計算科学における情報圧縮

Information Compression in Computational Science

2022.1.13

#13: General tensor network representations

理学系研究科　大久保　毅

Graduate school of scicence, **Tsuyoshi Okubo**

- This class is from 13:15 to 14:45 (90 min.).

Outline

- Application to data science
  - ~~Classification problem~~
  - Generative models
  - Compressing (deep) neural network
- Breakdown of MPS representation
  - Critical system
  - Higher dimensional system
- Tensor Network for critical systems
  - Multi-scale Entanglement Renormalization Ansatz (**MERA**)
- Tensor Network for higher dimensions
  - Tensor Product State (**TPS**)
- Appendix:Tensor network renormalization
  - Tensor network representation of a scalar
  - Tensor network renormalization
  - Tensor network renormalization around critical point

Application to data science

E. Miles Stoudenmire and D. J. Schwab, NIPS 2016

Z.-Y. Han et al, Phys. Rev. X **8**, 031012 (2018).

Z.-F. Gao et al, Phys. Rev. Research **2**, 023300 (2020).

Machine learning for classification problem

Problem: we want to classify an input vector by several labels

E.g Classification of handwriting images

Standard procedure:

First, input vector $\boldsymbol{x}$ is mapped onto higher dimensional space

$\vec{\psi}(\boldsymbol{x})$ (non-linear feature map)

Then it is transformed to labels through a linear map

$$f^l = W^l \vec{\psi}(\boldsymbol{x})$$

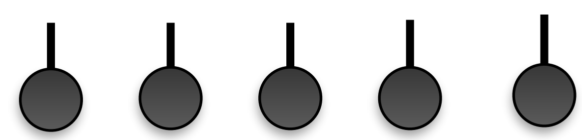
In the case of supervised machine learning, we optimize W based on the correct labels of a lot of input vectors.

MPS representation of the classification problem

E. Miles Stoudenmire and D. J. Schwab, NIPS 2016

If we select a "product state" as a feature map

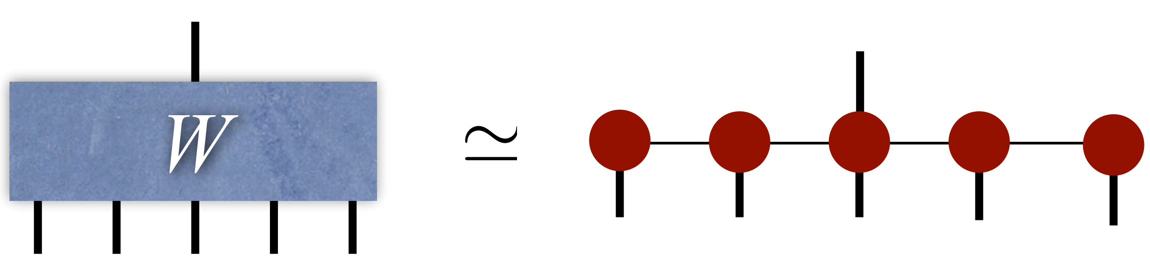
$$\psi_{i_1, i_2, \ldots, i_N}(\boldsymbol{x}) = \phi_{i_1}(x_1) \otimes \phi_{i_2}(x_2) \otimes \cdots \otimes \phi_{i_N}(x_N)$$

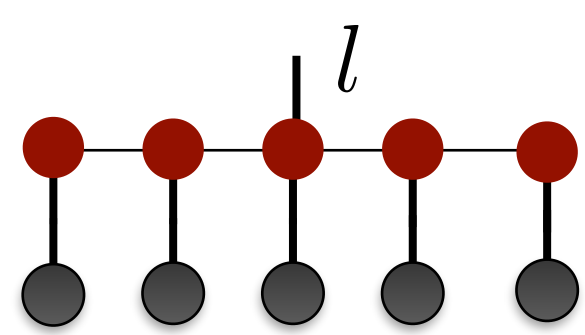
$\vec{\psi} =$ 

$\vec{\phi}(x_1) \quad \cdots \quad \vec{\phi}(x_N)$

The dimension of vector space is a^N

Then we can apply MPS approximation for W



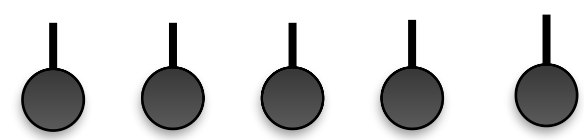
$f^l = W^l \vec{\psi}(\boldsymbol{x}) =$ 

MPS representation of the classification problem

Feature map

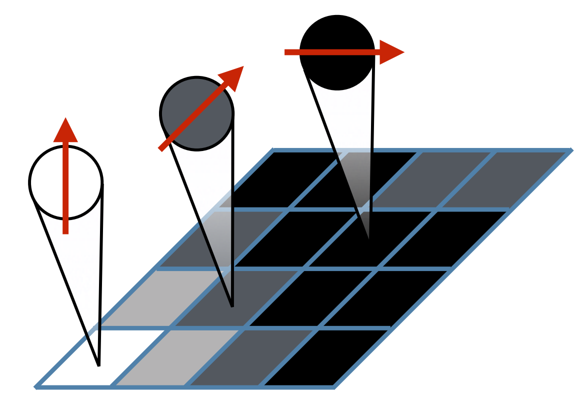
E. Miles Stoudenmire and D. J. Schwab, NIPS 2016
https://github.com/emstoudenmire/TNML

$$\psi_{i_1,i_2,\ldots,i_N}(\boldsymbol{x}) = \phi_{i_1}(x_1) \otimes \phi_{i_2}(x_2) \otimes \cdots \otimes \phi_{i_N}(x_N)$$



Their feature map:

$$\phi^{s_j}(x_j) = \left[\cos\left(\frac{\pi}{2}x_j\right),\ \sin\left(\frac{\pi}{2}x_j\right)\right]$$



For the case of grayscale image

Application to MNIST database of handwritten digits

(handwritten numbers from 0 to 9)

χ	Test set error	
10	~5%	500/10000
20	~2%	200/10000
120	~0.97%	97/10000

<1%

It is comparable with the state of the art!

Unsupervised Generative Modeling

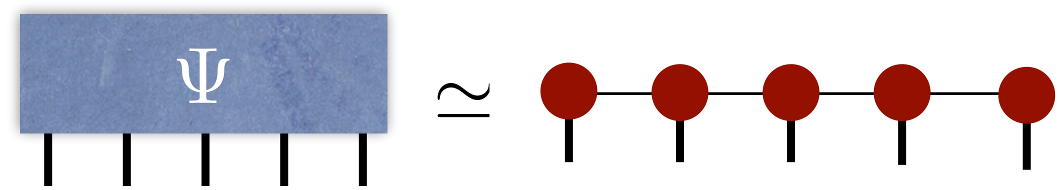
Z.-Y. Han et al, Phys. Rev. X **8**, 031012 (2018).

N pixel grey scale image → Binary data: $\vec{v} \in \mathbb{V} = \{0,1\}^{\otimes N}$

Same structure with qubits

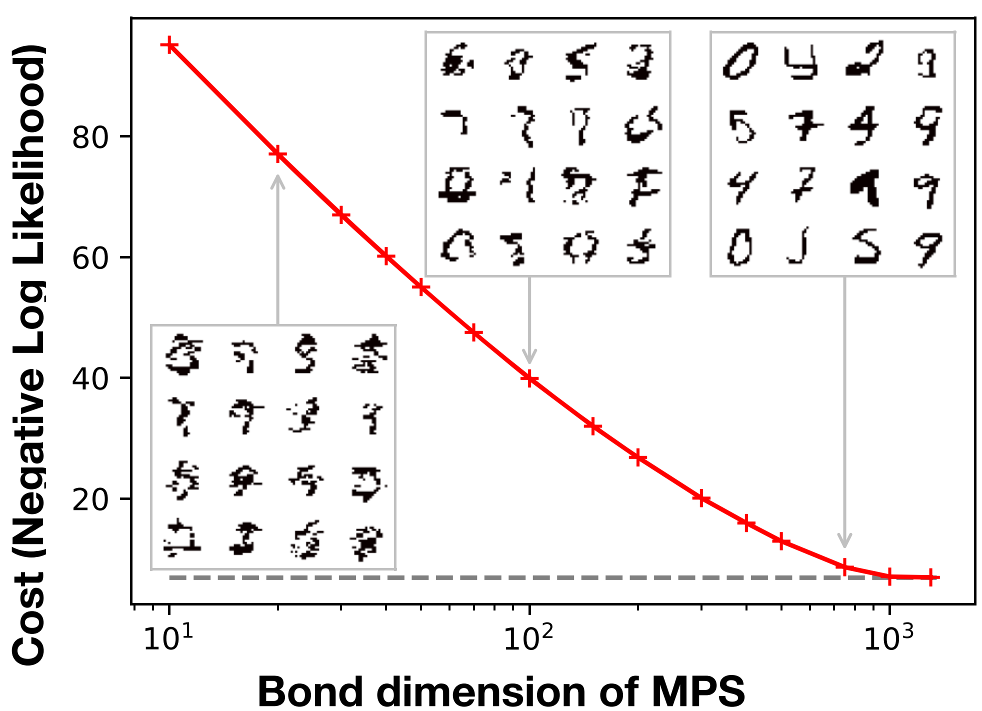
Probability distribution of images

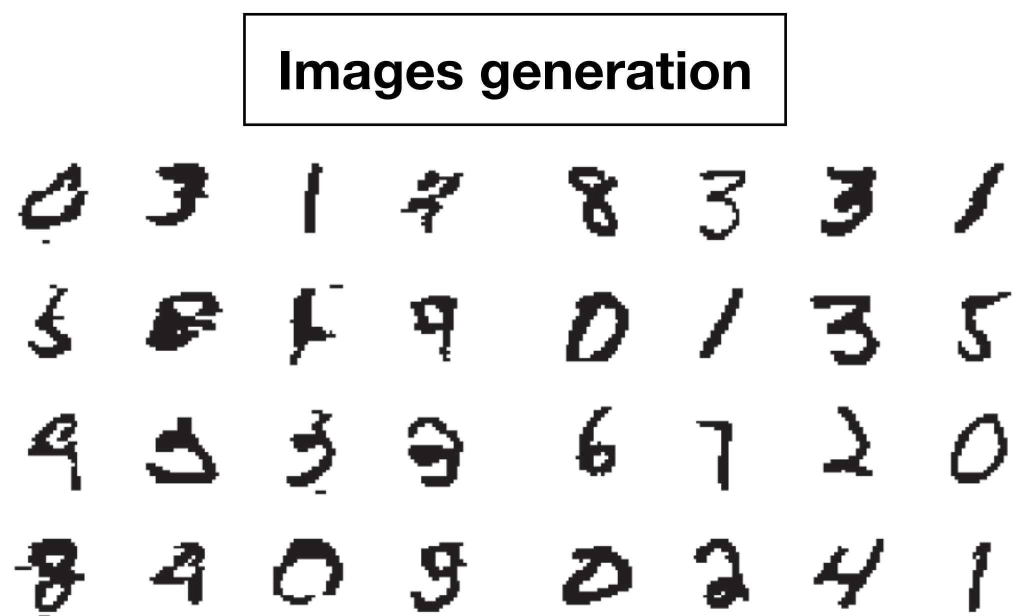
$$P(\vec{v}) = \frac{|\Psi(\vec{v})|^2}{Z} \qquad (Z = \sum_{\vec{v}} |\Psi(\vec{v}_i)|^2)$$



Find optimal MPS to minimize $F = -\frac{1}{\mathbb{T}} \sum_{\vec{v} \in \mathbb{T}} \ln P(\vec{v})$

$\mathbb{T}$: Set of training data





(a) Generated (b) Original

Compressing deep neural network

Z.-F. Gao et al, Phys. Rev. Research **2**, 023300 (2020).

