}_{a_2} t^{i_1}_{a_3} t^{i_2}_{a_4} + \frac{1}{2} \times A^{a_1a_2}_{i_1i_2} g^{i_1a_3}_{i_3i_4} t^{i_3}_{a_1} t^{i_4}_{a_2} t^{i_2}_{a_3})
\end{aligned}$$

dominant tensor contraction

$$\begin{aligned}
\langle 0|\hat{A}_3\bar{H}\,|0\rangle = (\frac{1}{24} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{i_1i_2}_{i_4i_5} t^{i_3i_4i_5}_{a_1a_2a_3} + -\frac{1}{4} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{i_1i_2}_{i_4a_1} t^{i_3i_4}_{a_2a_3} + \frac{1}{12} \times A^{a_1a_2a_3}_{i_1i_2i_3} f^{a_4}_{a_1} t^{i_1i_2i_3}_{a_2a_3a_4} + -\frac{1}{12} \times A^{a_1a_2a_3}_{i_1i_2i_3} f^{i_1}_{i_4} t^{i_2i_3i_4}_{a_1a_2a_3} + -\frac{1}{4} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{i_1a_4}_{a_1a_2} t^{i_2i_3}_{a_3a_4} + \\
-\frac{1}{4} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{i_1a_4}_{i_4a_1} t^{i_2i_3i_4}_{a_2a_3a_4} + \boxed{\frac{1}{24} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{a_1a_2} t^{i_1i_2i_3}_{a_3a_4a_5}} + -\frac{1}{4} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4i_5} t^{i_4}_{a_1} t^{i_5}_{a_2} t^{i_1}_{a_4} t^{i_2i_3}_{a_3a_5} + -\frac{1}{4} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4i_5} t^{i_4}_{a_1} t^{i_1}_{a_4} t^{i_2}_{a_5} t^{i_3i_5}_{a_2a_3} + \frac{1}{4} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4i_5} t^{i_4}_{a_4} t^{i_1i_5}_{a_1a_2} t^{i_2i_3}_{a_3a_5} + \\
-\frac{1}{8} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4i_5} t^{i_1}_{a_4} t^{i_4i_5}_{a_1a_2} t^{i_2i_3}_{a_3a_5} + \frac{1}{2} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4i_5} t^{i_1}_{a_4} t^{i_2i_4}_{a_1a_2} t^{i_3i_5}_{a_3a_5} + \frac{1}{2} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4i_5} t^{i_4}_{a_1} t^{i_1i_2}_{a_2a_4} t^{i_3i_5}_{a_3a_5} + -\frac{1}{8} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4i_5} t^{i_4}_{a_1} t^{i_1i_5}_{a_2a_3} t^{i_2i_3}_{a_4a_5} + \frac{1}{24} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4i_5} t^{i_1}_{a_4} t^{i_2}_{a_5} t^{i_3i_4i_5}_{a_1a_2a_3} + \\
\frac{1}{12} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4i_5} t^{i_1}_{a_4} t^{i_4}_{a_5} t^{i_2i_3i_5}_{a_1a_2a_3} + \frac{1}{24} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4i_5} t^{i_4}_{a_1} t^{i_5}_{a_2} t^{i_1i_2i_3}_{a_3a_4a_5} + -\frac{1}{4} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4i_5} t^{i_4}_{a_1} t^{i_1}_{a_4} t^{i_2i_3i_5}_{a_2a_3a_5} + \frac{1}{12} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4i_5} t^{i_4}_{a_1} t^{i_5}_{a_4} t^{i_1i_2i_3}_{a_2a_3a_5} + -\frac{1}{4} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4a_1} t^{i_1}_{a_4} t^{i_2}_{a_5} t^{i_3i_4}_{a_2a_3} + \\
-\frac{1}{2} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4a_1} t^{i_4}_{a_2} t^{i_1}_{a_4} t^{i_2i_3}_{a_3a_5} + -\frac{1}{2} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{i_1a_4}_{i_4i_5} t^{i_4}_{a_1} t^{i_2}_{a_4} t^{i_3i_5}_{a_2a_3} + -\frac{1}{4} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{i_1a_4}_{i_4i_5} t^{i_4}_{a_1} t^{i_5}_{a_2} t^{i_2i_3}_{a_3a_4} + -\frac{1}{4} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{a_1a_2} t^{i_1}_{a_4} t^{i_2i_3}_{a_3a_5} + -\frac{1}{2} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{i_1a_4}_{i_4a_1} t^{i_4}_{a_2} t^{i_2i_3}_{a_3a_4} \\
-\frac{1}{2} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{i_1a_4}_{i_4a_1} t^{i_2}_{a_4} t^{i_3i_4}_{a_2a_3} + \frac{1}{48} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4i_5} t^{i_1i_2}_{a_4a_5} t^{i_3i_4i_5}_{a_1a_2a_3} + \frac{1}{24} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4i_5} t^{i_1i_4}_{a_4a_5} t^{i_2i_3i_5}_{a_1a_2a_3} + \frac{1}{24} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4i_5} t^{i_4i_5}_{a_1a_4} t^{i_1i_2i_3}_{a_2a_3a_5} + \frac{1}{4} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4i_5} t^{i_1i_4}_{a_1a_4} t^{i_2i_3i_5}_{a_2a_3a_5} + \\
\frac{1}{8} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4i_5} t^{i_1i_2}_{a_1a_4} t^{i_3i_4i_5}_{a_2a_3a_5} + \frac{1}{8} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4i_5} t^{i_1i_4}_{a_1a_2} t^{i_2i_3i_5}_{a_3a_4a_5} + \frac{1}{48} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4i_5} t^{i_4i_5}_{a_1a_2} t^{i_1i_2i_3}_{a_3a_4a_5} + -\frac{1}{8} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{i_1a_4}_{i_4i_5} t^{i_2i_3}_{a_1a_4} t^{i_4i_5}_{a_2a_3} + \frac{1}{2} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{i_1a_4}_{i_4i_5} t^{i_2i_4}_{a_1a_4} t^{i_3i_5}_{a_2a_3} + \\
-\frac{1}{12} \times A^{a_1a_2a_3}_{i_1i_2i_3} f^{a_4}_{i_4} t^{i_4}_{a_1} t^{i_1i_2i_3}_{a_2a_3a_4} + -\frac{1}{12} \times A^{a_1a_2a_3}_{i_1i_2i_3} f^{a_4}_{i_4} t^{i_1}_{a_4} t^{i_2i_3i_4}_{a_1a_2a_3} + \frac{1}{12} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{i_1a_4}_{i_4i_5} t^{i_2}_{a_4} t^{i_3i_4i_5}_{a_1a_2a_3} + \frac{1}{12} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{i_1a_4}_{i_4i_5} t^{i_4}_{a_4} t^{i_2i_3i_5}_{a_1a_2a_3} + -\frac{1}{4} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{i_1a_4}_{i_4i_5} t^{i_4}_{a_1} t^{i_2i_3i_5}_{a_2a_3a_4} + \\
-\frac{1}{8} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4a_1} t^{i_1i_4}_{a_2a_3} t^{i_2i_3}_{a_4a_5} + \frac{1}{2} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4a_1} t^{i_1i_2}_{a_2a_4} t^{i_3i_4}_{a_3a_5} + -\frac{1}{4} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{i_1i_2}_{i_4i_5} t^{i_4}_{a_1} t^{i_3i_5}_{a_2a_3} + \frac{1}{4} \times A^{a_1a_2a_3}_{i_1i_2i_3} f^{a_4}_{i_4} t^{i_1i_4}_{a_1a_2} t^{i_2i_3}_{a_3a_4} + \frac{1}{12} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4a_1} t^{i_4}_{a_2} t^{i_1i_2i_3}_{a_3a_4a_5} + \\
-\frac{1}{4} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4a_1} t^{i_1}_{a_4} t^{i_2i_3i_4}_{a_2a_3a_5} + \frac{1}{12} \times A^{a_1a_2a_3}_{i_1i_2i_3} g^{a_4a_5}_{i_4a_1} t^{i_4}_{a_4} t^{i_1i_2i_3}_{a_2a_3a_5})
\end{aligned}$$

same example



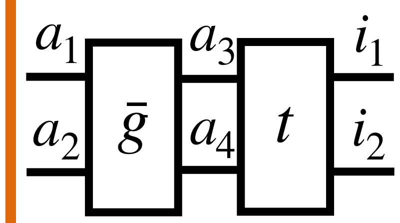
size(T3) = 2.6 PB

cost ~ 2.4 x 10^{20} FLOPs

Our Goal: Complexity Reduction by Factorizing Coulomb Tensor

O(N^6) => O(N^5) by factorizing g

$$\frac{1}{2}\bar{g}^{a_3 a_4}_{a_1 a_2} t^{i_1 i_2}_{a_3 a_4}$$

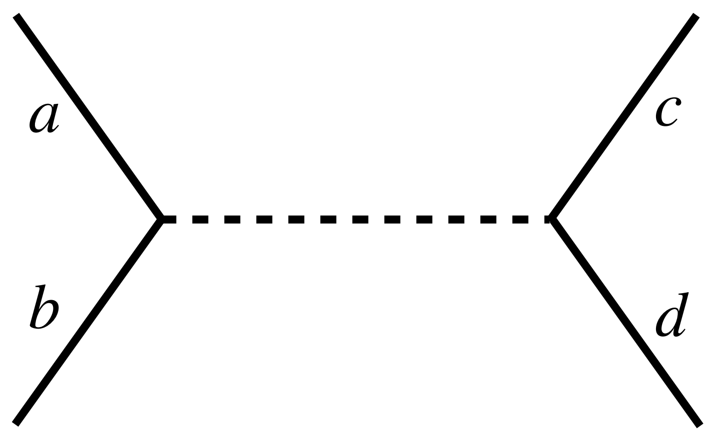


PPL, the dominant "diagram" in CCD, CCSD, and other models

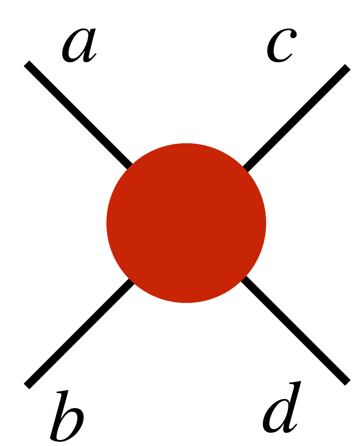
Coulomb Integral Tensor

$$g_{ab,cd} \equiv (ab\,|\,cd) \equiv \int \mathrm{d}\mathbf{r}\, a(\mathbf{r}) b(\mathbf{r}) \hat{V} c d(\mathbf{r}) \equiv \iint \mathrm{d}\mathbf{r}\ \mathrm{d}\mathbf{r}' a(\mathbf{r}) b(\mathbf{r})\,|\,\mathbf{r}-\mathbf{r}'\,|^{-1} c(\mathbf{r}') d(\mathbf{r}')$$

$$\hat{V} \equiv \int \frac{\mathrm{d}\mathbf{r}'}{|\,\mathbf{r}-\mathbf{r}'\,|} \times$$



