TN-Based Real Space Renormalization Group and Related Things

東京大学 物性研究所

Naoki KAWASHIMA (ISSP)

2025.07.30

Collaborators

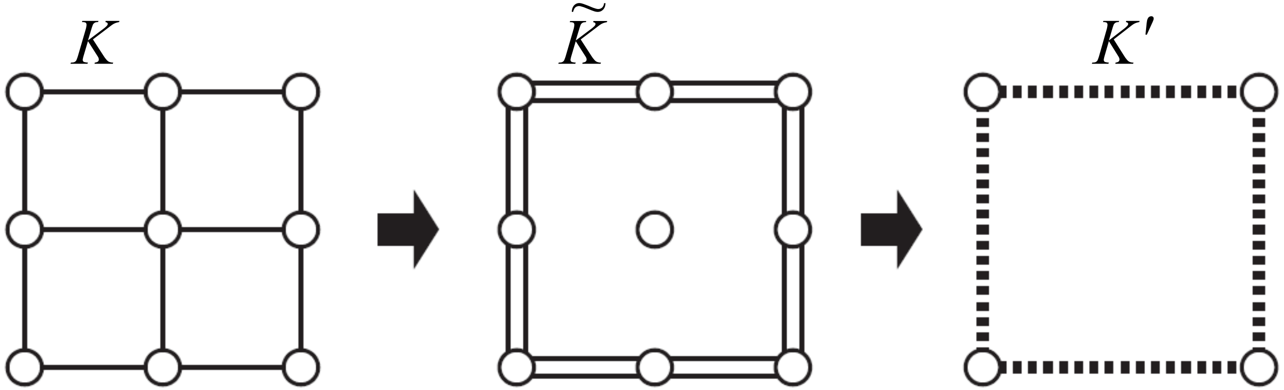
Kenji HOMMA (ISSP) — Nuclear norm regularization

Xinliang LYU (IHES, France) — Linearized TNRG, Entanglement filtering for 3D

Satoshi MORITA (Keio U.) — Impurity method

Tsuyoshi OKUBO (U Tokyo) — Nuclear norm regularization

Migdal-Kadanoff Real-Space RG

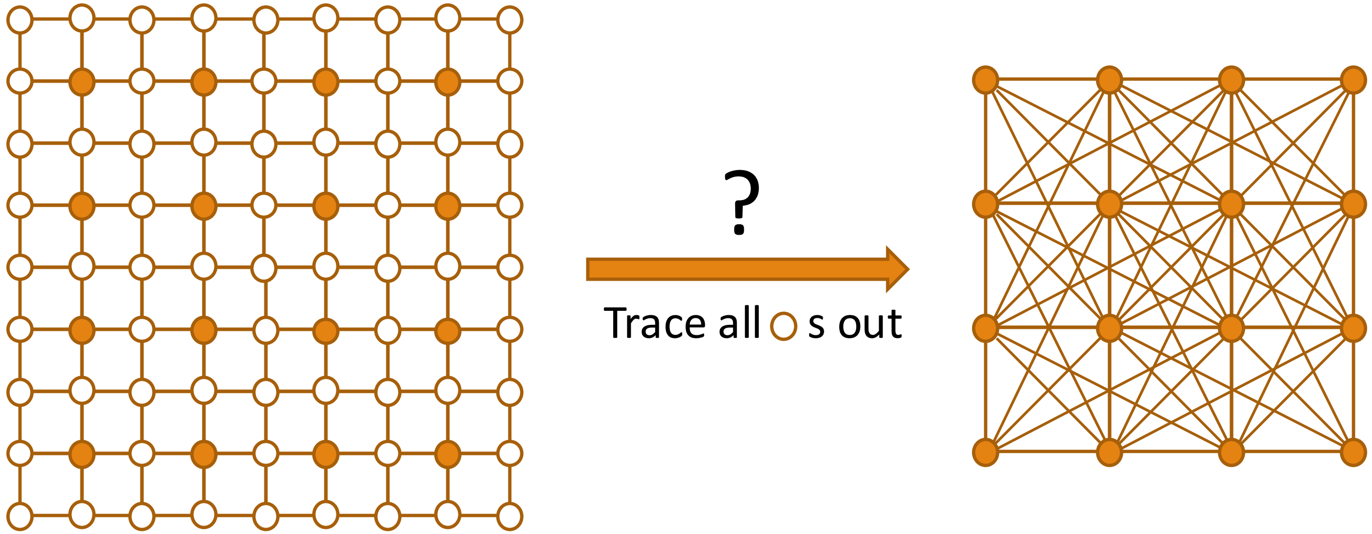


$$\widetilde{K} = 2K, \quad \tanh K' = \left(\tanh \widetilde{K}\right)^2$$
$$K' = \tanh^{-1}\left(\left(\tanh 2K\right)^2\right)$$

Simple, but not very accurate or controllable.

Even more fundamental problem in Hamiltonian formulation

We (or some of us, incl. me) often say in lectures that an RG transformation produces a lot of complicated interactions. **But is that all?**



van Enter, Fernandez and Sokal, JSP72, 879 (1993)

Can we define the renormalized Hamiltonian only with multi-body **interactions decaying exponentially as a function of distance?**

NO!

Fundamental Problem in Hamiltonian Formulation

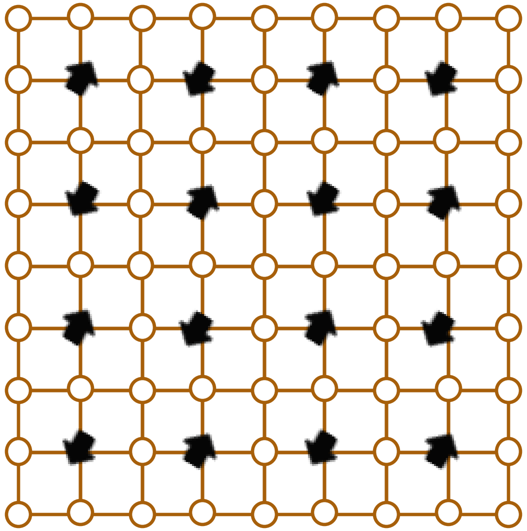
van Enter, Fernandez and Sokal, JSP72, 879 (1993)

Why not? --- Because there are some spin configurations after RG that allows the decimated spins undergoes a 1st order phase transition.

The ○ spins form a spin system on the decorated square lattice.

For the configuration of ■ spins in the right figure, they have no effects on ○ spins because of cancellation.

Therefore, below the Tc of the decorated square Ising model, the effect of the boundary (infinitely far away) controles deep inside.



It means that the local thermal distribution of any cluster in the renormalized system can NOT be uniquely determined by fixing a finite number of ■ spins surrounding it.

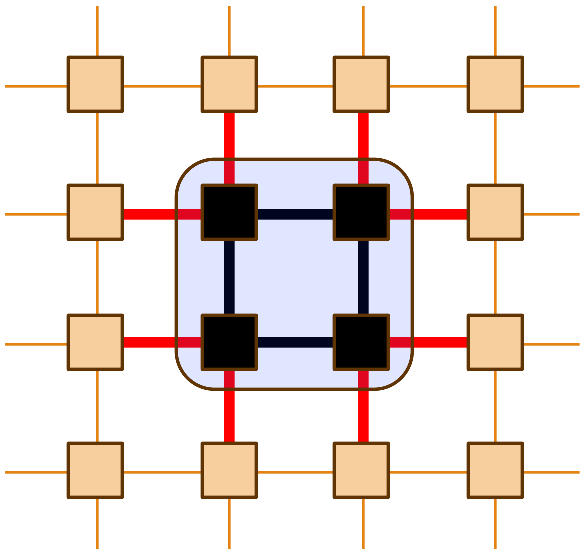
No regular Hamiltonian can produce such a property

Markov Property of TN

In contrast to the Hamiltonian RG, in TNRGs, the correlation between inside and outside of a cluster can flow only through the boundary indices.

(By fixing the boundary indices, we can uniquely determine the state inside the cluster, independent of the state of any other part of the environment.)

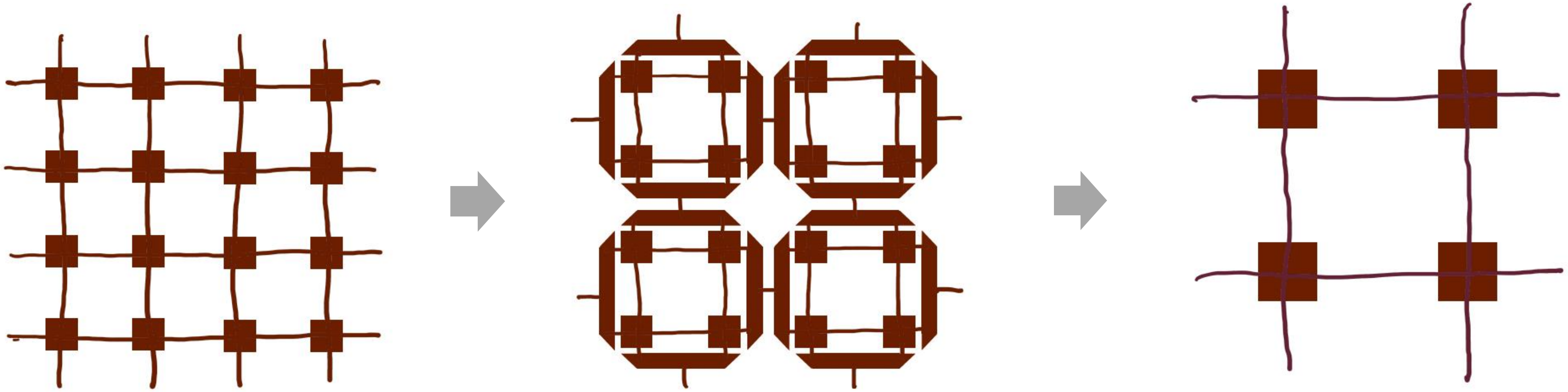
The RSRG can be well defined in terms of TN!