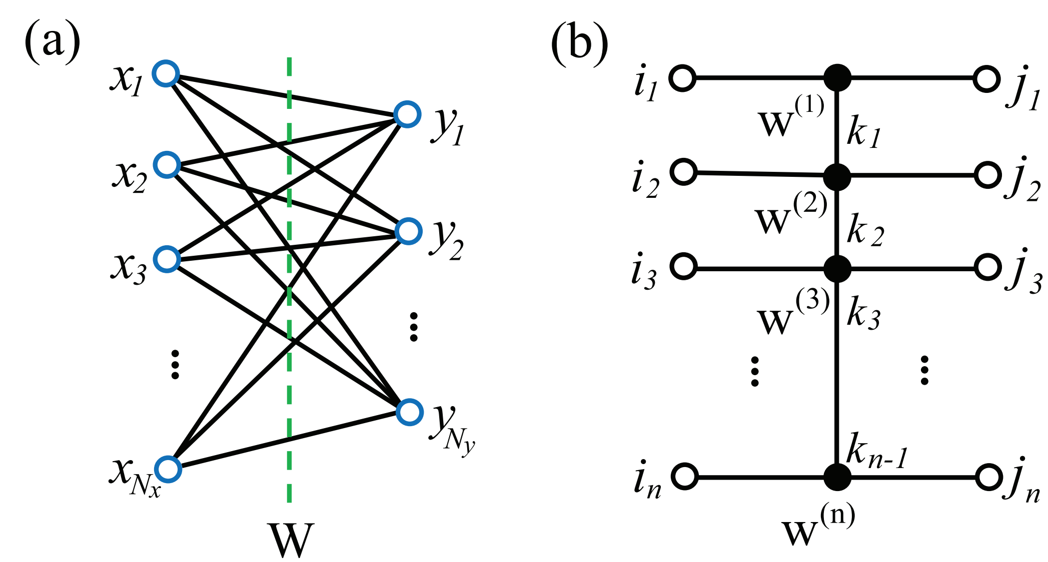
MPO approximation of the weight matrix

x_i: input neuron (pixel)

y_i: output neuron

W_{ij}: weight matrix connecting x and y

→ MPO approximation of W



Example: application to classification problems

TABLE I. Test accuracy a and compression ratios ρ obtained in the original and MPO representations of LeNet-5 on MNIST and VGG on CIFAR-10.

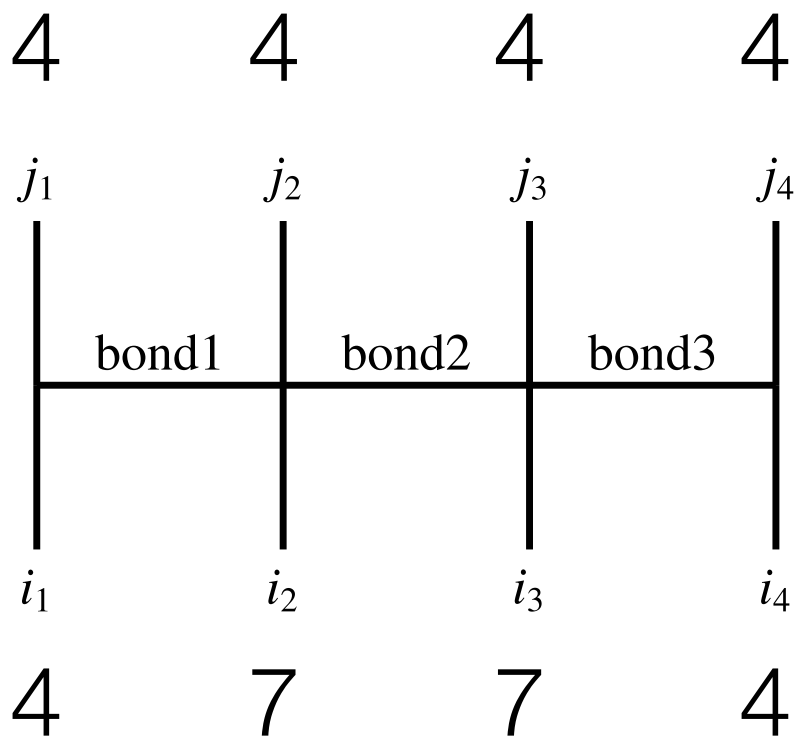
Data set	Network	Original Rep a (%)	MPO-Net a (%)	MPO-Net ρ
MNIST	LeNet-5	99.17 ± 0.04	99.17 ± 0.08	0.05
CIFAR-10	VGG-16	93.13 ± 0.39	93.76 ± 0.16	~ 0.0005
	VGG-19	93.36 ± 0.26	93.80 ± 0.09	~ 0.0005

a:**accuracy (%)**

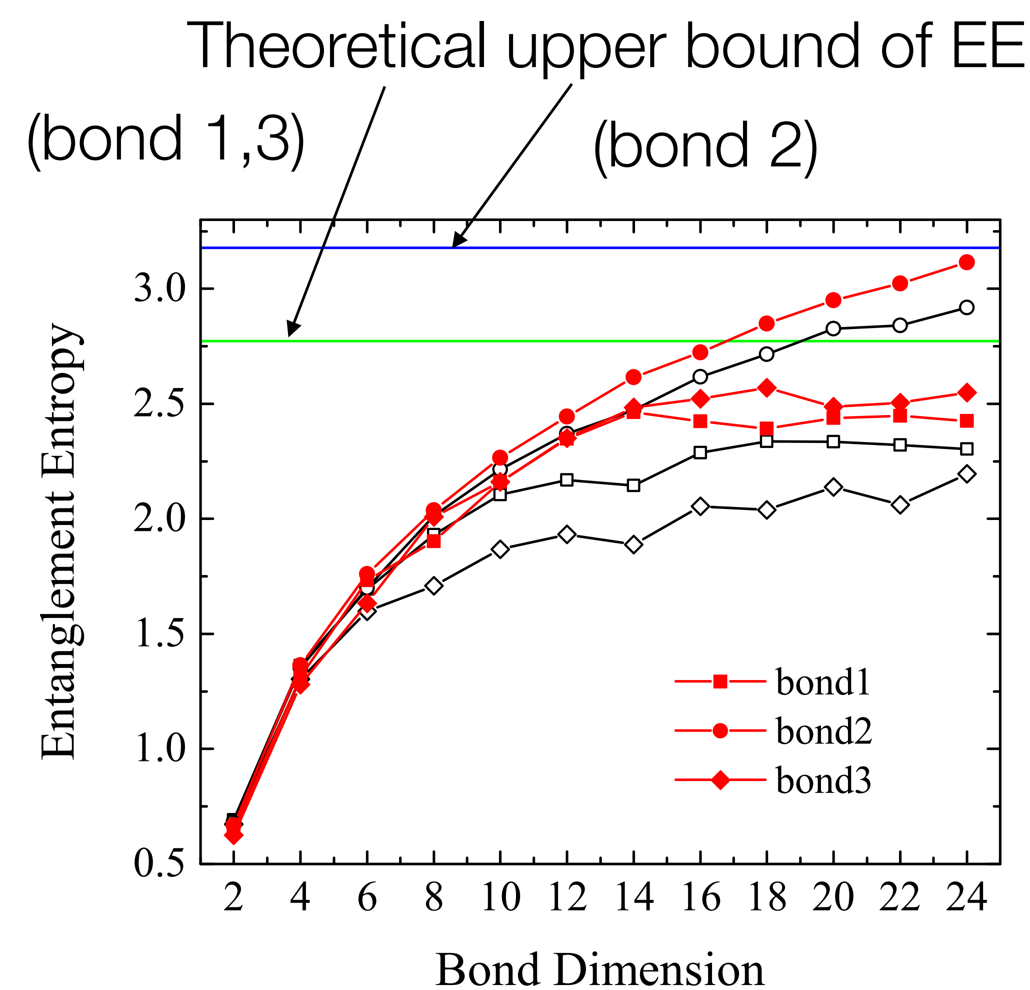
ρ:**compression ratio**

Entanglement entropy of (trained) MPO

Z.-F. Gao et al, Phys. Rev. Research **2**, 023300 (2020).



input: 28 x 28 pixel image



- Fashion-MNIST gives larger EE
  - It might be related to the difficulty of the dataset
- For bond 1 and 3, EE saturated smaller values than the theoretical upper bound
  - It indicate we can use MPO for approximation

Black: MNIST
Red: Fashion-MNIST
(more complicated)

References for application to machine learning

Low-Rank Tensor Networks for Dimensionality Reduction and Large-Scale Optimization Problems: Perspectives and Challenges PART 1

A. Cichocki, N. Lee, I.V. Oseledets, A.-H. Phan, Q. Zhao, D. Mandic

Foundations and Trends in Machine Learning, vol. 9, no. 4-5, pp. 249-429, 2016 (arXiv.1609.00893)

Tensor Networks for Dimensionality Reduction and Large-Scale Optimizations. Part 2 Applications and Future Perspectives

A. Cichocki, A-H. Phan, Q. Zhao, N. Lee, I.V. Oseledets, M. Sugiyama, D. Mandic

Foundations and Trends in Machine Learning: Vol. 9: No. 6, pp 431-673, 2017 (arXiv.1708.09165)

Topics:

- Supervised Learning with Tensors
- Tensor Train Networks for Selected Huge-Scale Optimization Problems
- Tensor Networks for Deep Learning
- ...

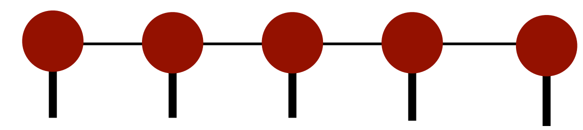
Outline

- Application to data science
    - ~~Classification problem~~
    - Generative models
    - Compressing (deep) neural network
- Breakdown of MPS representation
    - Critical system
    - Higher dimensional system
- Tensor Network for critical systems
    - Multi-scale Entanglement Renormalization Ansatz (**MERA**)
- Tensor Network for higher dimensions
    - Tensor Product State (**TPS**)
- Appendix:Tensor network renormalization
    - Tensor network representation of a scalar
    - Tensor network renormalization
    - Tensor network renormalization around critical point

Breakdown of MPS representation

Required bond dimension in MPS representation

$$\boxed{S_A = -\mathrm{Tr}\, \rho_A \log \rho_A \leq \log \chi}$$



The upper bound is independent of the "length".

length of MPS ⇔ size of the problem

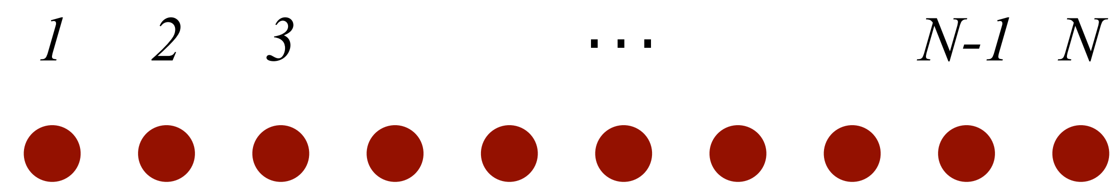
N $\qquad\qquad$ a^N

EE of the original vector	Required bond dimension in MPS representation
$S_A = O(1)$	$\chi = O(1)$
$S_A = O(\log N)$	$\chi = O(N^\alpha)$
$S_A = O(N^\alpha)$	$\chi = O(c^{N^\alpha})$

Phase transition

Transverse field Ising chain:

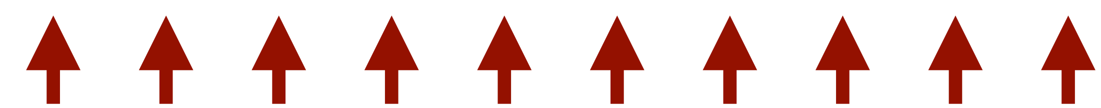
$$\mathcal{H} = -\sum_{i=1}^{N-1} S_i^z S_{i+1}^z - h\sum_{i=1}^{N} S_i^x$$



$(|\uparrow\rangle, |\downarrow\rangle)$

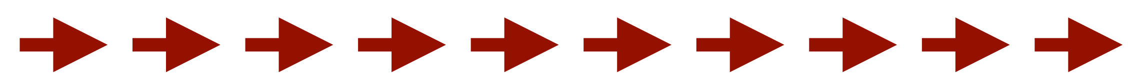
Ground state $|\Psi\rangle$

$h = 0$:Ferromagnetic state



$h \to \infty$:Disordered state

(Field induced ferro)



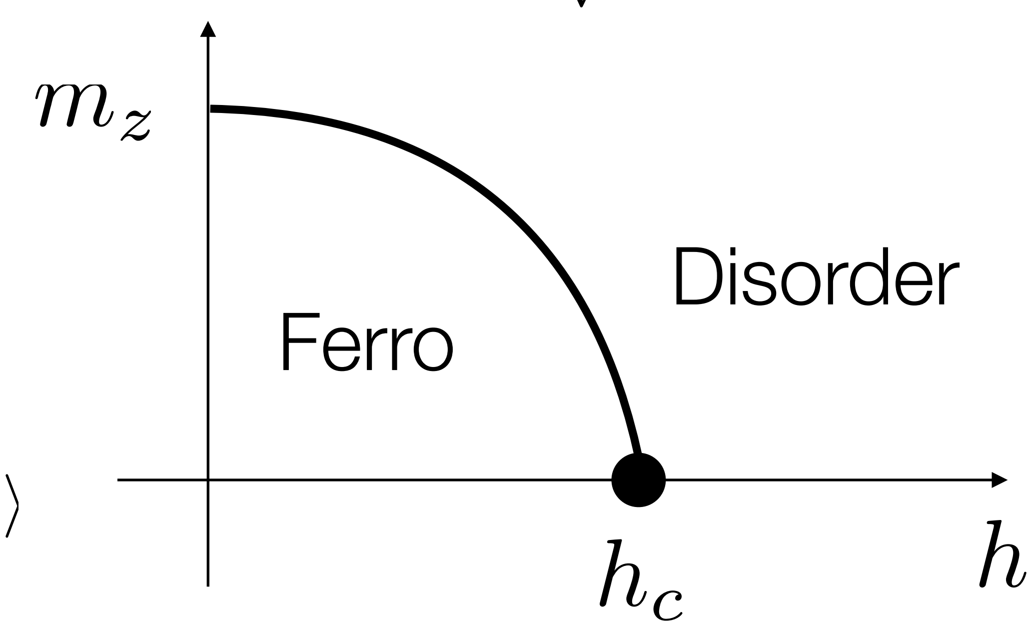
$$= \frac{1}{\sqrt{2}}(|\uparrow\rangle + |\downarrow\rangle)$$

In between these two limits, there is a phase transition.