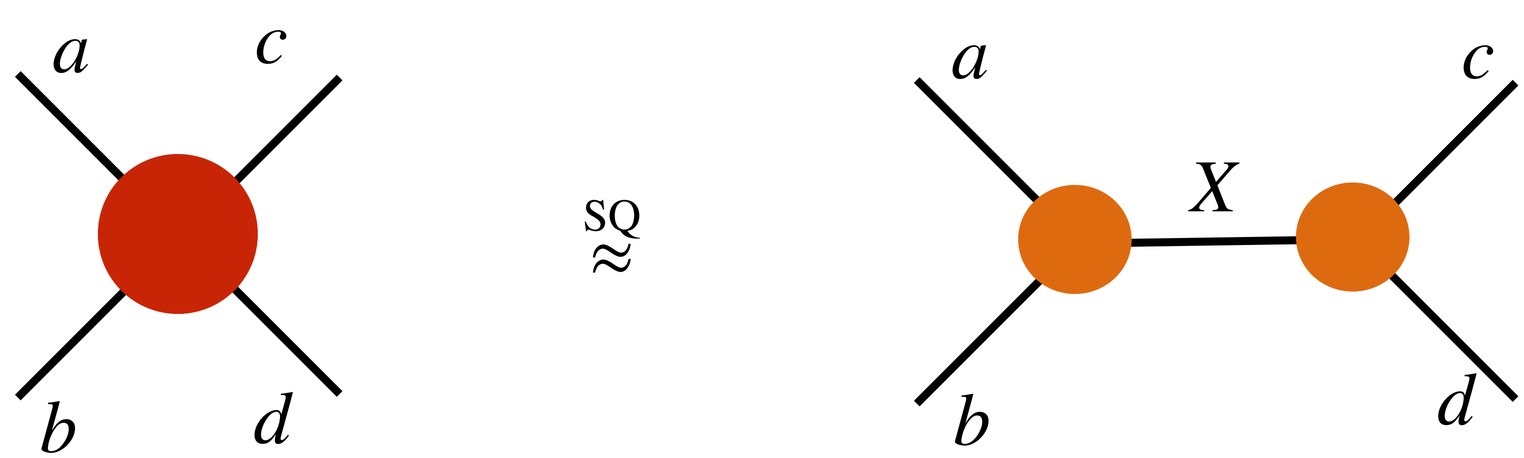
or



typically never computed explicitly, reconstructed on the fly from factorized form

“Square Root” Approximations of Coulomb Integrals

$$g_{ab,cd} \overset{\mathrm{SQ}}{\approx} \sum_X B_{ab,X} B_{cd,Y}$$



factors computed via physics-based or math-based approximation

(global) density fitting (DF)
aka resolution-of-identity (RI)

needs empirically-optimized AO basis

Boys, Shavitt, Whitten, Baerends, Ahlrichs, Almlof, Dunlap, …

Cholesky Decomposition (CD)

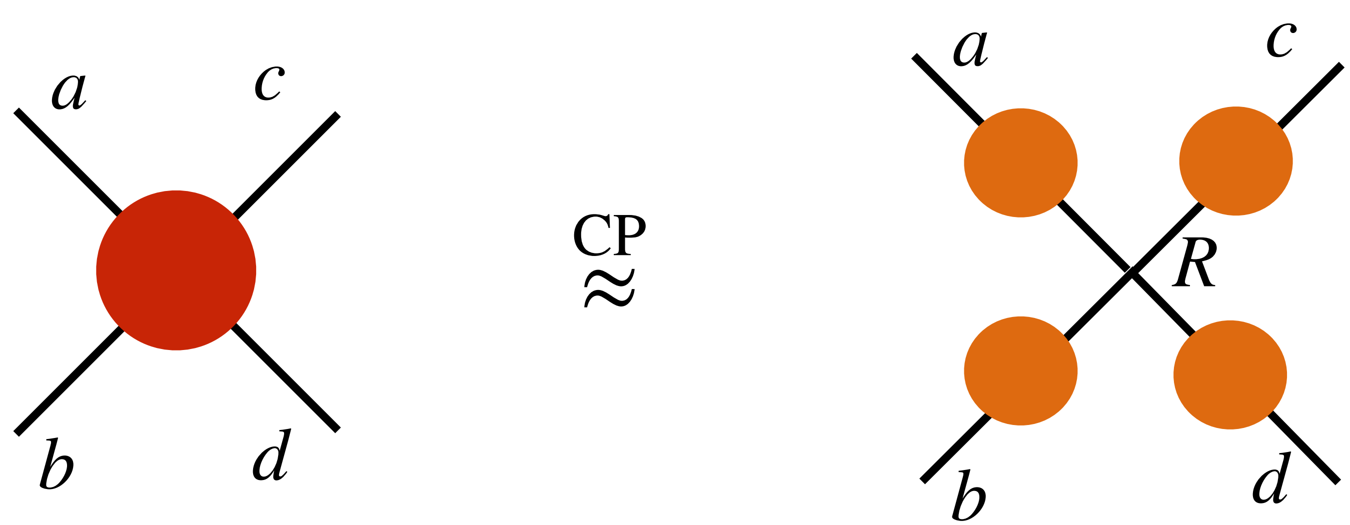
better error control than DF

Beebe, Linderberg, Lowdin + Koch, Lindh, Aquilante, …

does it make sense to use CP directly?

CP Approximation of Coulomb Integrals

$$g_{ab,cd} \overset{\mathrm{CP}}{\approx} (\mathbf{U}^{(1)})_{a,R}(\mathbf{U}^{(2)})_{b,R}(\mathbf{U}^{(3)})_{c,R}(\mathbf{U}^{(4)})_{d,R}$$

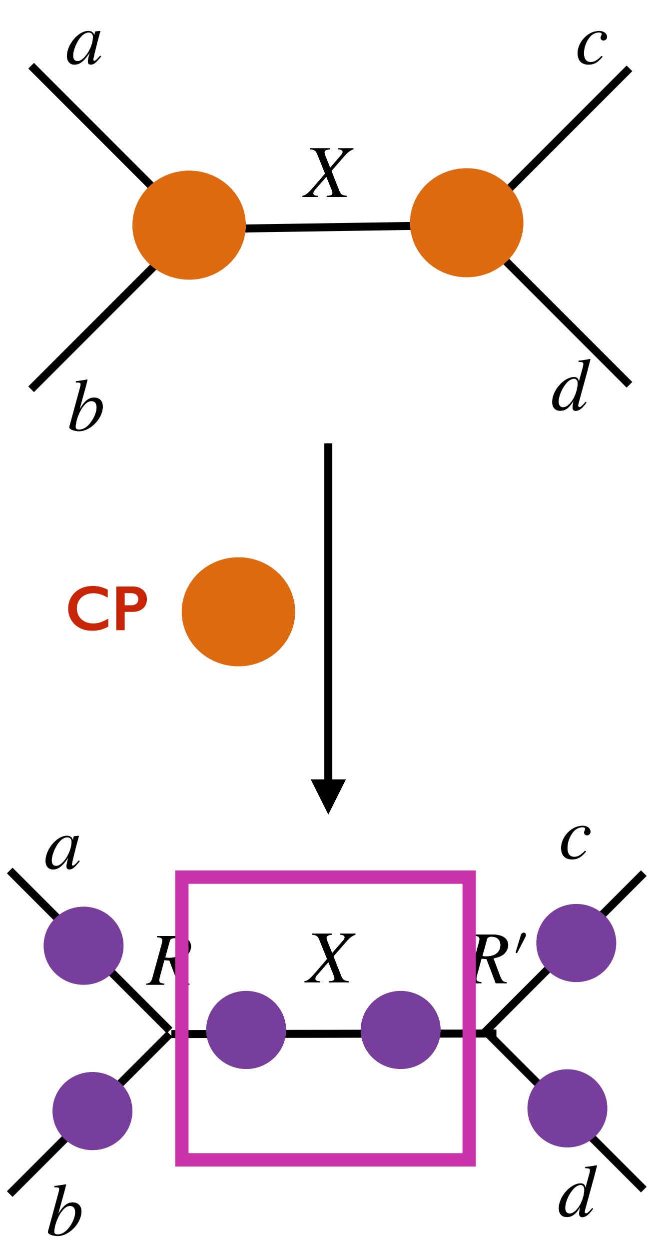


compute by Alternating Least Squares (ALS) or gradient-based methods at *O(N⁵)* cost

CP in QC is barely used: Auer, Espig, et al.; Khoromskaya, Khoromskij, et al.

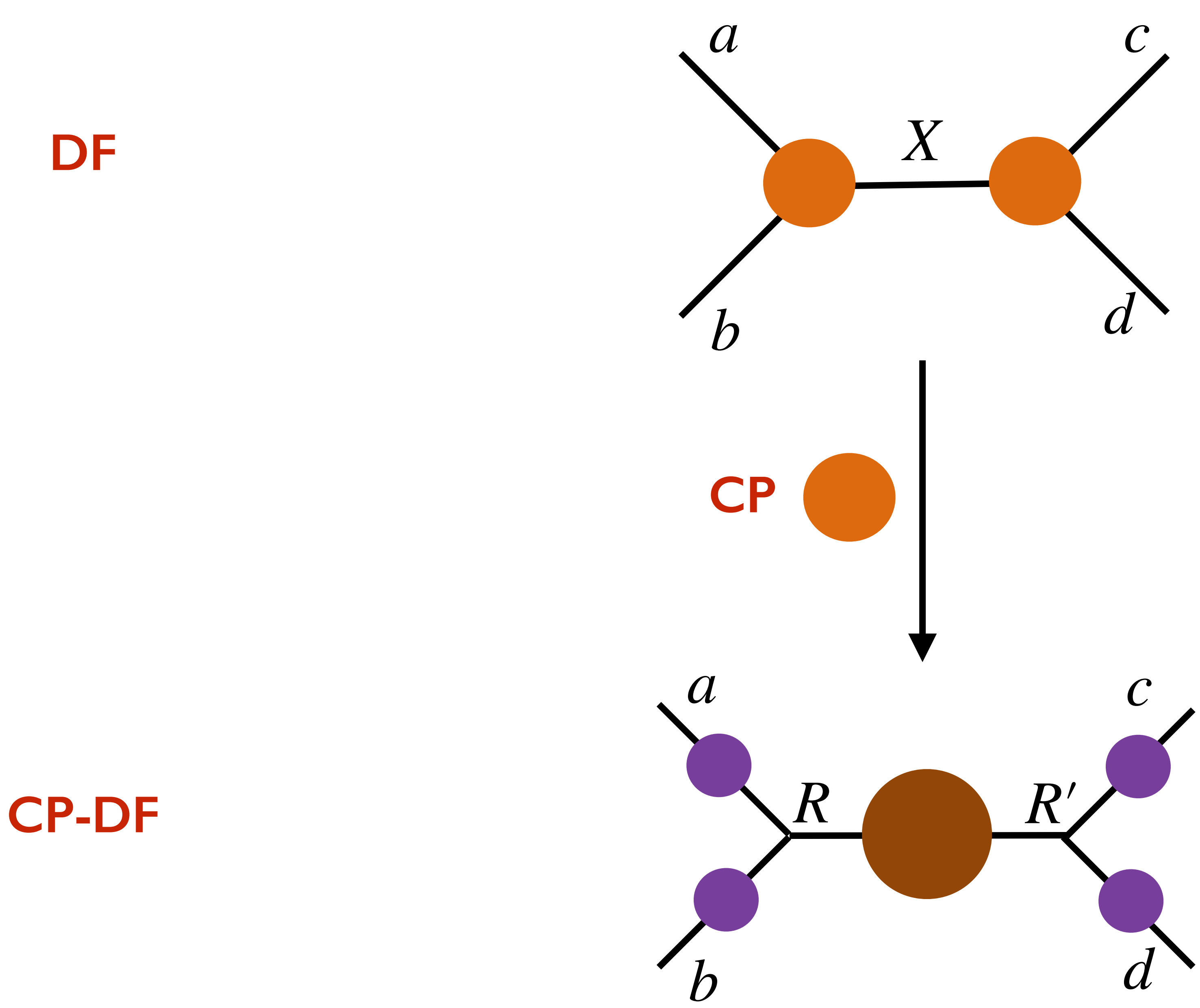
In This Talk: Only CP of SQ-Factorized Coulomb Integrals

