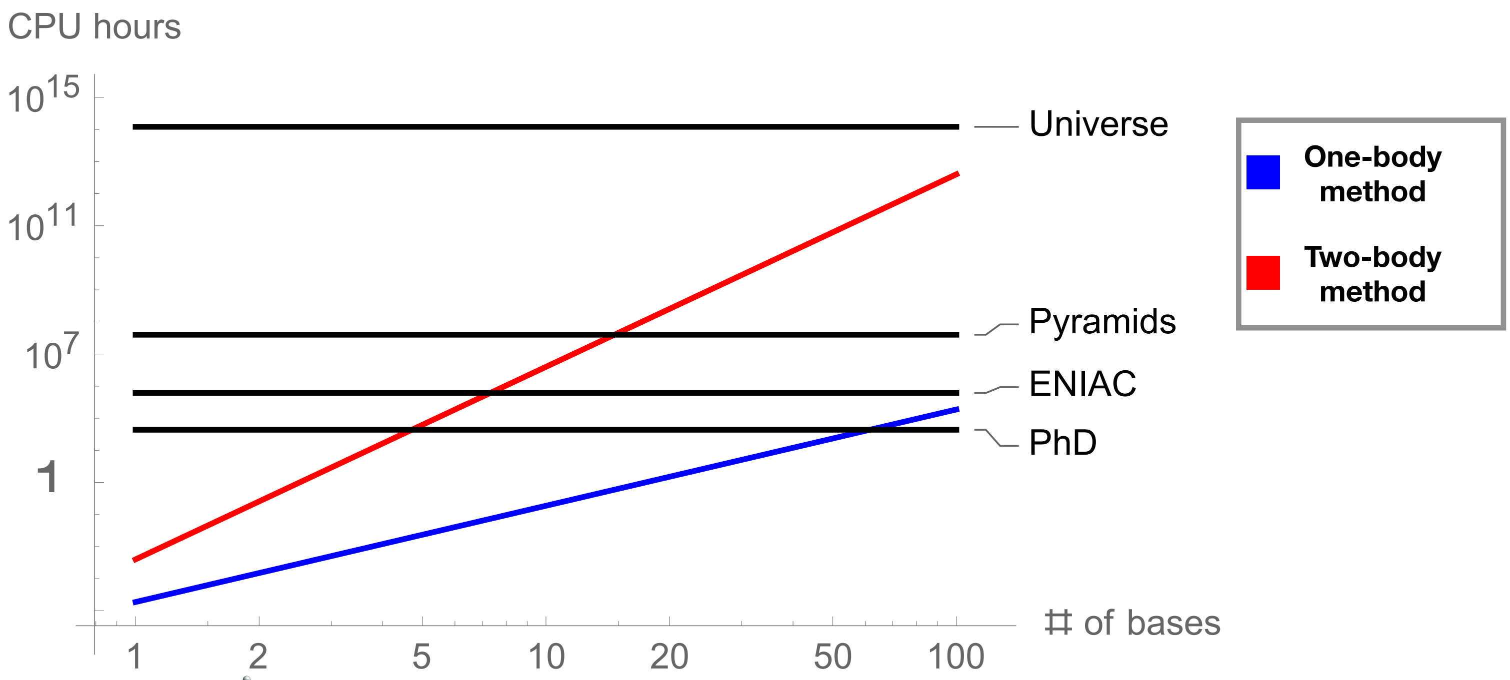
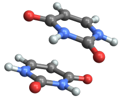
FLATIRON INSTITUTE
Center for Computational Quantum Physics

APPROXIMATING TENSOR CONTRACTIONS VIA A MATRIX-FREE TENSOR DECOMPOSITION

KARL PIERCE

The cost of molecular electronic structure methods





FLATIRON INSTITUTE
Center for Computational Quantum Physics

The correlation between storage and computational cost

	Computational Scaling	Storage	Element of highest order tensor
CCSD	$\mathcal{O}(N^6)$	$\mathcal{O}(N^4)$	t_{ab}^{ij}
CCSDT	$\mathcal{O}(N^8)$	$\mathcal{O}(N^6)$	t_{abc}^{ijk}
CCSD(T)	$\mathcal{O}(N^7)$	$\mathcal{O}(N^4)$	t_{ab}^{ij}

Purvis, G. D.; Bartlett, R. J. *The Journal of Chemical Physics* 1982, *76*, 1910–1918.

Raghavachari, K.; Trucks, G. W.; Pople, J. A.; Head-Gordon, M . *Chemical Physics Letters* 1989, *157*, 479–483.

Noga, J.; Bartlett, R. J. *J. Chem. Phys.* 1987, *86*, 7041.