At the phase transition, order parameter becomes zero.
（秩序変数）

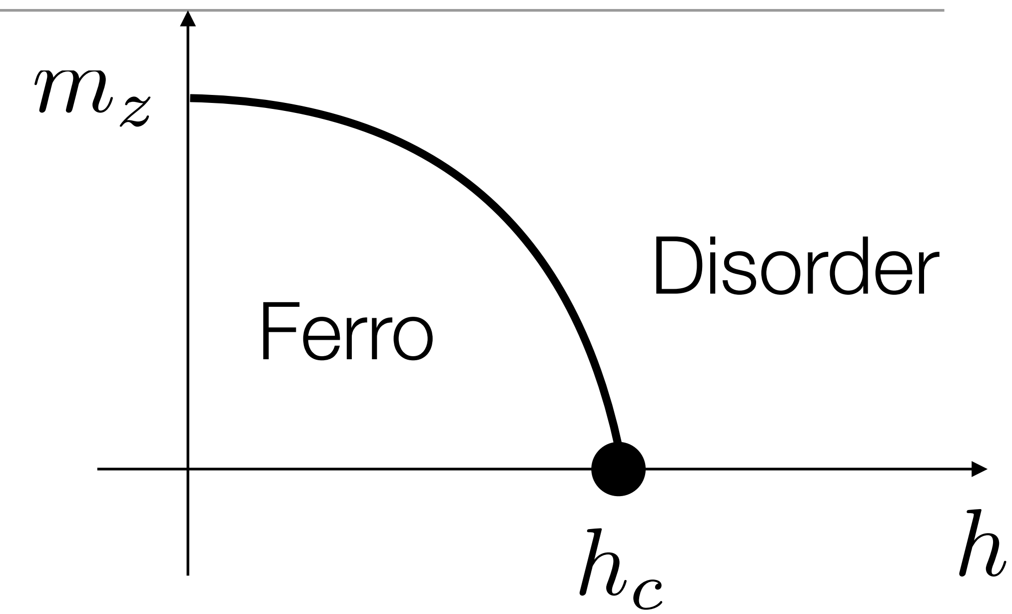
(Spontaneous) Magnetization （自発磁化）

$$m_z = \frac{1}{N}\sum_i \langle\Psi|S_i^z|\Psi\rangle$$



Critical point and correlation length

$h = h_c$:Critical point （臨界点）



Behavior of a correlation function:

$0 \leq h < h_c$:Ferromagnetic state

$$\langle \Psi | S_i^z S_{i+r}^z | \Psi \rangle \sim C e^{-\frac{r}{\xi}} + m_z^2$$

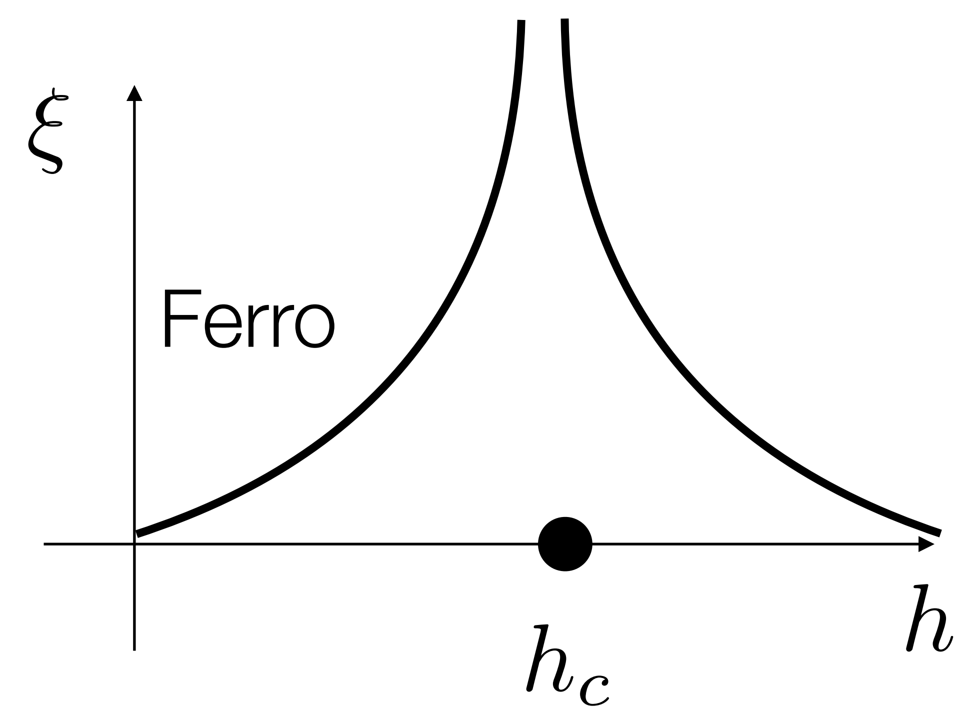
$h_c < h$:Disordered state

$$\langle \Psi | S_i^z S_{i+r}^z | \Psi \rangle \sim e^{-\frac{r}{\xi}}$$

$h = h_c$:Critical point

$$\langle \Psi | S_i^z S_{i+r}^z | \Psi \rangle \sim r^{-2p}$$

$\xi = \xi(h)$:Correlation length (相関長)



Correlation length diverges at critical point!

Scale invariance at the critical point

$h = h_c$:Critical point　（臨界点）

$$C(r) \equiv \langle \Psi | S^z_i S^z_{i+r} | \Psi \rangle \sim r^{-2p}$$

Power low decay!

After a scale transformation $r' = br$

$$C(r') = C(br) = b^{-2p} C(r)$$

Change in the correlation function is only a constant factor.

If we scale spins as $\tilde{S}^z_i = b^p S^z_i$

the correlation function becomes

$$\tilde{C}(r') \equiv \langle \Psi | \tilde{S}^z_i \tilde{S}^z_{i+r'} | \Psi \rangle = C(r)$$

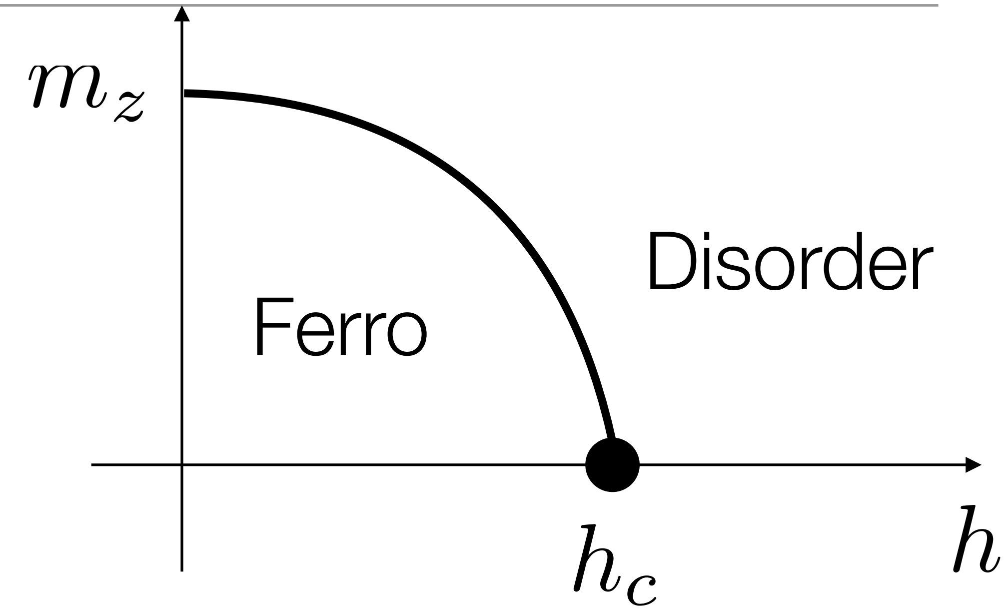
This property is called as "scale invariance".（スケール不変性）

Physics (properties) in different scale is essentially same.

DMRG (variational MPS) calculation of TFI model

Ö. Legeze, and G. Fáth, Phys. Rev. B **53**, 14349 (1996)

Errors of the ground and the 1st excited states energies varying system size N.



For a fixed dimension m,

Ferro and disordered states:

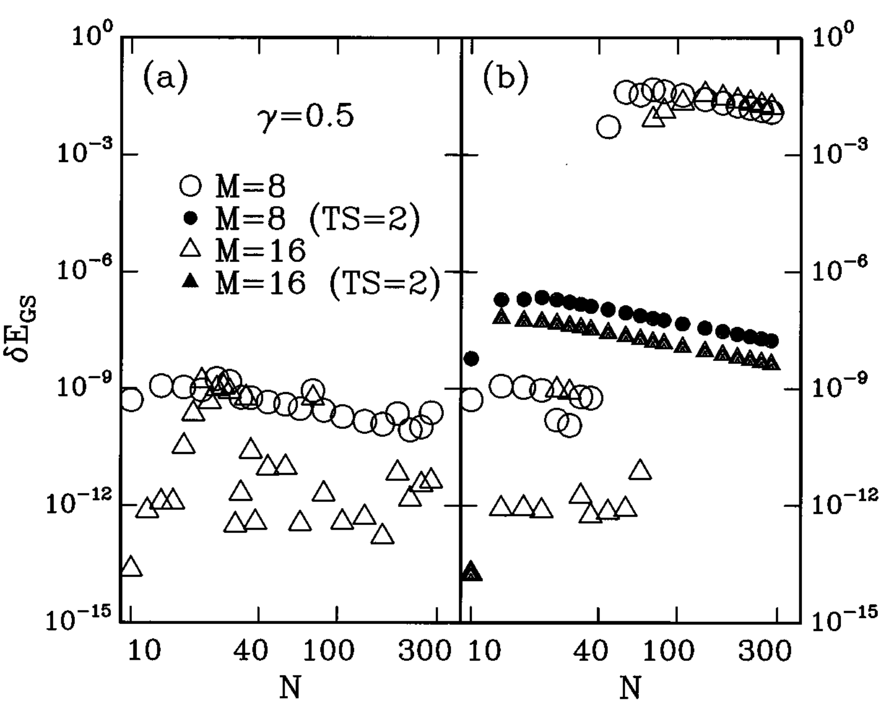
The errors are almost independent of N.

Critical point:

The errors gradually increases as increase N.

