Tensor decomposition techniques in nuclear theory

GDR workshop

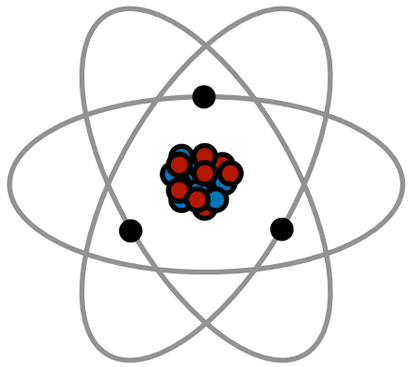
February 2021

TECHNISCHE UNIVERSITÄT DARMSTADT

SFB 1245

Alexander Tichai

Technische Universität Darmstadt



Overview

I Introduction - perspectives on heavy nuclei

II Low-rank properties of nuclear interaction

- Core ideas of data compression
- Tensor formats in nuclear theory

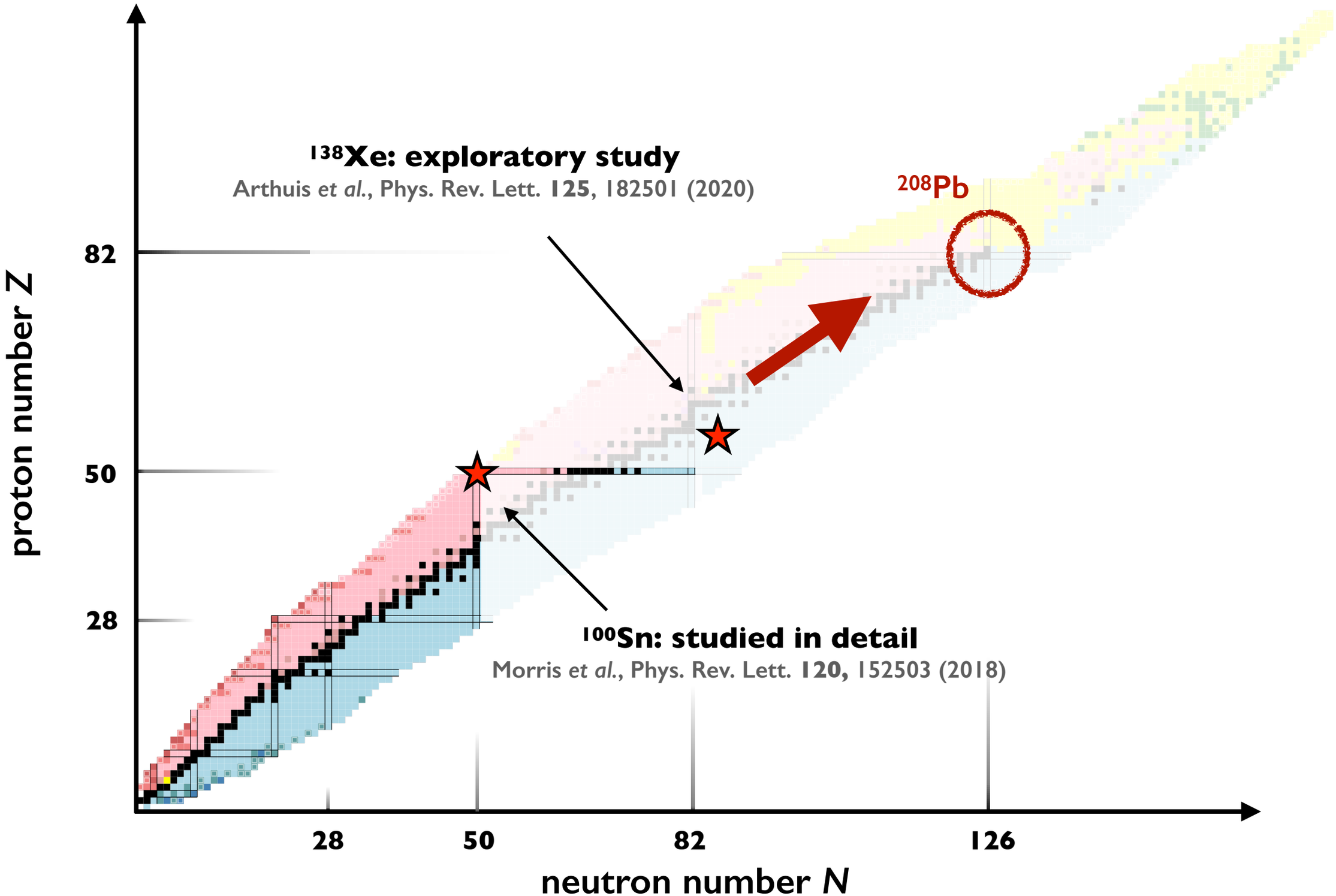
III Tensor-structured many-body theory

- Scaling advantages
- Future perspectives

Introduction

Perspectives on heavy nuclei

The quest for heavy nuclei



The setting

- **<u>Goal</u>: solve quantum many-body problem over the entire nuclear chart ($A \gg 100$)**

$$H|\Psi_n\rangle = E_n|\Psi_n\rangle$$

- **Converged calculations require very large single-particle model spaces**

$$V_{3\mathrm{N}} = \frac{1}{(3!)^2} \sum_{pqrstu} \bar{w}_{pqrstu} c_p^\dagger c_q^\dagger c_r^\dagger c_u c_t c_s$$

- **Three-body forces: introduction of additional cut on quantum numbers**

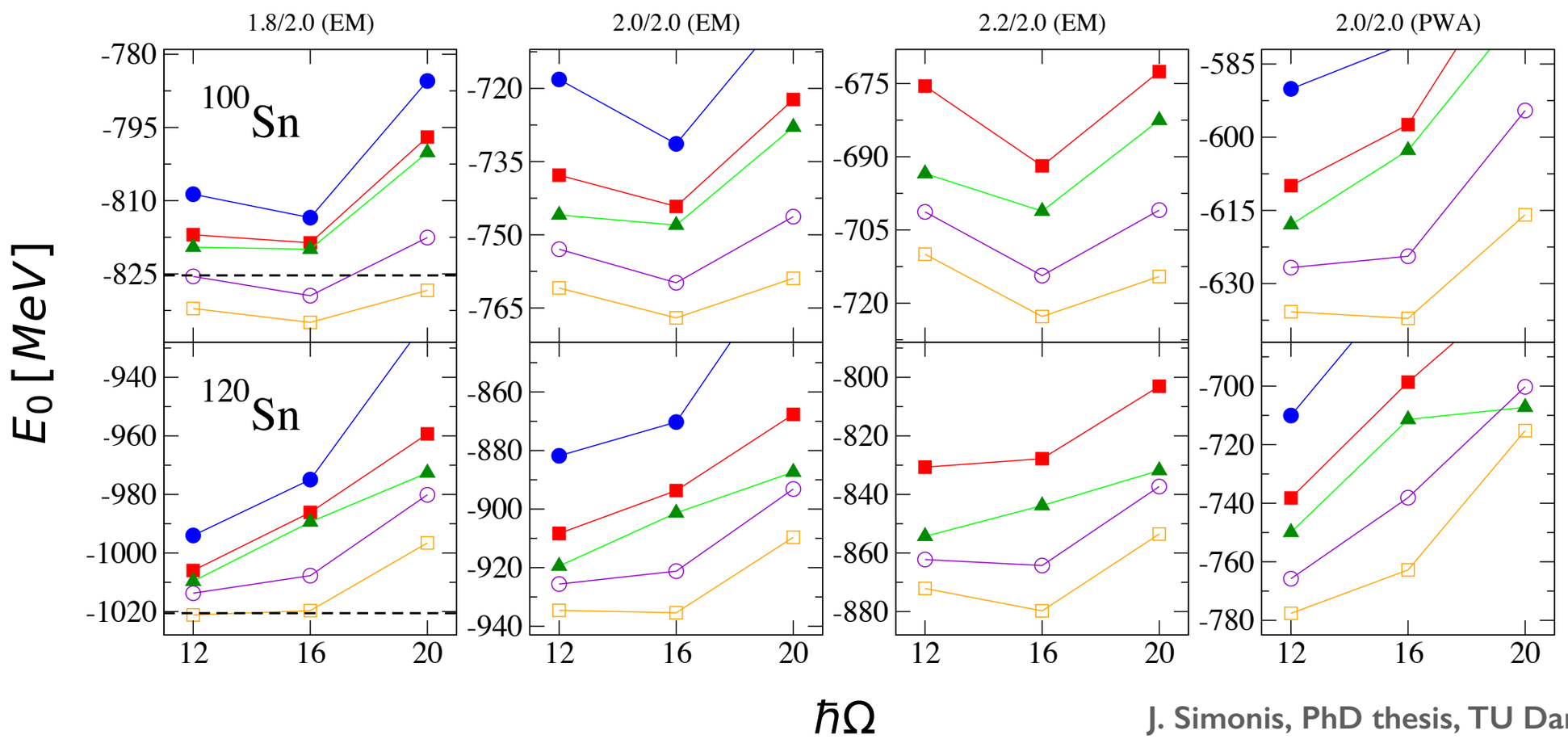
$$e_p + e_q + e_r \leq E_{3\mathrm{max}} \qquad\qquad e_u + e_t + e_s \leq E_{3\mathrm{max}}$$

- **Current capabilities are insufficient to reach convergence beyond $A=100$**

sizeable dependence on $E_{3\mathrm{max}}$

E_{3max} in practice

IMSRG(2) calculations in A=100/120 systems

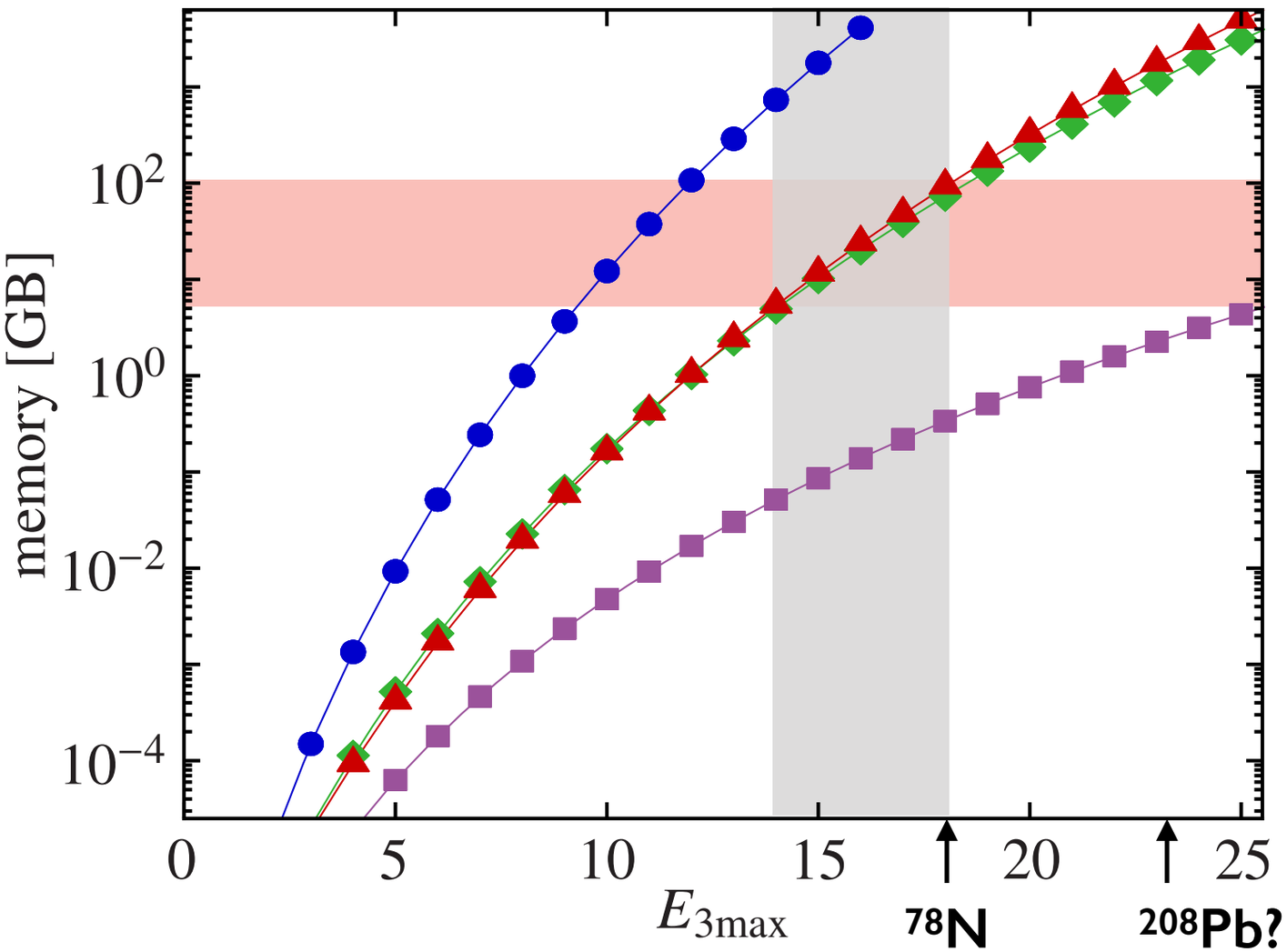


J. Simonis, PhD thesis, TU Darmstadt (2017)

e_{max}/E_{3max}:	●	■	▲	○	□
	10/14	12/14	14/14	14/16	14/18
					(the best we can do!)