Tensor compression techniques

Matrix Decomposition

Singular Value Decomposition

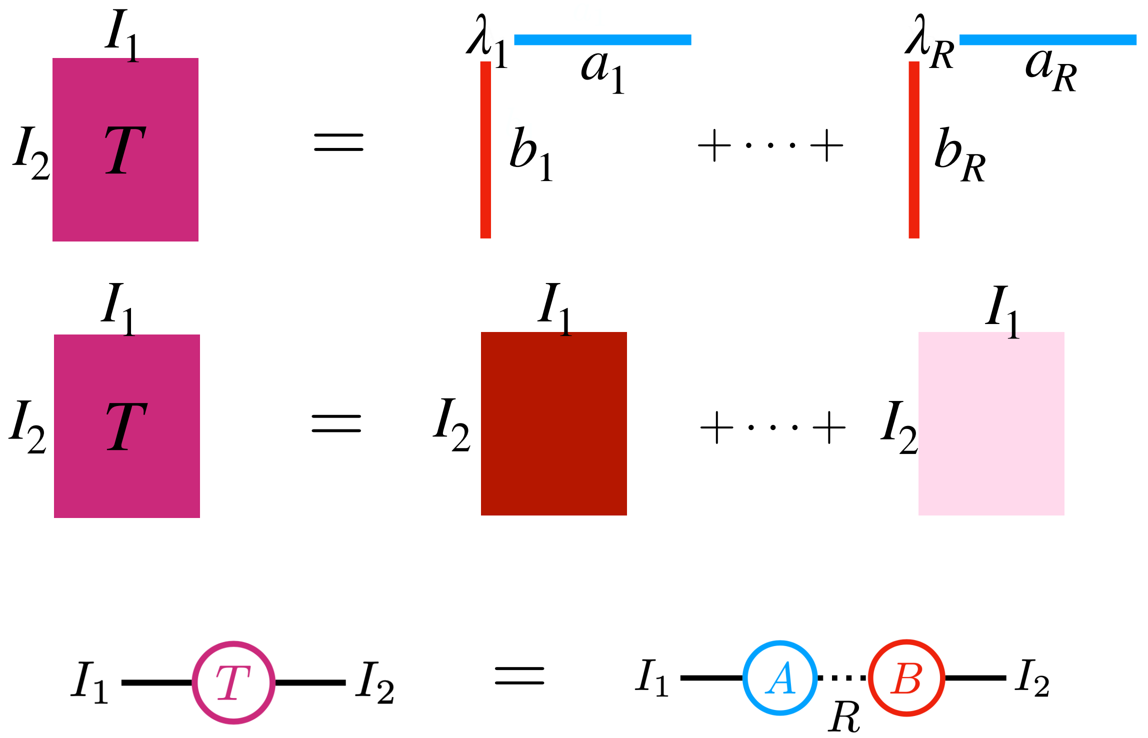
Cholesky Decomposition

Tensor Decomposition

Canonical Polyadic Decomposition

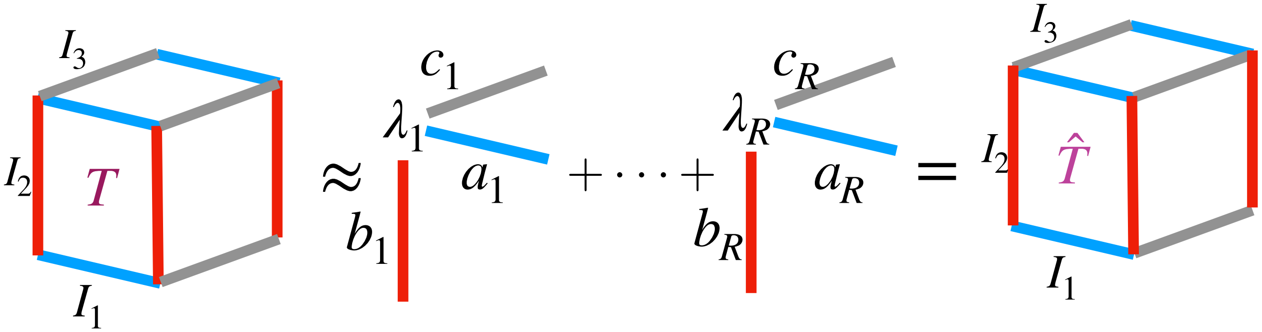
Tucker Decomposition

Singular value decomposition (SVD)

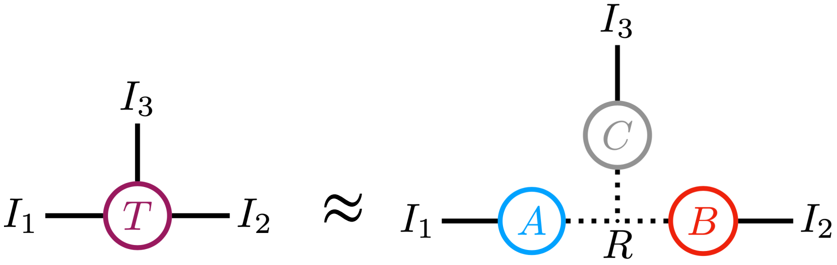


FLATIRON INSTITUTE
Center for Computational Quantum Physics

Canonical Polyadic Decomposition (CPD)



$$A = \{a_r\} \in \mathscr{R}^{I_1 \times R}$$



$$\mathcal{O}(I^3) \rightarrow \mathcal{O}(3IR)$$

Harshman, R. UCLA Work. Pap. Phonetics 1970, 16 (10), 1–84.
Carroll, J. D.; Chang, J. J. Psychometrika 1970, 35 (3), 283–319.

Coupled Cluster Optimization

Compute the coupled cluster wavefunction

$$|\Psi_{cc}\rangle = e^{\hat{T}}|\Psi_0\rangle$$

$$\hat{T} = \sum_{ia} t_i^a a_a^i + \sum_{ijab} t_{ij}^{ab} a_{ab}^{ij} + \dots$$

Truncate the cluster operator at nth-excitation

FLATIRON INSTITUTE
Center for Computational Quantum Physics

Coupled Cluster Optimization

Compute the coupled cluster wavefunction

$$|\Psi_{cc}\rangle = e^{\hat{T}}|\Psi_0\rangle$$

$$\hat{T}_{ccsd} = \sum_{ia} t_i^a a_a^i + \sum_{ijab} \boxed{t_{ij}^{ab}} a_{ab}^{ij}$$

Goal: Optimize the n excitation tensors

FLATIRON INSTITUTE
Center for Computational Quantum Physics

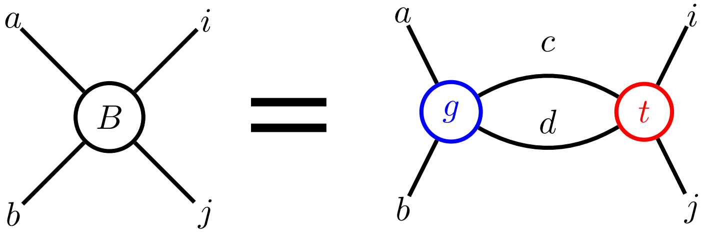
CCSD Optimization

t_{ij}^{ab} tensor optimization involves

$$^{n}R_{ij}^{ab} = A_{ij}^{ab} + B_{ij}^{ab} + C_{ij}^{ab} + D_{ij}^{ab} \qquad\qquad {}^{n}t = {}^{n-1}t + \alpha({}^{n}R)$$

Particle-Particle Ladder Diagram

$$B_{ij}^{ab} = \sum_{cd} g_{cd}^{ab} t_{ij}^{cd}$$



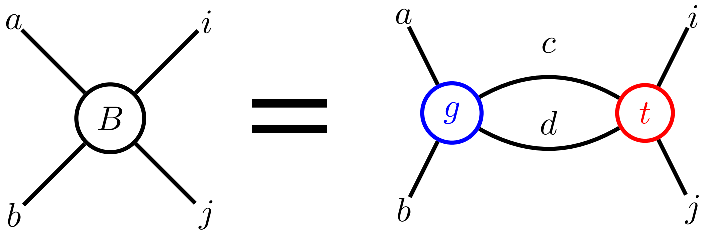
G. D. Purvis and R. J. Bartlett,
J. Chem. Phys. 76, 1910–1918 (1982).