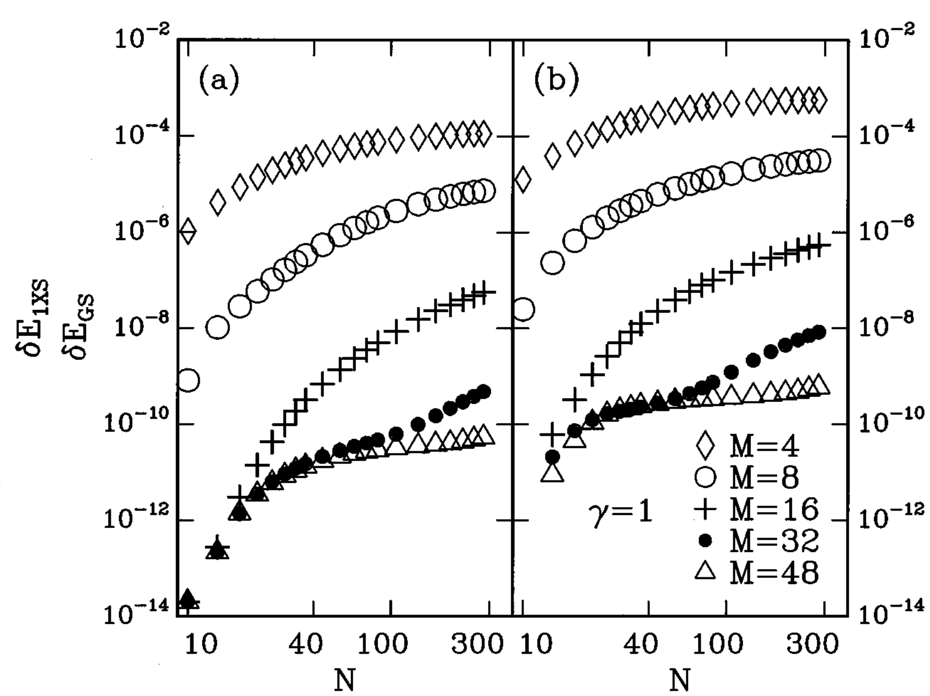
$0 \leq h < h_c$

$h = 0.25$



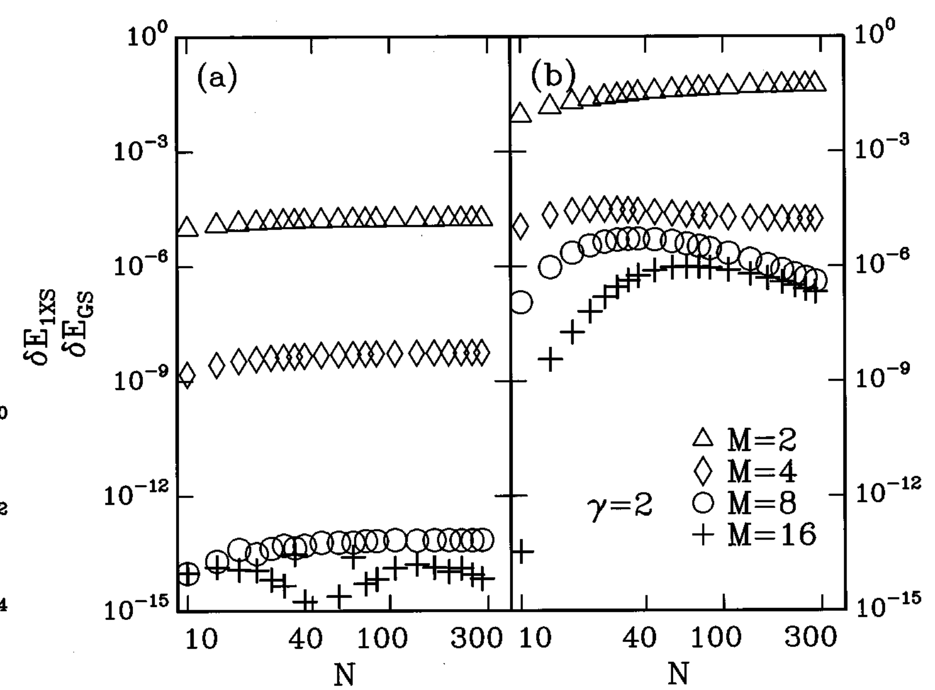
$h = h_c$

$h = 0.5$



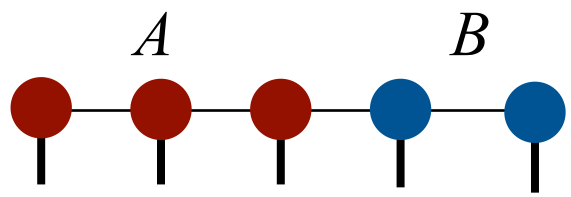
$h_c < h$

$h = 1.0$



Entanglement entropy of TFI model

Entanglement entropy: $\boxed{S_A = -\mathrm{Tr}\ \rho_A \log \rho_A}$



State	EE of the original vector	Required bond dimension
Ferro or Disordered	$S_A = O(1)$	$\chi = O(1)$
Critical	$S_A = O(\log N)$	$\chi = O(N^\alpha)$

We need polynomially large bond dimension for critical system!

More efficient tensor network for critical systems?

Key point: Scale invariance of the system

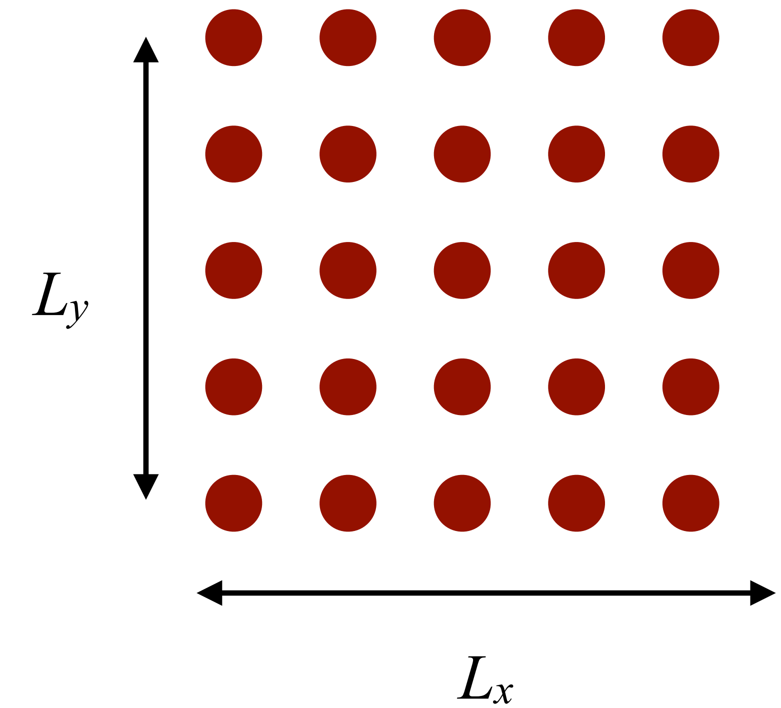
Higer dimensional system

Transverse field Ising model on square lattice:

$$\mathcal{H} = -\sum_{\langle i,j \rangle} S_i^z S_j^z - h \sum_{i=1}^{N} S_i^x$$

$\sum_{\langle i,j \rangle}$:Summation over the nearest neighbor pair

Two-dimensional array



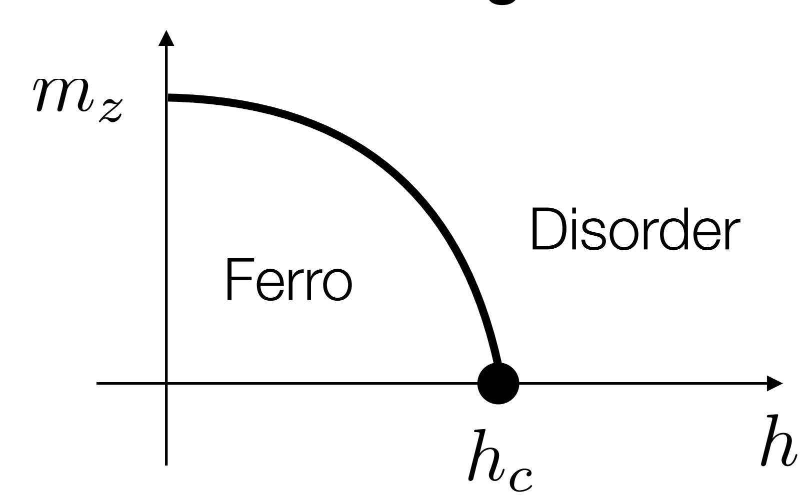
$$N = L_x \times L_y$$

Area law

Even in ferro and disordered phases,
the entanglement entropy depends on size N.

$$S_A \sim \sqrt{N} = L$$

Phase diagram



MPS for two-dimensional system

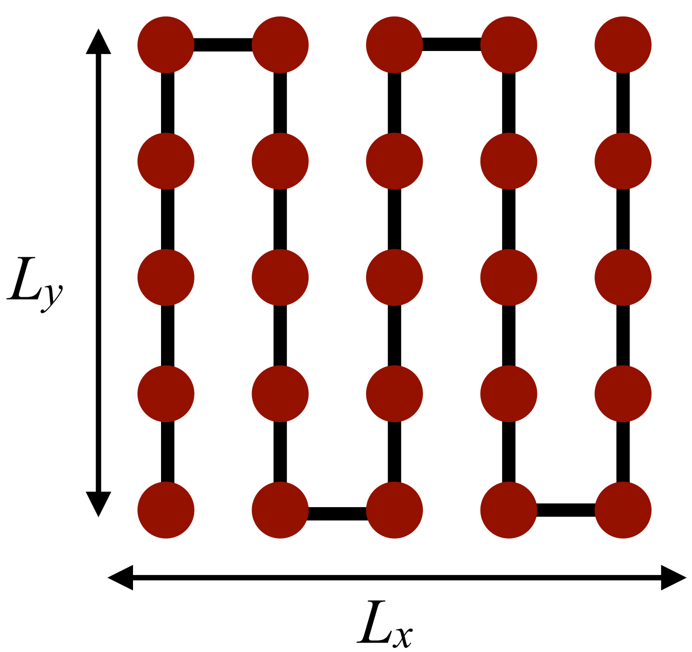
When we apply MPS representation for a square lattice system:

Setting **(1)** $S_A \leq L_x \log \chi$:Satisfying area law?

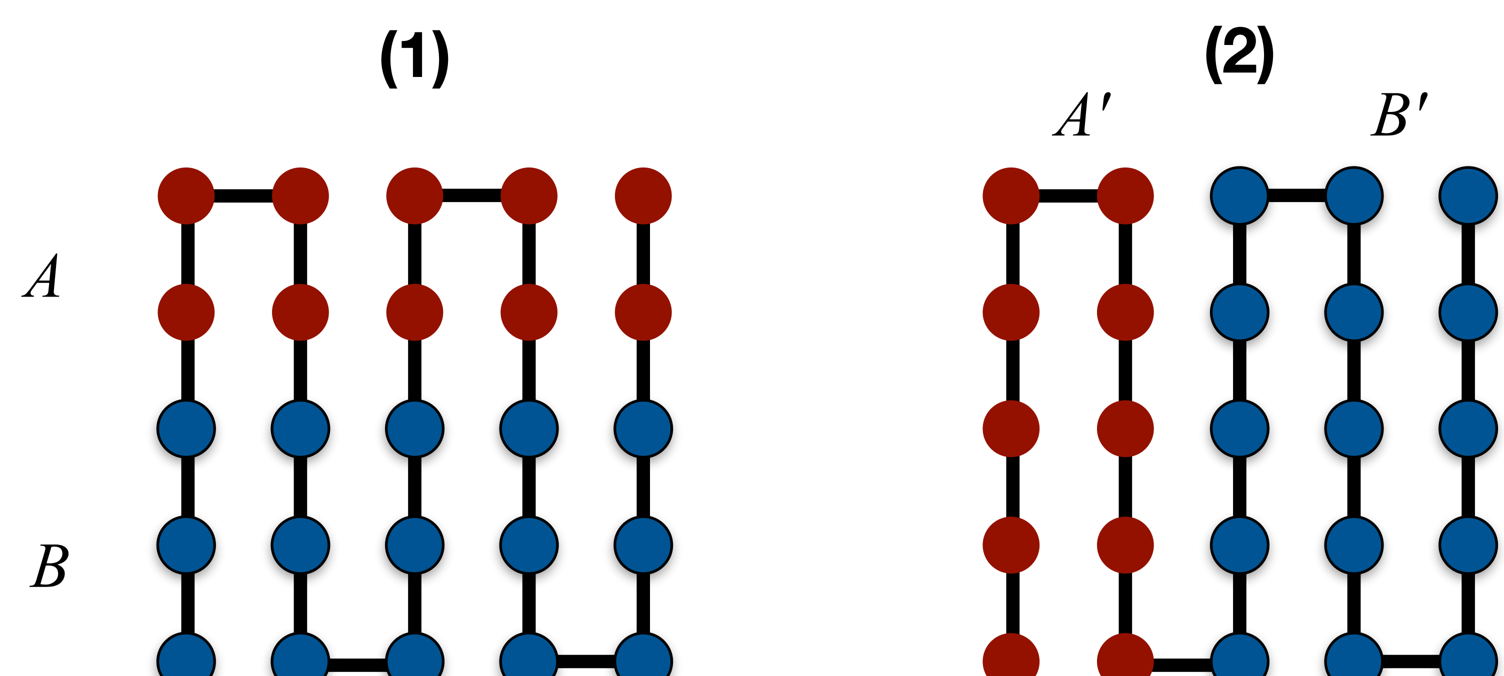
Setting **(2)** $S_{A'} \leq \log \chi$:Break down of the area law!