Dimensionalities

Roth *et al.*, PRC **90**, 024325



Storage schemes

- M-scheme
- J-scheme
- JT-scheme
- Jacobi basis

(Hartree-Fock)

Low-rank properties of nuclear interactions

What is the origin of the complexity?

Operator → $\mathcal{O}(100)$ **topologies**

↓ momentum discretization

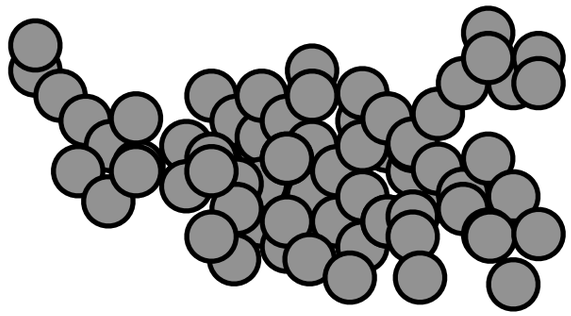
Momentum basis (matrix/tensor) → $\mathcal{O}(10^6)$ **matrix elements**

↓ Talmi-Moshinsky + HO transformation

Single-particle basis (matrix/tensor) → $\mathcal{O}(10^9)$ **matrix elements**

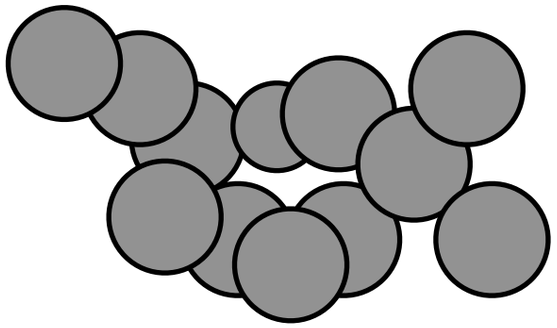
Concepts of data compression

Complex object



… but lower
resources required

loss of detailed
information …



Simpler object

Concepts of data compression

Example from image processing

791 · 640 pixels

Original picture

data compression algorithm
(singular value decomposition)

30 %
(of initial size)

9 %

3 %

Basics of tensor decomposition

- **Mode-*N* tensors are multi-variate data arrays depending on *N* indices**

$$T_{i_1 \cdots i_N}$$

- **Storage requirements depends on index ranges of individual indices**

$$I_1 \cdots I_N$$

- **Tensor factorization: rewrite tensor as sum of products that is *'close'* in some sense**

$$\tilde{T}_{i_1 \cdots i_N} \approx T_{i_1 \cdots i_N}$$

- **Relative approximation error typically accessed via Frobenius norm**

$$\Delta T = \frac{\|\tilde{T} - T\|}{\|T\|} \qquad \|T\| = \sqrt{\sum_{i_1 \cdots i_N} T_{i_1 \cdots i_N} T^{\star}_{i_1 \cdots i_N}}$$

- **Important: tensor norms measure quality of non-observable object (e.g. potential)**

What matters is performance on observables!

Basics of tensor decomposition

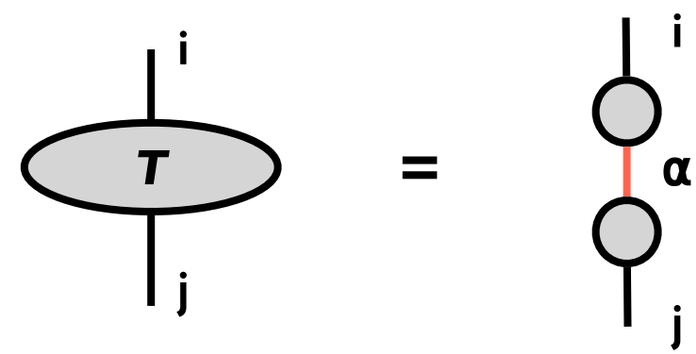
- **Tensor format**: specification of its factors and contraction scheme

rank

factors

$$\tilde{T}_{ij} = \sum_{\alpha}^{R} X_{i\alpha} Y_{\alpha j}$$

auxiliary index



Feynman-style notation

- Rank controls both **complexity** and **accuracy** of the decomposition
- In the limit of an infinite rank the tensor decomposition becomes exact

$$\tilde{T} \xrightarrow{R \to \infty} T$$

- There is a finite **format-specific critical rank** at beyond which no error occurs

$$\Delta T = 0 \quad \forall R \geq R_{\text{crit}}$$

Critical rank

- **Example: reconsider tensor format with index range: *i,j = 1,…100***

$$\tilde{T}_{ij} = \sum_{\alpha}^{R} X_{i\alpha} Y_{\alpha j}$$

- **Make an educated guess for the decomposition factors:** $X_{ij} = T_{ij}$ $Y_{ij} = \delta_{ij}$

$$\tilde{T}_{ij} = \sum_{\alpha}^{R} T_{i\alpha} \delta_{\alpha j} = T_{ij}$$

$R_{\text{crit}} = 100$
(upper bound)

- **General alternative: count the degrees of freedom on both sides**

$$\tilde{T}_{ij} = \sum_{\alpha}^{R} X_{i\alpha} Y_{\alpha j}$$

100 · 100

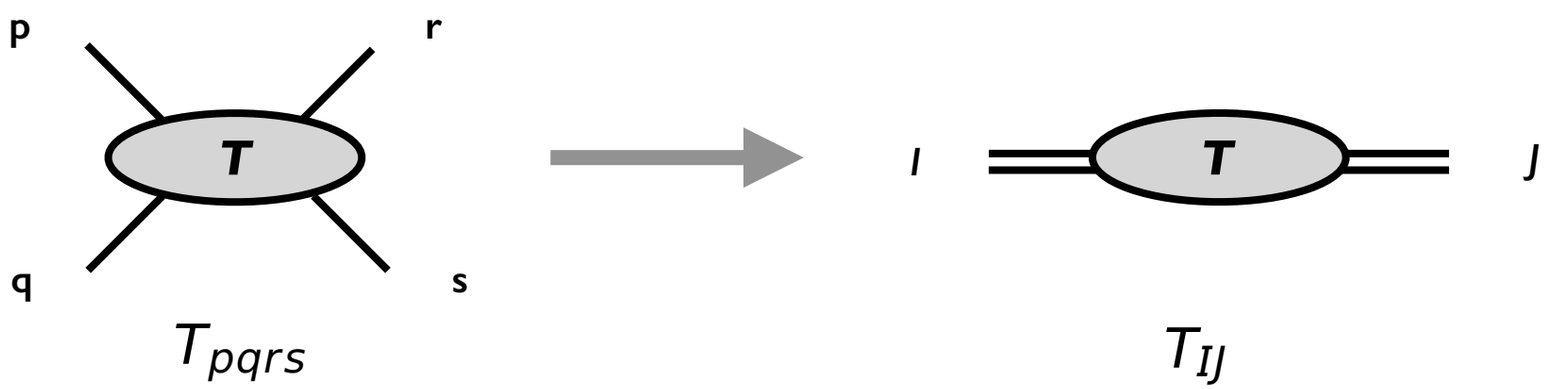
2 · 100 · R

- **Again: the critical rank will be different if the tensor format is different!**

Reshaping operations

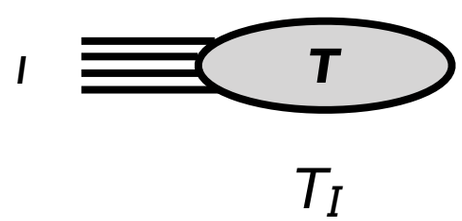
- Most tensors appearing in quantum many-body theory have mode greater than two

- It is often useful (and necessary) to interpret them as matrices (mode-2 tensors)

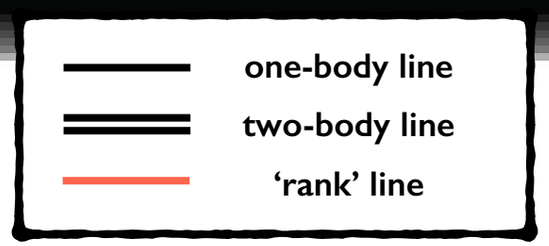


- Reshaping imposes no approximation but is rather data rearrangement

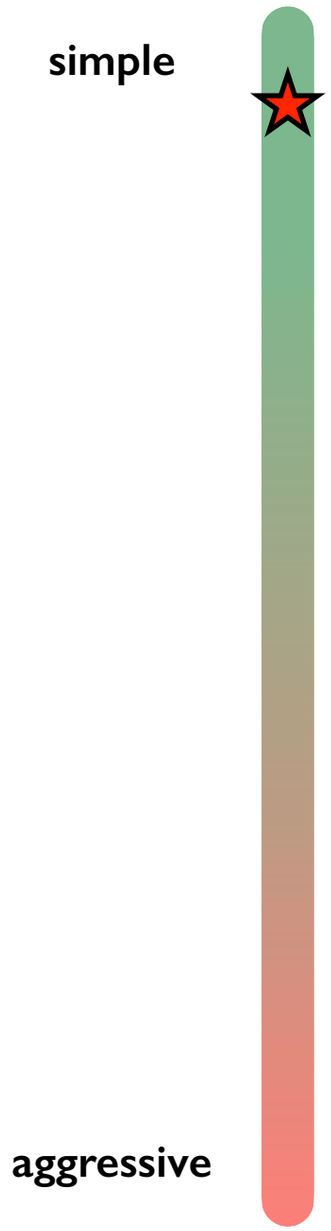
- Very natural since computationally the data is a vector either way