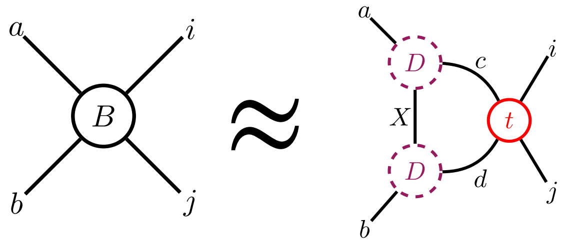
It is standard to approximate tensors in a network

$$B_{ij}^{ab} = \sum_{cd} g_{cd}^{ab} t_{ij}^{cd}$$

Density Fitting Approximation

$$B_{ij}^{ab} \approx \sum_{cdX} D_c^{aX} D_d^{bX} t_{ij}^{cd}$$



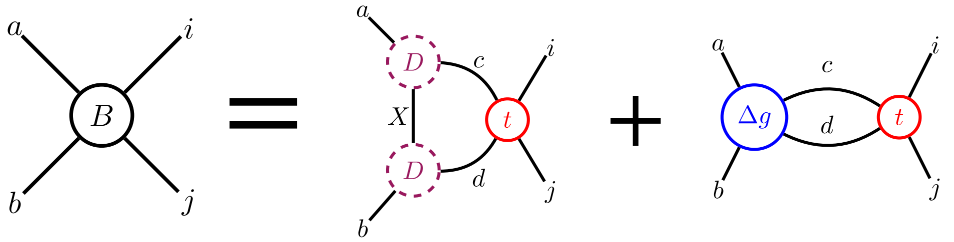


FLATIRON INSTITUTE
Center for Computational Quantum Physics

Error associated with approximated tensors in a network is amplified.

DF approximation error

$$B_{ij}^{ab} = \sum_{cdX} D_c^{aX} D_d^{bX} t_{ij}^{cd} + \sum_{cdX} \Delta g_{cd}^{ab} t_{ij}^{cd}$$



Center for Computational Quantum Physics

Error associated with approximated tensors in a network is amplified.

CPD approximation

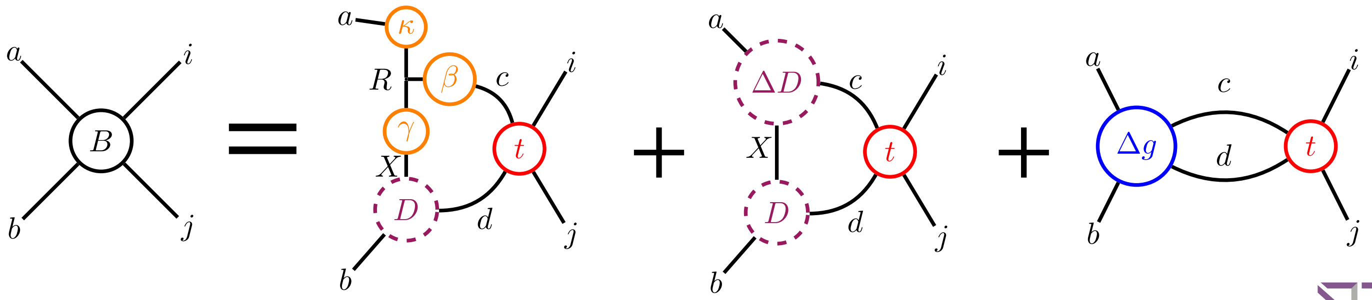
$$\sum_{cd} g_{cd}^{ab} t_{ij}^{cd} \approx \sum_{cd} \sum_{X} \hat{D}_c^{aX} D_d^{bX} t_{ij}^{cd}$$

FLATIRON INSTITUTE
Center for Computational Quantum Physics

Error associated with approximated tensors in a network is amplified.

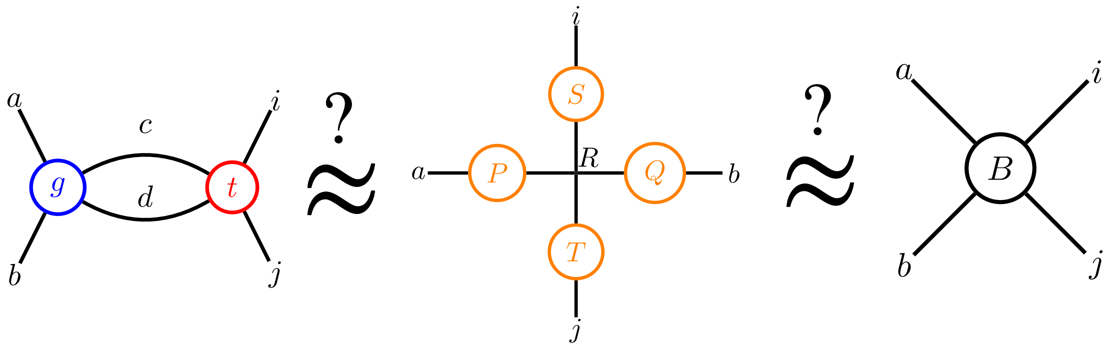
CPD approximation error

$$\sum_{cd} g_{cd}^{ab} t_{ij}^{cd} = \sum_{cd} \Big(\sum_{X} \hat{D}_c^{aX} D_d^{bX} + \sum_{X} \Delta D_c^{aX} D_d^{bX} + \Delta g_{cd}^{ab} \Big) t_{ij}^{cd}$$



FLATIRON INSTITUTE
Center for Computational Quantum Physics

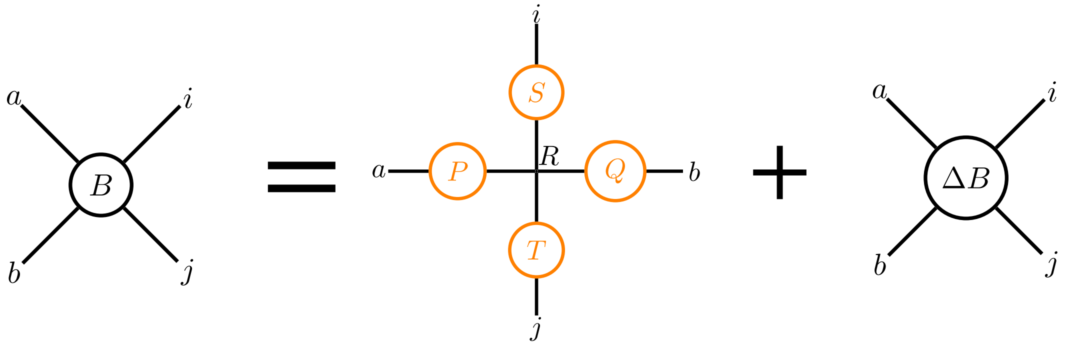
Is it possible to approximate the tensor contraction as a network?



$$f(x) = \frac{1}{2} \| \sum_{cd} g_{cd}^{ab} t_{ij}^{cd} - \hat{B}_{ij}^{ab}(x) \|_2^2 \qquad \hat{B}_{ij}^{ab}(x) = \sum_{r}^{R_{cp}} P_r^a Q_r^b S_i^r T_j^r$$

FLATIRON INSTITUTE
Center for Computational Quantum Physics

Is it possible to approximate the tensor contraction as a network?