DF
could use any SQ



CP-DF

CP-DF is related to THC

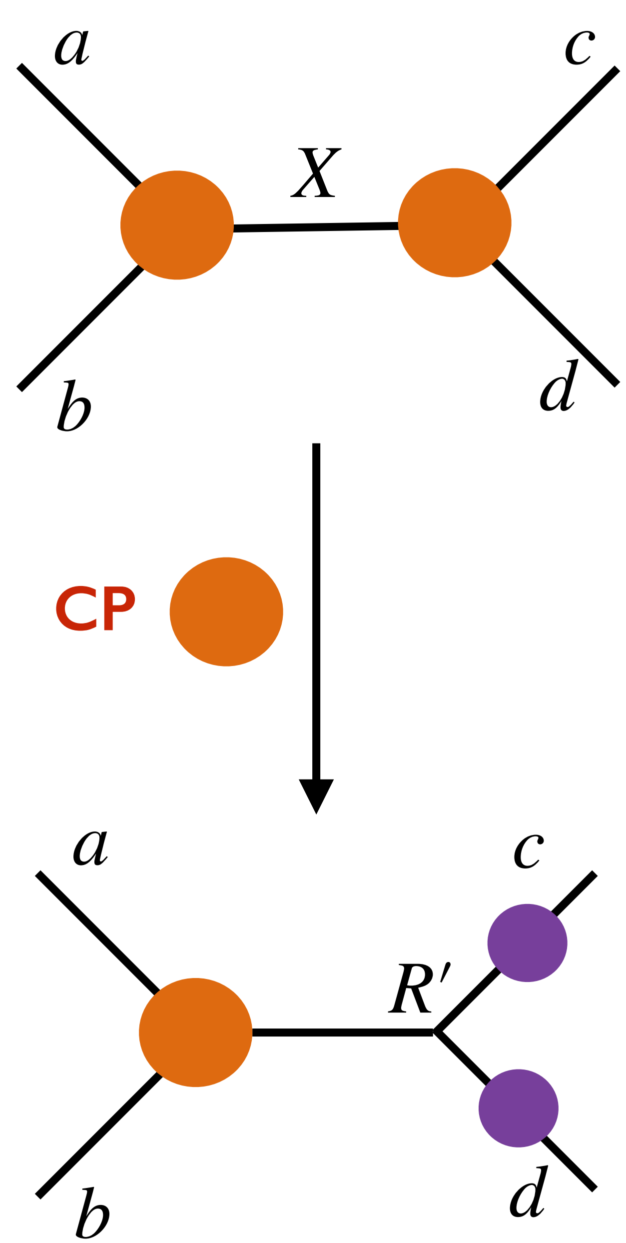


same topology as <u>Tensor Hypercontraction</u> (THC), but purely algebraic!

Related THC work: Hohenstein et al., (2012); Shenvi et al (2014); Hummel et al (2017); Schutski et al (2017)

Another Way: CP-PS

DF



CP-PS

same topology as pseudospectral (PS) factorization, but purely algebraic

Related PS work: Friesner (1985), Carter, Martinez, Ten-no (in F12), Neese (COSX), Klopper, Ochsenfeld (sn-K), …

Questions

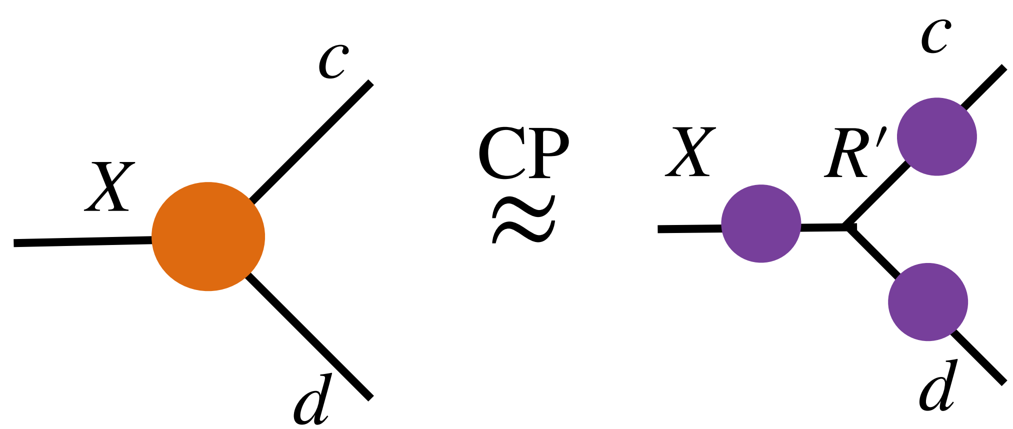
How to best use CP: 1 vertex (PS), 2 vertices (THC), or smth else?
I.e. what's the relationship between errors of PS and THC at same rank?

How large are CP ranks for reasonable errors?
I.e. can they be made smaller than the most compact PS/THC grids? (roughly, 10 x M)

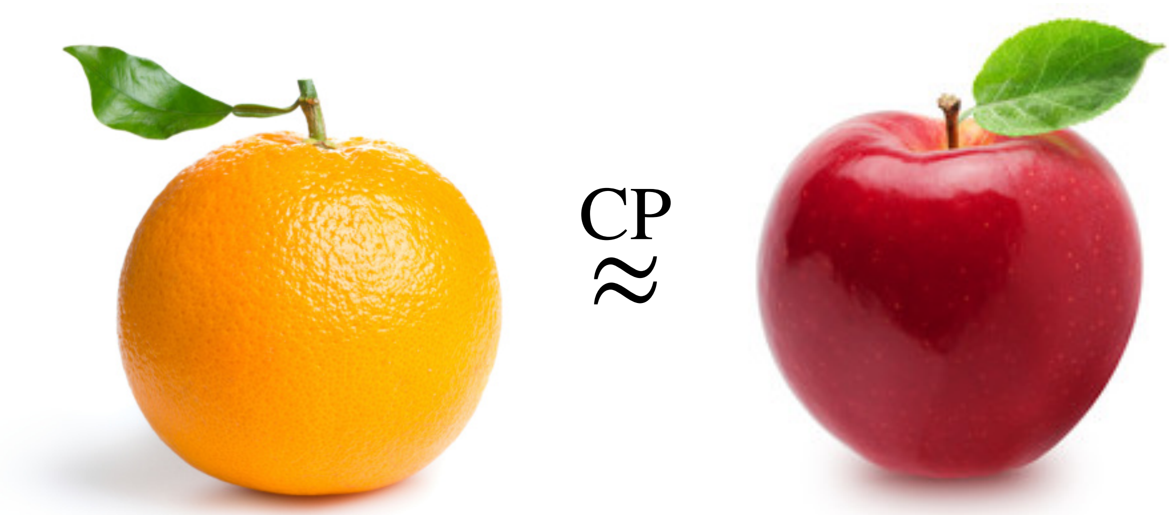
How practical are CP solvers?
Notoriously slow convergence typical of simple solvers (ALS)?
How prone to local minima?
Costs?