Overview of tensor formats

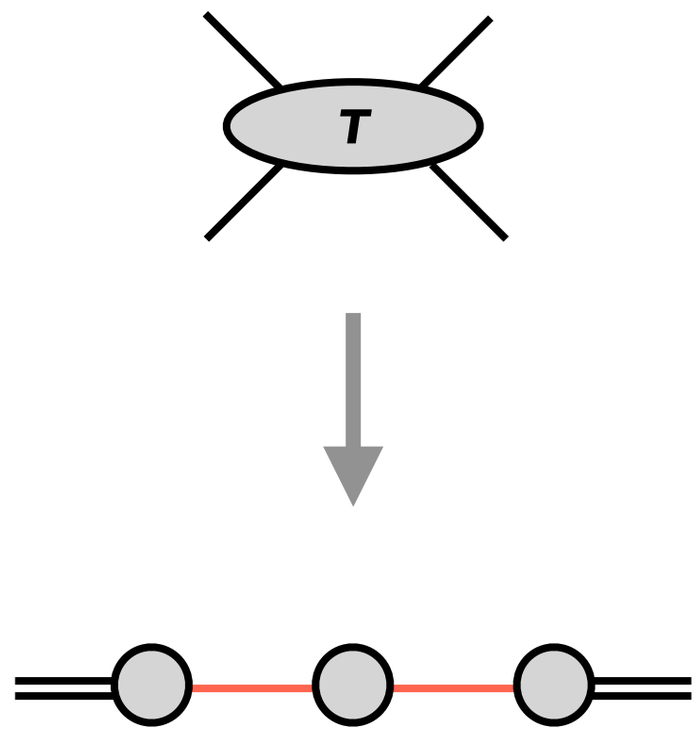


The 'spectrum' of tensor formats



tSVD

truncated singular value decomposition



Matrix decompositions

- **Prototype of a matrix factorization: singular value decomposition (SVD)**

singular values (non-negative)

$$M = U\Sigma V^T \quad \text{with} \quad \Sigma = \text{diag}(s_1, \ldots, s_n)$$

left/right singular vectors

- **Truncated singular valued decomposition: keep only largest singular values**

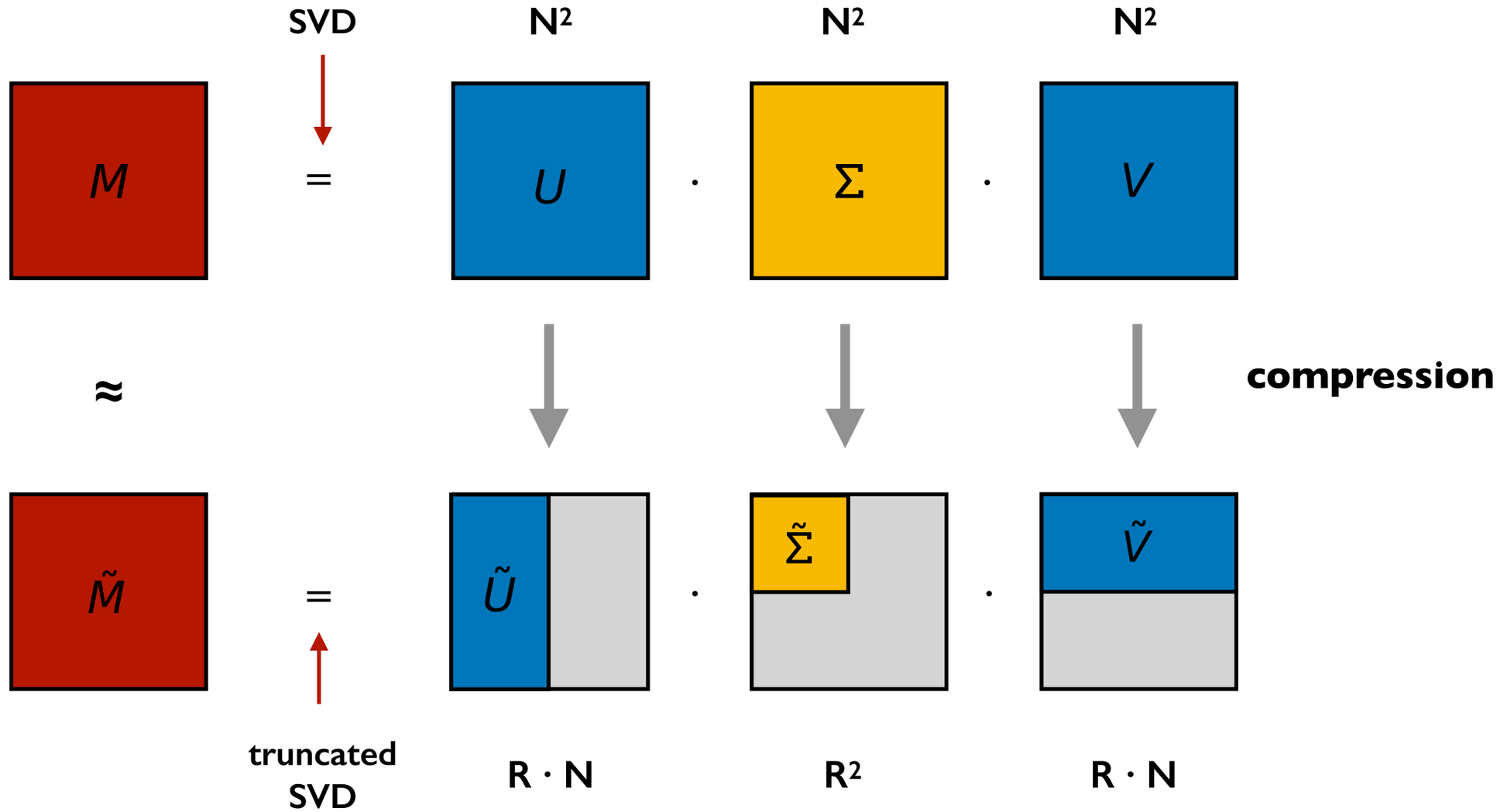
$$\tilde{M} = \tilde{U}\tilde{\Sigma}\tilde{V}^T \quad \text{with} \quad \tilde{S} = \text{diag}(s_1, \ldots, s_{R_{\text{SVD}}}, 0, \ldots, 0)$$

- **Versatility: can be applied to non-square and non-Hermitian matrices**

- **Eckart-Young theorem: tSVD provides best rank-R approximation to matrix**

- **Practical advantage: fast algorithms implemented in any good library**

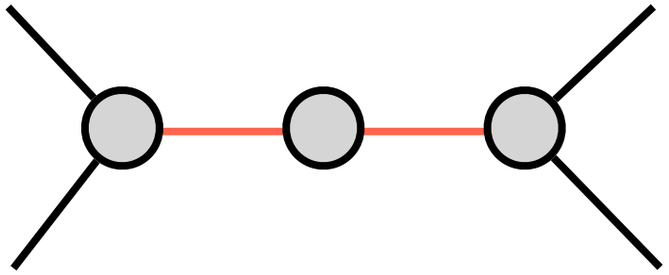
How to approximate SVD



Why is SVD so natural?

- Size of singular values is a natural measure for importance of information

- Standard reshaping operation respects Fermionic permutation symmetries



- Historically nuclear physicists extensively studied separable potentials

$$V(q', q) = g \cdot f(q') \cdot f(q)$$

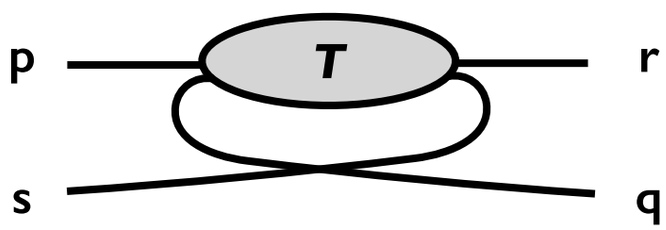
- A rank-1 tSVD yields a separable representation of a potential …

- … and higher ranks in the decomposition correct for the non-separable character

Results on SVD

- Singular values are rapidly suppressed in standard ordering ('1234')

- For unnatural grouping ('1423') almost no compression

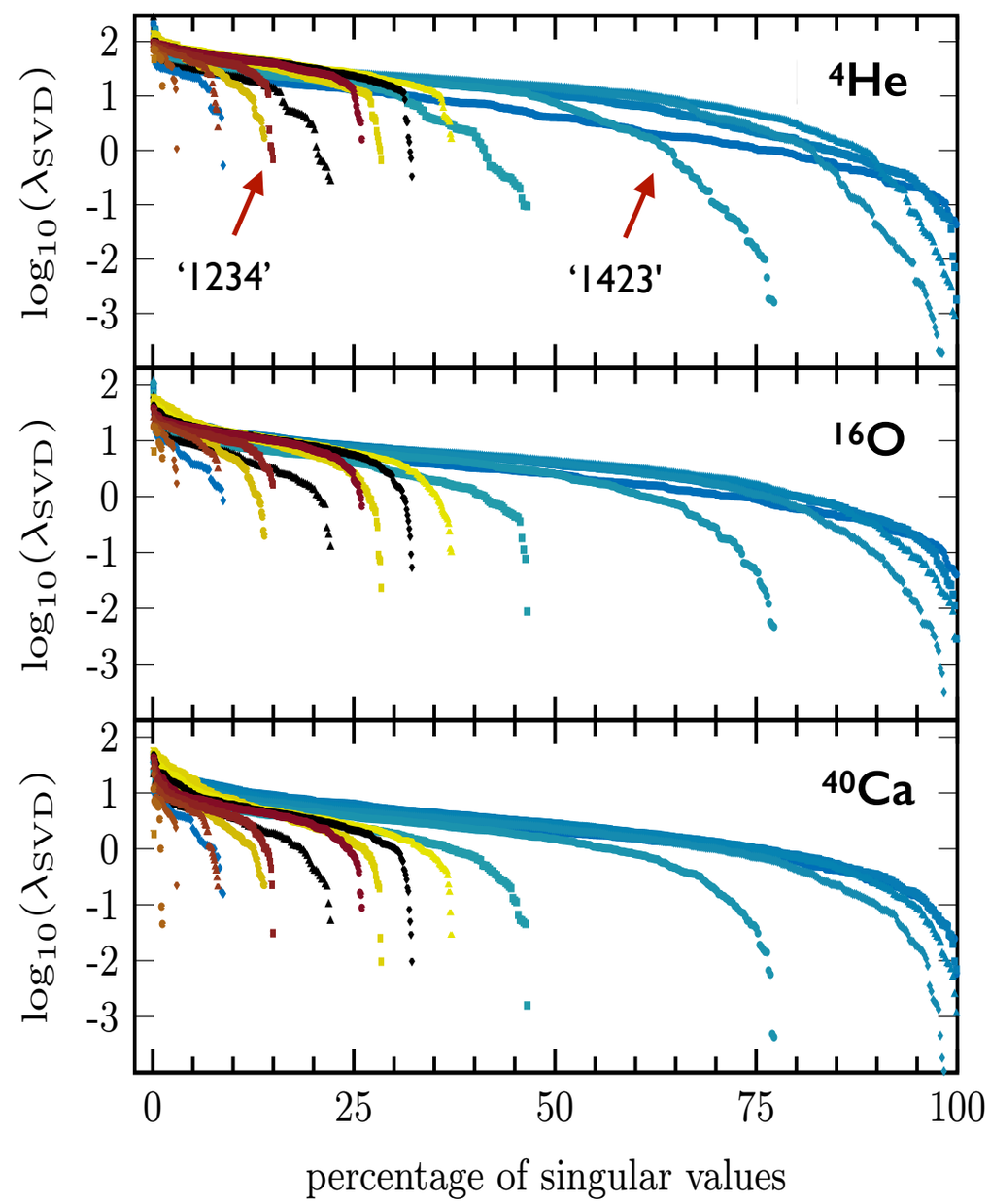


'1423' grouping

- Strong dependence on coupled two-body angular momentum *J*

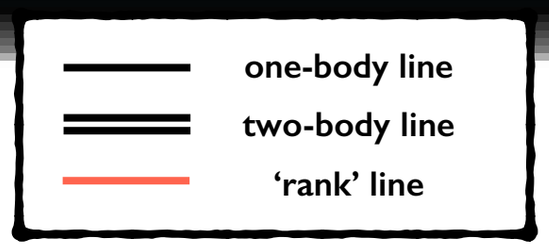
$$\langle \tilde{p}\tilde{q}J|H|\tilde{r}\tilde{s}J\rangle$$