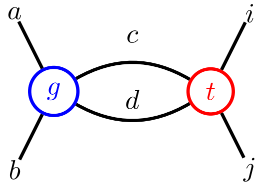


$$B_{ij}^{ab} = \hat{B}_{ij}^{ab} + \Delta_{CPD} B_{ij}^{ab}$$

$\Delta_{CPD}B$ is directly controllable via the CPD

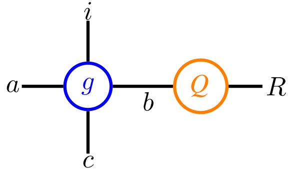
FLATIRON INSTITUTE
Center for Computational Quantum Physics

Side note: the CPD optimization of B is cheaper than the canonical contraction



$$\mathcal{O}(N^6)$$

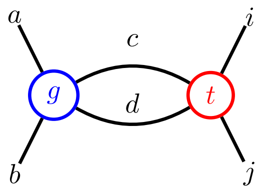
$$\frac{\partial f(x)}{\partial P} = \sum_{cd} \boxed{\Big(\sum_{b} g^{ab}_{cd} Q^{b}_{r}\Big)} \Big(\sum_{ij} t^{cd}_{ij} S^{r}_{i} T^{r}_{j}\Big) - P^{a}_{r}(Q^{b}_{r} Q^{b}_{r'}) * (S^{r}_{i} S^{r'}_{i}) * (T^{r}_{j} T^{r'}_{j})$$



$$\mathcal{O}(N^4 R)$$

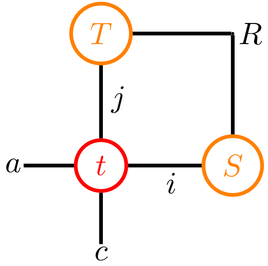
FLATIRON INSTITUTE
Center for Computational Quantum Physics

Side note: the CPD optimization of B is cheaper than the canonical contraction



$$\mathcal{O}(N^6)$$

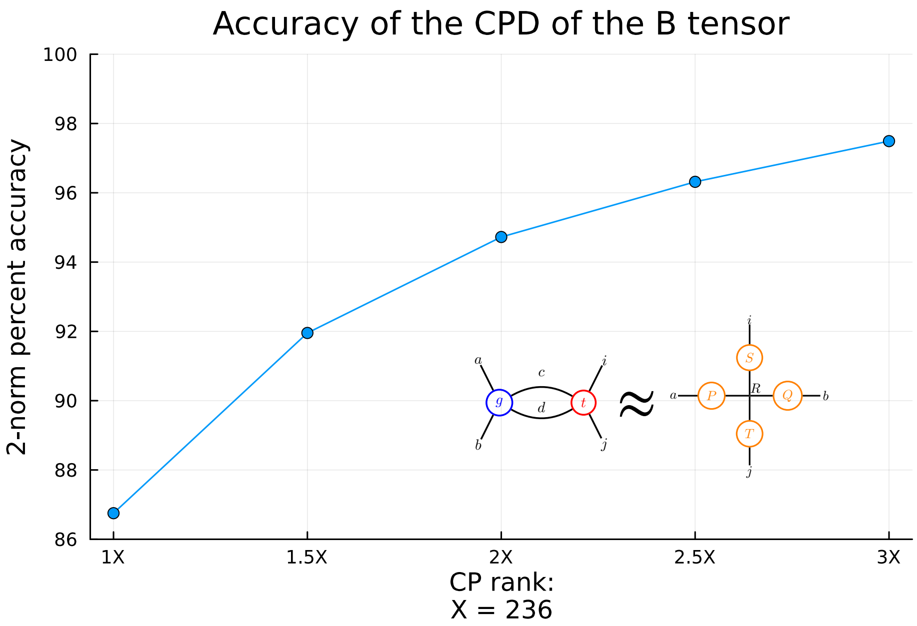
$$\frac{\partial f(x)}{\partial P} = \sum_{cd} g_{cd}^{ar}\left(\boxed{\sum_{ij} t_{ij}^{cd} S_i^r T_j^r}\right) - P_r^a (Q_r^b Q_{r'}^b) * (S_i^r S_i^{r'}) * (T_j^r T_j^{r'})$$

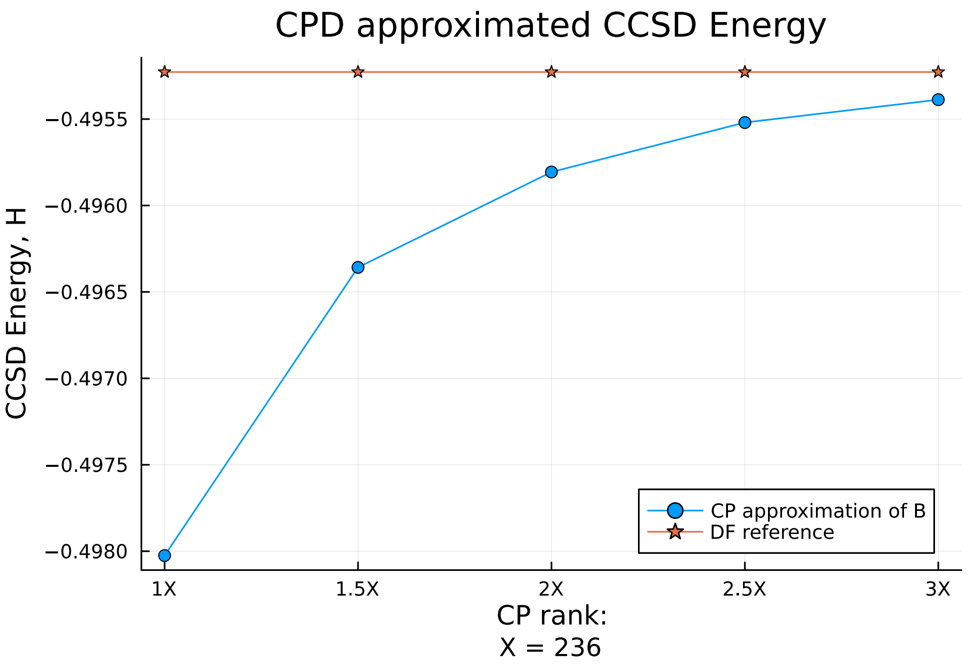


$$\mathcal{O}(N^4 R)$$

CP approximation of PPL tensor network

$(H_2O)_2$ TIP4P geometry
DZ-F12/aug-DZ-RI





$$\text{percent accuracy} = (1.0 - \|B - \hat{B}\|_2 / \|B\|_2) * 100$$

How does one choose a stopping condition?