Are there practical benefits?
Only consider CCSD here

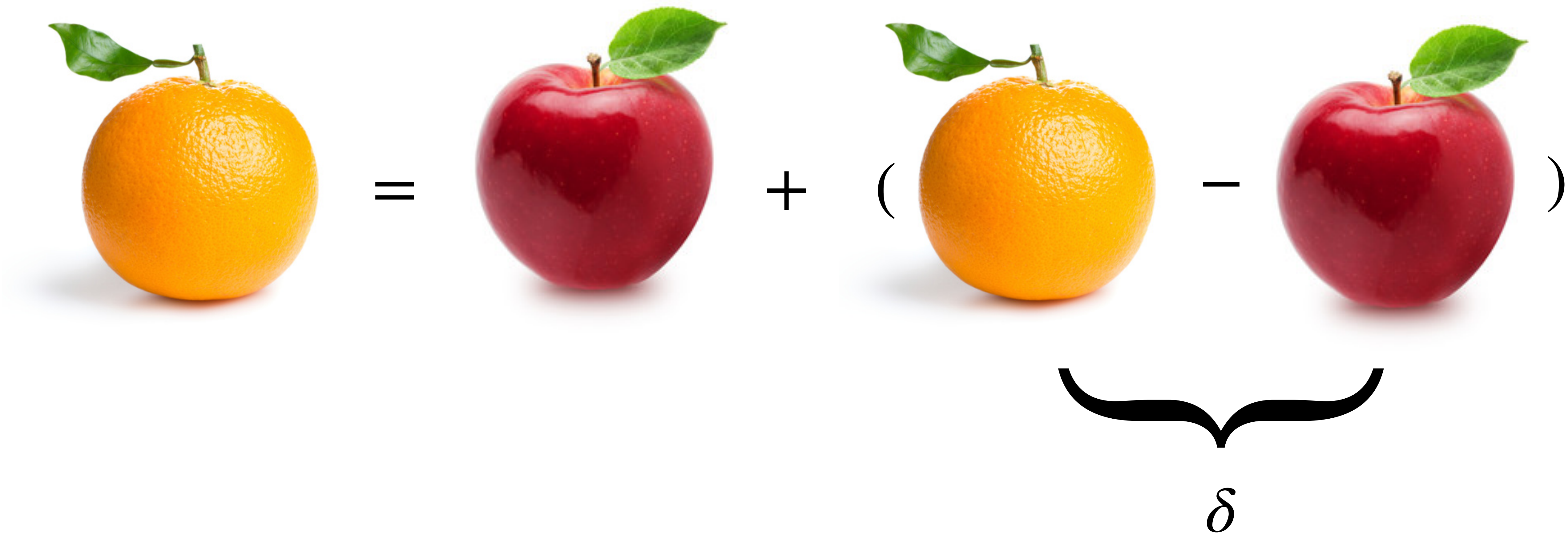
Error Analysis: PS vs THC



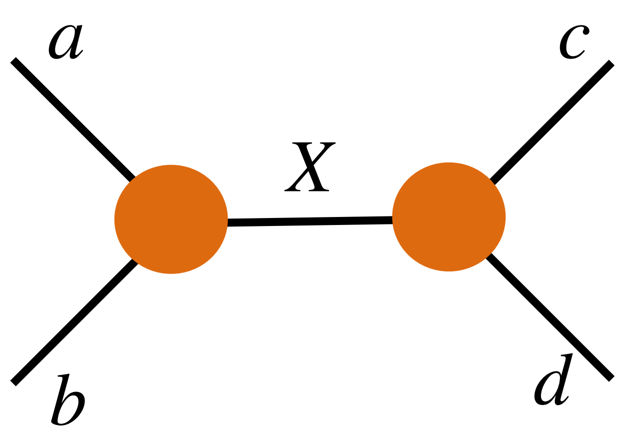
or



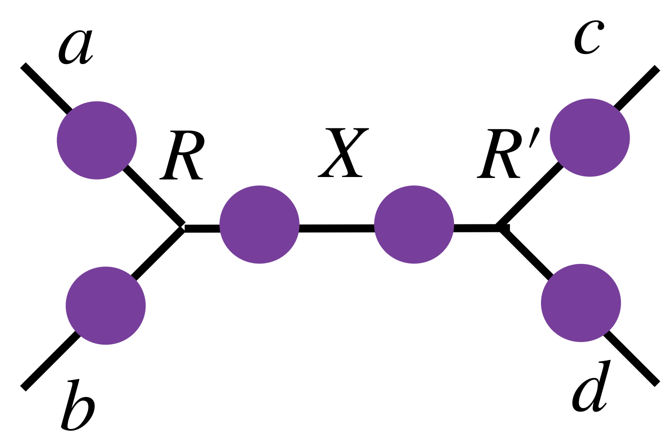
Error Analysis: PS vs THC



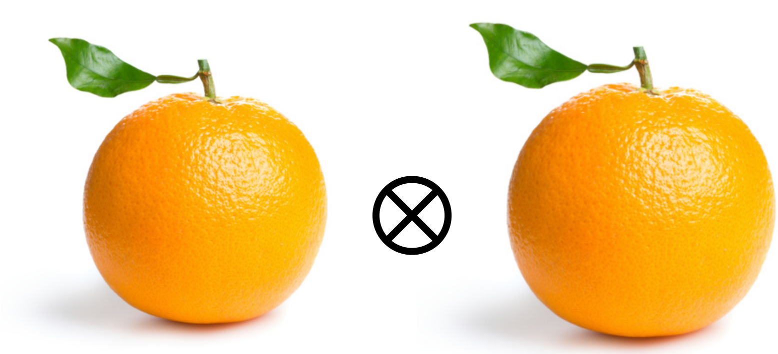
Error Analysis: PS vs THC



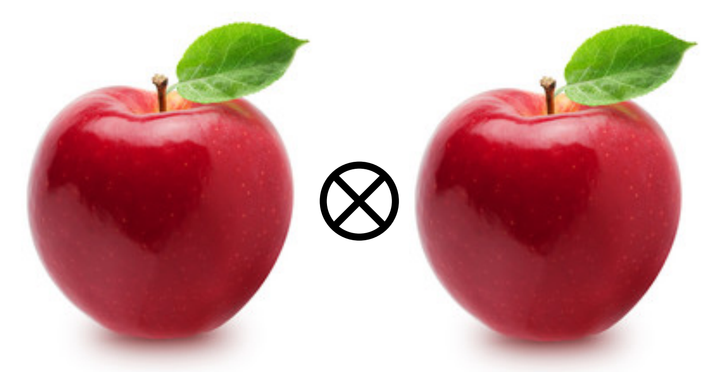