singular values of HF matrix elements

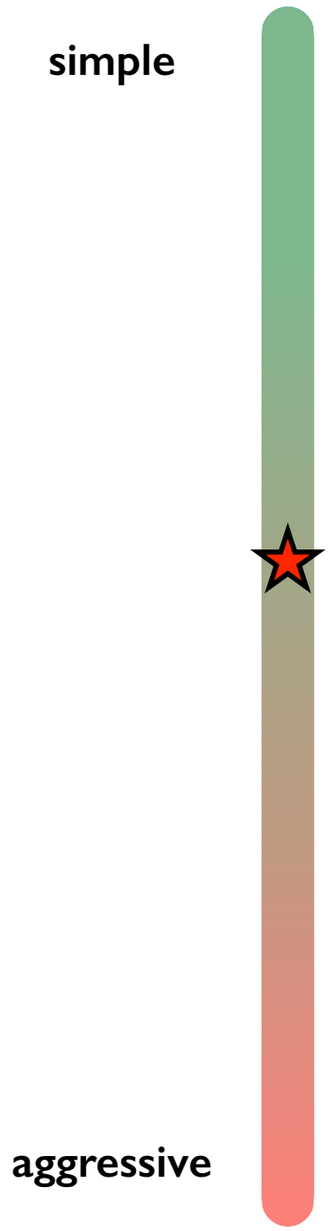


Tichai, Schutski, Scuseria, Duguet, PRC **99**, 034320

Overview of tensor formats

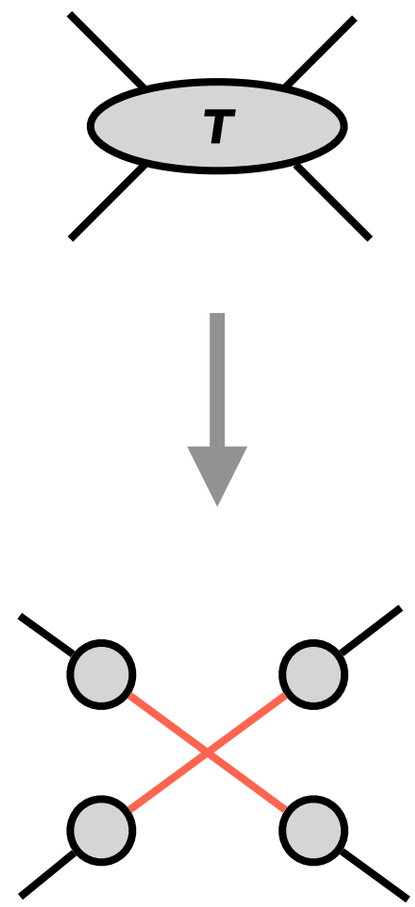


The 'spectrum' of tensor formats



CPD

canonical polyadic decomposition



Canonical polyadic decomposition

critical rank
$R_{\text{crit}} = N^{d-1}$

- **More aggressive factorization that decouples all external indices**

$$\tilde{T}_{k_1 k_2 k_3 k_4} = \sum_{\alpha}^{R_{\text{CPD}}} X^1_{k_1 \alpha} X^2_{k_2 \alpha} X^3_{k_3 \alpha} X^4_{k_4 \alpha}$$

- **CPD tensor format naturally extends to arbitrary higher modes …**

- **… but numerical computation is quite challenging for larger dimensions**

$$\mathcal{O}(N^{d-1} \cdot R_{\text{CPD}} \cdot n_{\text{iter}})$$

- **Example: two-body interaction (*d=4*) with target rank $R=N^{2.5}$**

$$\mathcal{O}(N^{5.5} \cdot n_{\text{iter}})$$

- **Decomposition factors obtained from least-square minimisation of norm difference**

$$\min_{X^1,\ldots,X^N} \|\tilde{T} - T\|^2$$

Results on CPD

- Rather slow convergence of error as function of CPD rank

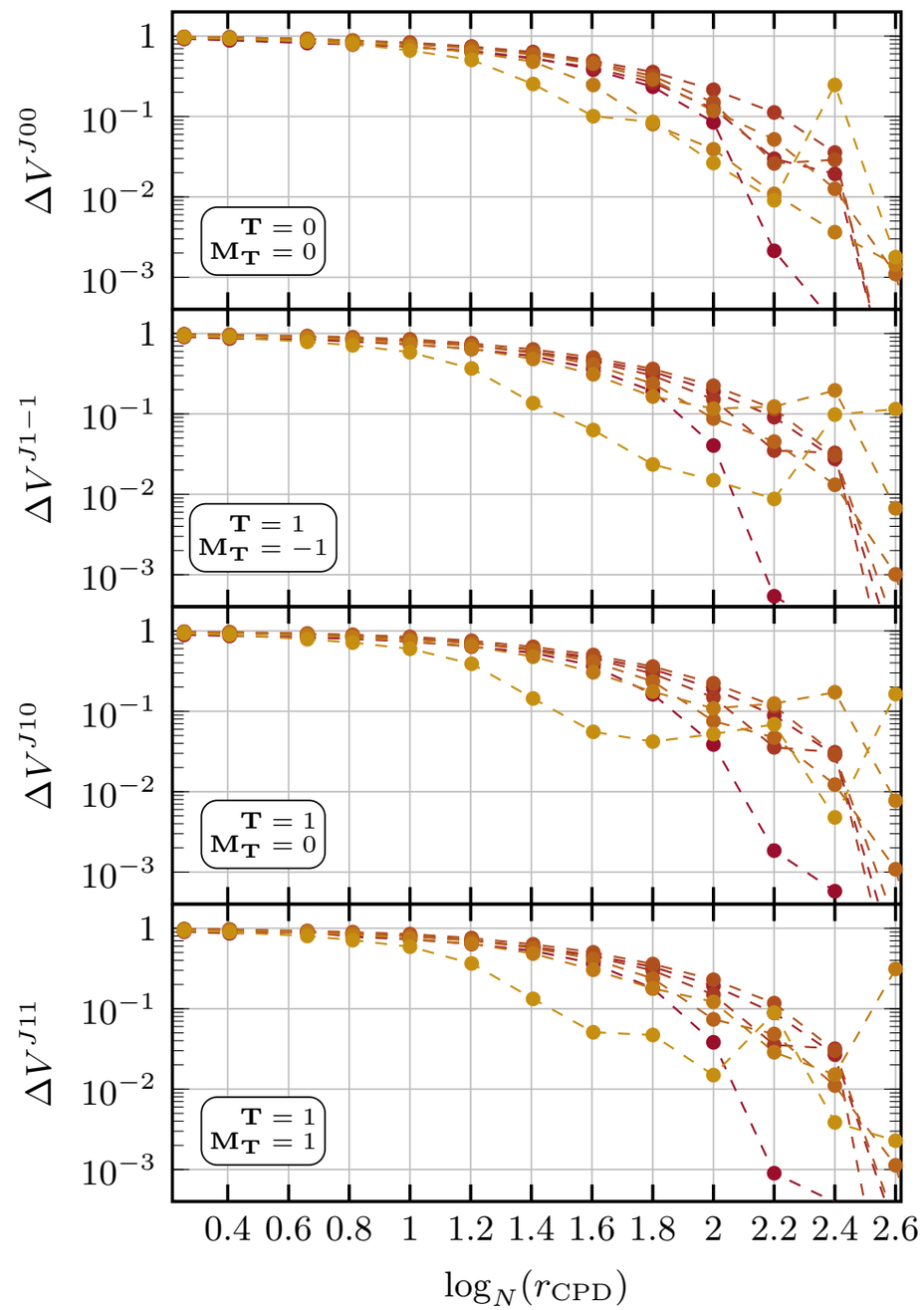
$$\langle \check{k}_1 \check{k}_2 JT | V_{\mathrm{NN}} | \check{k}_3 \check{k}_4 JT \rangle$$

- Non-monotonic error hints at numerical issues in minimisation

- When approaching the critical rank the error goes to zero

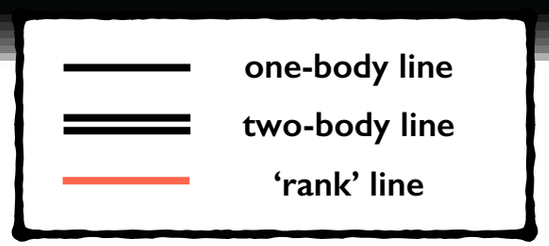
- Tensor format seems inappropriate since bra/ket treated on equal footing

CPD decomposition of HO matrix elements

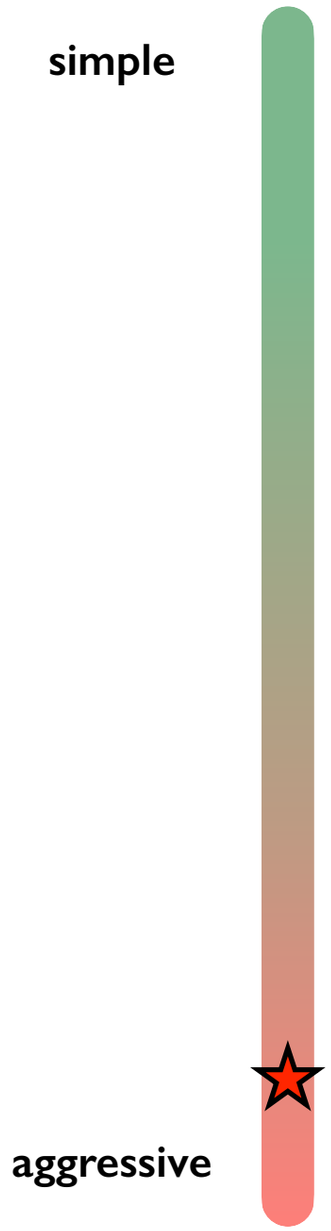


Tichai, Schutski, Scuseria, Duguet, PRC **99**, 034320

Overview of tensor formats

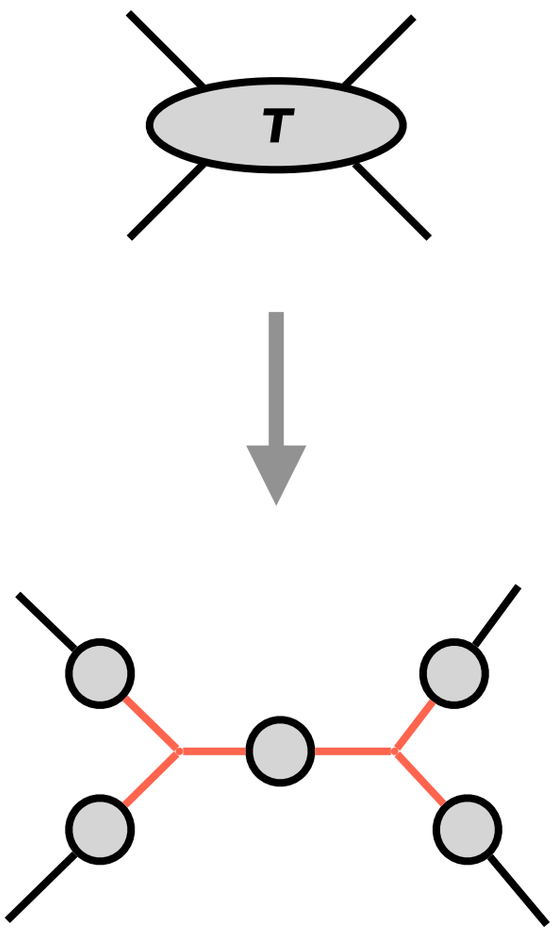


The 'spectrum' of tensor formats



THC

tensor hypercontraction



Tensor hypercontraction

Hohenstein, Parrish, Martinez, Schutski, Scuseria, …

- **Decoupling of external indices only among bra and ket indices**

critical rank
$R_{\text{crit}} = N^2$

$$\tilde{T}_{k_1 k_2 k_3 k_4} = \sum_{\alpha\beta}^{R_{\text{THC}}} X^1_{k_1\alpha} X^2_{k_2\alpha} W_{\alpha\beta} X^3_{k_3\beta} X^4_{k_4\beta}$$