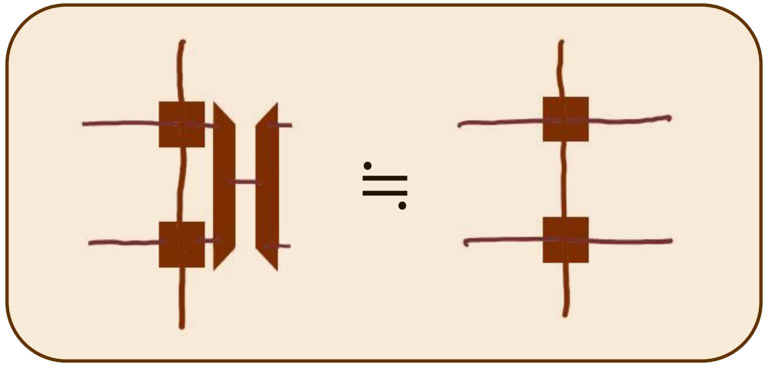


By choosing the indices on ▬ s, the state of the central cluster is uniquely determined.

HOTRG

Xie, et al. Phys. Rev. B **86**, 045139(2012)

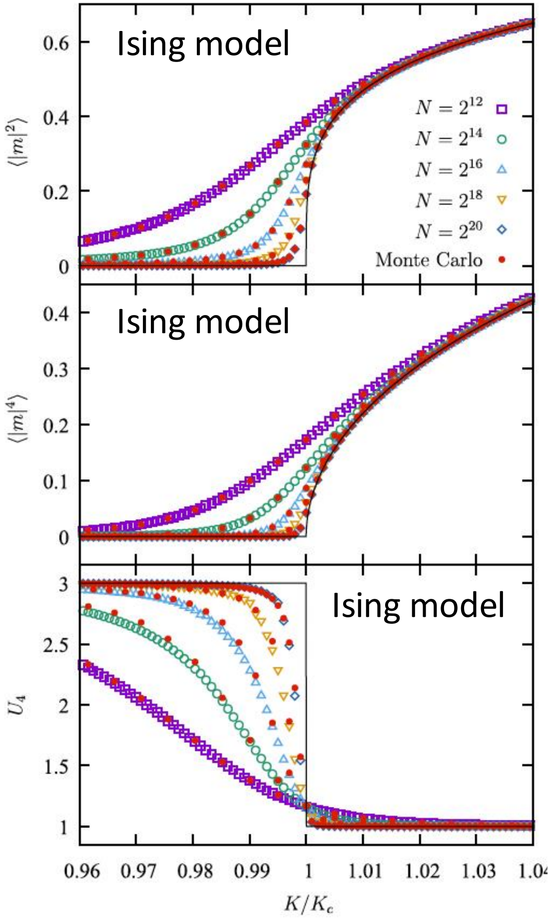


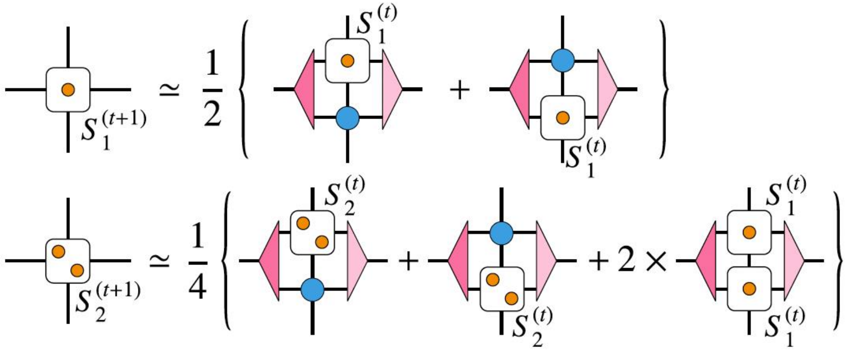


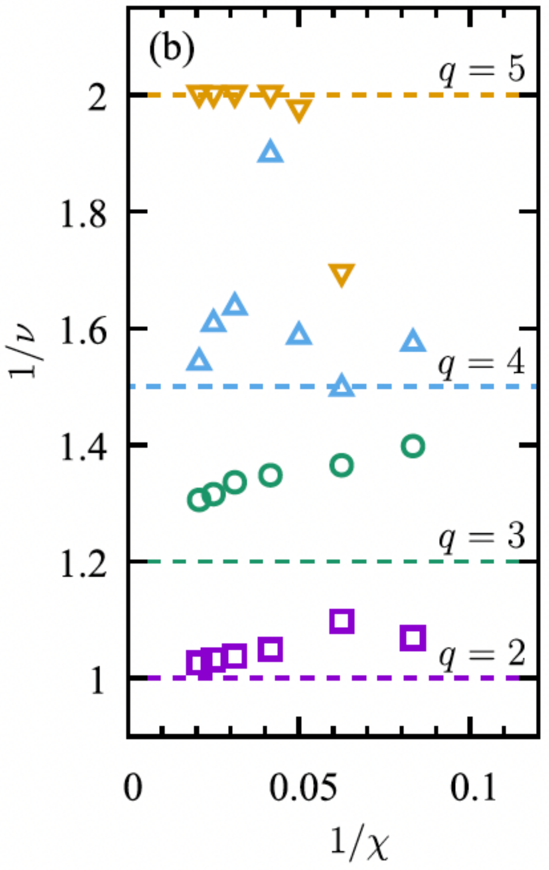
※ These diagrams are just for the concept. In Xie, et al (2012), one direction is contracted at each time.

Estimation of correlation functions
--- Impurity method ---

TNRG is not only good at critical properties but accurately reproduces the finite-size corrections as well.







TNRG successfully identify the 1st order nature of the transition of 5-Potts in 2D.

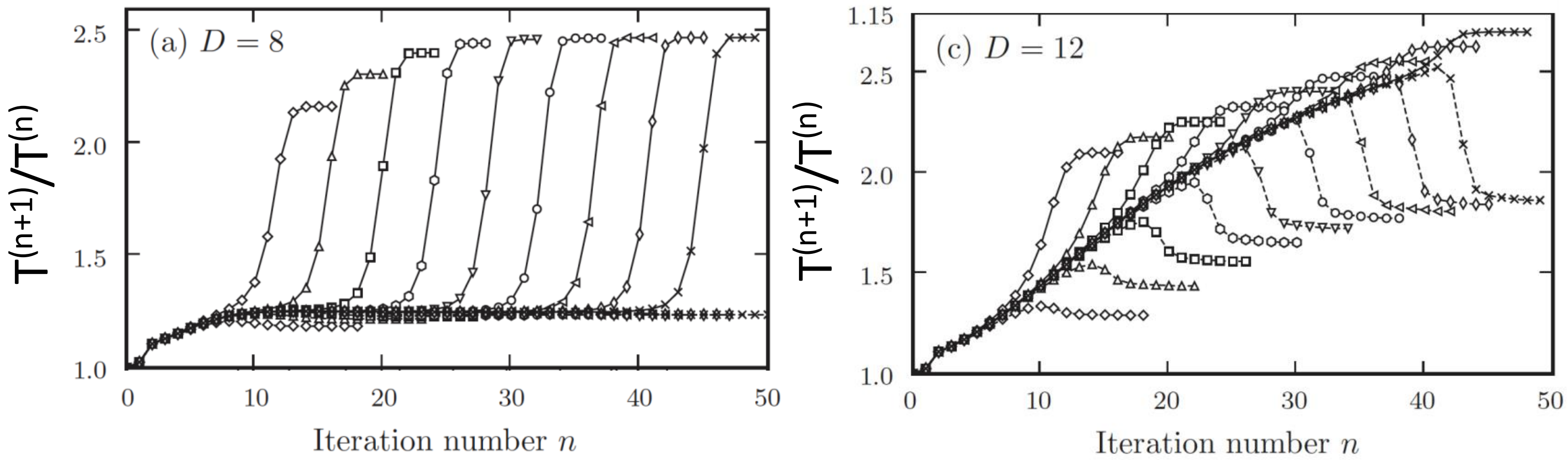
Morita and NK: Comp. Phys. Comm. 236 65 (2019); PRB111, 054433 (2025)

Hinczewski and Berker: PRE 77 011104 (2008)

Convergence to fixed point

--- TRG fails to converge, not just inaccurate ---

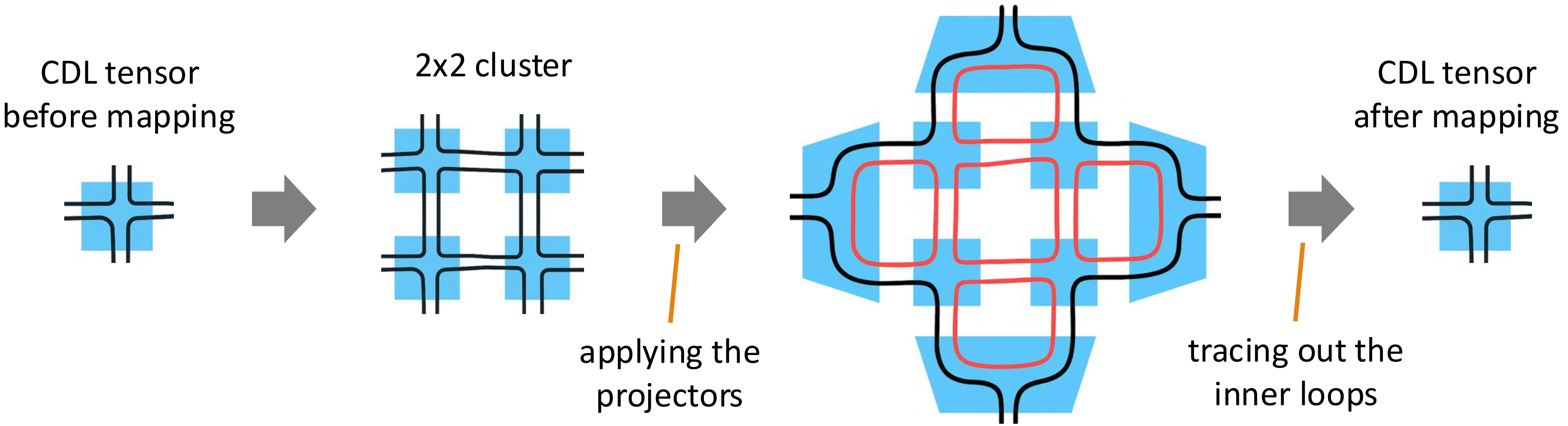
TRG (tensor network renormalization a la Levin and Nave) for triangular lattice Ising