CP−DF ≈



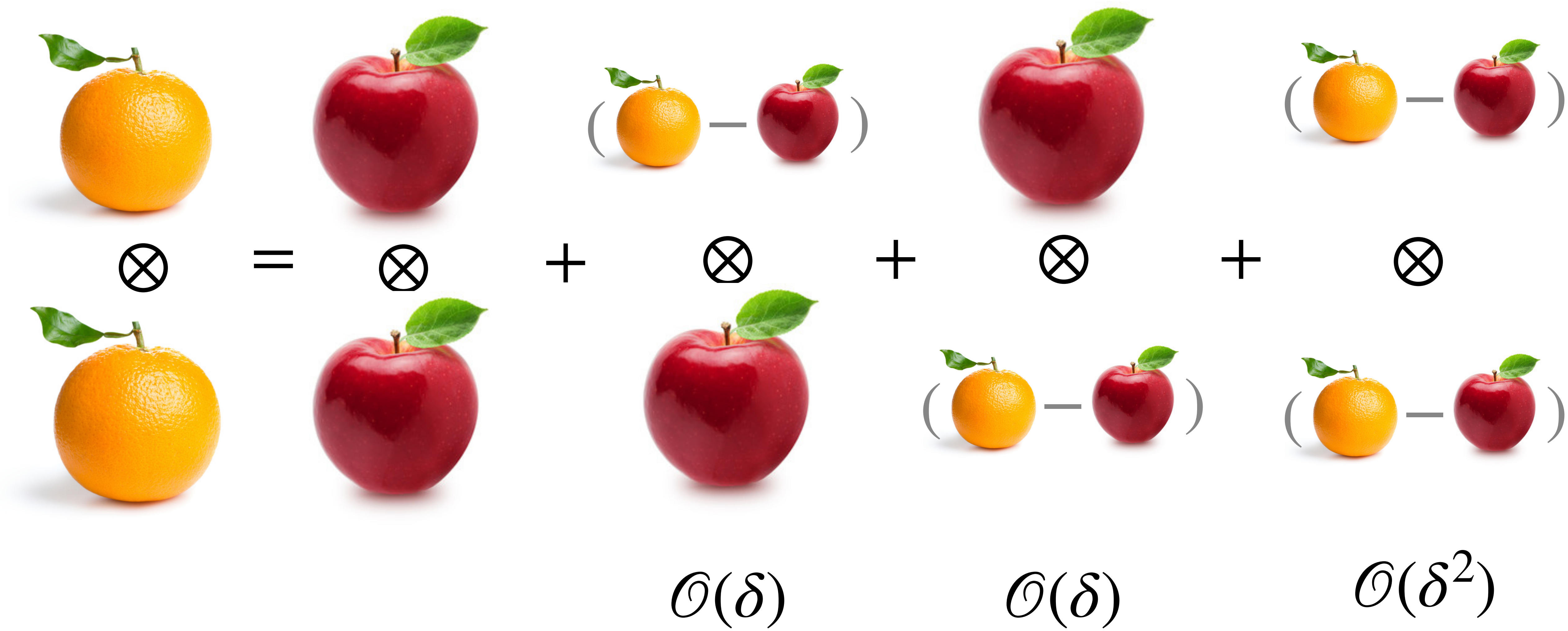
or



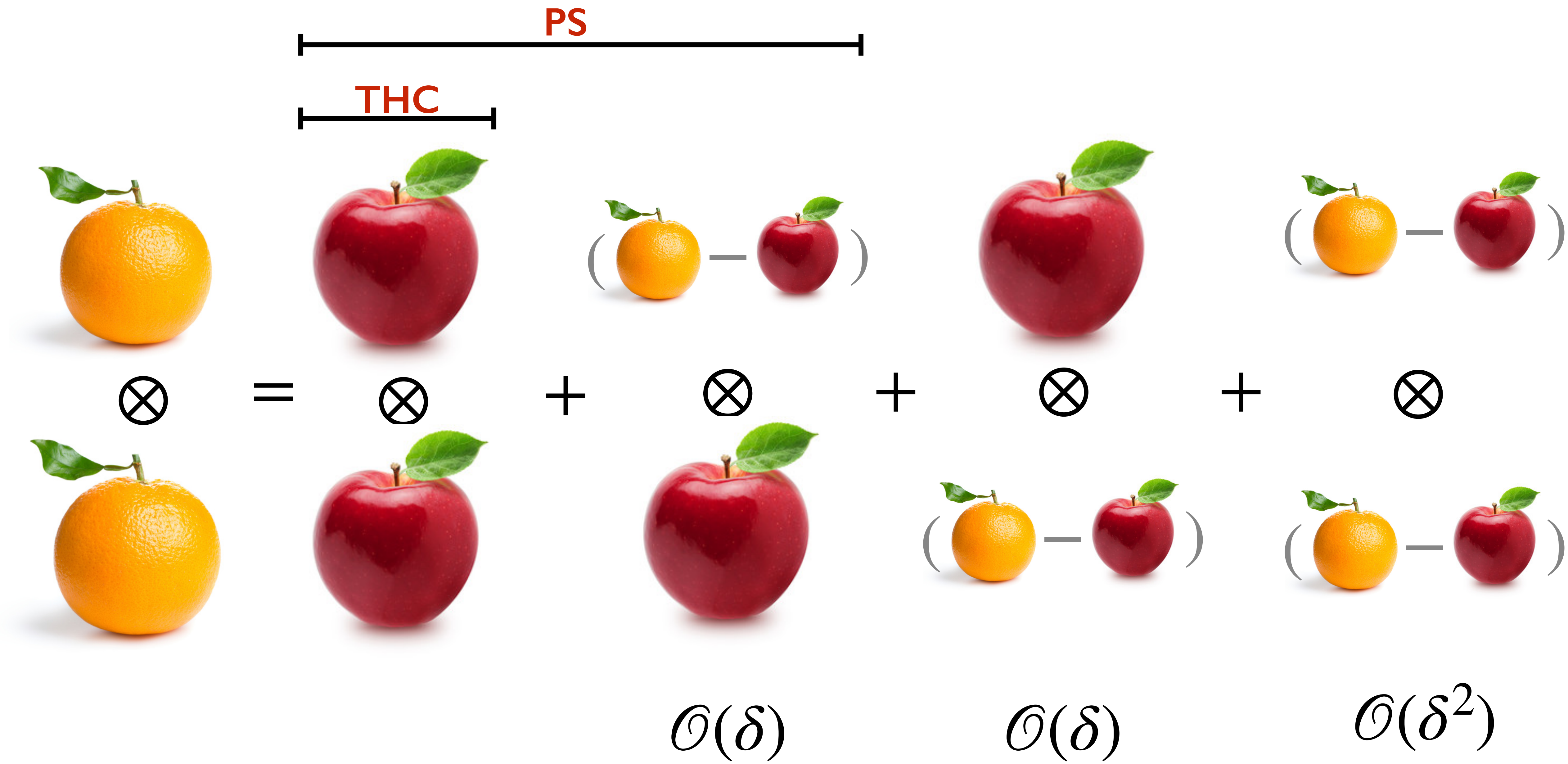
CP−DF ≈



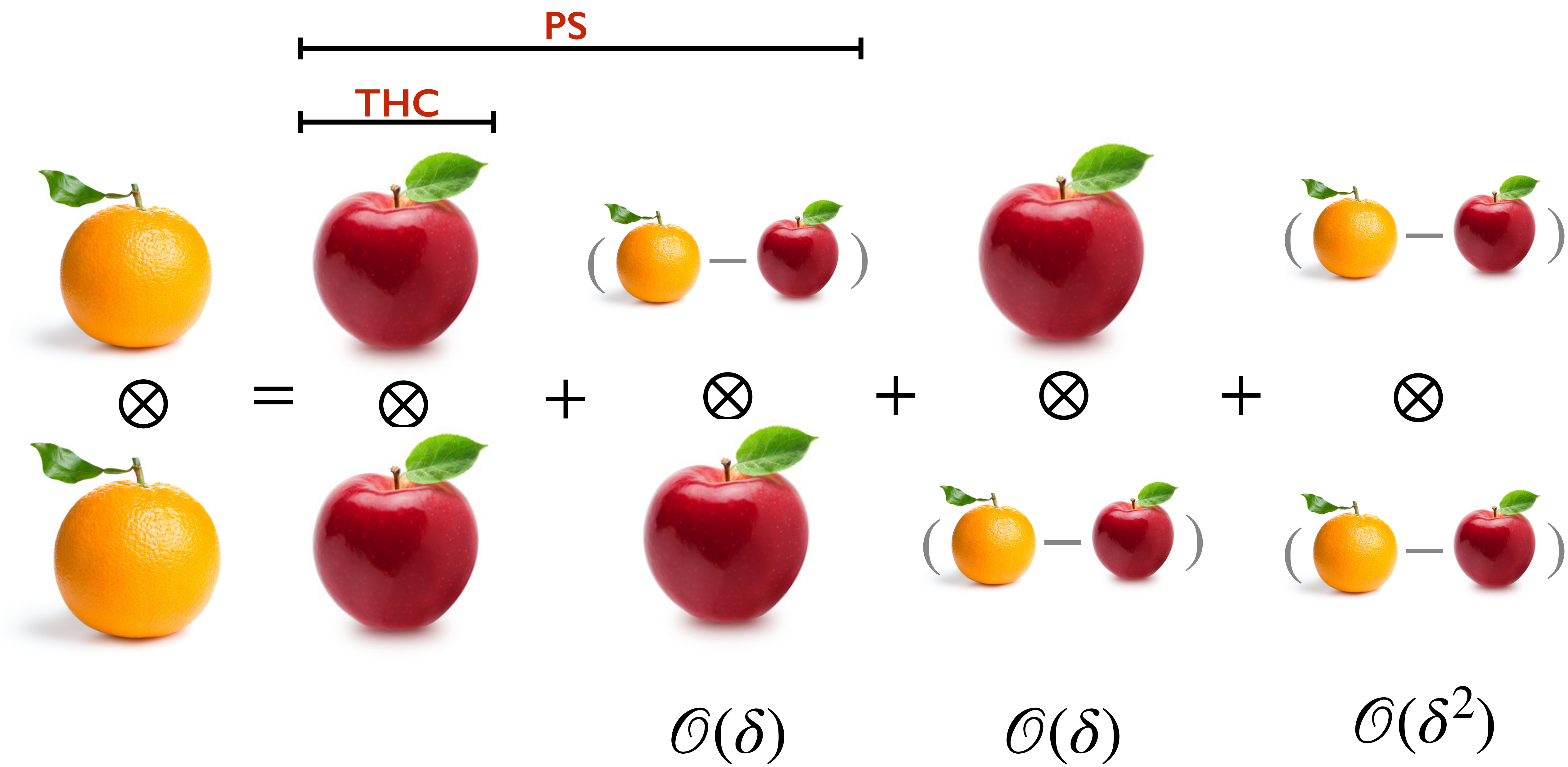
Error Analysis: PS vs THC



Error Analysis: PS vs THC

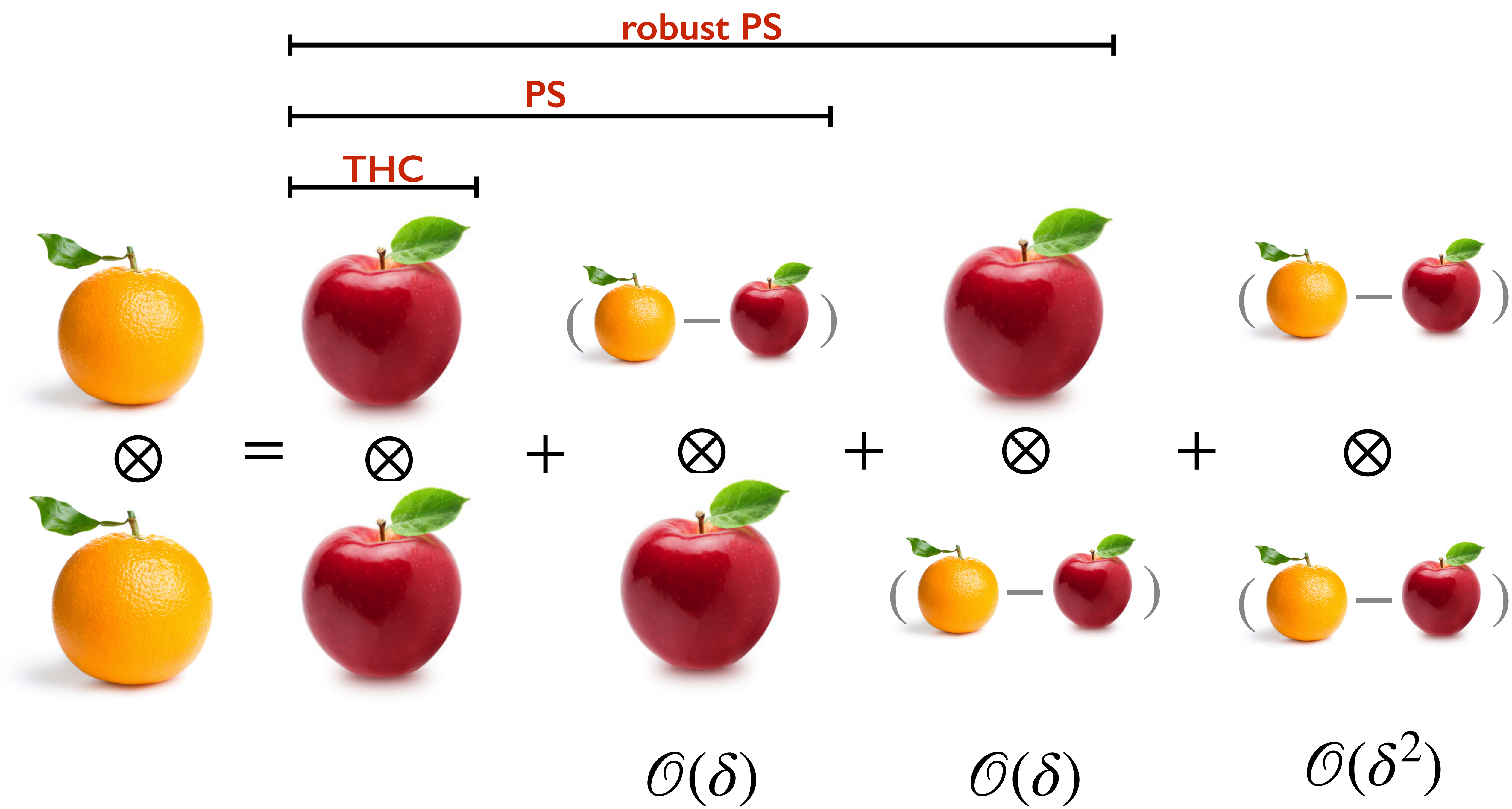


Error Analysis: PS vs THC



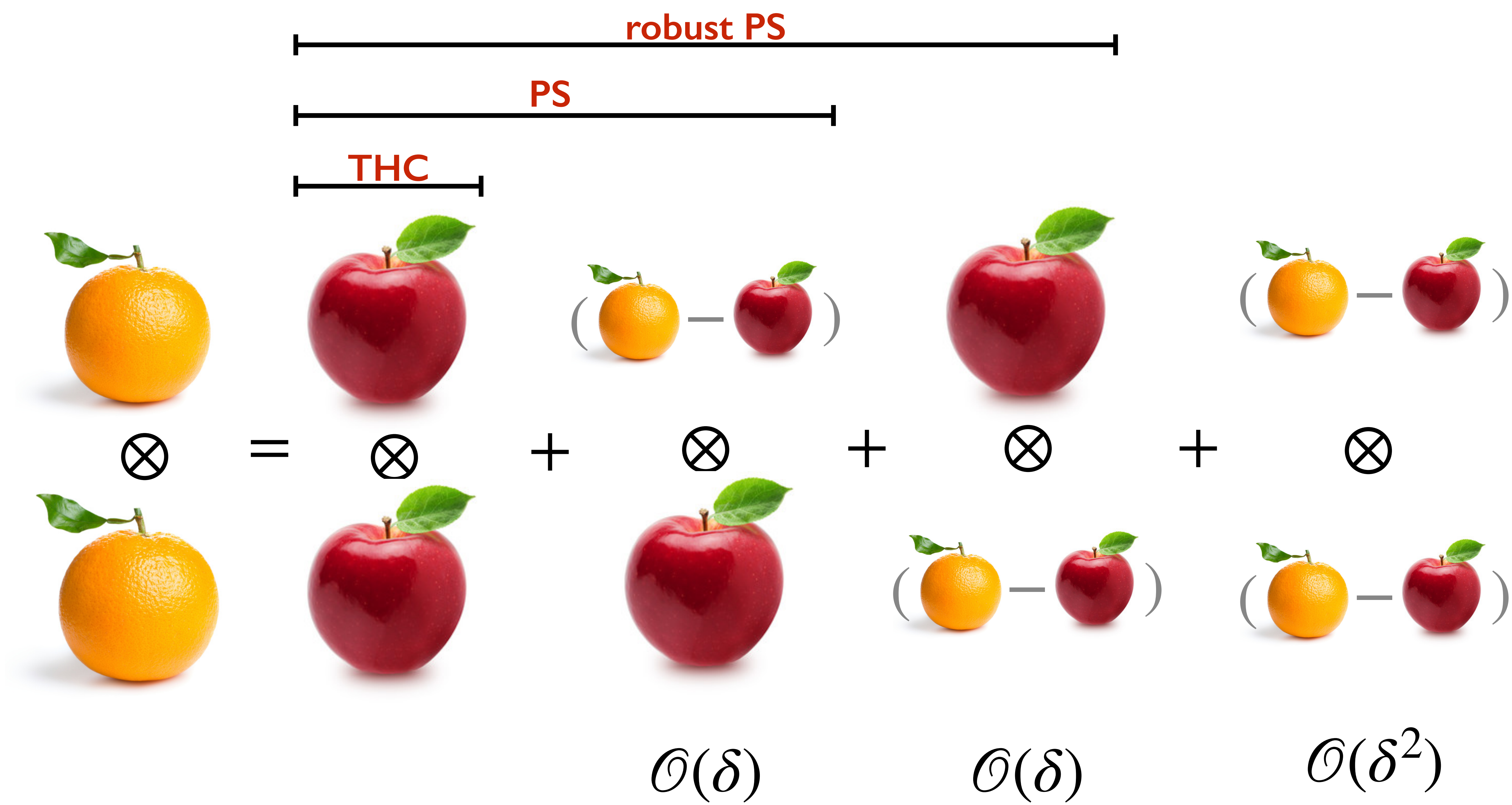
error(THC) ~ 2 x error(PS)