core tensor

- **Hybrid tensor format merging central ideas from tSVD and CPD**

- **Storage requirements scale quadratically with respect to tensor rank**

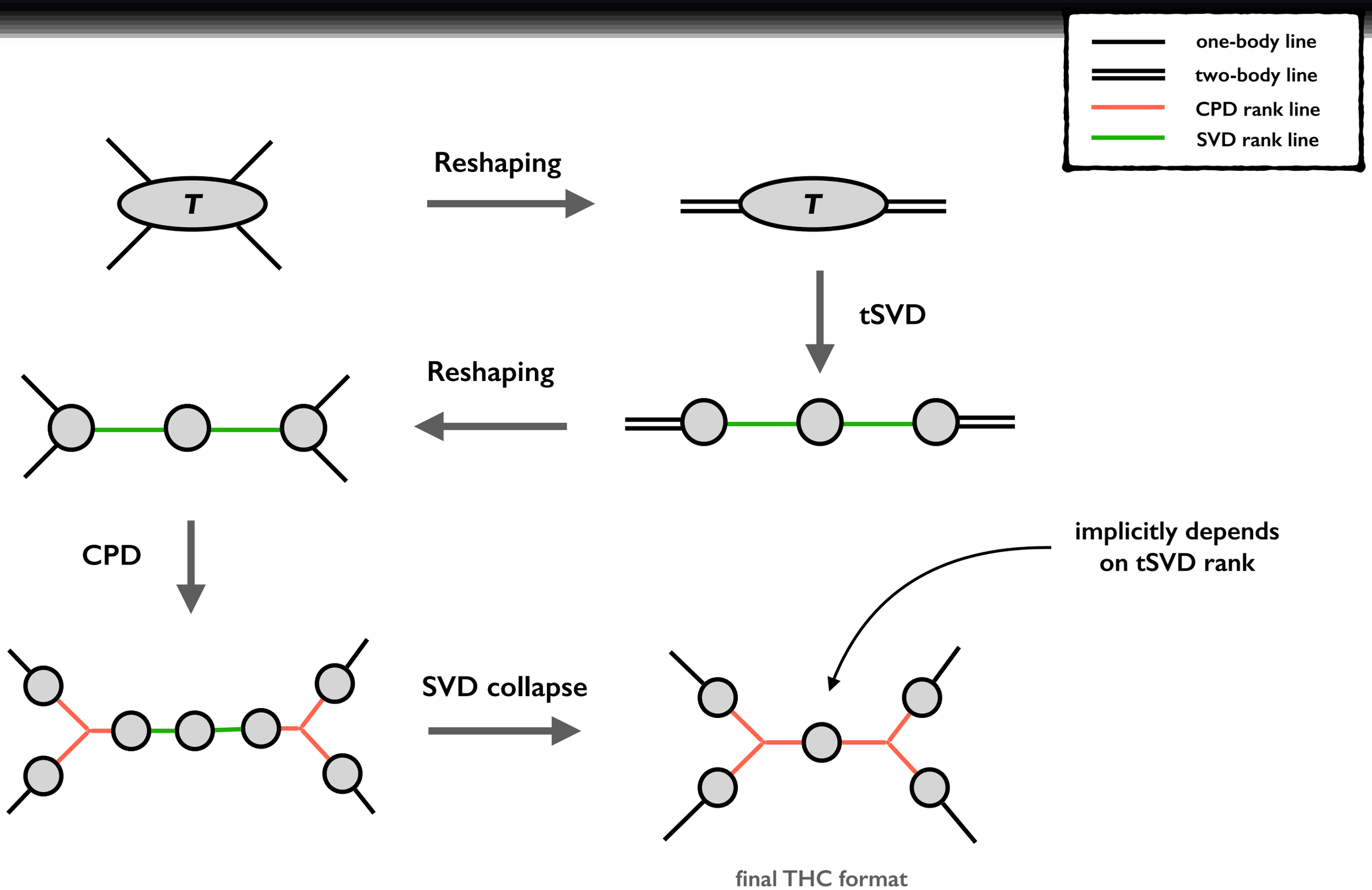
$$4 \cdot N \cdot R_{\text{THC}} + R^2_{\text{THC}}$$

- **Very successfully applied in various quantum chemistry applications**

$$R_{\text{THC}} = \mathcal{O}(N)$$

- **Practically obtained from a multi-step procedure involving CPD and tSVD**

How to obtain the THC format



THC results

- **Fast convergence for most channels**

$$\langle \tilde{k}_1 \check{k}_2 J | H_{\text{intr.}} | \tilde{k}_3 \tilde{k}_4 J \rangle$$

- **Intermediate values of two-body angular moment converge slower**

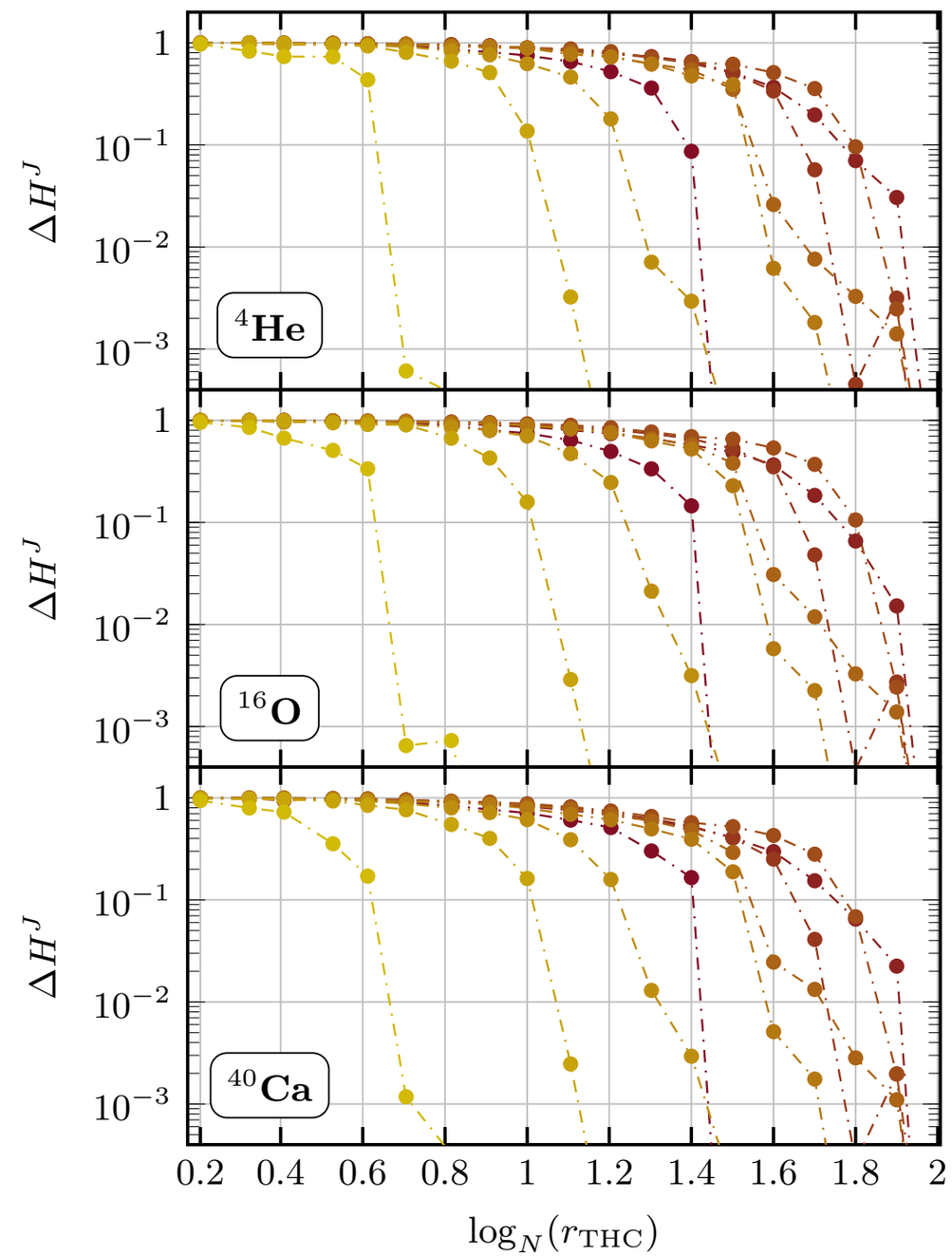
- **Rapid decrease of decomposition error near critical rank**

- **Computationally cheaper than CPD**

- **Less good than in quantum chemistry**

$$R_{\text{THC}} = \mathcal{O}(N^{1.4-1.8})$$

THC decomposition of HF matrix elements



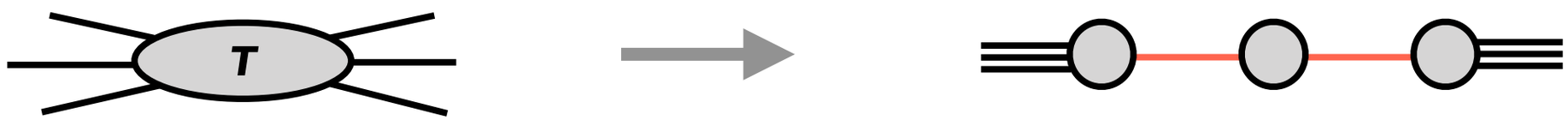
Tichai, Schutski, Scuseria, Duguet, PRC **99**, 034320

Perspectives on three-body operators

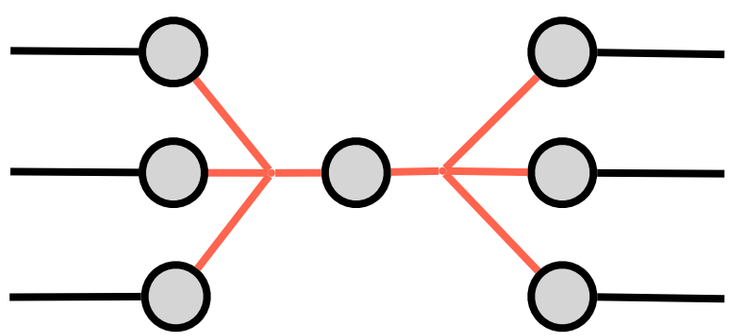
- Low-mode tensors are part of extensive investigations in applied mathematics

... but structured mode-6 tensors are exotic!

- Naive approach: application of truncated SVD in three-body space



- Design of more aggressive factorizations is less straightforward



Idea: THC for mode-6 tensor

Main theorem of tensor calculus

"Employ factorization techniques as early as possible in your workflow. Once your tensors are decomposed you should NEVER (!!!) reconstruct them!"

Factorization of matrix elements

see talk on Tuesday!

Momentum space
(system independent)

Talmi-Moshinsky +
HO transformation

Harmonic oscillator
(system independent)

isospin decoupling +
s.p. basis transformation

Hartree-Fock
(system dependent)

s.p. basis transformation

Natural orbitals
(system dependent)

currently factorization applied here!

Factorization of matrix elements

see talk on Tuesday!

Momentum space
(system independent)

Future:
apply factorization here!