We don’t reach the fixed-point tensor for large bond dimensions.
(Something similar happens to HOTRG.)

artificial

CDL tensor is a fixed point of TNRG



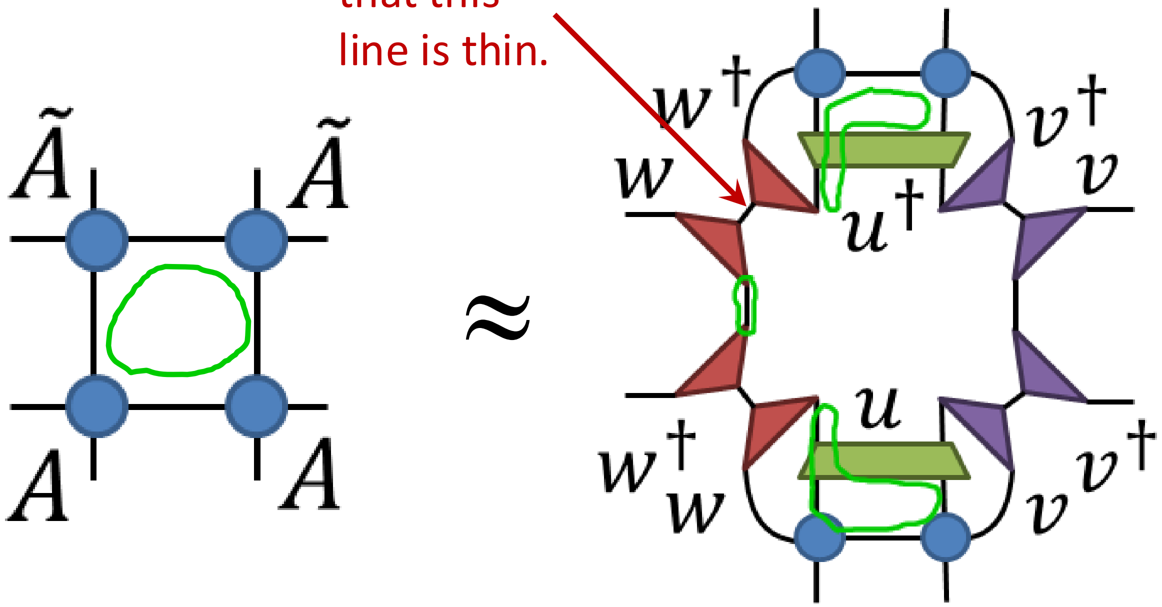
HOTRG does not change the CDL tensor.

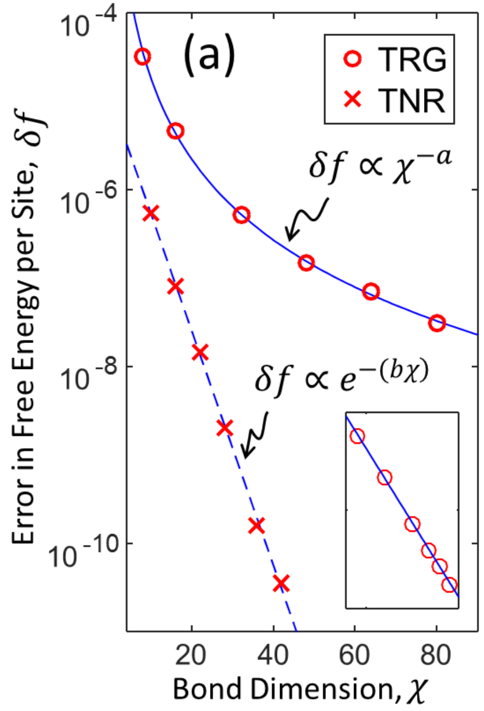
Typically, TNRG at criticality is more complicated. Still, the CDL structure may arise, and the tensor may become a tensor product of the meaningless CDL and the core that carries the essence of the critical properties. Once the CDL arises, it stays there and cramping the core.

The 1st approach

Disentangler removes correlation loops

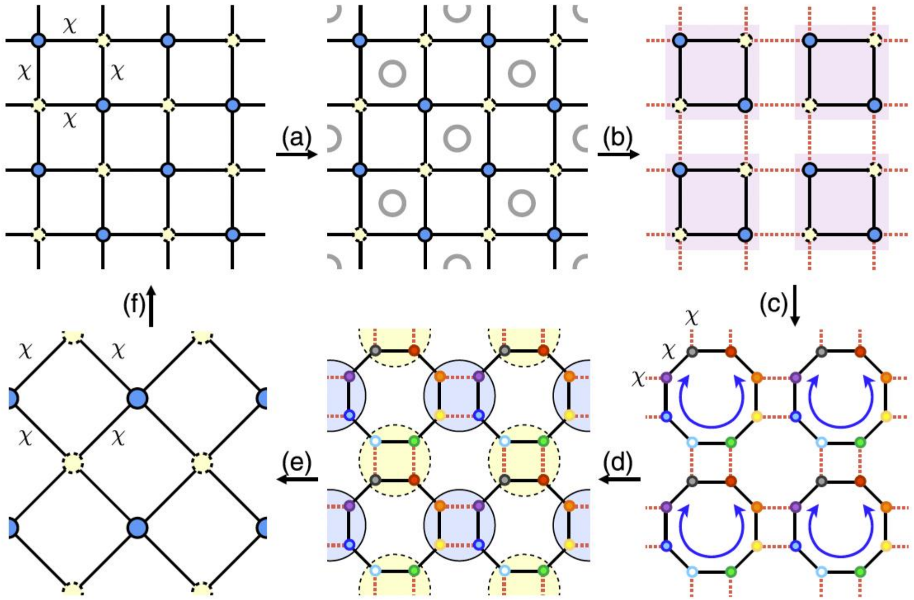
It is essential that this line is thin.

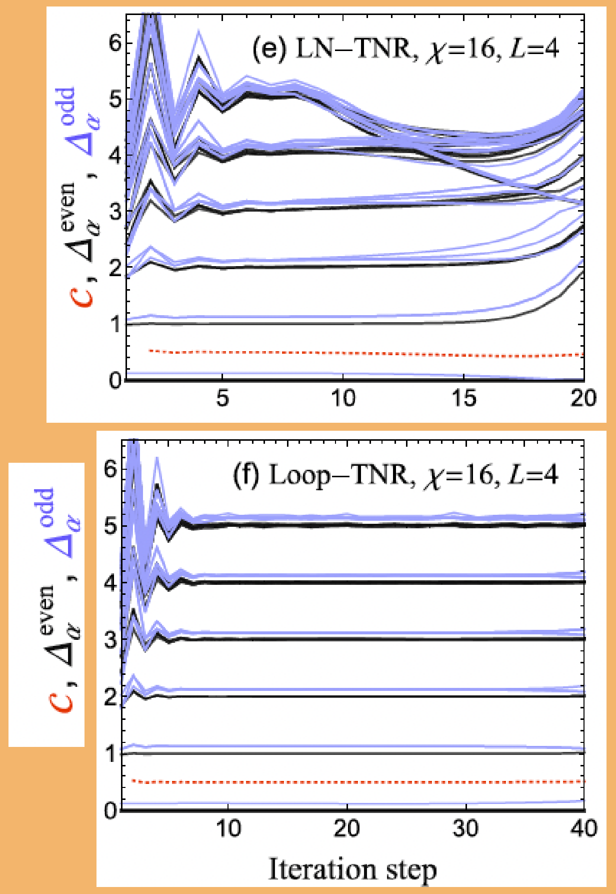




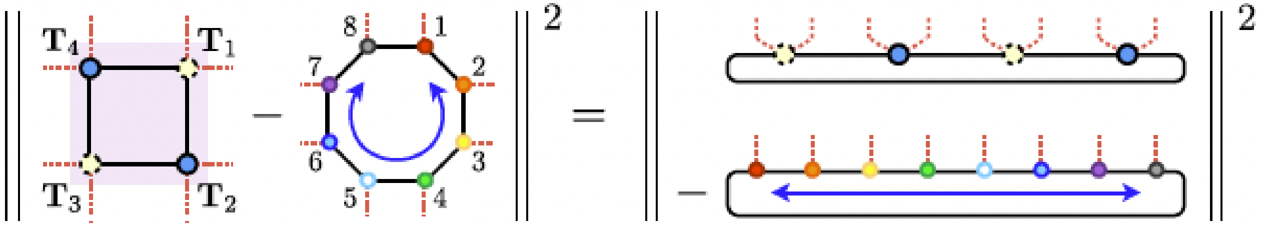
By pinching the "information path", we can split the remaining loop, and remove them at the next contraction.

Loop TNR





Loop Optimization

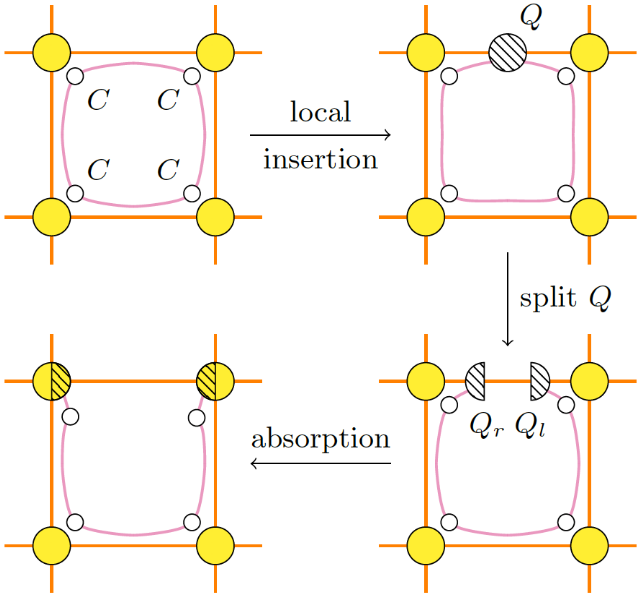


※ *The result of the TM method with L=4*

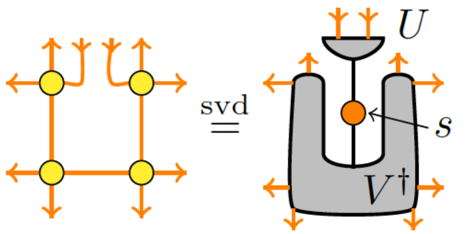
Removal of the short-range correlation

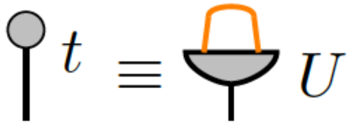
GILT

A general method for constructing a projector "Q" that cuts the redundant entanglement loops, if it exists.