Canonical stopping condition:
2-norm error

$$\Delta_n = \|B_{ij}^{ab} - \hat{B}_{ij}^{ab}(x_n)\|$$

$$\Delta_n - \Delta_{n-1} < \epsilon_{ALS}$$

Computational complexity: $\mathcal{O}(N^6)$

Proposed stopping condition:
2-norm error

$$\Upsilon_n = \|\hat{B}_{ij}^{ab}(x_n) - \hat{B}_{ij}^{ab}(x_{n-1})\|$$

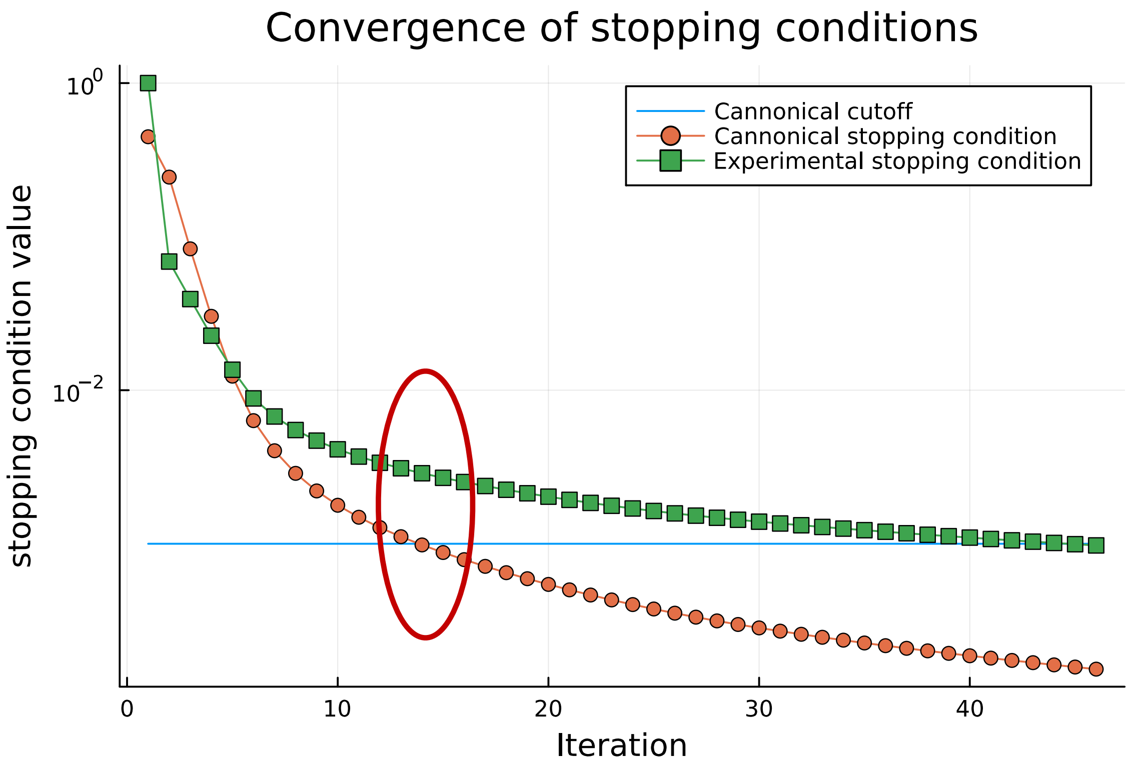
$$\Upsilon_n < \epsilon_{ALS}$$

Computational complexity: $\mathcal{O}(N^3)$

FLATIRON INSTITUTE
Center for Computational Quantum Physics

Comparing stopping conditions

$(H_2O)_2$ TIP4P geometry
DZ-F12/aug-DZ-RI



FLATIRON INSTITUTE
Center for Computational Quantum Physics

Result

It is feasible to approximate a tensor network contraction as a matrix free tensor decomposition

Question: Can we make this CPD "contraction" more efficient?

FLATIRON INSTITUTE
Center for Computational Quantum Physics

Is there a way to reduce the cost of the CPD?

$$ {}^{n}R_{ij}^{ab} = A_{ij}^{ab} + B_{ij}^{ab} + C_{ij}^{ab} + D_{ij}^{ab} \qquad {}^{n}t = {}^{n-1}t + \alpha({}^{n}R) $$

Iteration 1: $${}^{1}B_{ij}^{ab} = g_{cd}^{ab}({}^{1}t_{ij}^{cd})$$

Iteration 2: $${}^{2}B_{ij}^{ab} = g_{cd}^{ab}({}^{2}t_{ij}^{cd}) \qquad {}^{2}B_{ij}^{ab} \approx g_{cd}^{ab}({}^{1}t_{ij}^{cd} + \delta {}^{1}t_{ij}^{cd})$$

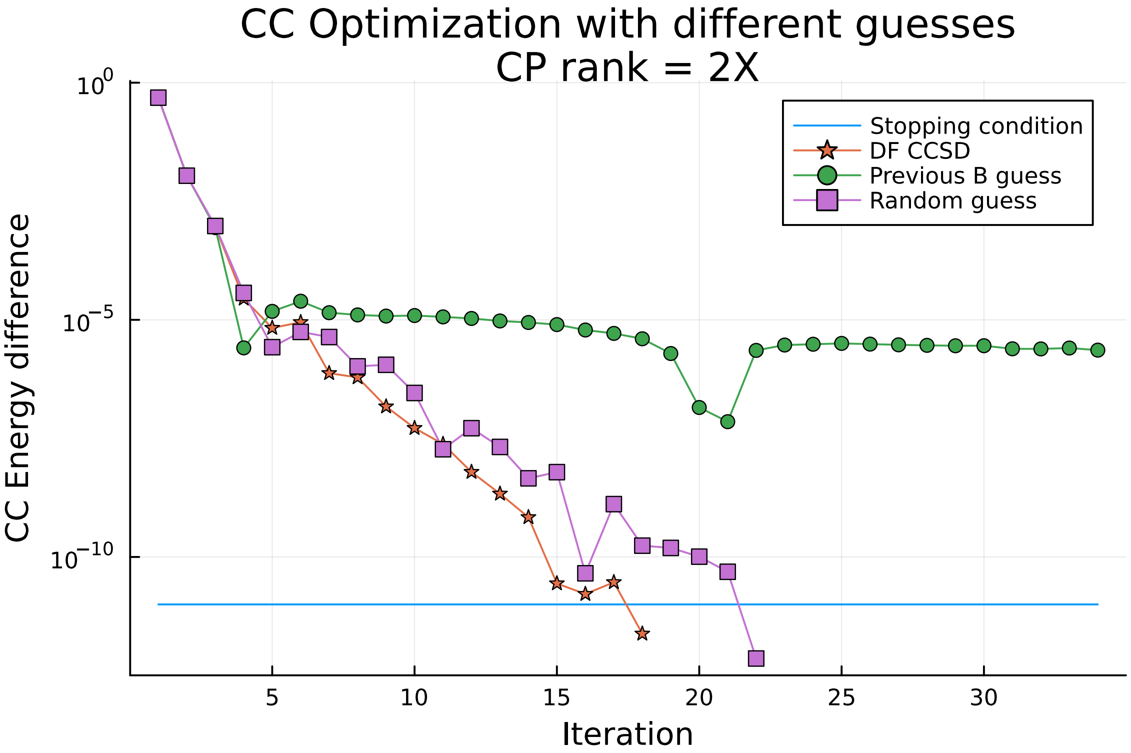
Idea:

The optimized ${}^{n-1}\hat{B}$ factors might be a good initial guess for the nth CC iteration's CPD

Flatiron Institute
Center for Computational Quantum Physics

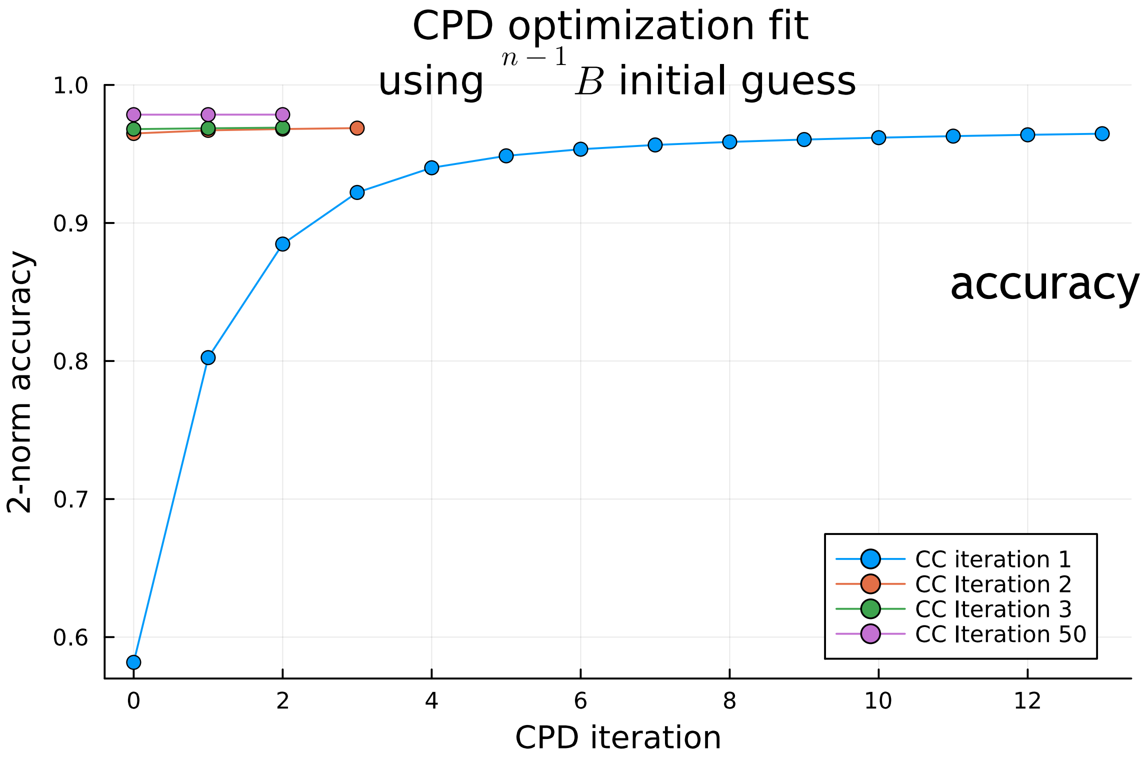
Is there a way to reduce the cost of the CPD?

$(H_2O)_2$ TIP4P geometry
DZ-F12/aug-DZ-RI



Perhaps this is border-rank optimization issue.

$(H_2O)_2$ TIP4P geometry
DZ-F12/aug-DZ-RI



$$\text{accuracy} = 1.0 - \frac{\|B - \hat{B}\|_2}{\|B\|_2}$$

FLATIRON INSTITUTE
Center for Computational Quantum Physics

Perhaps this is border-rank optimization issue.

TAMARA G. KOLDA AND BRETT W. BADER

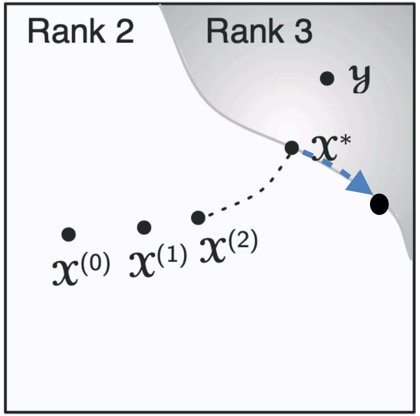


Fig. 3.2 *Illustration of a sequence of tensors converging to one of higher rank* [144].

FLATIRON INSTITUTE
Center for Computational Quantum Physics

Perhaps this is border-rank optimization issue.

Strategy:
If CPD does 8+ iterations,
Save the factors of iteration 7
else,
Construct $\hat{B}$ and throw factors away

TAMARA G. KOLDA AND BRETT W. BADER

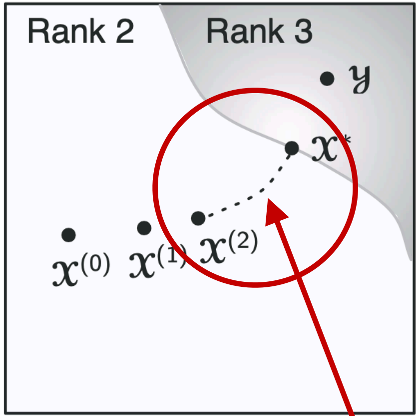


Fig. 3.2 *Illustration of a sequence of tensors converging to one of higher rank* [144].