Talmi-Moshinsky +
HO transformation

Harmonic oscillator
(system independent)

isospin decoupling +
s.p. basis transformation

Hartree-Fock
(system dependent)

s.p. basis transformation

Natural orbitals
(system dependent)

Factorization of basis transformation

- Many-body theory requires extensive use of single-particle basis transformation

$$T^{\mathcal{B}'}_{k_1k_2k_3k_4} = \sum_{l_1l_2l_3l_4} C_{l_1k_1}C_{l_2k_2}C_{l_3k_3}C_{l_4k_4}T^{\mathcal{B}}_{l_1l_2l_3l_4}$$

- THC decomposition enables transforming factor matrices only (same for CPD)

$$Y^i_{k_1\alpha} = \sum_{l_i} C_{l_ik_i}X^i_{l_i\alpha}$$

- Core tensor remains unchanged since it only depends on auxiliary indices

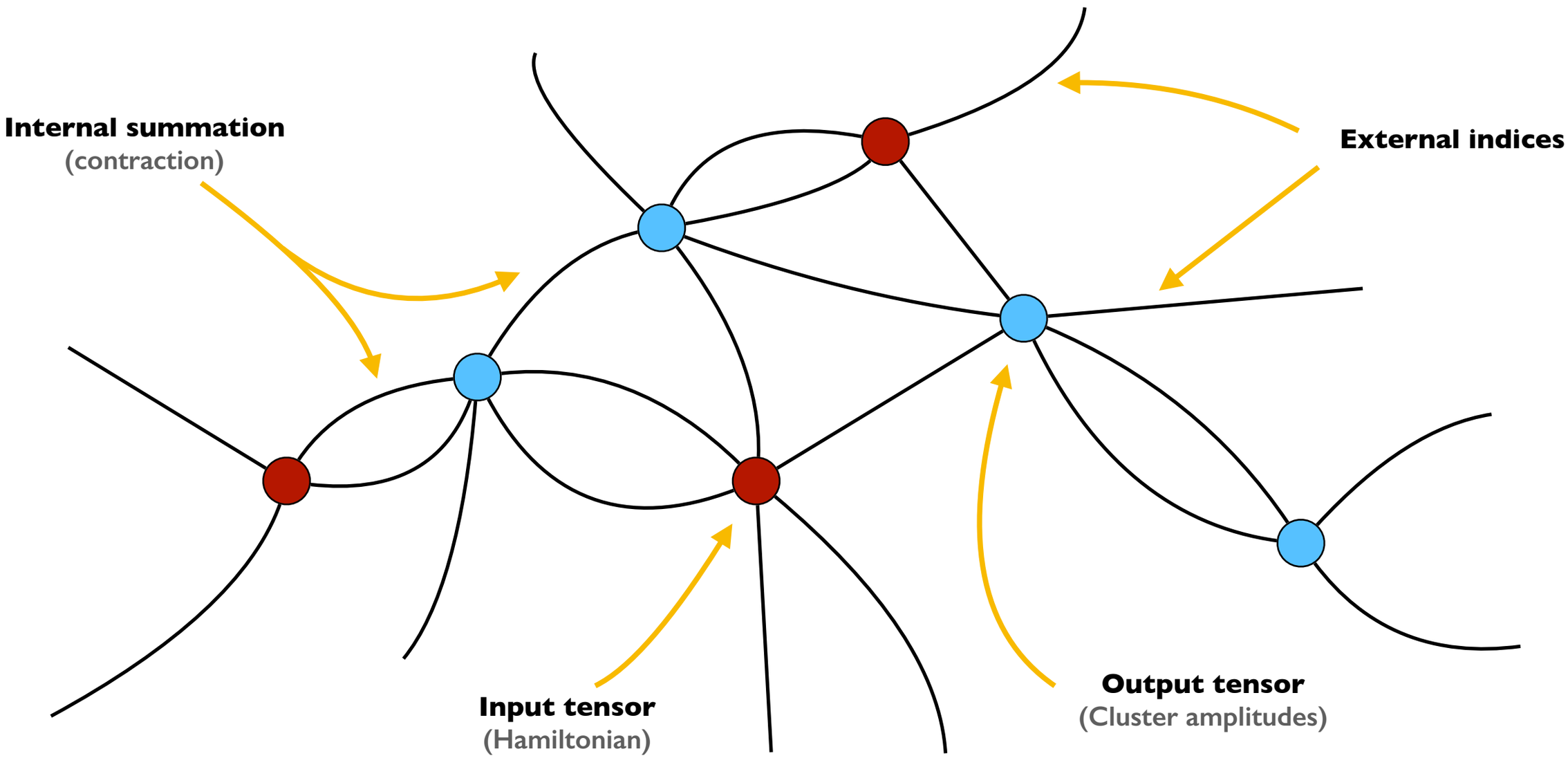
$$\tilde{T}^{\mathcal{B}'}_{k_1k_2k_3k_4} = \sum_{\alpha\beta}^{R_{\mathrm{THC}}} Y^1_{k_1\alpha}Y^2_{k_2\alpha}W_{\alpha\beta}Y^3_{k_3\beta}Y^4_{k_4\beta}$$

- Reduced computational complexity when operating on decomposition factors

$$\mathcal{O}(N^5) \xrightarrow{\mathrm{THC}} \mathcal{O}(N^2R_{\mathrm{THC}})$$

Tensor-structured many-body theory

The general problem

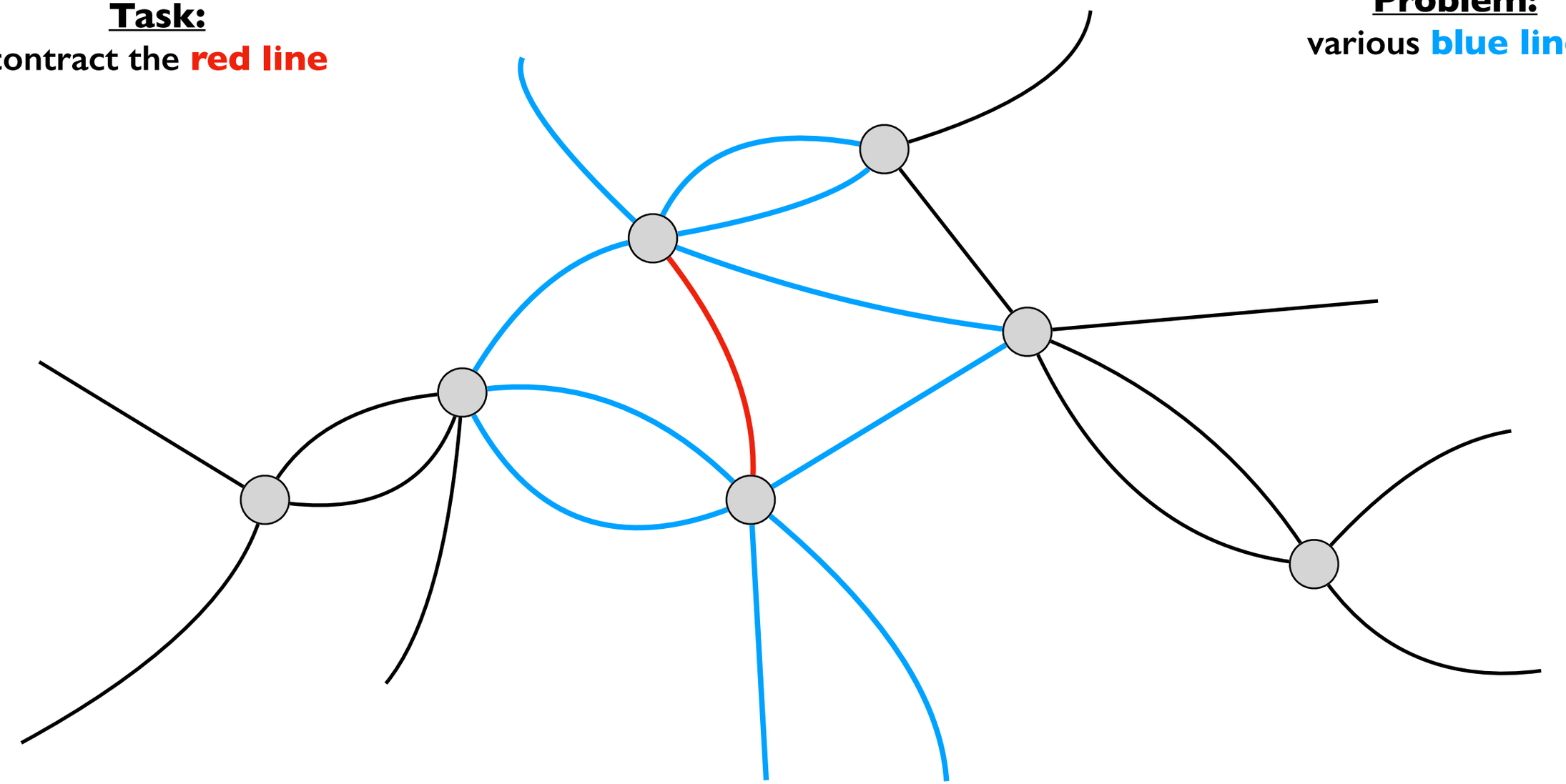


***Ab initio* many-body theory is (to a large extent)**
the theory of efficiently processing tensor contractions!

Why is many-body theory so hard?

Task:
contract the **red line**

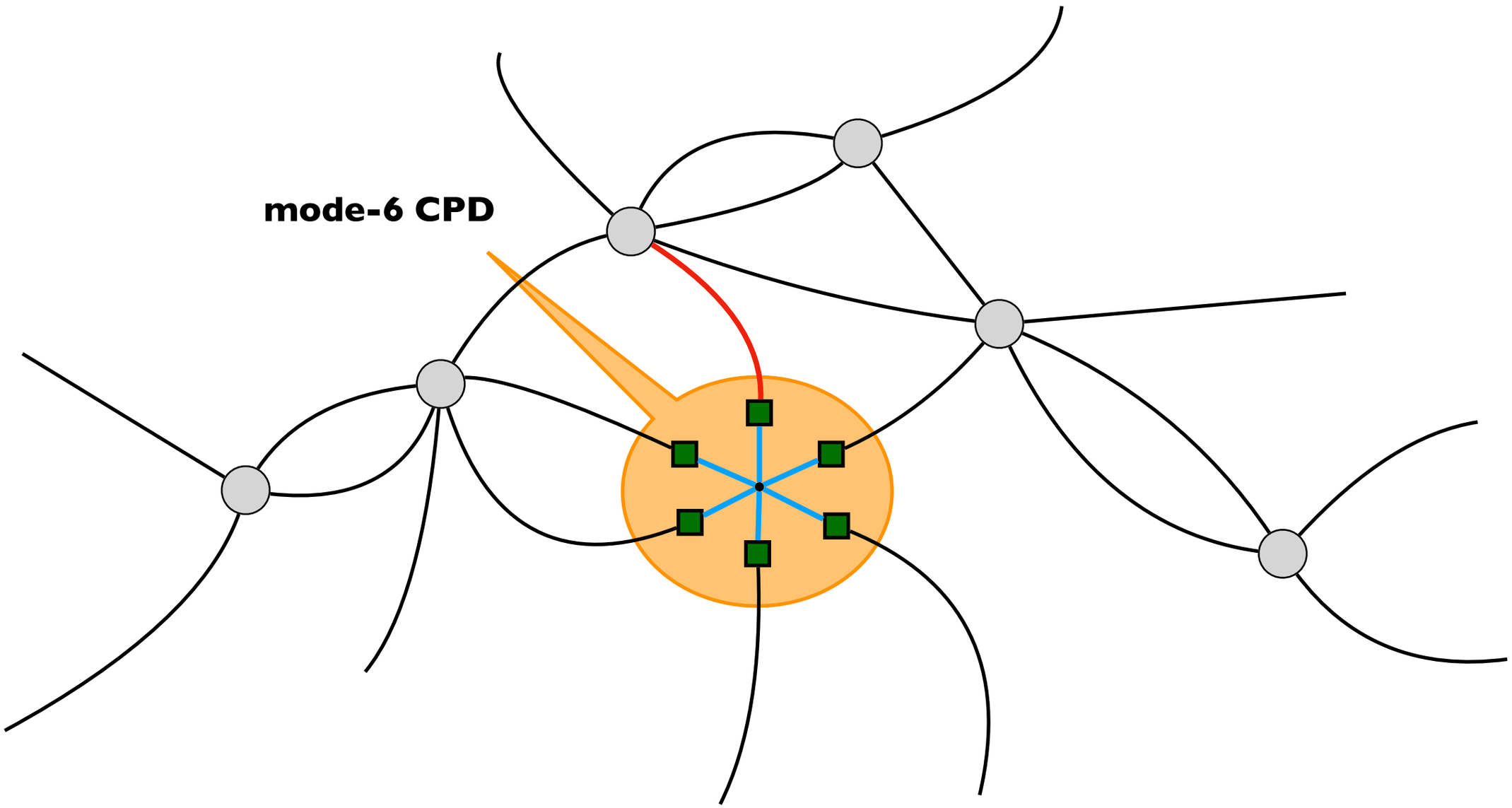
Problem:
various **blue lines**



A contraction needs to be done
for every combination of external indices!