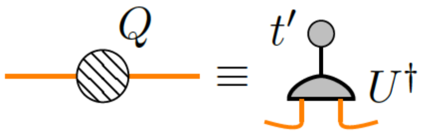


Q filters out the redundant loops of short-range correlations





$$t'_i = t_i \frac{s_i^2}{s_i^2 + \epsilon_{\rm gilt}^2}$$

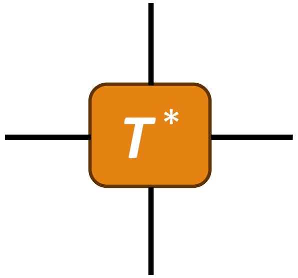


Nuclear Norm Regularization (NNR)

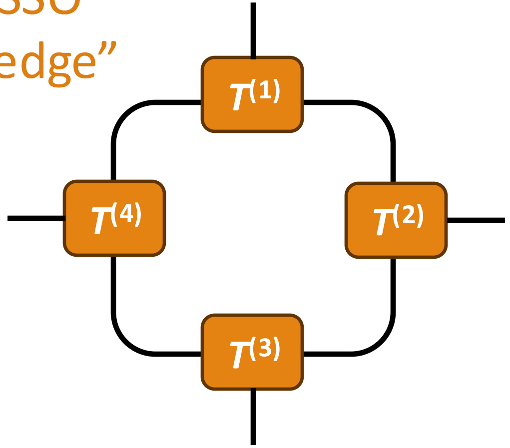
$$C(T^{(1)}, T^{(2)}, \cdots, T^{(d)}) = |\mathcal{T} - \mathcal{T}^*|^2 + \lambda \sum_{i=1}^{d} \sum_{\alpha} \left| T^{(i)}_{(\alpha)} \right|_*^{1}$$

linear norm ...
similar to LASSO
with "sharp edge"

$$(\text{Target}) : \mathcal{T}^*$$



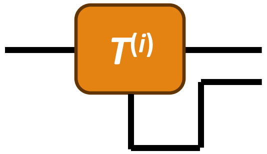
$$(\text{Ansatz}) : \mathcal{T} \equiv \text{Cont} \bigotimes_{i=1}^{d} T^{(i)}$$



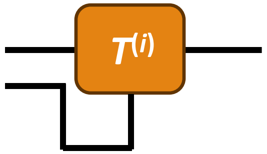
$$X_{(\alpha)} \equiv (\text{The } \alpha\text{th mode slice matrix of } X)$$

$$|M|_* \equiv \sum_{\mu} p_{\mu} \quad (p_{\mu} \equiv (\text{the } \mu\text{th sng. val. of } M))$$

$$T^{(i)}_{(1)} \equiv$$

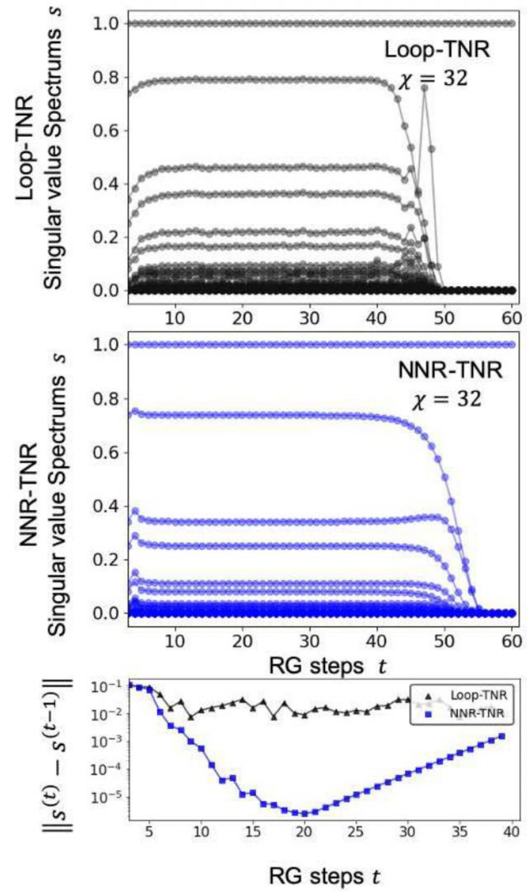


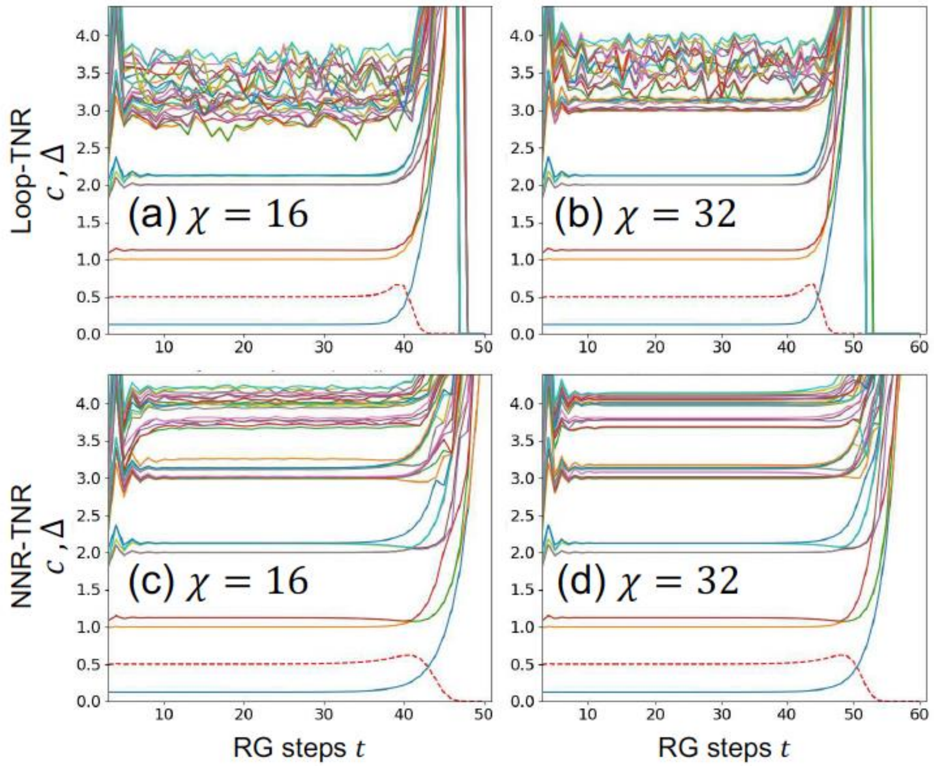
$$T^{(i)}_{(3)} \equiv$$



Very similar to entropic bias, but the bias term is LINEAR in the singular value. (CF: LASSO)

NNR-TNR (numerical stability)





"Loop TNR"

"NNR TNR"
(= Levin-Nave TRG with NNR)

※ *The result of the TM method with L=2*

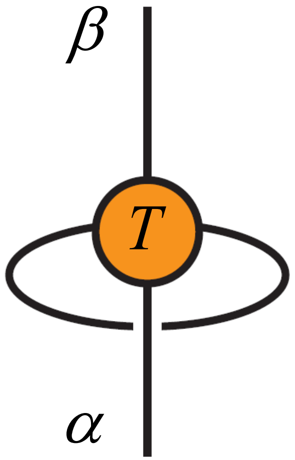
NNR-TNR seems more stable than loop-TNR

Homma, Okubo and NK: PRR6, 043102 (2024)

Gu-Wen: PRB80, 155131 (2009)

Transfer matrix method

--- 2dCFT-aided method for scaling dimensions ---



$$\lambda_\mu = e^{-2\pi\left(\Delta_\mu - \frac{c}{12}\right)}$$

eigenvalues of the partially contracted scale invariant tensor

$$\zeta_s \equiv e^{-f_s} = \sum_\mu e^{-2\pi\left(\Delta_\mu - \frac{c}{12}\right)\mathrm{Im}\tau + i\sigma_\mu \mathrm{Re}\tau}$$

$$\Delta_\mu \equiv h_\mu^R + h_\mu^L \ , \quad \sigma_\mu \equiv h_\mu^R - h_\mu^L$$

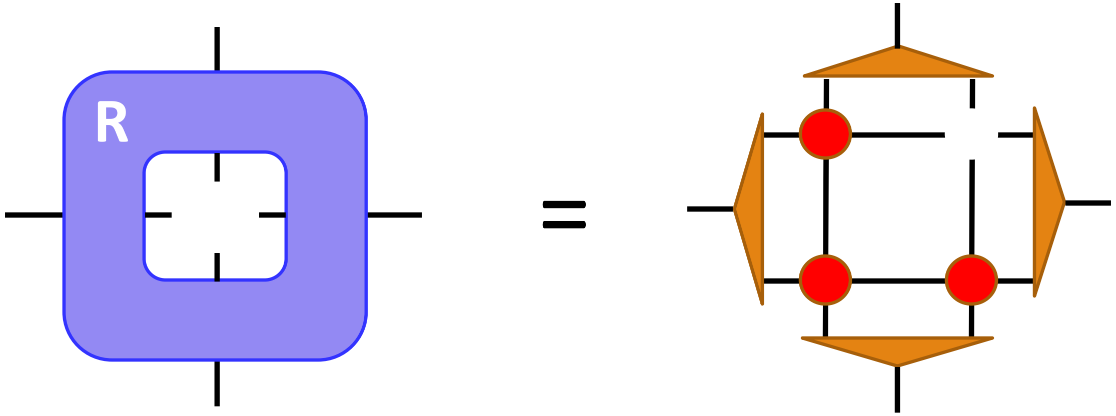
$$\tau \equiv (\text{complex aspect ratio parameter})$$

Cardy: Nucl. Phys. B 270 (1986)

It is not clear how to generalize to higher dimensions.