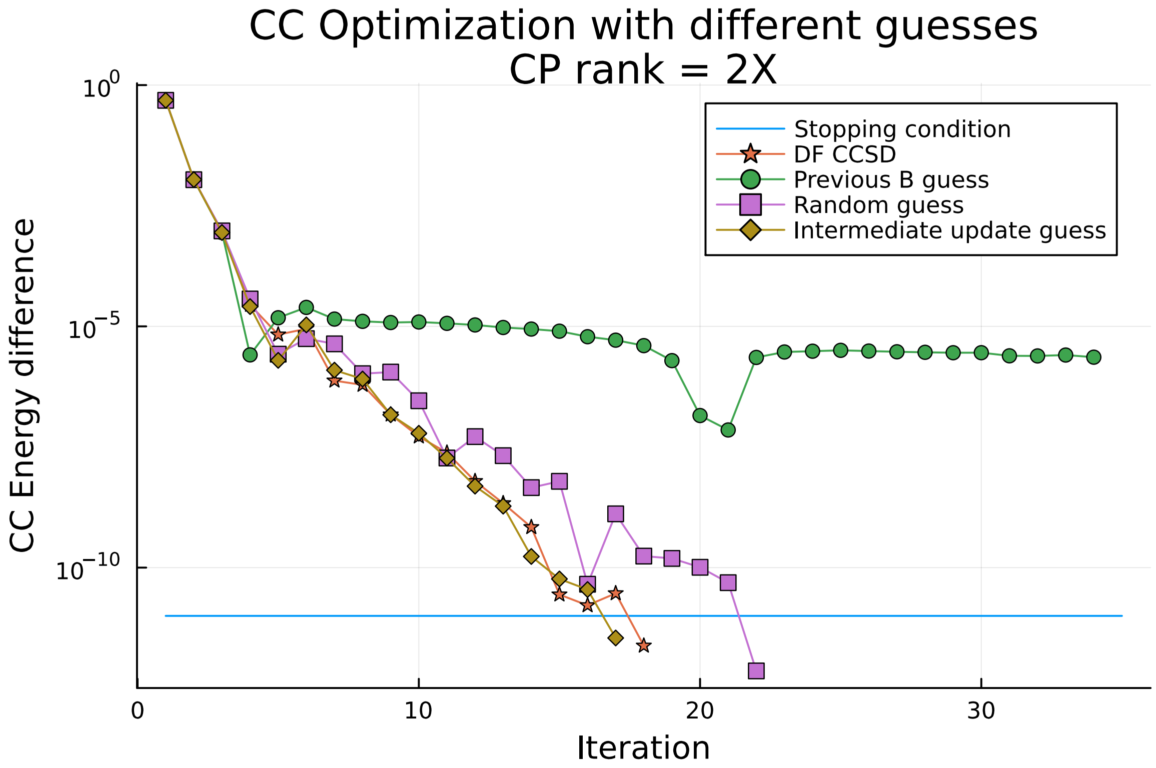
Better idea:
Take advantage of the optimizations velocity to put us at a "good" answer

Use a set of factors from one of these iterations

FLATIRON INSTITUTE
Center for Computational Quantum Physics

Is there a way to reduce the cost of the CPD?

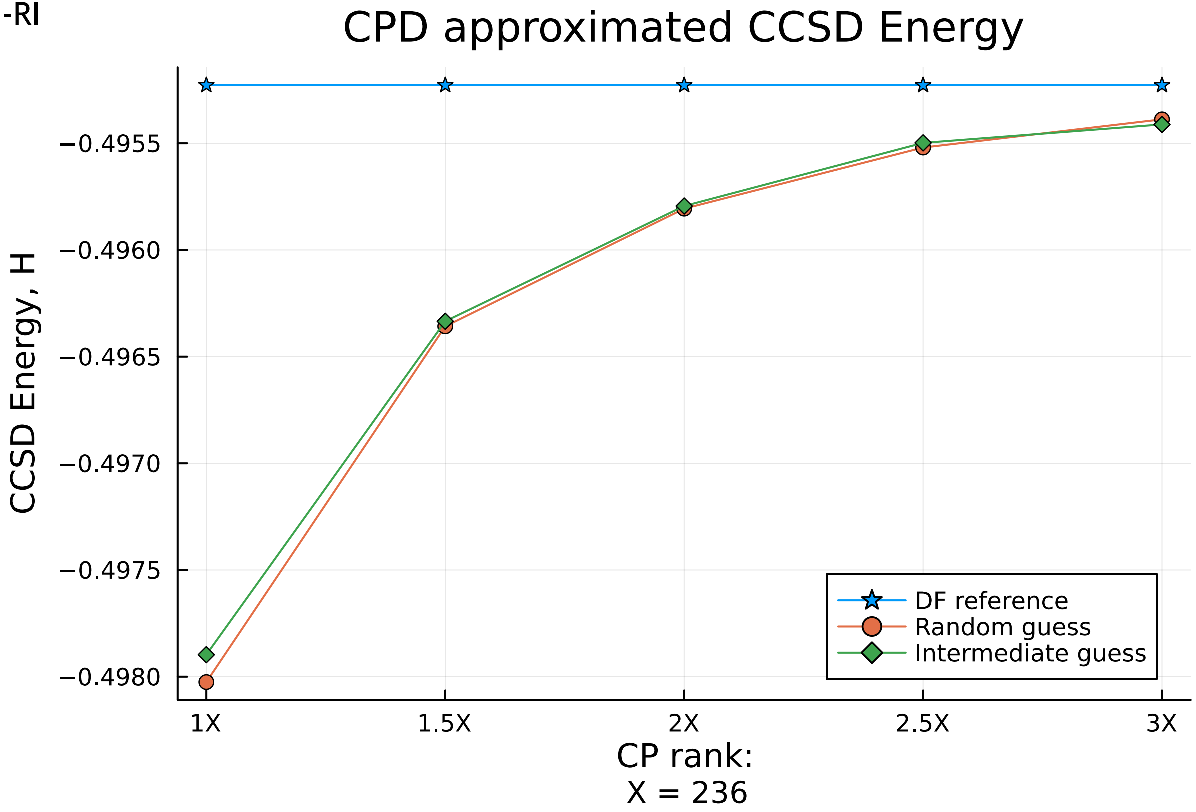
$(H_2O)_2$ TIP4P geometry
DZ-F12/aug-DZ-RI