MPS cannot cover the area law of the entanglement entropy in higher (d =2,3, ...) dimensions.

Possible MPS (Snake form)



Two settings of system and environment



MPS for two-dimensional system: comment

MPS can treat "**rectangular**" or "**quasi one dimensional**" lattice.

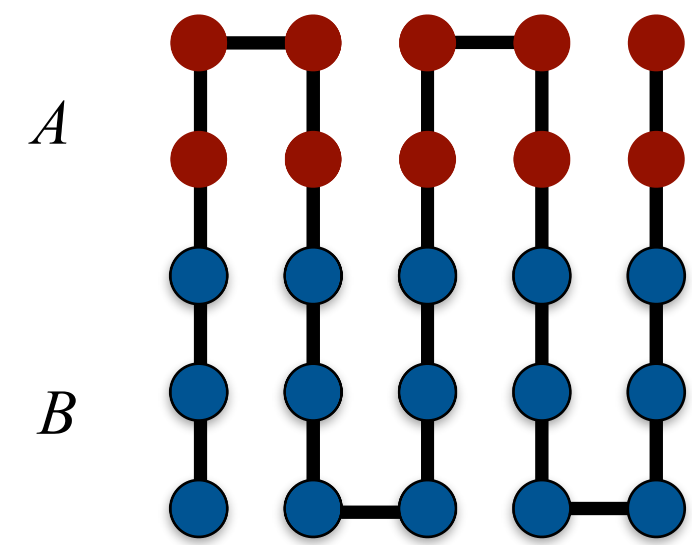
In setting (1), MPS can satisfy the area low partially.

➡ We can increase L_x easily with keeping L_y constant.

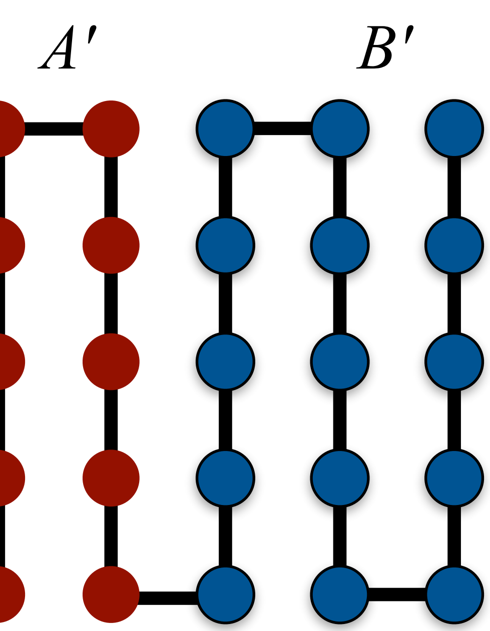
$$\chi = O(e^{L_y})$$

$$L_y \lesssim 10, L_x \gg L_y$$

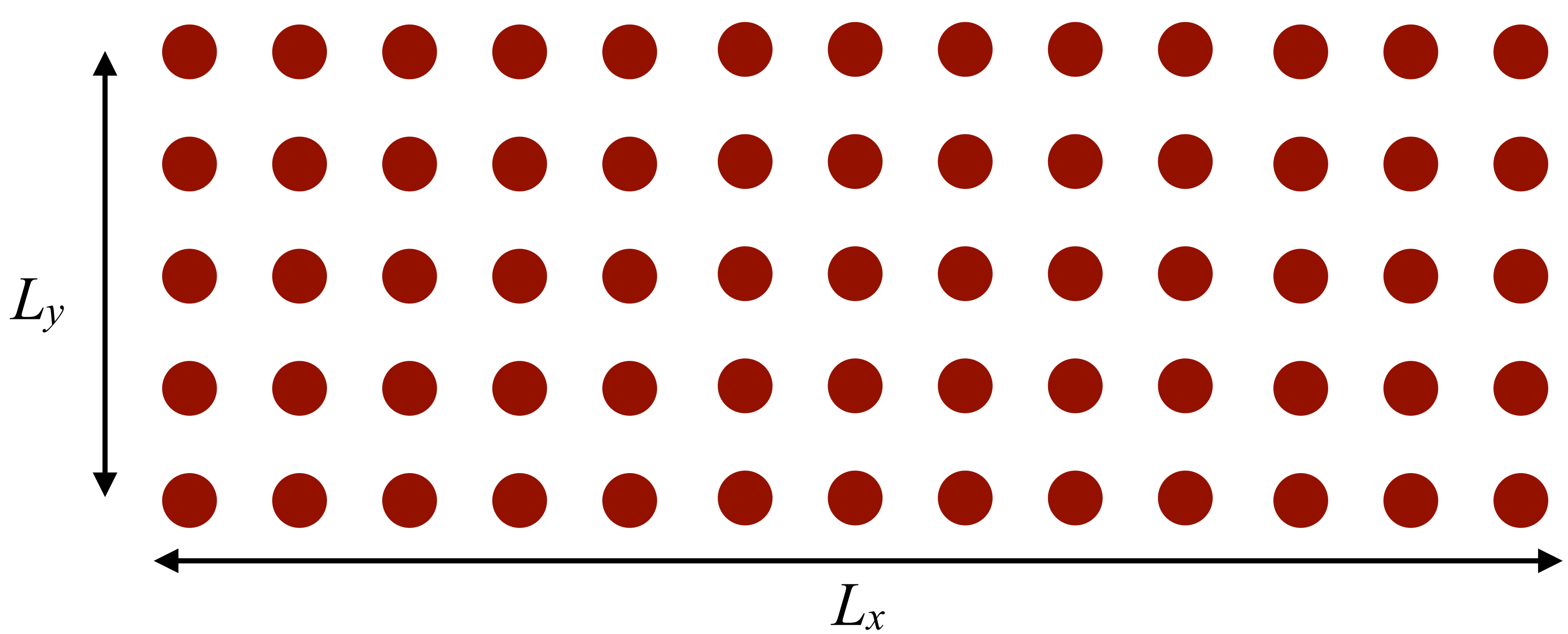
(1) $S_A \leq L_x \log \chi$



(2) $S_{A'} \leq \log \chi$



Quasi one dimensional system ("strip" or "cylinder")



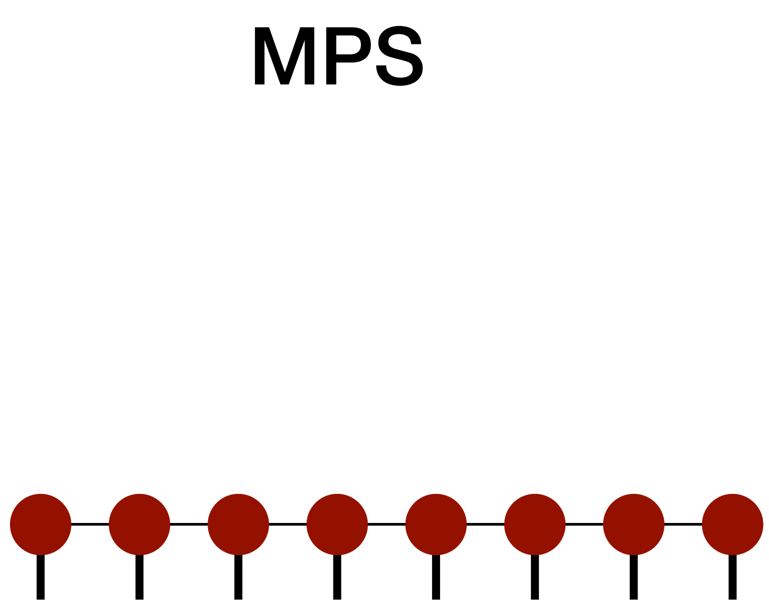
Tensor network for critical systems:

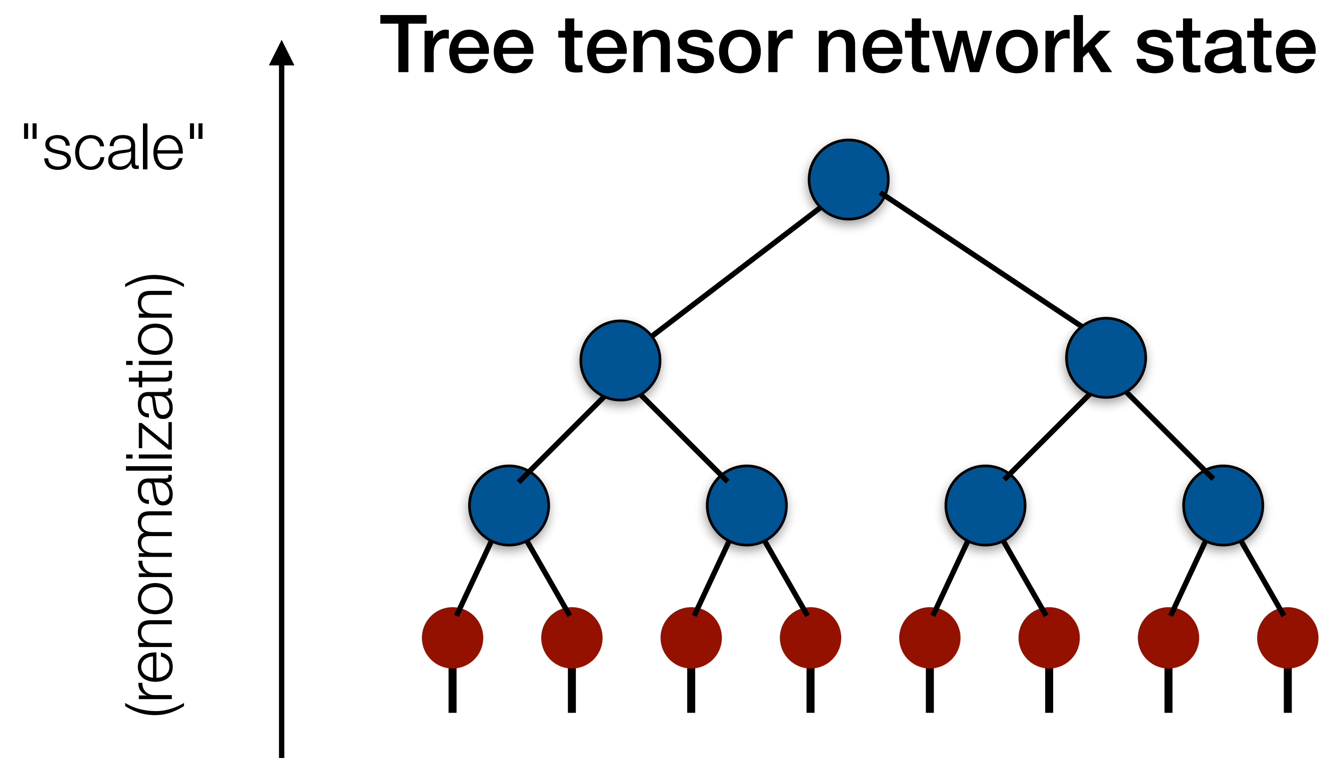
Multi-scale Entanglement Renormalization Ansatz

Hierarchical structure: tree tensor network

Critical system ⟺ Scale invariance

A simple scale invariant tensor network: tree tensor network



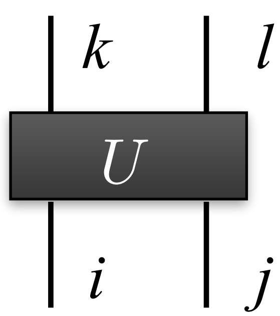


Notice:

Unitary tensors

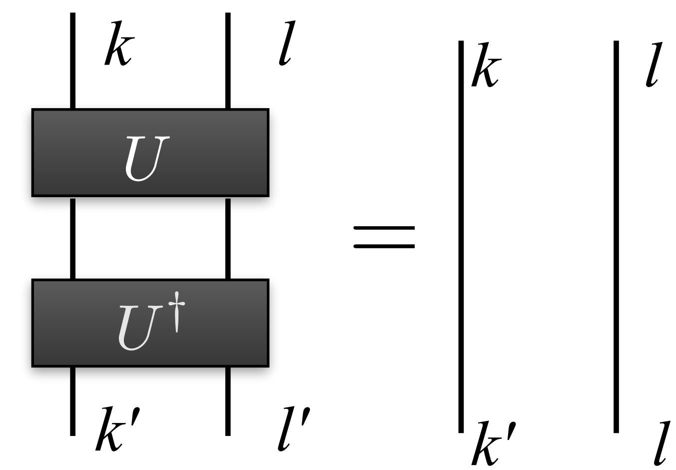
Unitary tensor

$$U_{ij}^{kl} =$$

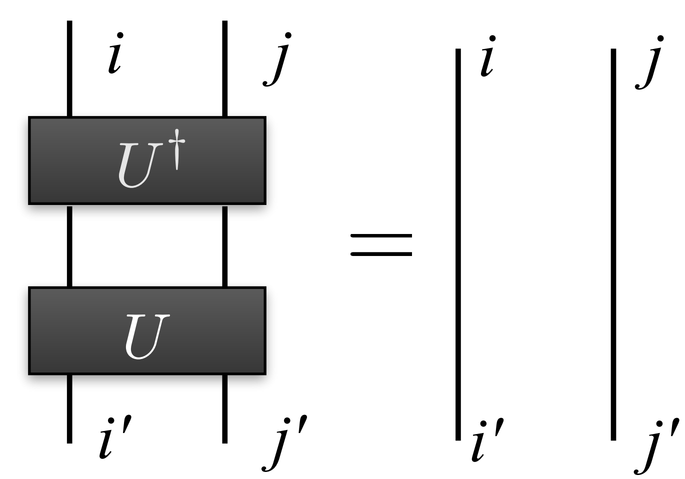


$$\boxed{(U^\dagger)_{kl}^{ij} = (U_{ij}^{kl})^*}$$

$$\sum_{i,j} U_{ij}^{kl} (U^\dagger)_{k'l'}^{ij} = \delta_{kk'}\delta_{ll'}$$



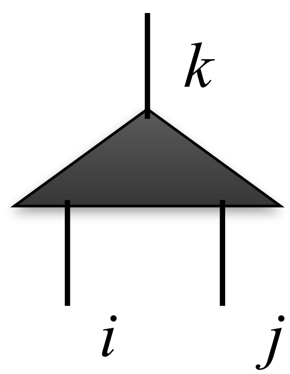
$$\sum_{k,l} (U^\dagger)_{kl}^{ij} U_{i'j'}^{kl} = \delta_{ii'}\delta_{jj'}$$



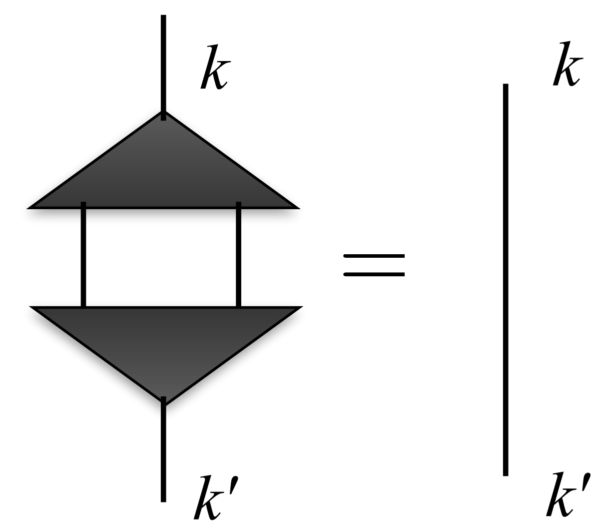
Isometric tensors

Isometric tensor (half unitary tensor) = Isometry

$$W^{k}_{ij} =$$



$$\sum_{i,j} W^{k}_{ij} (W^{\dagger})^{ij}_{k'} = \delta_{kk'}$$



Unitarity condition only for "bottom" legs.

Isometry works as a "projector" from the bottom space to the top space.

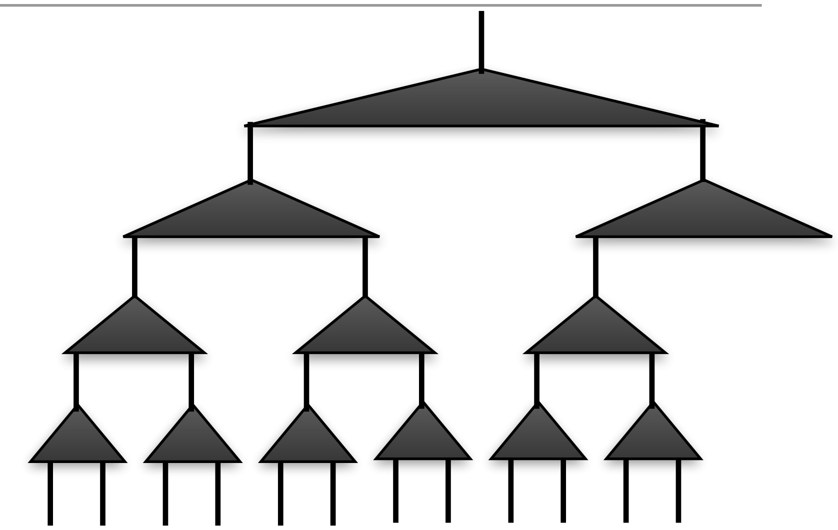
$$\mathrm{dim(bottom)} \geq \mathrm{dim(top)}$$

It is also related to the "renormalization" of degree of freedoms.

We pick up "important" degree of freedoms by isometries.

Isometric tree tensor network and its scale invariance

Consider an (infinite) tree tensor network consists of identical isometries as a wave function.

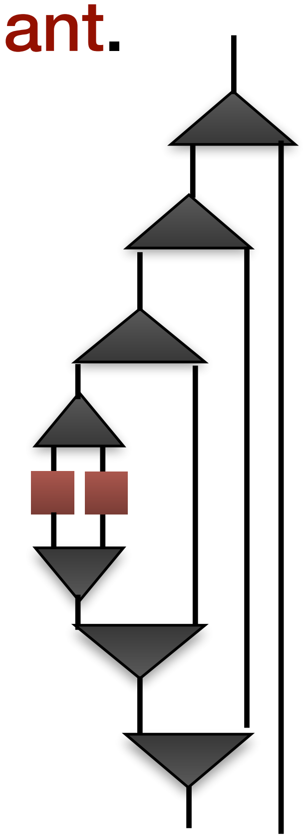
$|\Psi\rangle =$ 

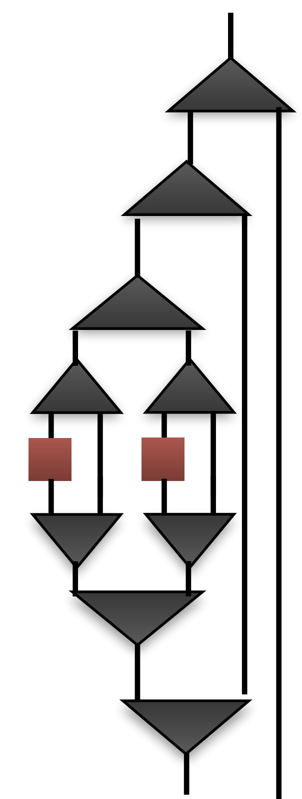
Properties:

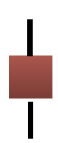
1. It is normalized as $\langle\Psi|\Psi\rangle = 1$

(Trivial from the definition of the isometry)

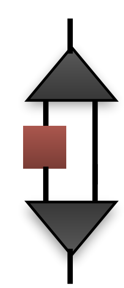
2. It can be scale invariant.

$C(1) \equiv \langle\Psi|S_1^z S_2^z|\Psi\rangle =$  $\cdots$

$C(2) \equiv \langle\Psi|S_1^z S_3^z|\Psi\rangle =$  $\cdots$

$S_i^z =$ 

spin

$(S_i^z)' =$ 

"renormalized" spin

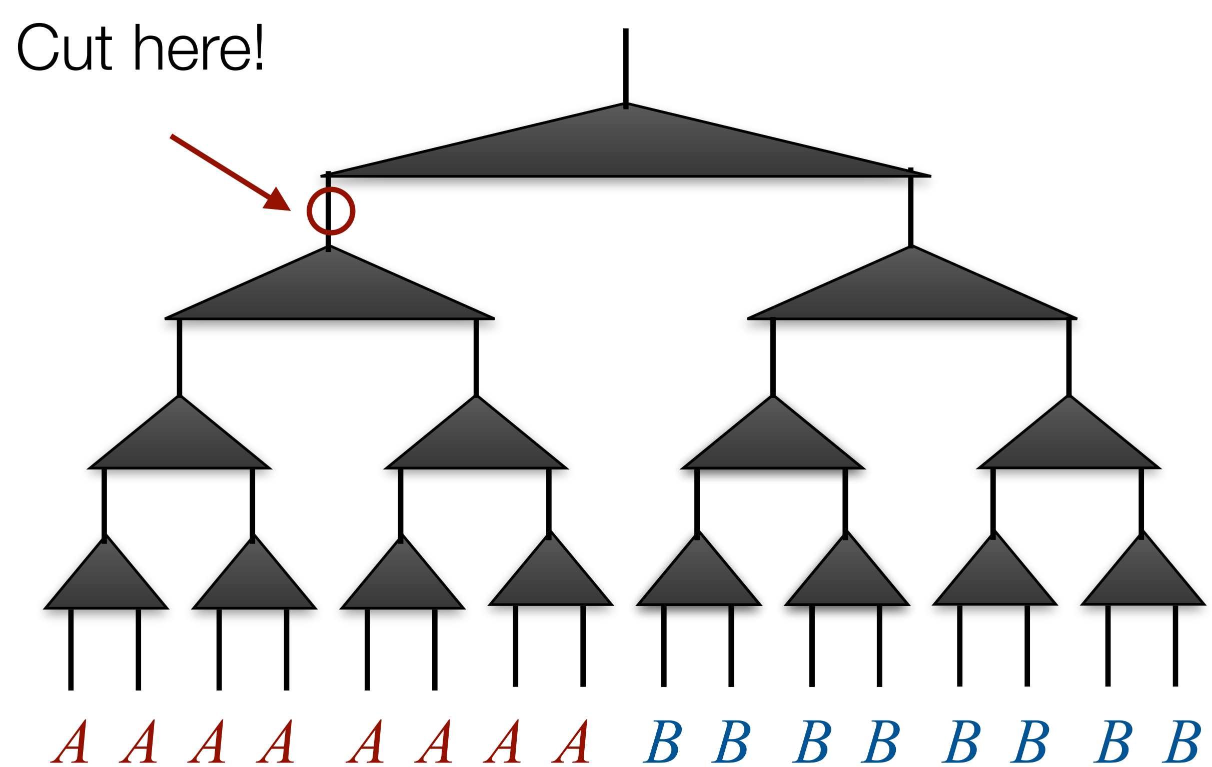
If $(S_i^z)' = 2^{-p} S_i^z$, then $C(2r) = 2^{-2p} C(r)$ $\boxed{\text{Scale invariant!}}$