What about d>=3 ?
--- Textbook RG program with HOTRG ---

HOTRG ... easier to extend for higher dimensions



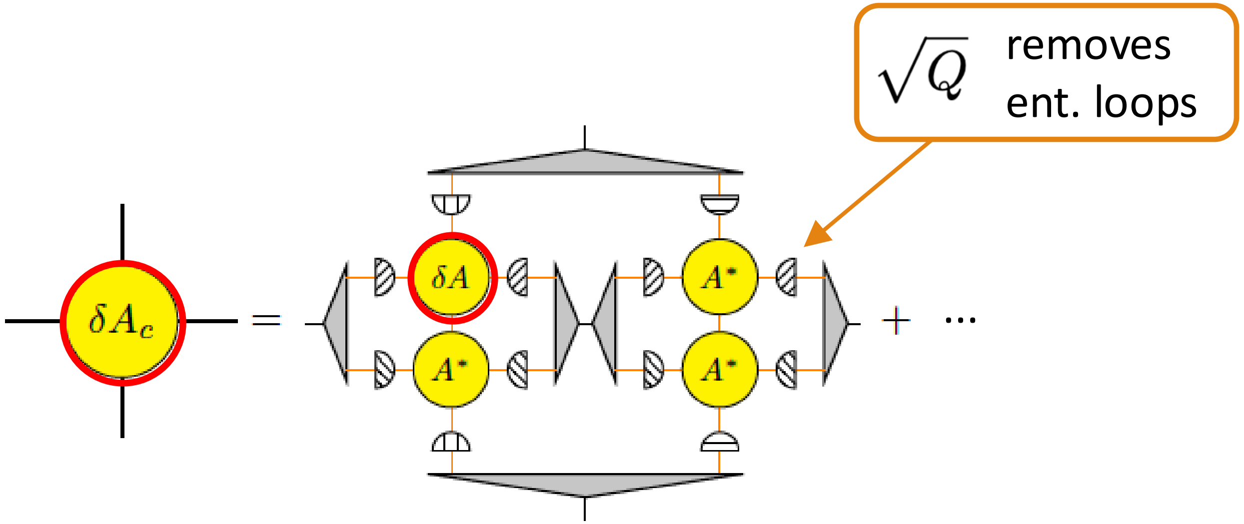
✓ Short-range correlation stays no matter how much we repeat RG steps.

Linearized Super-operator

$$A' = R(A)$$

$$\delta A' = S\,\delta A$$

$$S \equiv \frac{\partial R(A)}{\partial A}$$



In differentiating R(A) w.r.t. A, we did not take the variation in the projectors into account, i.e., the projectors are fixed.

This treatment relates S to the dilatation operator.

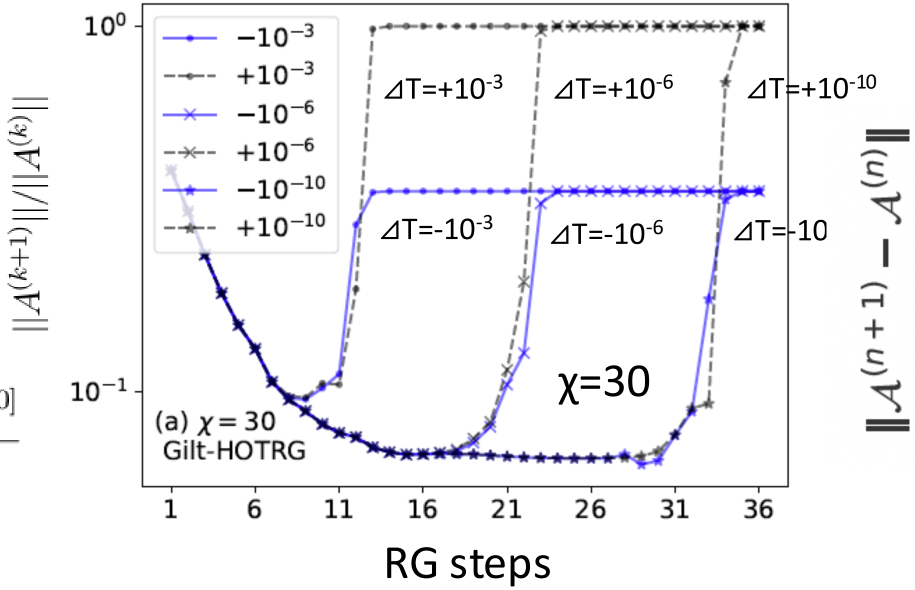
Lyu, Xu and NK: PRR 3 023048 (2021)

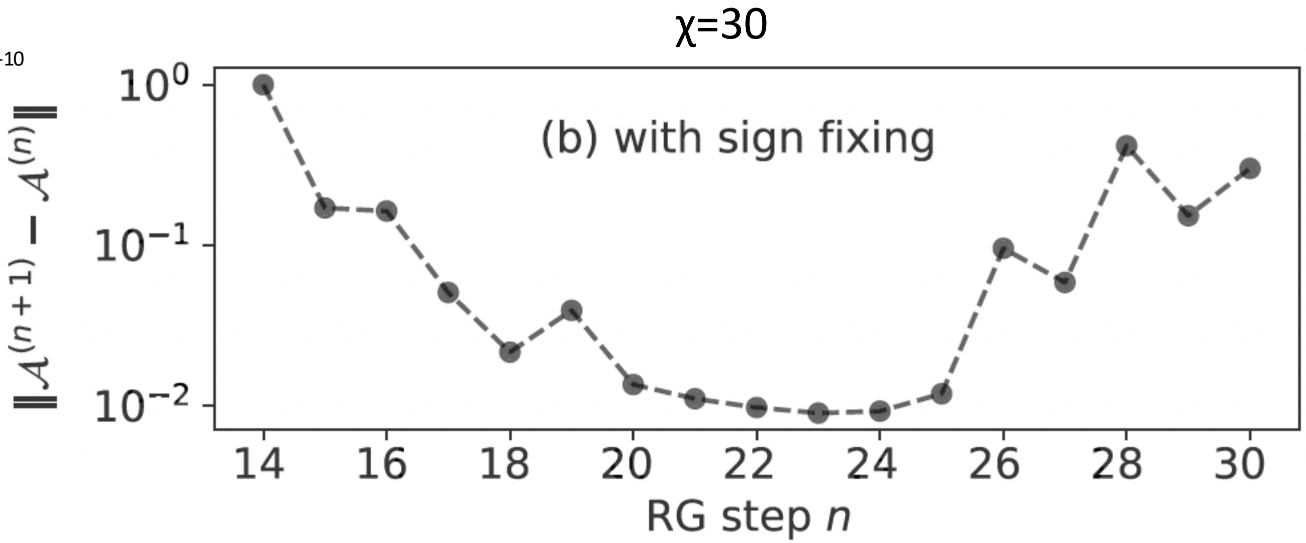
Benchmark (2d Ising)

Convergence of the tensor

- Depending on the temperature, tensor norm ratio curves start to deviate from the critical curve.
- At χ=30, it stays flat at the bottom, indicating approach to the true fixed point as we increase χ.
- The difference in the normalized tensor is small.

$$\Delta T \equiv \frac{T - T_c^{[\chi=30]}}{T_c}$$



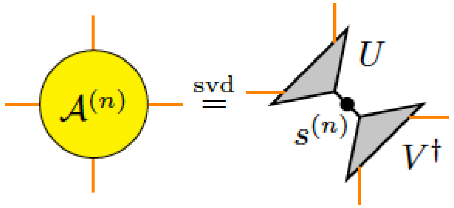


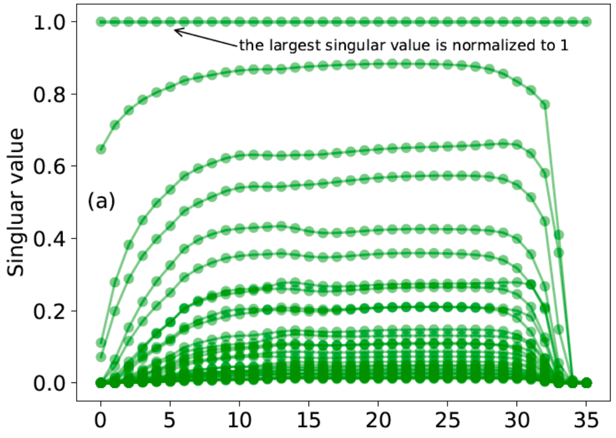
Lyu, Xu and NK: PRR 3 023048 (2021)

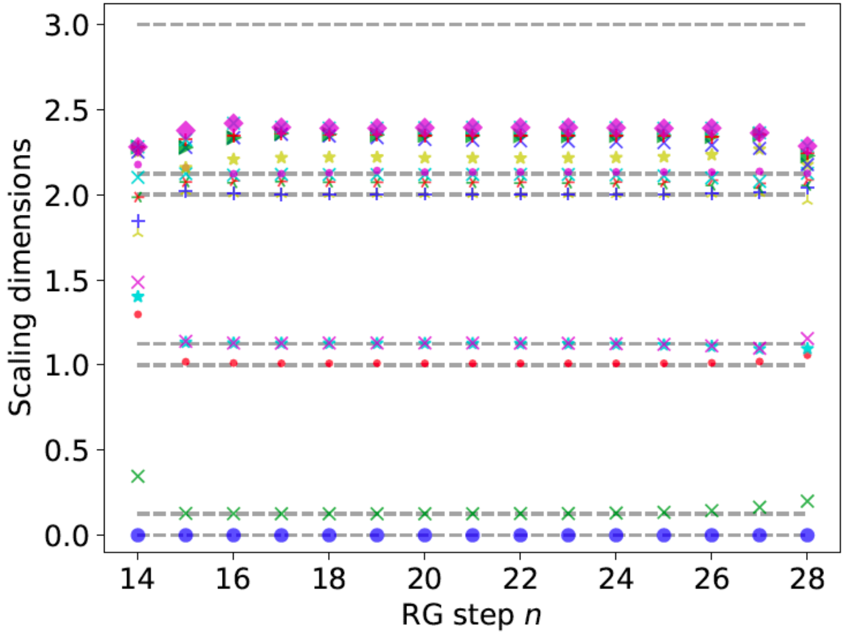
Benchmark (2d Ising)

- The method yields at least 4 digits of scaling dimensions for the most relevant ones.