Why is many-body theory so hard?

Now with tensor decomposition!



Ultimately the many-body approach is expressed in terms of <u>one-body operators</u> only!

Why tensor decomposition?

- Decompositions enable for flexible contractions schemes for tensor networks

$$\tilde{M} \cdot \tilde{M} = \sum_{p}^{n} \tilde{M}_{k_1 p} \tilde{M}_{p k_2} = \sum_{p}^{n} \sum_{\alpha\beta\gamma\delta}^{R} (\tilde{U}_{k_1 \alpha} \tilde{\Sigma}_{\alpha\beta})(\tilde{V}^T_{\beta p} \tilde{U}_{p\gamma})(\tilde{\Sigma}_{\gamma\delta} V^T_{\delta k_2})$$

n^3 operations

nR^2 operations

- Tensor contractions can be performed within different complexity class

$$N^p \longrightarrow N^{p'} \qquad p' < p$$

- Example: tSVD of matrix with dimension *N=100* and SVD rank *R=20*

$$(R = N^{0.65})$$

$$N^3 \longrightarrow N^{2.3}$$

- Problem: not all tensor contractions are simple matrix products!

Energy denominators

- Perturbation theory expressions naturally involve energy denominators

$$D_{abij} = \frac{1}{\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j}$$

- Analytical CPD factorization can be obtained via inverse Laplace transform

$$D_{abij} = \int_0^\infty e^{-t(\epsilon_a + \epsilon_b - \epsilon_i - \epsilon_j)} dt$$

- Decomposition factors are obtained via numerical quadrature

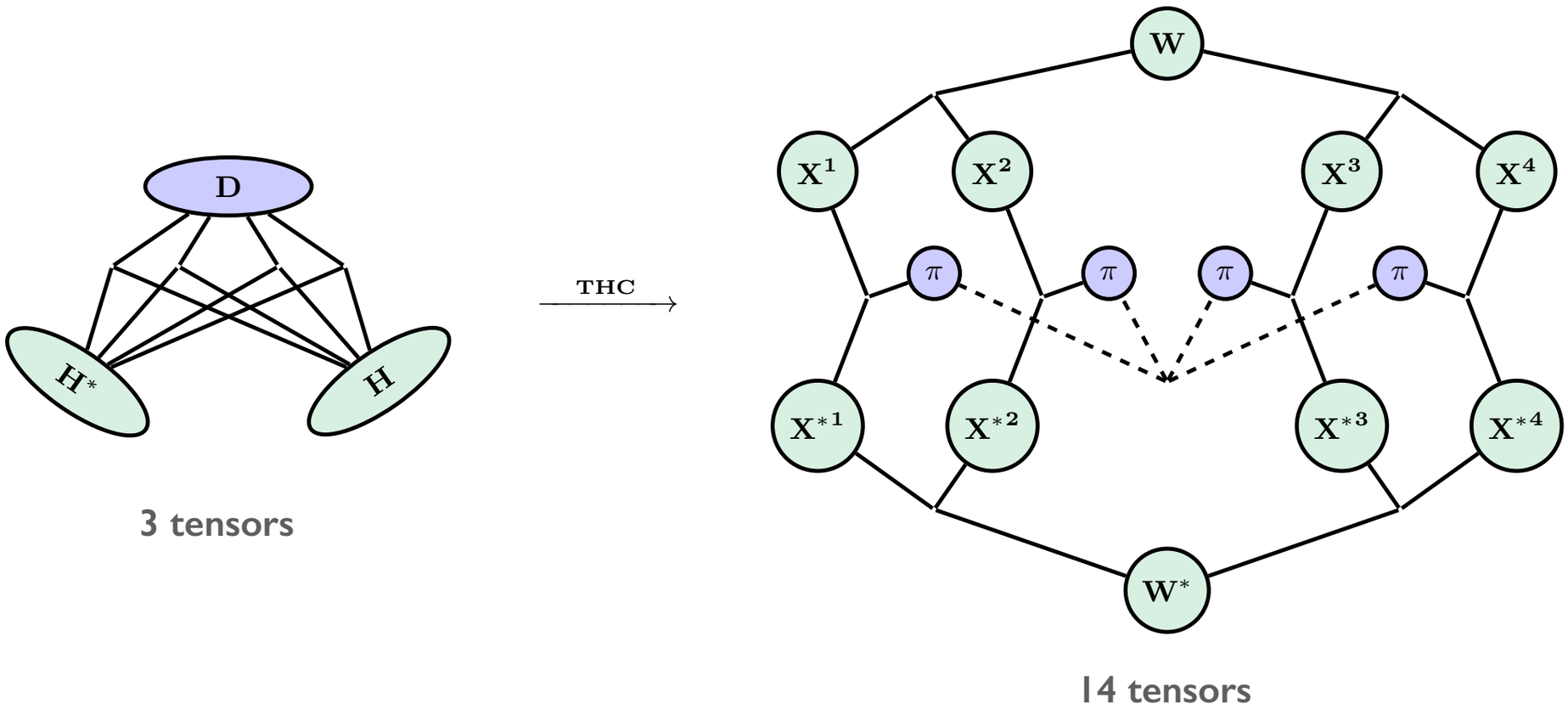
$$\tilde{D}_{abij} = \sum_s \pi_{as} \pi_{bs} \omega_s \pi_{is} \pi_{js}$$

Braess, Hackbusch,
IMA J. Numer. Anal. 25, 685 (2005)

- Integration with very high precision using constant (system-independent) mesh size

- Extension to higher-mode tensors can be done in the same way

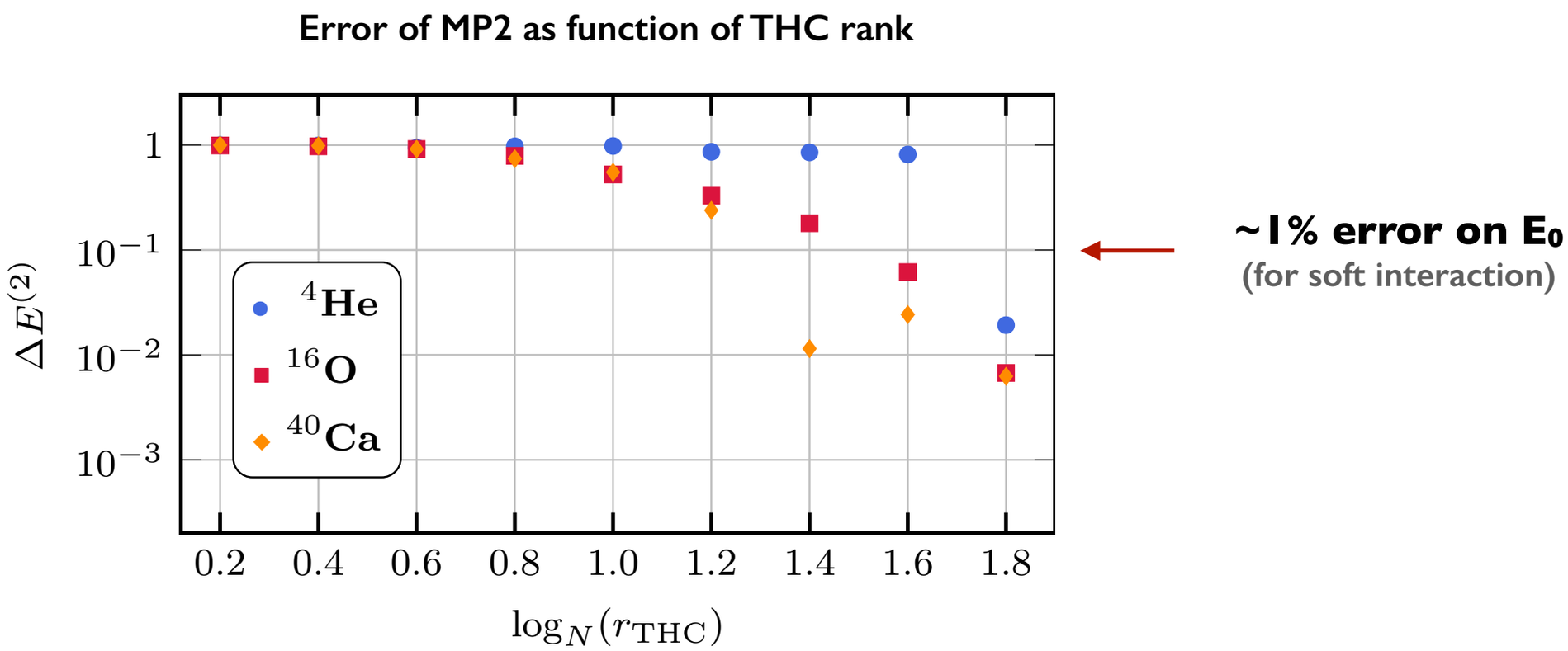
THC-factorized MP2



- Simple MP2 tensor network replaced by more complicated factorized topology
- Factorization leaves more freedom in optimising the contraction order
- General feature: many-body frameworks becomes more involved