Error Analysis: PS vs THC



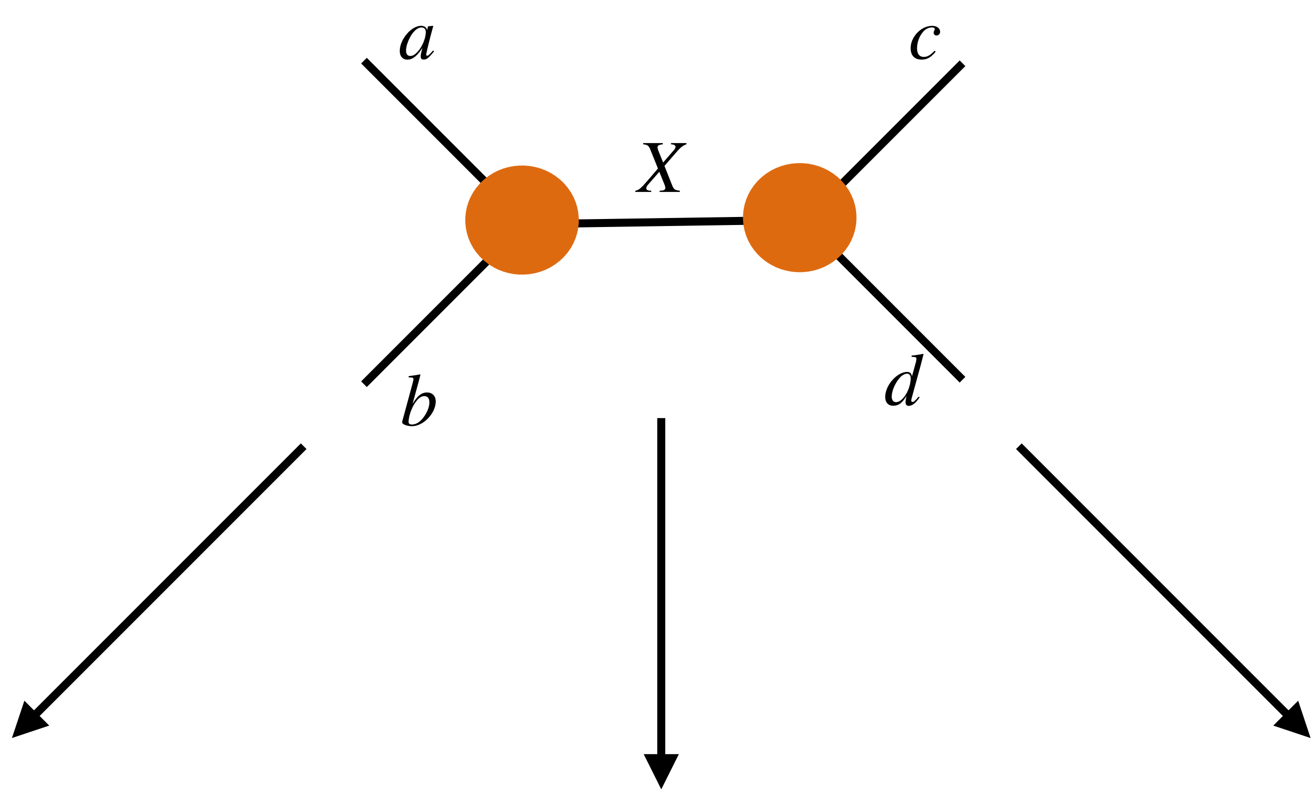
“robust” as in robust DF: Dunlap (2000)

Error Analysis: PS vs THC



error(THC) ~ 2 x error(PS) >> error(robust PS)