Exact	0.125	1	1.125	1.125	2	2	2	2
RG pres.	0.127	1.009	1.125	1.128	2.002	2.004	2.068	2.073
Trans. mat.	0.125	1.009	1.130	1.148	1.313	1.457	1.558	1.654







Lyu, Xu and NK: PRR 3 023048 (2021)

You can get OPE coefficient, too.

$$\left\langle T_\alpha \middle| R \middle| \begin{matrix} T_\beta & T_\gamma \\ T & T \end{matrix} \right\rangle = \frac{c^{\alpha}_{\beta\gamma}}{a^{\Delta_\beta+\Delta_\gamma-\Delta_\alpha}}$$

↑ Just the essence. In actual computation, a more sophisticated formula were used.

... The good thing is, we can use the same method for 3D or higher since we don't rely on any formula specific to 2D CFT!

	Exact	lTRG
Δ_σ	1/8	0.127
Δ_ε	1	1.002
$C_{\sigma\sigma\sigma}$	0	2.4×10^{-7}
$C_{\sigma\sigma\varepsilon}$	1/2	0.512
$C_{\varepsilon\varepsilon\sigma}$	0	1.1×10^{-6}
$C_{\varepsilon\varepsilon\varepsilon}$	0	0.168
$C_{\sigma\varepsilon\sigma}$	1/2	0.409
$C_{\sigma\varepsilon\varepsilon}$	0	1.5×10^{-7}

Newton method with Jacobian

$$F(T) \equiv T - R(T)$$ ← We want make this 0.

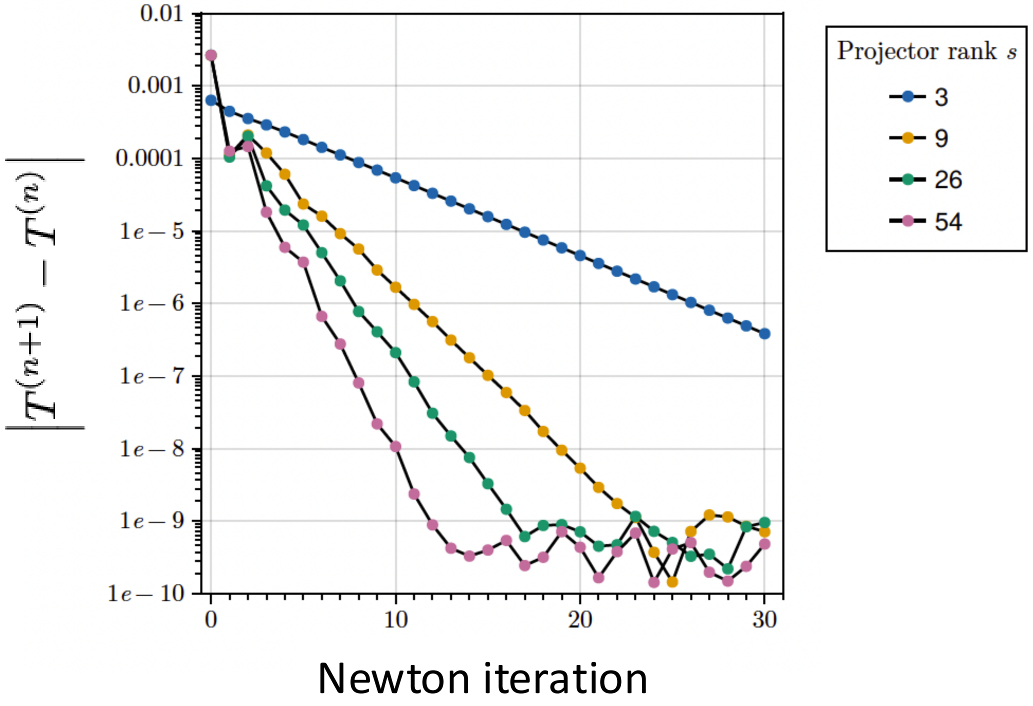
$$G(T) \equiv T - \left(\frac{dF}{dT}\right)^{-1} F$$ ← Newton map

$$T^{(n+1)} = G(T^{(n)})$$ ← Newton iteration

To make the burden lighter, the gradient is replaced as

$$\frac{dF}{dT} = I - \frac{dR}{dT} \quad \Rightarrow \quad I - P_s \left.\frac{dR}{dT}\right|_{T=T^{(0)}} P_s$$

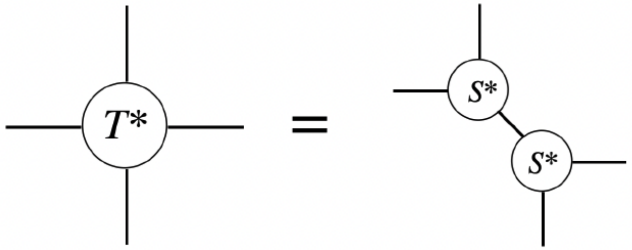
※ Rotation is necessary to avoid zero eigen value



While other methods (i.e., "shooting methods") tend to go away from the fixed point, this method go toward it by design.)

Elements of the universal tensor in 2D

The elements of fixed-point tensor are the universal correlation function.



$$\frac{S^*_{\alpha\beta\gamma}}{S^*_{111}} = \langle \phi_\alpha(-x_S)\phi_\beta(ix_S)\phi_\gamma(0)\rangle_{\text{pl}}$$

$$x_S = e^{\pi/4}$$

$$x_T = e^{\pi/2}/2$$

$$\frac{T^*_{\alpha\beta\gamma\delta}}{T^*_{1111}} = \langle \phi_\alpha(-x_T)\phi_\beta(ix_T)\phi_\gamma(x_T)\phi_\delta(-ix_T)\rangle_{\text{pl}}$$

※ The counterpart of 3D or higher is not known.

Toward 3D

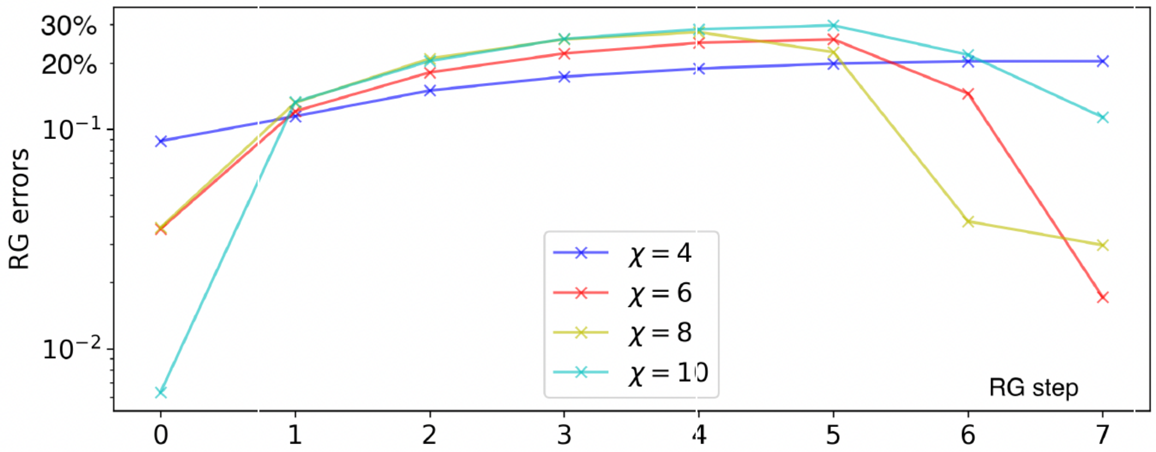
The information mediated by the corner (2D) or the edge (3D) caused by the CDL structure.

$$S \propto L^0 = b^0 \to \text{const} \qquad (2\text{D})$$

$$S \propto L^1 = b^g \to \infty \qquad (3\text{D})$$

While the loop filtering is luxury in 2D, it's necessity in 3D.

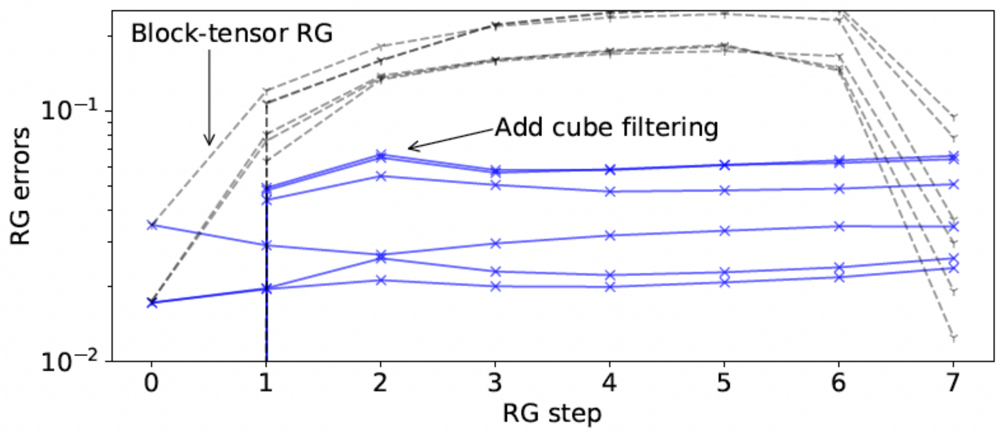
Relative truncation error in RG steps. (T=Tc(χ))



Around n=5 with χ=10, elements of the tensor with almost 30% in amplitude are discarded.