FLATIRON INSTITUTE
Center for Computational Quantum Physics

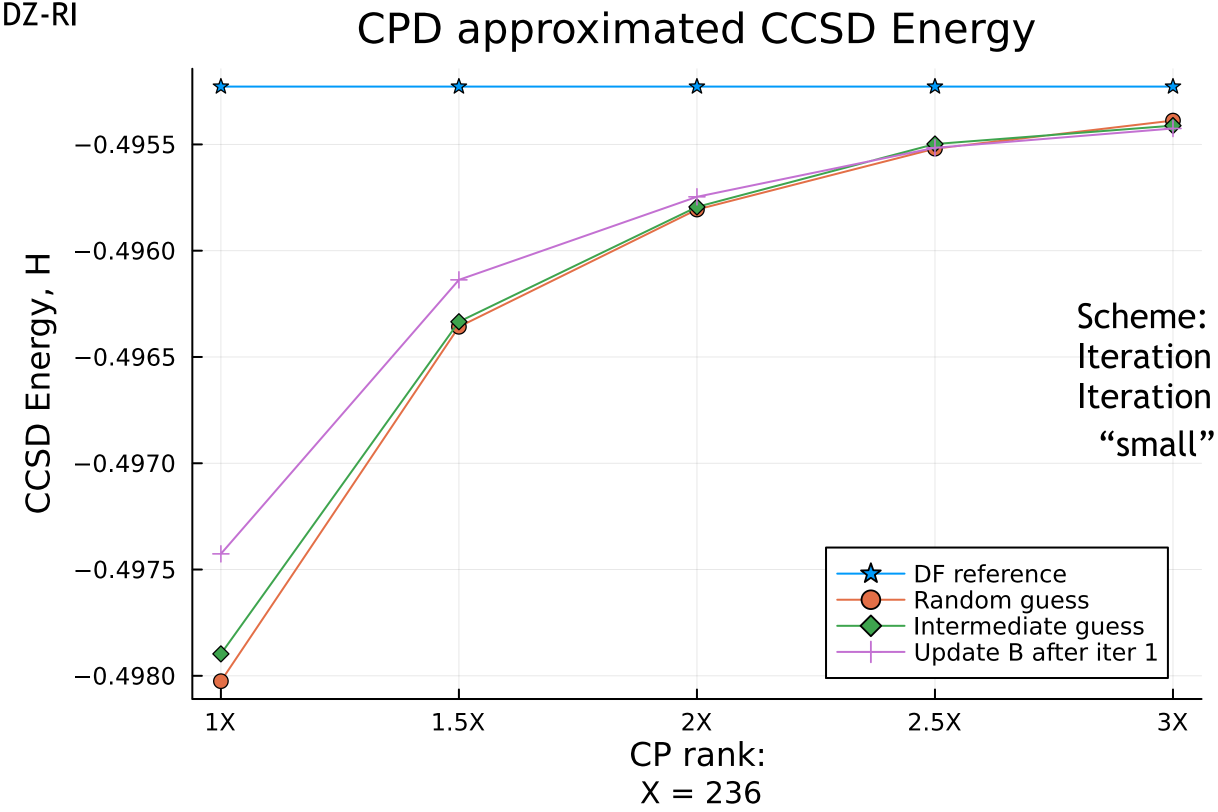
Is there a way to reduce the cost of the CPD?

$(H_2O)_2$ TIP4P geometry
DZ-F12/aug-DZ-RI



Is it necessary to recompute $\hat{B}$ every CC iteration?

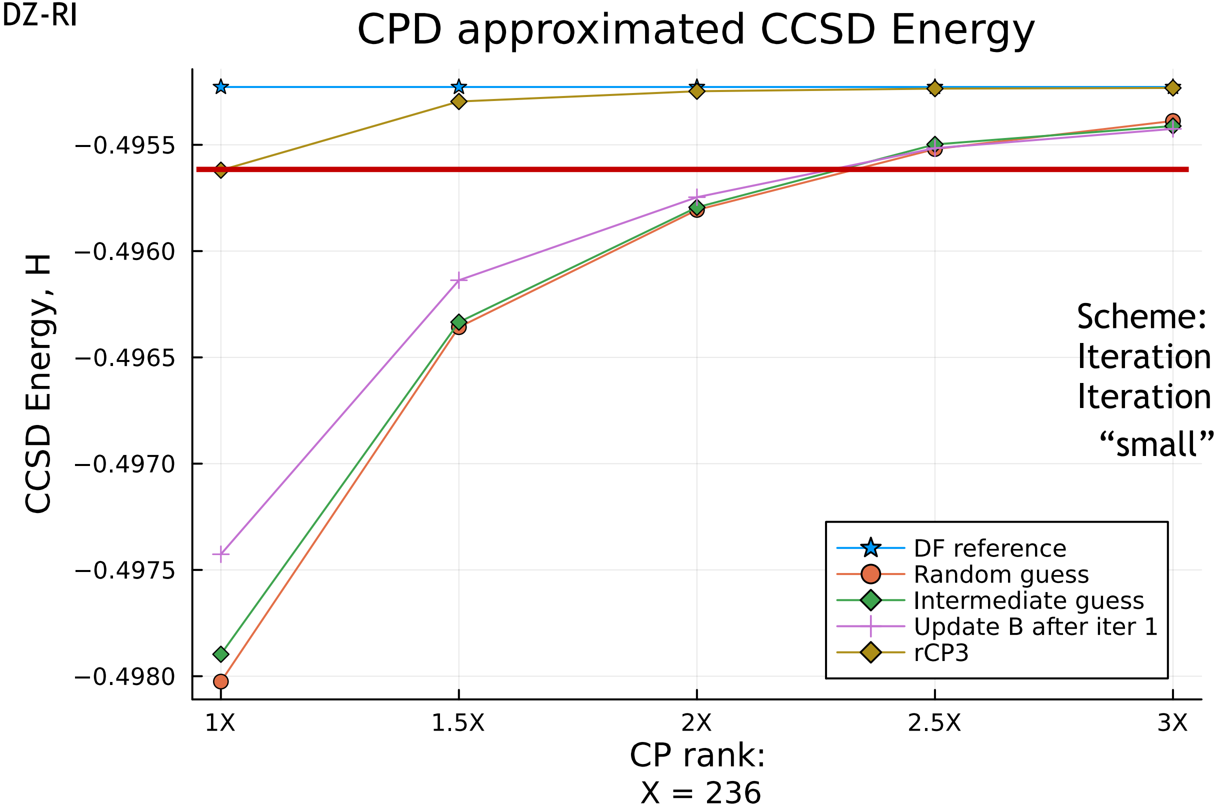
$(H_2O)_2$ TIP4P geometry
DZ-F12/aug-DZ-RI



Scheme:
Iteration 1: use full CPD rank
Iteration 2+: use fixed, "small" CPD rank (here I use X)

Is it necessary to recompute $\hat{B}$ every CC iteration?

$(H_2O)_2$ TIP4P geometry
DZ-F12/aug-DZ-RI



Accuracy threshold from previous work

Scheme:
Iteration 1: use full CPD rank
Iteration 2+: use fixed, "small" CPD rank (here I use X)

FLATIRON INSTITUTE
Center for Computational Quantum Physics

Summary

Mina Mandic
CCQ summer student

- It is possible to replace contractions with matrix-free tensor approximations.
- We can contrive some reliable and inexpensive way to stop matrix-free optimizations.
- Finding a way to improve the CPD accuracy for sparse tensors will directly improve accuracy of this tensor contraction.
- It could be possible to use randomize algorithms to further reduce the cost of the tensor optimization.

FLATIRON INSTITUTE
Center for Computational Quantum Physics

FLATIRON INSTITUTE

Center for Computational Quantum Physics