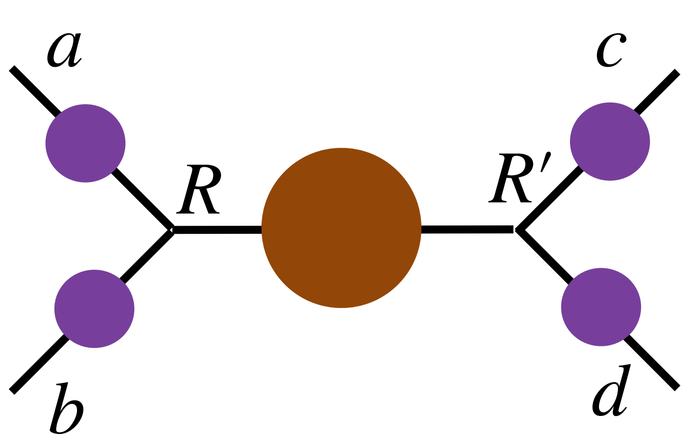
Error Analysis: PS vs THC

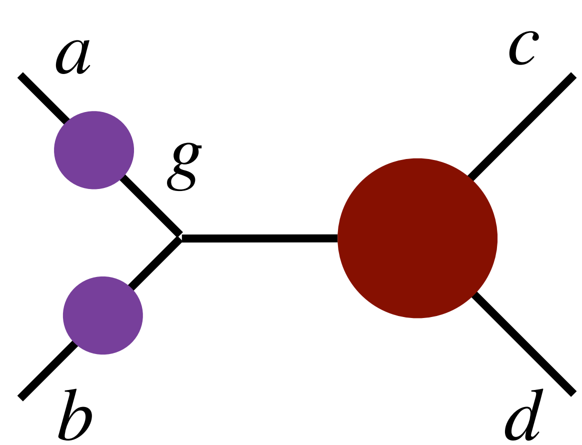


CP-DF: THC-like

CP-PS: PS-like

rCP-DF

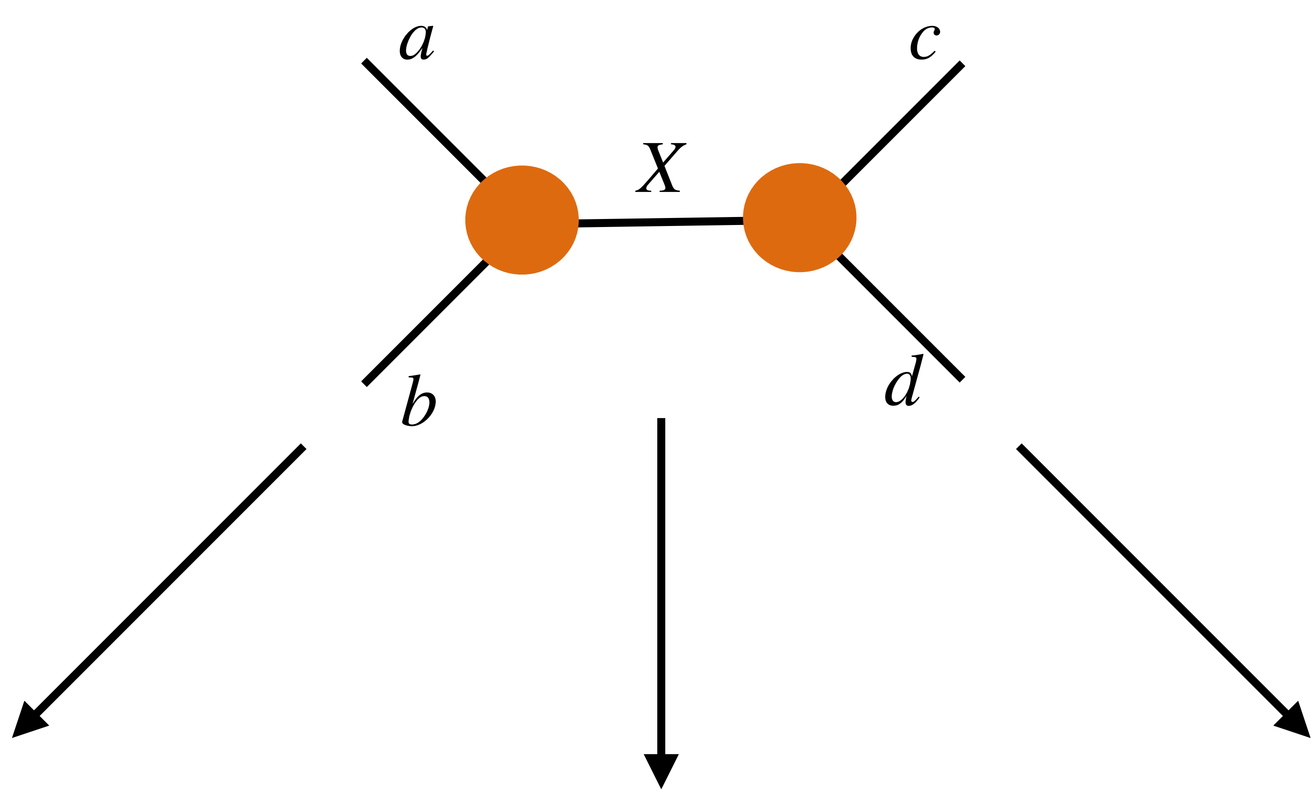




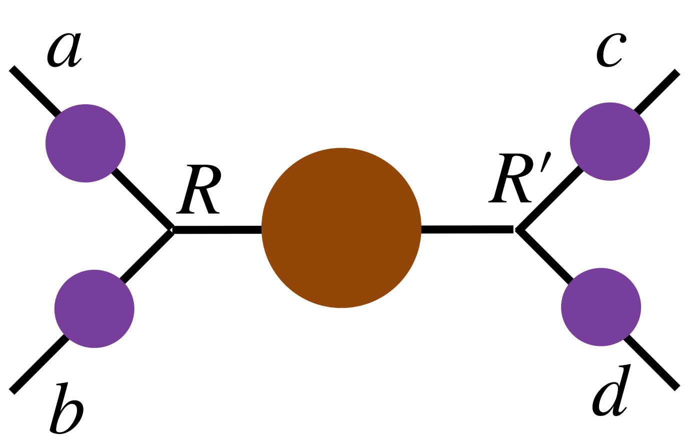
2 CP-PS - CP-DF

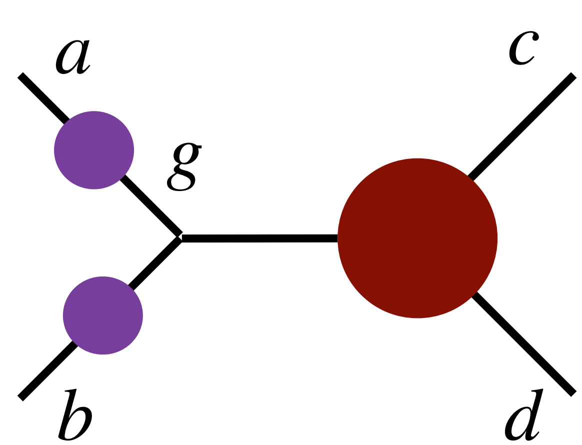
zero extra cost vs CP-DF/THC

Error Analysis: PS vs THC



CP-DF: THC-like	CP-PS: PS-like	rCP-DF



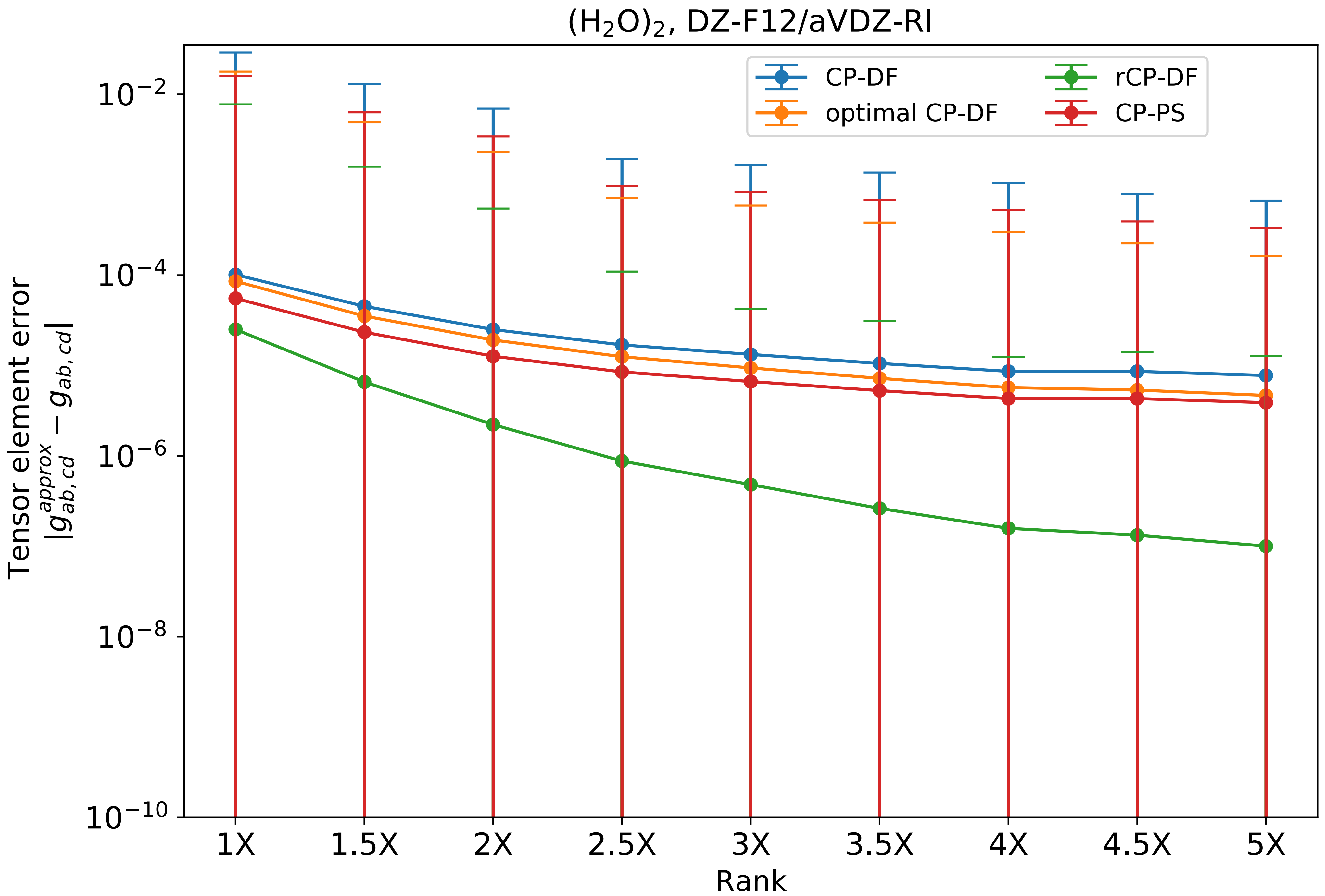


2 CP-PS - CP-DF

zero extra cost vs CP-DF/THC

N.B. can improve by relaxing the CP factors to minimize the error in the <u>full</u> network

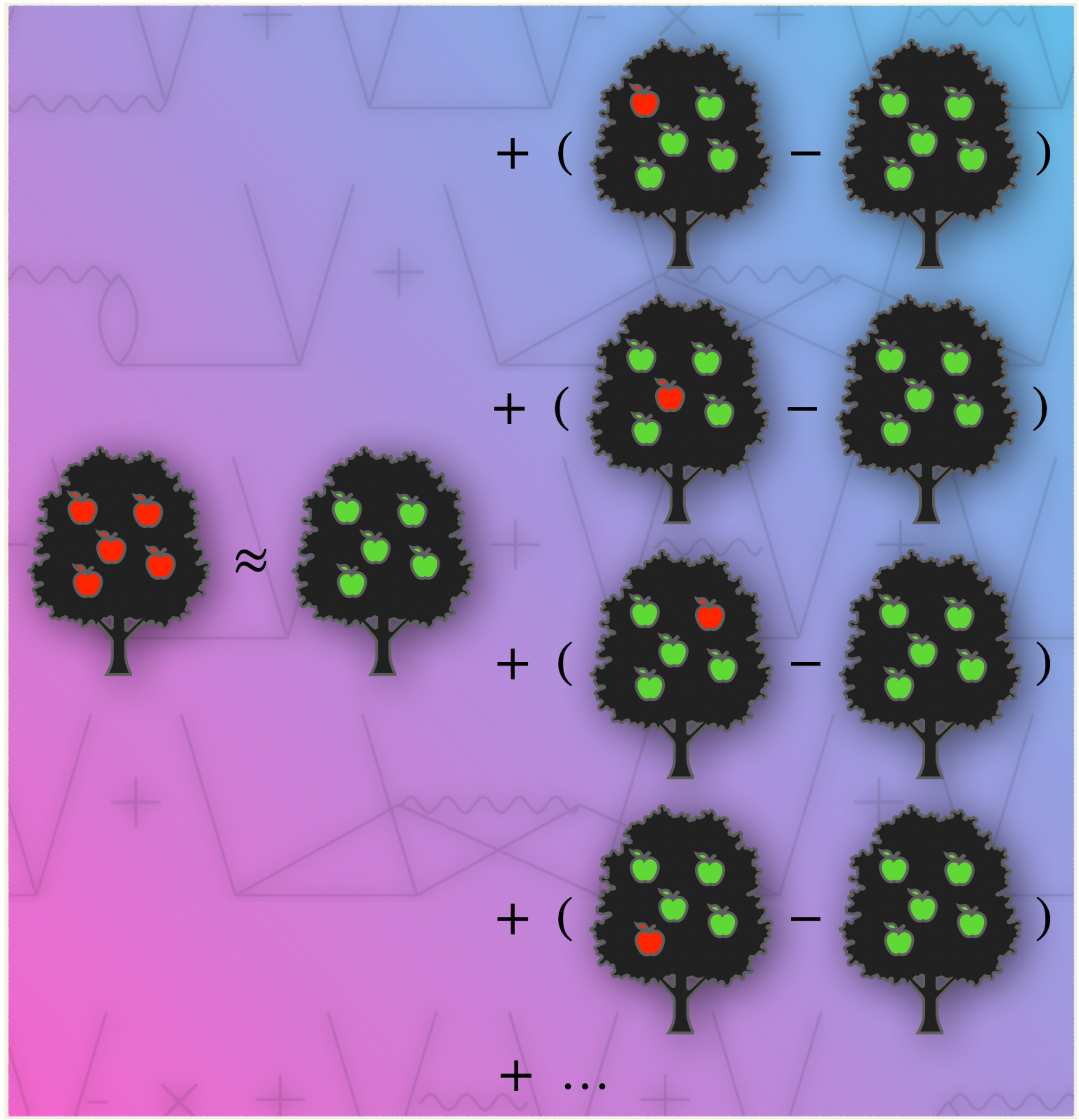
Error Analysis: PS vs THC



as expected: error(THC) ~ 2 x error(PS) >> error(robust PS)

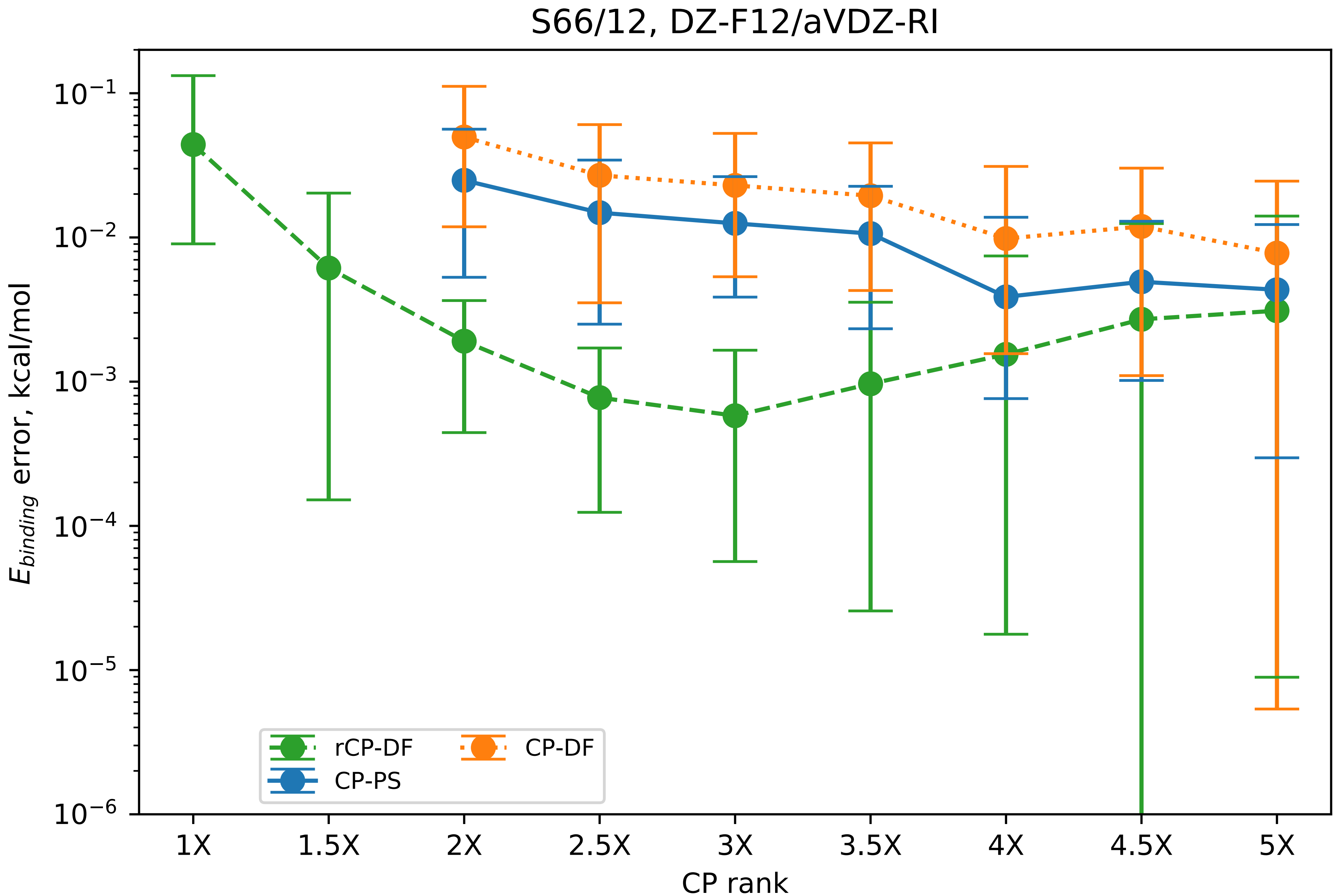
only minor improvement from relaxing CP factors in the CP-DF tensor network