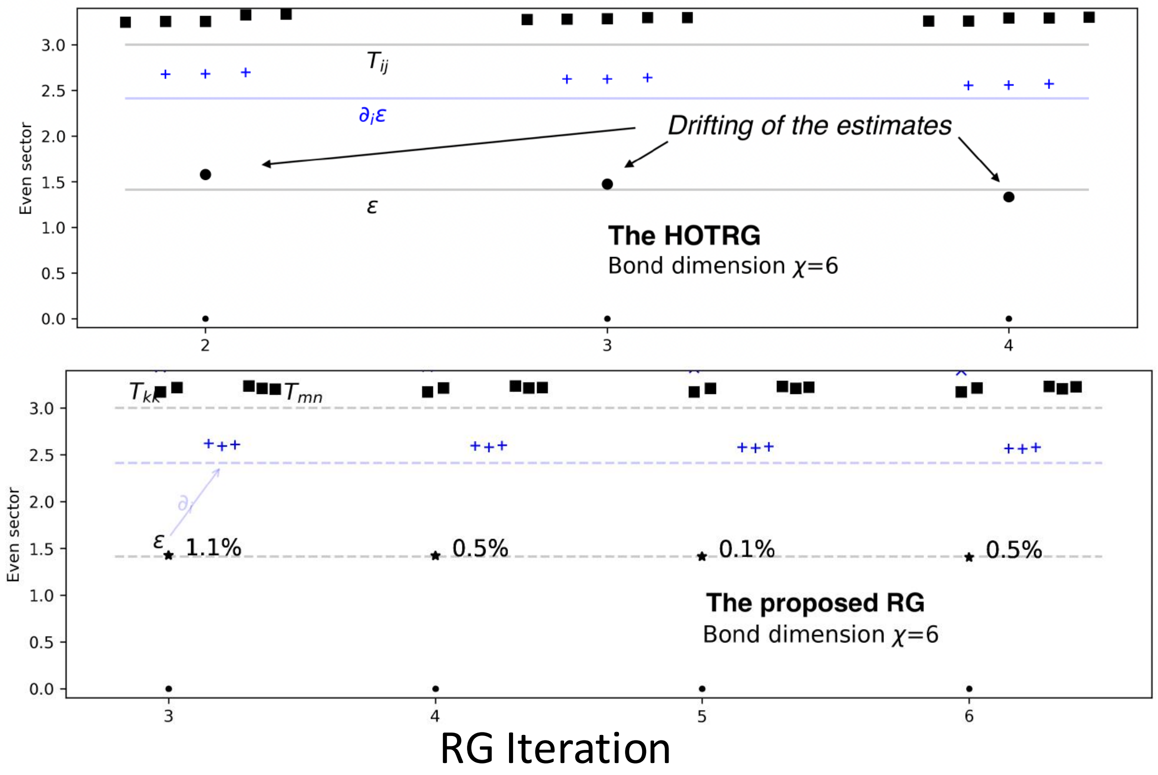
X. Lyu (IHES), NK: 2311.05891v2; PRE 111 (5), 054140 (2025)

Scaling Dimensions in 3D



Entanglement filtering is necessary.

Scaling dimensions in the even parity sector
(obtained by the linearized RG)



X. Lyu (IHES), NK: PRE 111 (5), 054140 (2025)

Intermediate Summary for RSRG

- The power of TN methods is the most prominent in its application to real-space RG, and the TN community has made a big progress in real-space RG in the last decade.
- Recently developed methods do not rely on the formula specific to 2D-CFT.
- Data scientific method is quite useful (CF: nuclear norm regularization)

Tensor Network Method --> Data Science

Collaborators

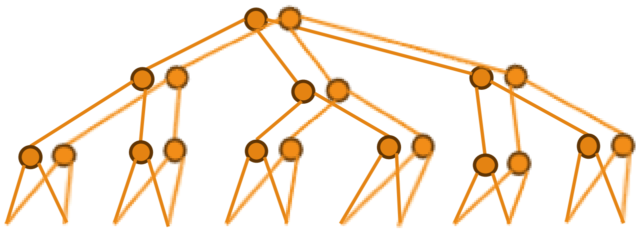
Katsuya AKAMATSU
(ISSP)

Kenji HARADA
(Kyoto)

Tsuyoshi OKUBO
(Tokyo)

Tensor Trees for Learning Distributions (ATT = Adaptive Tensor Tree)

BMATT (Born machine on ATT)



HARADA, et al.: Mach. Learn.: Sci. Technol. (2025)

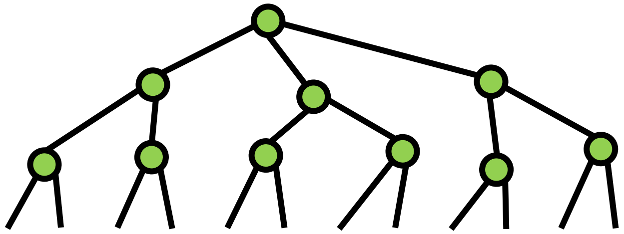
$$P(S) = |\Psi(S)|^2 \geq 0 \quad \Psi(S) = \mathrm{Tr}\left(\bigotimes_i T_i(S)\right)$$

T_i doesn't have to be real and non-negative

Optimization with no constraint on Ti.

Subnetworks are not interpretable.

NATT (non-negative ATT)



AKAMATSU, et al., arXiv:2504.06722

$$P(S) = \mathrm{Tr}\left(\bigotimes_i T_i(S)\right) \geq 0$$

T_i is real and non-negative

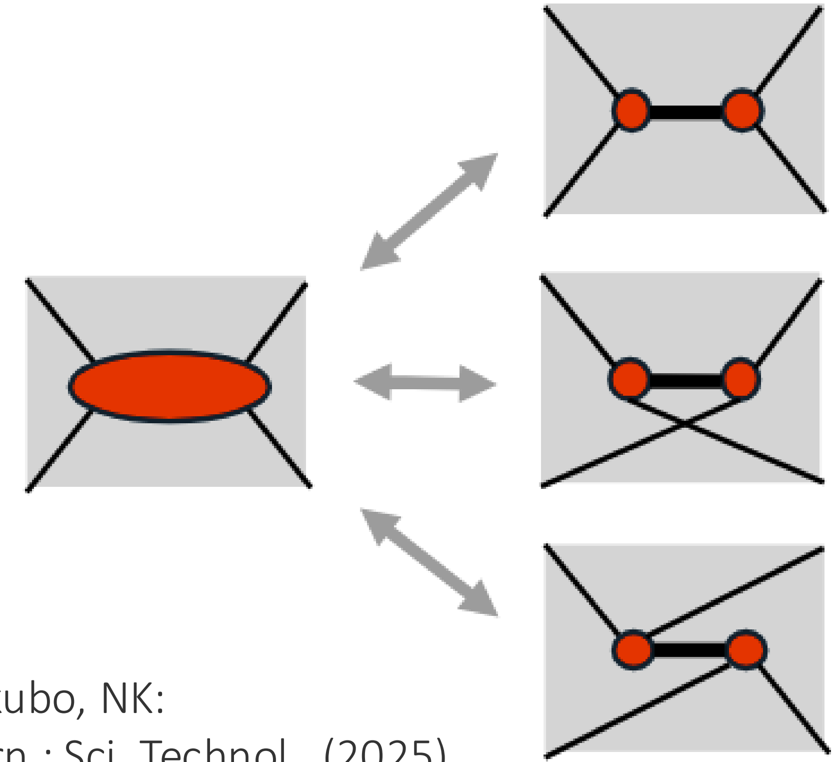
Optimization with positivity constraint.

Subnetworks are interpretable.

The computational complexity is of the same order.

Branch reconnection according to BMI

(BMI=bond mutual information)



Harada, Okubo, NK:
Mach. Learn.: Sci. Technol. (2025)

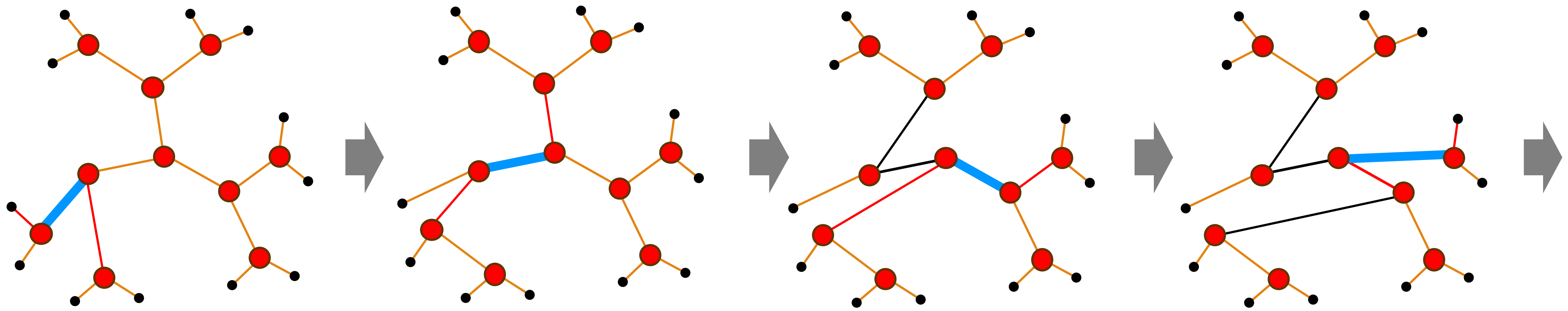
For each bond, we bind the two tensors and SVD in one of the three ways. We choose the SVD with the smallest BMI defined by

$$I_{AB} = \sum_{X_A, X_B} P_{AB}(X_A, X_B) \log \frac{P_{AB}(X_A, X_B)}{P_A(X_A) P_B(X_B)}$$

$$\approx \frac{1}{m_{\mathrm{batch}}} \sum_{\alpha \in \mathrm{batch}} \log \frac{P_{AB}(X_A^{(\alpha)}, X_B^{(\alpha)})}{P_A(X_A^{(\alpha)}) P_B(X_B^{(\alpha)})} \equiv I'_{AB}$$

※ In [Hikihara, et al, PRR5, 013031 (2023)], the tensor tree wave function was discussed for simulating quantum systems. There, the entanglement entropy was adapted as the cost function to be minimized.

Adaptive Tensor Tree (ATT)



After updating a bond, among its neighbors, the oldest one (the one that has been untouched for the longest time) is updated next.

How to make TT learn the target distribution

Typically, the only we have is just samples (observations). So, we minimize the KL divergence between the sample distribution P_{sample} and the tensor tree P_T .