THC-MP2 results

Tichai, Schutski, Scuseria, Duguet, PRC **99**, 034320

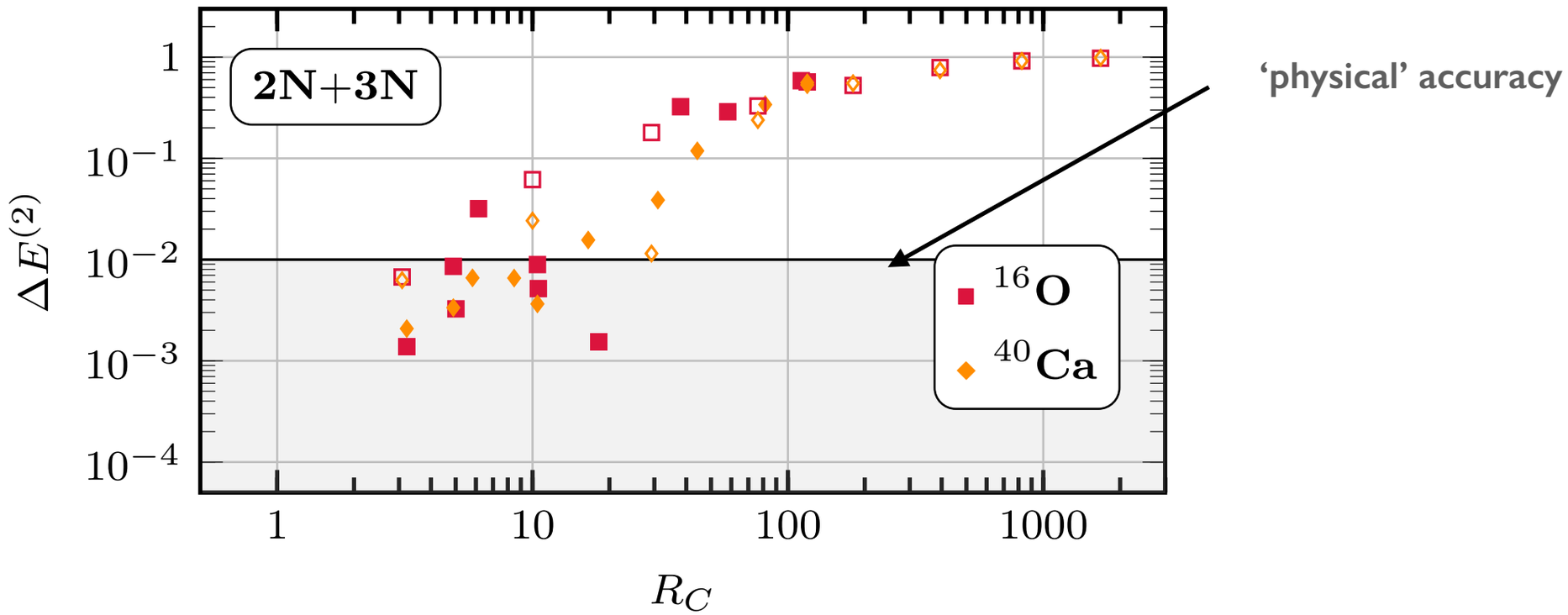


- **Improved accuracy** of correlation energy for higher decomposition ranks
- **Monotonic behaviour** sometimes broken and ^{4}He involves jumps
- Correlation error vanishes when approaching the **critical THC rank** of $R_{\mathrm{THC}} = N^2$

Compression rates

Modelspace:
5 oscillator shells
140 basis functions

Correlation between MP2 error and compression rate



- **General trend: higher precision corresponds to lower compression rates**

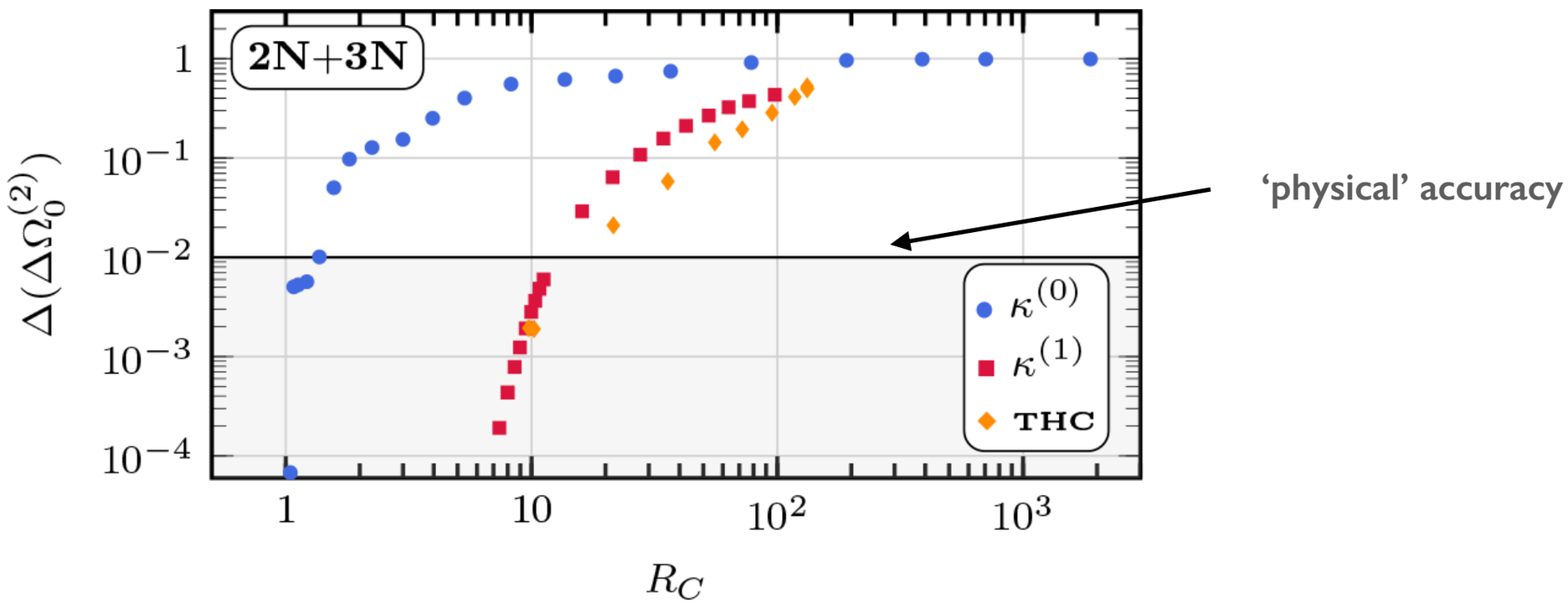
$$R_C = \frac{\text{full storage}}{\text{compressed storage}}$$

- **Significant compression rates obtained in high-precision regime (ΔE < 1%)**
- **Much larger compression rates expected in large model spaces**

Factorization vs. Selection

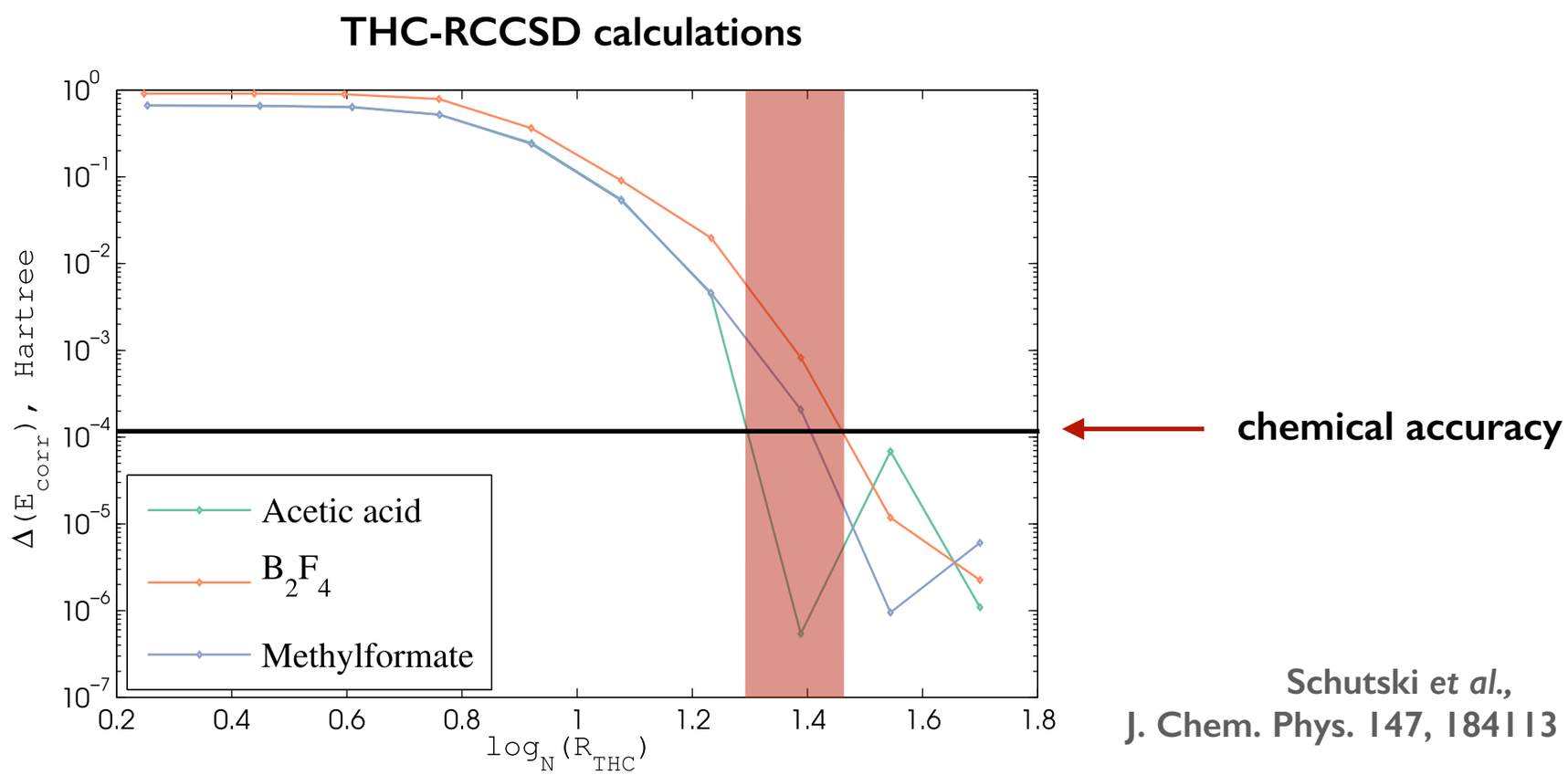
Modelspace:
5 oscillator shells
140 basis functions

Comparison of IT and THC for BMBPT2 error



- Efficiency of IT and tensor factorization is very similar in schematic applications
- Advantage of THC: pre-processing enables for lower-scaling many-body methods
- Advantage of IT: conceptually simpler since no (expensive) decomposition needed

Tensor-structured coupled cluster theory



Schutski *et al.*,
J. Chem. Phys. 147, 184113 (2017)

- **Chemical accuracy** in THC-RCCSD application with (almost) linear rank
- **Non-monotonic dependence** of error as function of decomposition rank
- Generation of source code **cannot be implemented manually** anymore

> 10.000 equations!

(>500 pages of appendices)

Wrap-up

Tensor decompositions

- Novel exciting tool to lower computational resources in many-body theory
- Tensor formats can easily adapt to various situations/symmetries
- First applications show promising performance in nuclear physics theory

Future work

- Development of new tensor formats specific to nuclear theory applications
- Implementation of large-scale codes to reach larger model spaces
- Adaption of many-body toolchain to factorized tensor representations

Further reading

- **General introduction to tensor calculus**

  Tensor review: Kolda, Bader, *SIAM Rev.*, 51(3), 455–500.

- **(Selected!) Applications in quantum chemistry**

  SVD-CCSD: Kinoshita, Hino, Bartlett, J. Chem. Phys. 119, 7756 (2003)

  SVD-CCSD-T1: Hino, Kinoshita, Bartlett, J. Chem. Phys. 121, 1206 (2004)

  THC: Parrish, Hohenstein, Schunck, Sherill, Martinez, Phys. Rev. Lett. 111, 132505 (2013)

  Rank-reduced CCSD: Parrish *et al.*, J. Chem. Phys. 150, 164118 (2019)

  Rank-reduced EOM-CCSD: Hohenstein *et al.*, J. Chem. Phys. 151, 164121 (2019)

  THC series: Hohenstein/Parrish/Martinez/Kokkila/Sherrill (2013-2017)

  THC-CCSD: Schutski *et al.*, J. Chem. Phys. 147, 184113 (2017)

- **Applications in nuclear physics**

  DMRG in shell model: Papenbrock, Dean, J. Phys. G 31, 1377 (2005)

  CPD/THC & MP2: Tichai, Schutski, Scuseria, Duguet, Phys. Rev. C 99, 034320 (2019)

  THC & BMBPT: Tichai, Ripoche, Duguet, Eur. Phys. Jour. A 55: 90 (2019)

- **Software:**

  Tensorlab 3.0 [Matlab]: Vervliet, Debals, Sorber, Van Barel, De Lathauwer, `https://www.tensorlab.net/`