Bigger Picture: Robust Approximation of Tensor Networks



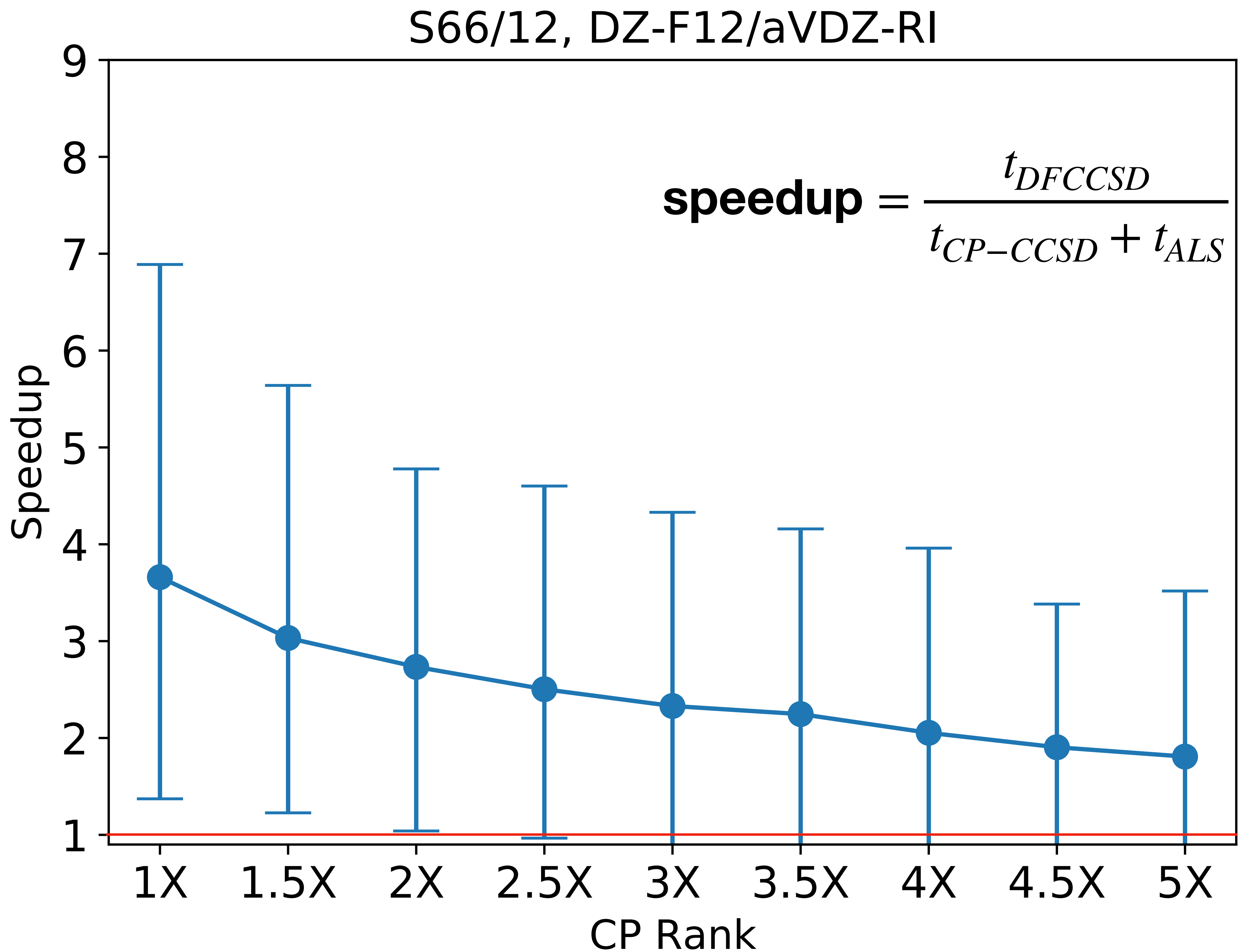
application to PPL diagram in CCSD

CP PPL CCSD: Binding Energy Errors vs CCSD



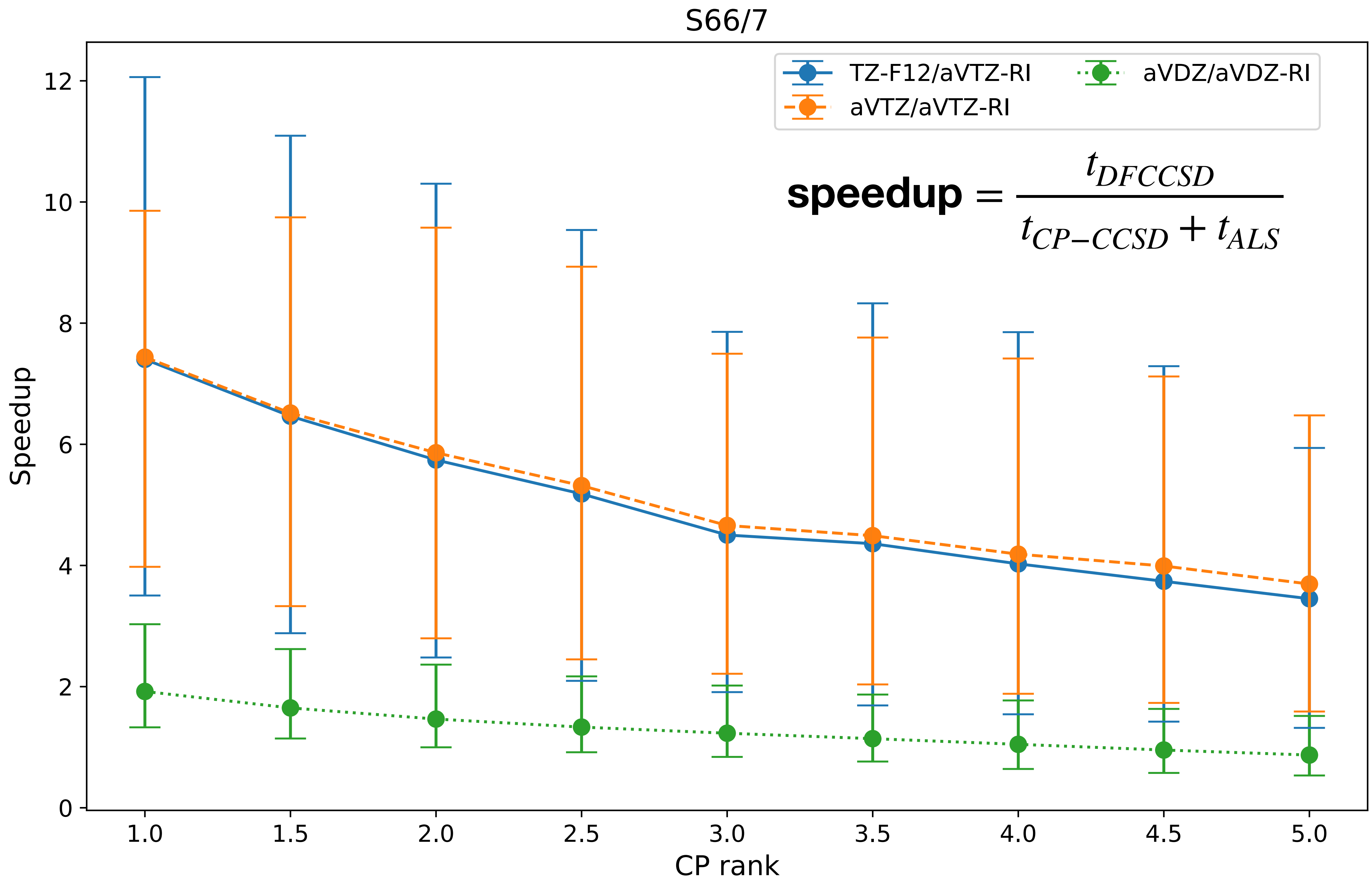
Robust CP-DF can use CP rank 3-4 times smaller than that used in THC/PS

rCP-DF: speedup vs DF-CCSD



with relevant CP ranks speedup even for a water molecule

rCP-DF: speedup vs DF-CCSD



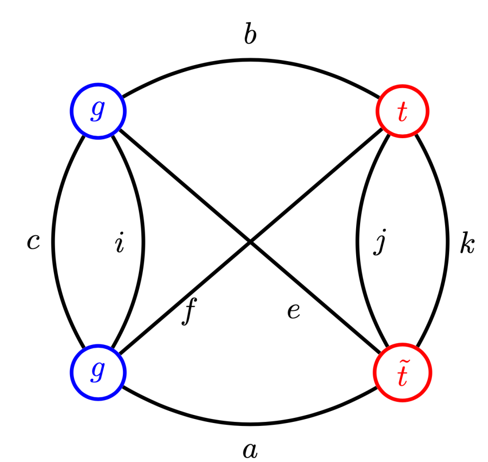
greater speedups for larger bases

Summary

Robust approximation of tensor networks
compensates for the leading-order errors due to each factor

Robust CP approximation of square-root factorized Coulomb integrals
practical and more accurate at lower ranks than comparable techniques (PS, THC)

More generally applicable, e.g.
in quantum chemistry: CCSD(T) and CCSDT
eigenvalue solvers using MPS/TT and other tensor networks



Open-Source C++ ALS solvers: github.com:ValeevGroup/BTAS

The Valeev Group

- Dr. Andrei Asadchev
- Dr. Adam Holmes
- Dr. Karl Pierce
- Nakul Teke
- Sam Slattery
- Bimal Gaudel
- Conner Masteran
- Samuel Powell
- Jaden Yon
- Ashawini Thakur

Details

Pierce, Rishi, and Valeev, *J Chem Theor Comp* 2021

JCTC Journal of Chemical Theory and Computation

pubs.acs.org/JCTC Article

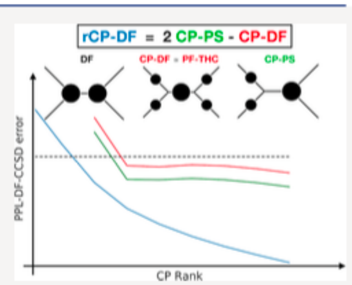
Robust Approximation of Tensor Networks: Application to Grid-Free Tensor Factorization of the Coulomb Interaction

Karl Pierce, Varun Rishi, and Edward F. Valeev*

Cite This: https://doi.org/10.1021/acs.jctc.0c01310 Read Online

ACCESS | Metrics & More | Article Recommendations | Supporting Information

ABSTRACT: Approximation of a tensor network by approximating (e.g., factorizing) one or more of its constituent tensors can be improved by canceling the leading-order error due to the constituents' approximation. The utility of such robust approximation is demonstrated for robust canonical polyadic (CP) approximation of a (density-fitting) factorized two-particle Coulomb interaction tensor. The resulting algebraic (grid-free) approximation for the Coulomb tensor, closely related to the factorization appearing in pseudospectral and tensor hypercontraction approaches, is efficient and accurate, with significantly reduced rank compared to the naive (nonrobust) approximation. Application of the robust approximation to the particle−particle ladder term in the coupled-cluster singles and doubles reduces the size complexity from $O(N^6)$ to $O(N^5)$ with robustness ensuring negligible errors in chemically relevant energy differences using CP ranks approximately equal to the size of the density-fitting basis.



Resources