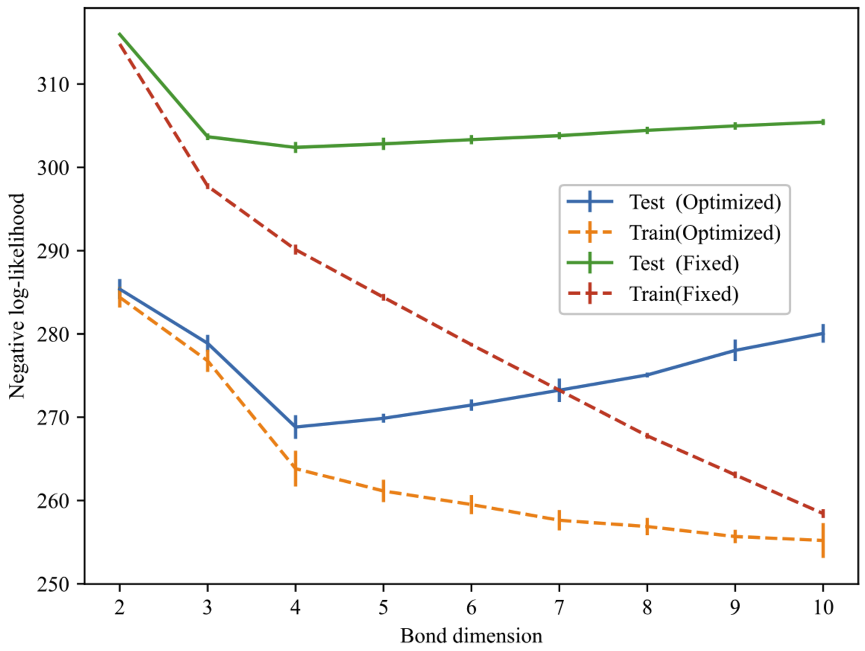
$$P_{\mathbf{sample}} \equiv \frac{1}{m}\sum_{\alpha=1}^{m}\delta_{X,X^{(\alpha)}} \qquad (\text{Sample set}) = (X^{(1)}, X^{(2)}, \cdots, X^{(m)})$$

The cost function to be minimized is the negative log likelihood (NLL):

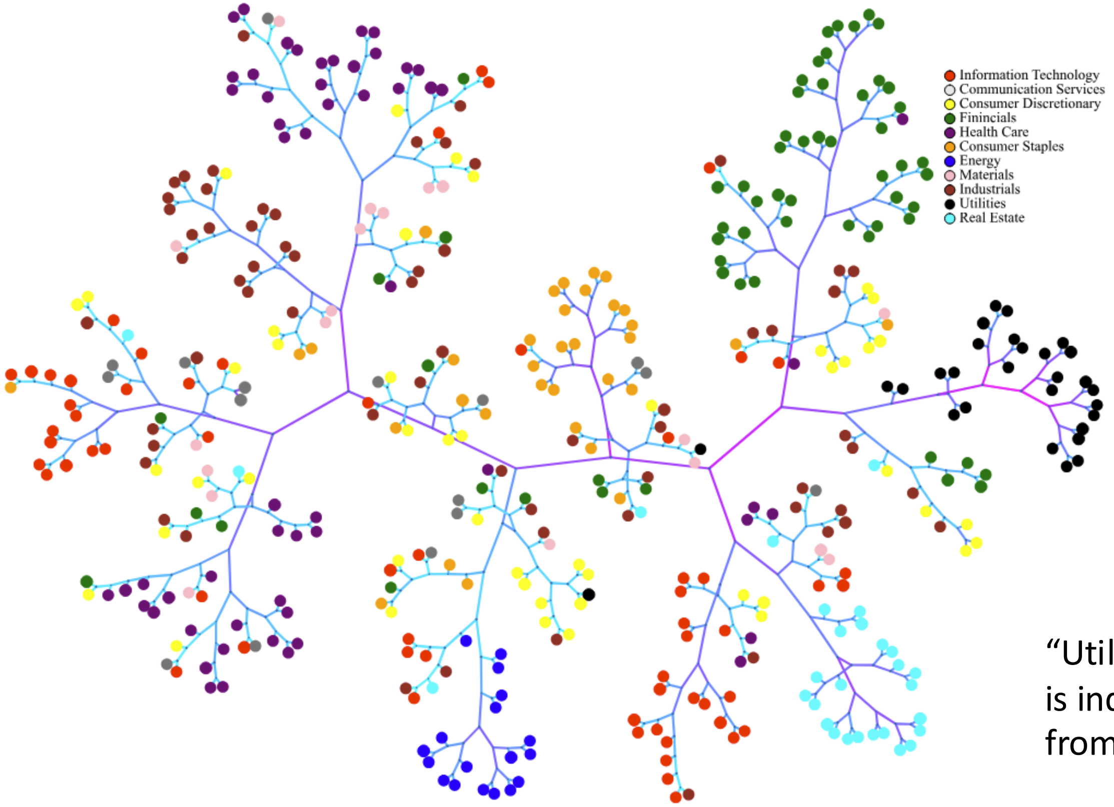
$$\mathrm{NLL}(T) = D_{\mathrm{KL}}(P_{\mathbf{sample}}||P_T) = \sum_{\alpha=1}^{m} \log \frac{1}{P_T(X^{(\alpha)})}$$

Example 1 (BMATT): Stock market (S&P 500)

436 companies in 10 years.
Compressed into bit variables (1=rise/0=fall).



$m_{total}=2m_{train}=2m_{batch}=3,589$, $N_{ite} = 2,000,000$



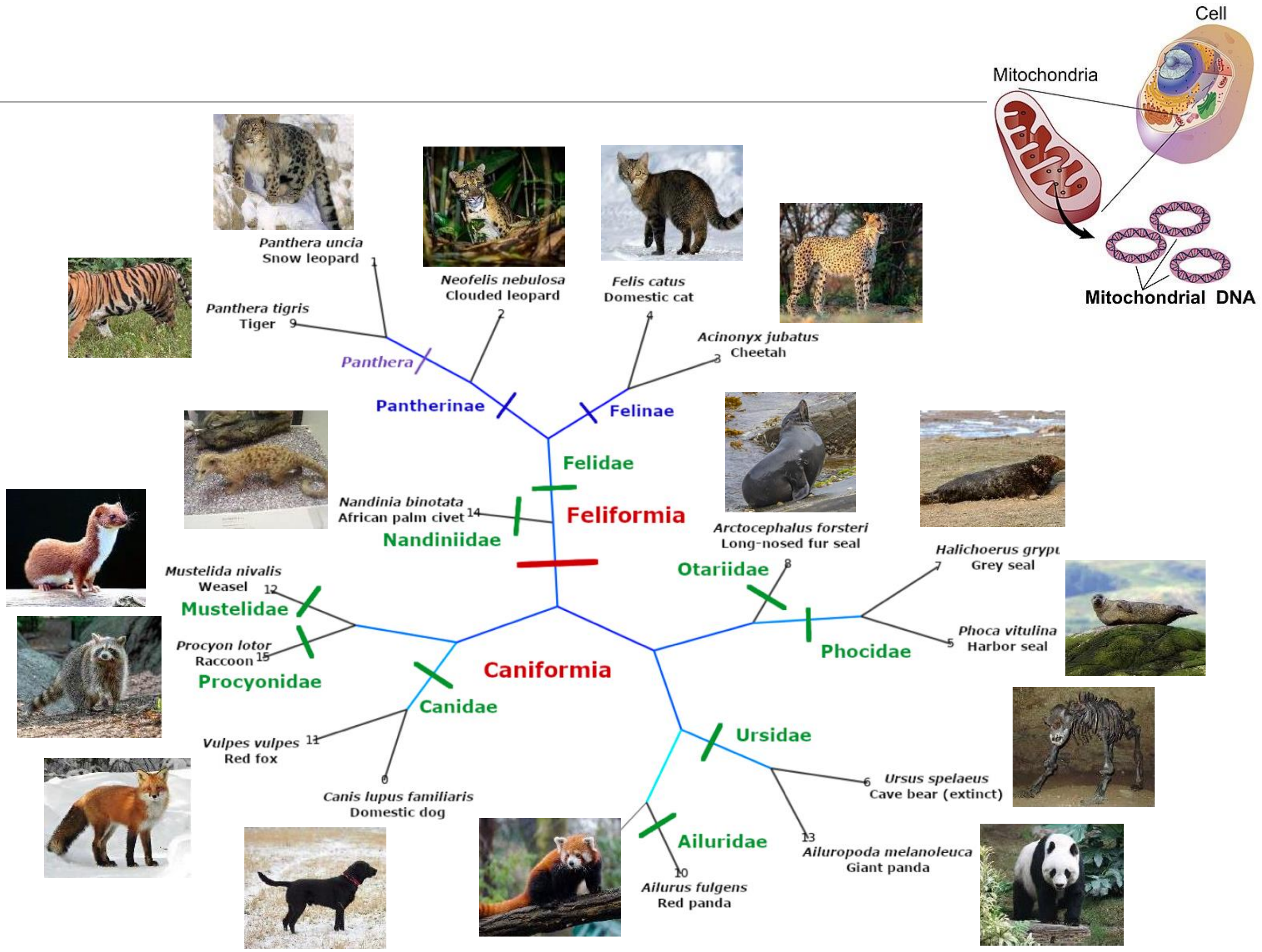
"Utilities" sector is independent from the others

"11 sector" structure emerged spontaneously.

Example 2 (NATT): Categorization of cats and dogs

- We consider sequence from the cytochrome B (cytB/COB) gene in mitochondrial DNA for 16 different animals in the taxonomic order Carnivora (dogs and cats).
- This gene always consists of 1140 base pairs and show clear interspecies variation.
- 16 species are represented by 16 nodes of our tree.
- 1140 bps are regarded as 1140 sample data.

Since we are not assuming any evolutionary model, the tree is not supposed to be accurate about the temporal order of branching. However, the tree seems to be close to the widely accepted evolutionary tree.



AKAMATSU, et al., arXiv:2504.06722

Summary

- TN methods benefit a lot from the data scientific consideration.
- TN are good for both finite-scale physics and critical properties.
- TN provides a way to construct flexible frameworks for machine learning.

Data Science ⟷ (TN) ⟷ Physics