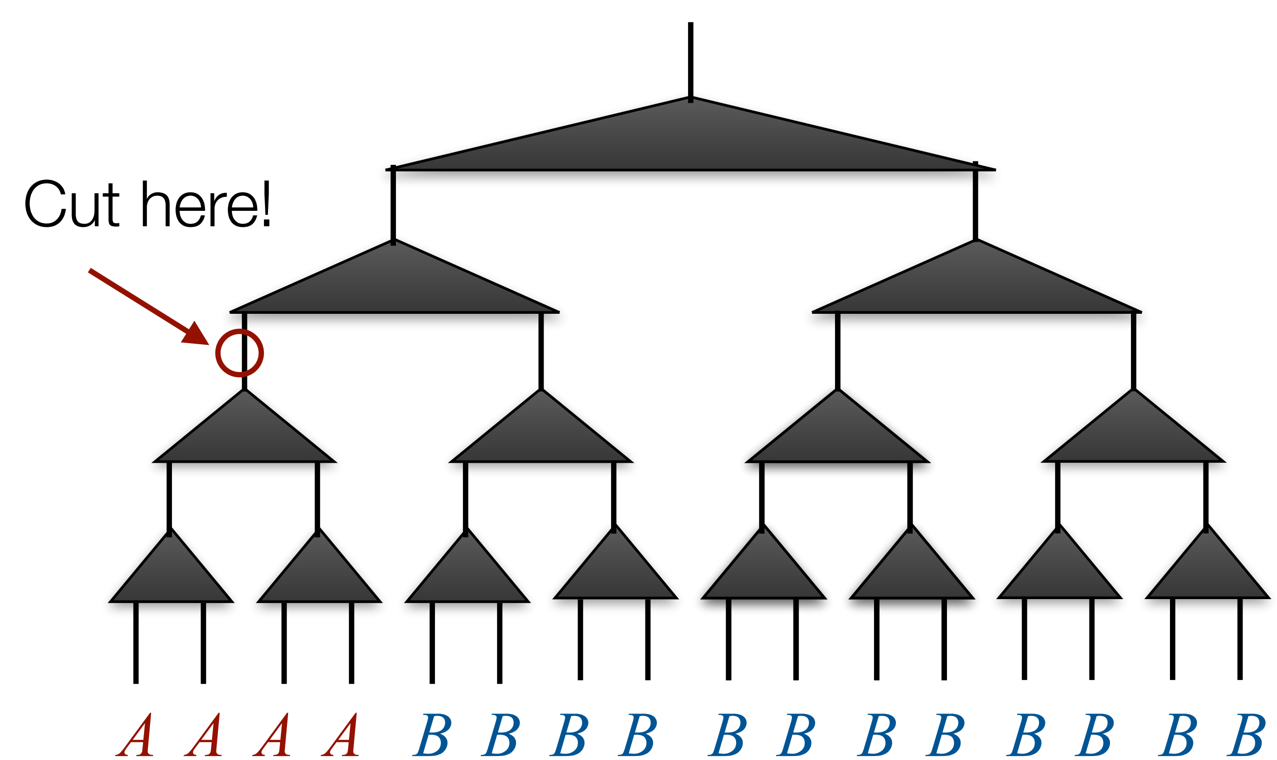
Entanglement entropy of TTN

Entanglement entropy of tree tensor networks (TTN):

Due to the tree structure, two regions are connected by only "one bond".
(or a few)

$$\boxed{S_A = -\mathrm{Tr}\ \rho_A \log \rho_A \leq \log \chi}$$



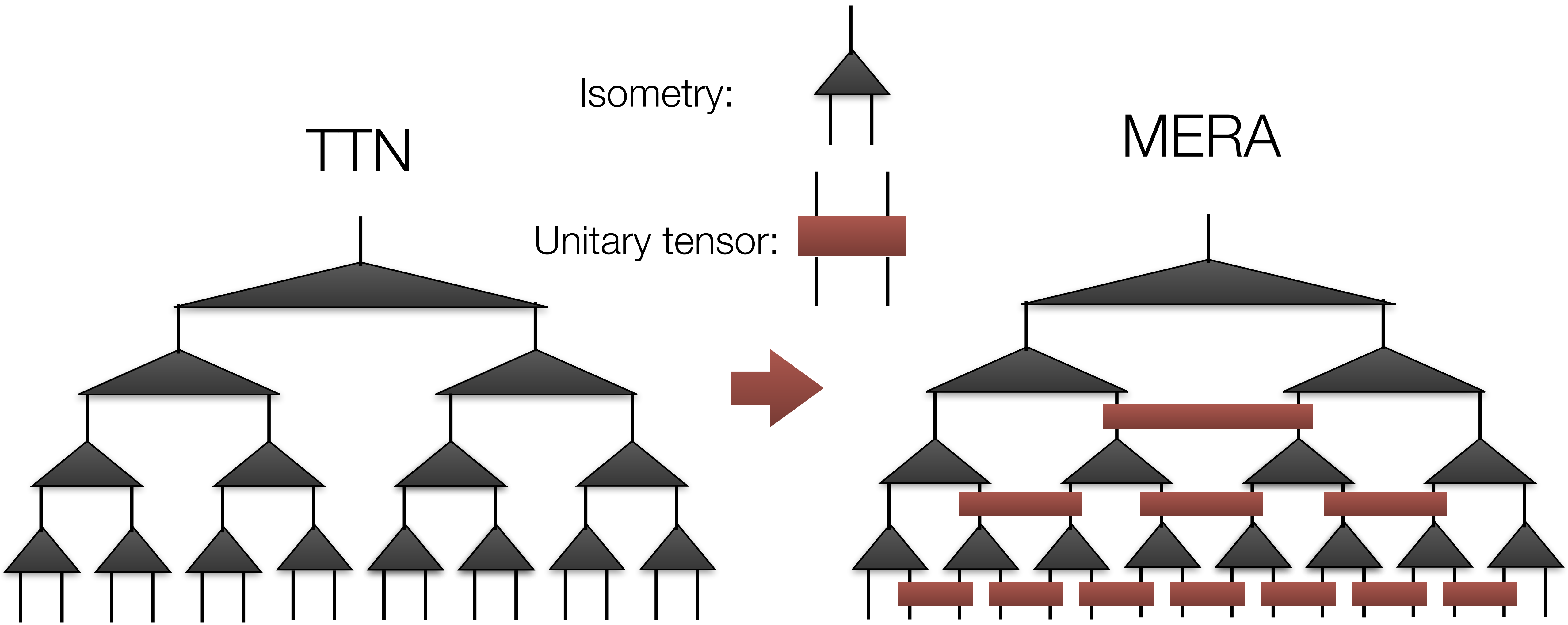


MERA

(G. Vidal, Phys. Rev. Lett. **99**, 220405 (2007))
(G. Vidal, Phys. Rev. Lett. **101**, 110501 (2008))

Multi-scale Entanglement Renormalization Ansatz (**MERA**)

Before applying isometry, insert a unitary tensor.



- [x] Normalization
- [x] Scale invariance (if we set the identical tensors)

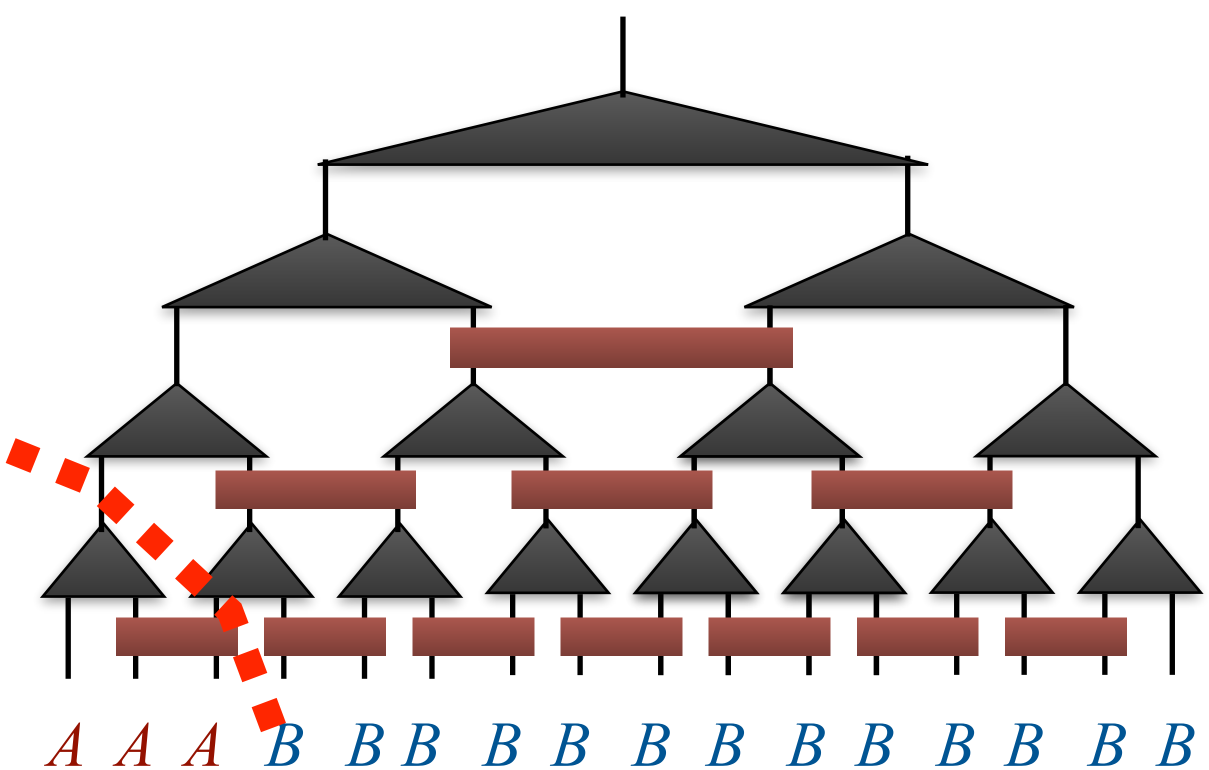
Entanglement entropy of MERA

Due to the unitary matrices, # of bonds connecting two regions logarithmically increase.

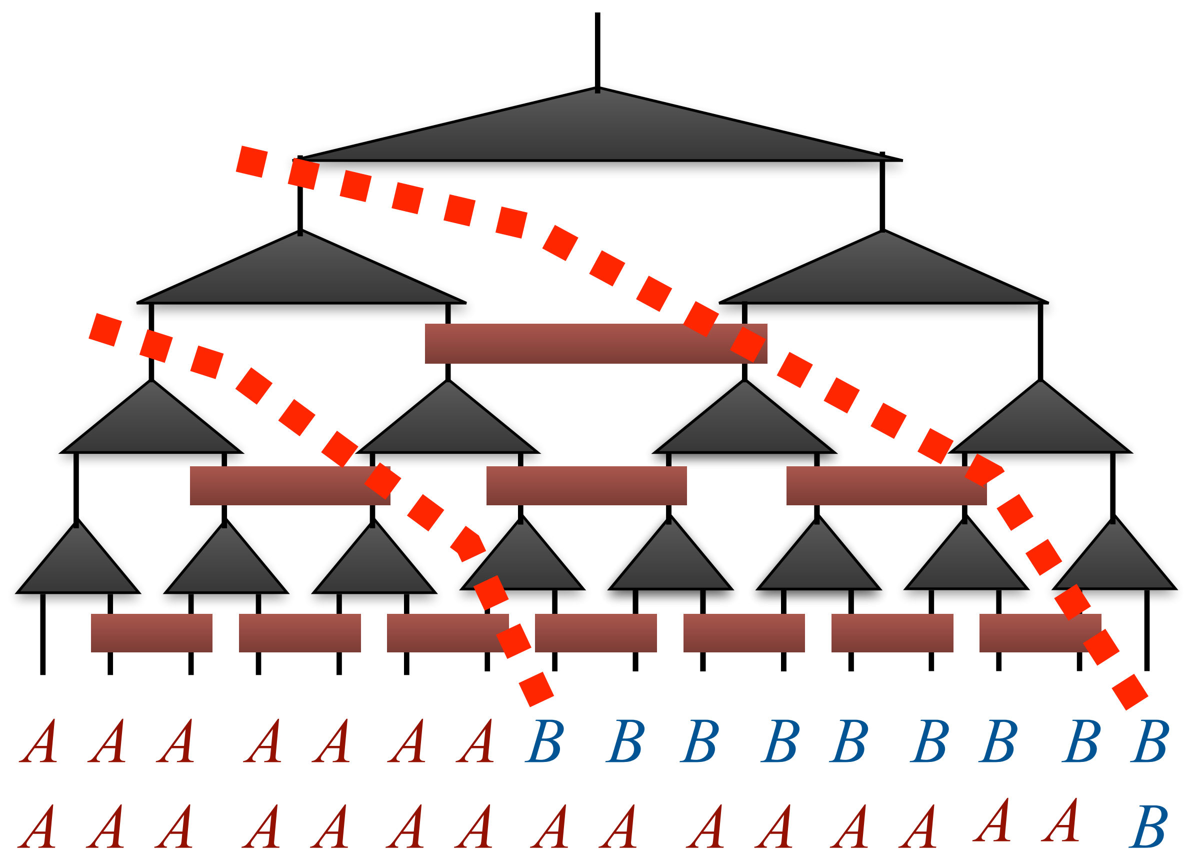
$$\text{rank}\ \rho_A \leq \chi^{N_c(N)} \sim \chi^{\log N}$$
$$S_A = -\text{Tr}\ \rho_A \log \rho_A \leq (\log \chi) \log N$$

$N_c(N)$:# of minimum cut for a N-site region

$$N = 2^{N_c} - 1$$
$$N_c \sim \log_2 N$$



Minimum # of cuts = 2



Minimum # of cuts = 3

Minimum # of cuts = 4

Application of MERA

Transverse field Ising chain:

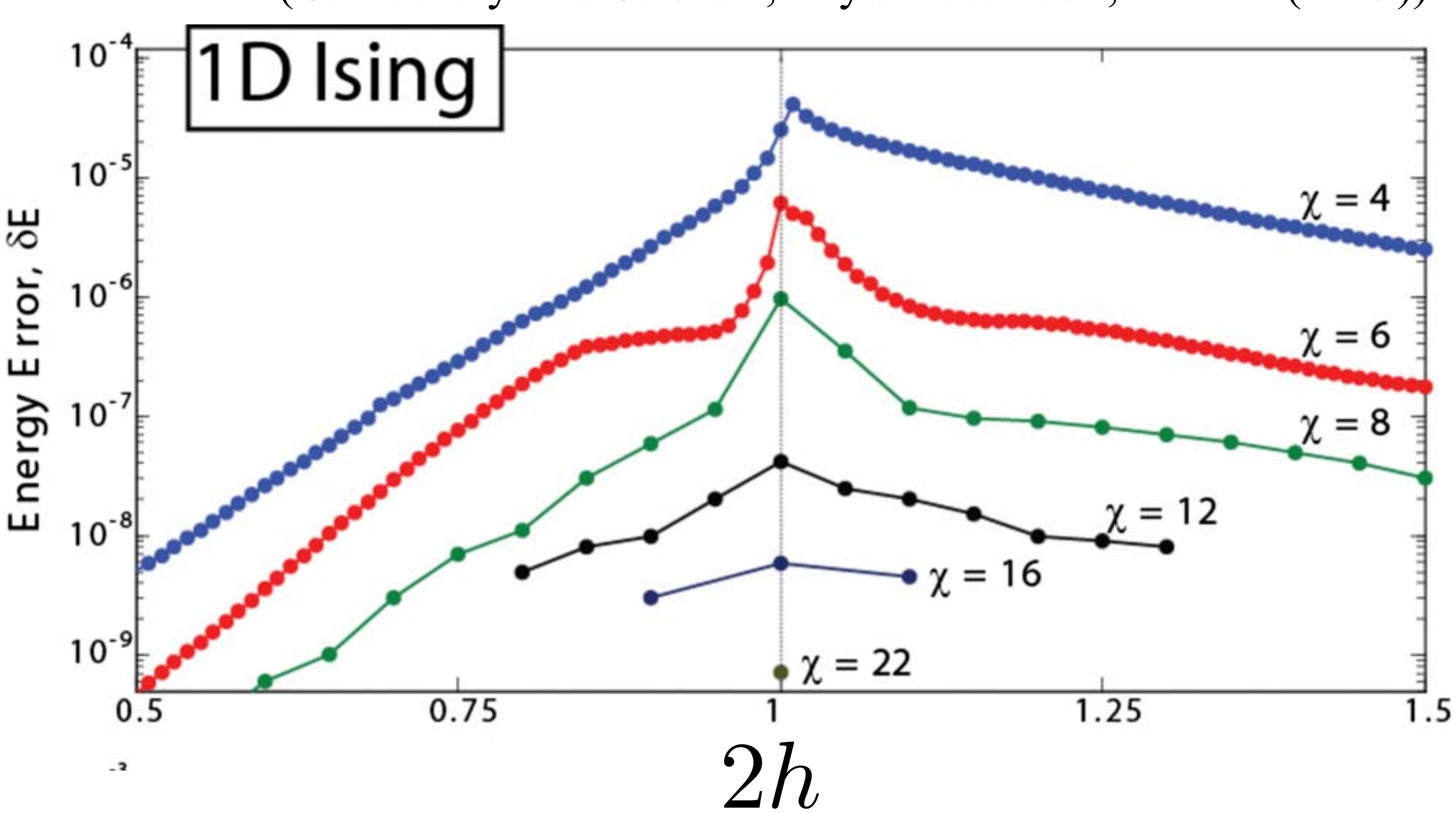
$$\mathcal{H} = -\sum_{i=1}^{N-1} S_i^z S_{i+1}^z - h \sum_{i=1}^{N} S_i^x$$

MERA can represent very large (Infinite) critical system!

Energy errors:

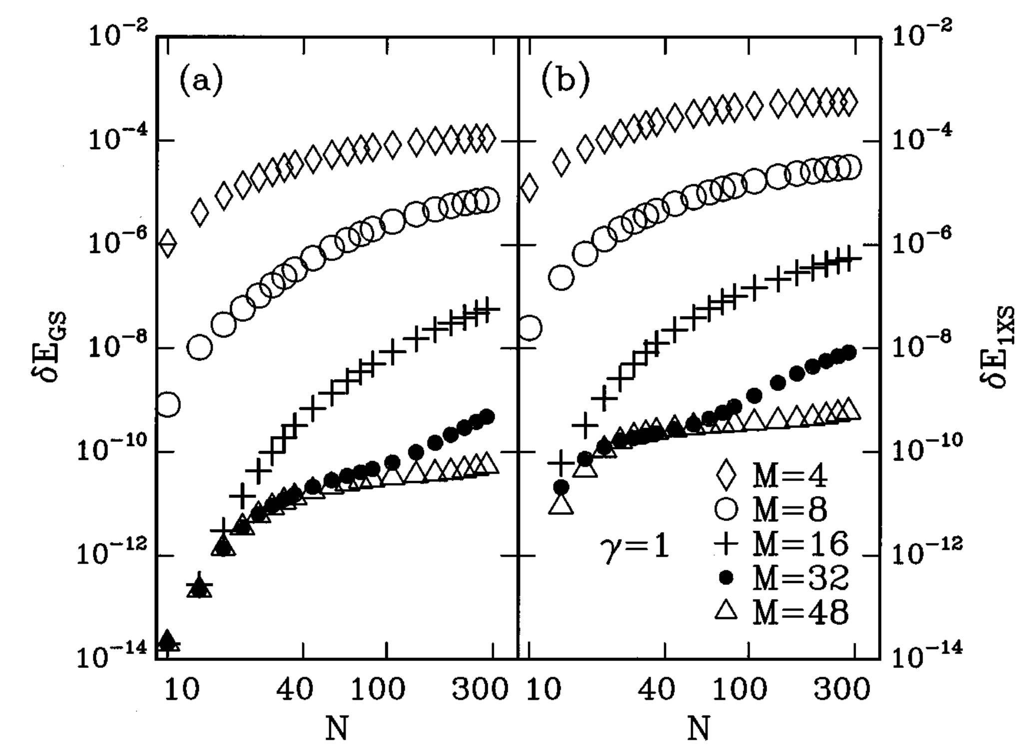
MERA (Infinite chain)

(G. Evenbly and G. Vidal, Phys. Rev. B. **79**, 144108 (2009))



DMRG (finite chain)

Ö. Legeze, and G. Fáth (1996)



$$h = h_c$$

Interesting topics related to MERA

- By using scale invariance of MERA, we can calculate properties of critical system accurately. (G. Evenbly and G. Vidal, Phys. Rev. B. **79**, 144108 (2009)) (R.N.C. Pfeifer, (G. Evenbly and G. Vidal, Phys. Rev. A. **79**, 040301(R) (2009))
  - Critical exponents and Operator product expansion coefficients in the Conformal Filed Theory (CFT)
- We can consider MERA in higher dimensions
  - It is scale invariant but satisfies the area low (G. Evenbly and G. Vidal, Phys. Rev. Lett. **102**, 180406 (2009))
  - For the system with logarithmic correction in the EE, such as metal, "branching MERA" has been proposed. (G. Evenbly and G. Vidal, Phys. Rev. Lett. **112**, 220502 (2014)) (G. Evenbly and G. Vidal, Phys. Rev. B. **89**, 235113 (2014))
- Relation between MERA and other fields
  - Wavelet transform (G. Evenbly and S. R. White, Phys. Rev. Lett. **112**, 140403 (2016))
  - AdS/CFT (quantum gravity, black hole) (M. Nozaki, S. Ryu, and T. Takayanagi, J. High Energy Phys. **10**, 193 (2012))

Tensor network for higher dimensional systems:

Tensor Product State

(Projected Entangled Pair State)

Entanglement entropy in higher dimensions

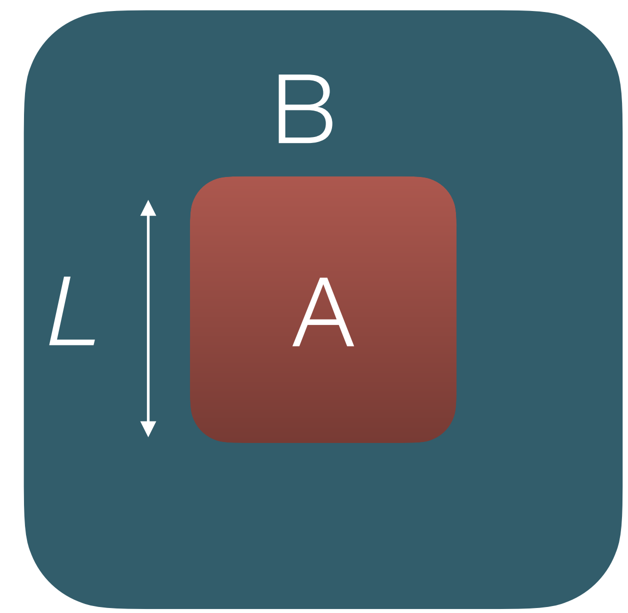
Ground state wave functions:

For a lot of ground states, EE is proportional to its area.

J. Eisert, M. Cramer, and M. B. Plenio, Rev. Mod. Phys, 277, **82** (2010)

Area low:

$$S = -\mathrm{Tr}\left(\rho_A \log \rho_A\right) \propto L^{d-1}$$



In d=1, MPS satisfies the area low.

Q. What is a simple generalization of MPS to d > 1?

A. It is Tensor Product State (TPS)!

Tensor Product State (TPS)

TPS (Tensor Product State) (AKLT, T. Nishino, K. Okunishi, …)

PEPS (Projected Entangled-Pair State)

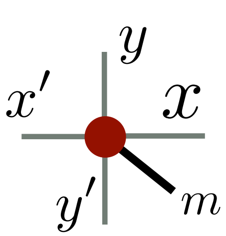
(F. Verstraete and J. Cirac, arXiv:cond-mat/0407066)

d-dimensional tensor network representation
for the wave function of a d-dimensional quantum system

$$|\Psi\rangle = \sum_{\{m_i=1,2,\cdots,m\}} \mathrm{Tr}\ A_1[m_1]A_2[m_2]\cdots A_N[m_N]|m_1 m_2 \cdots m_N\rangle$$

Tr: tensor network "contraction"

$A_{x_i x_i' y_i y_i'}[m_i]$: Rank 4+1 tensor

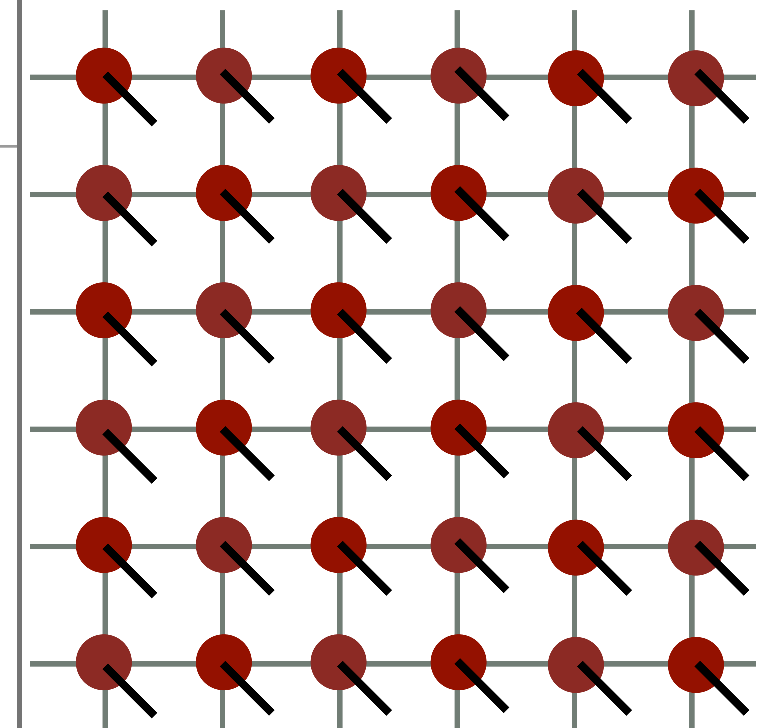


$x, y, x', y' = 1, 2, \ldots D$ $D =$ "bond dimension"

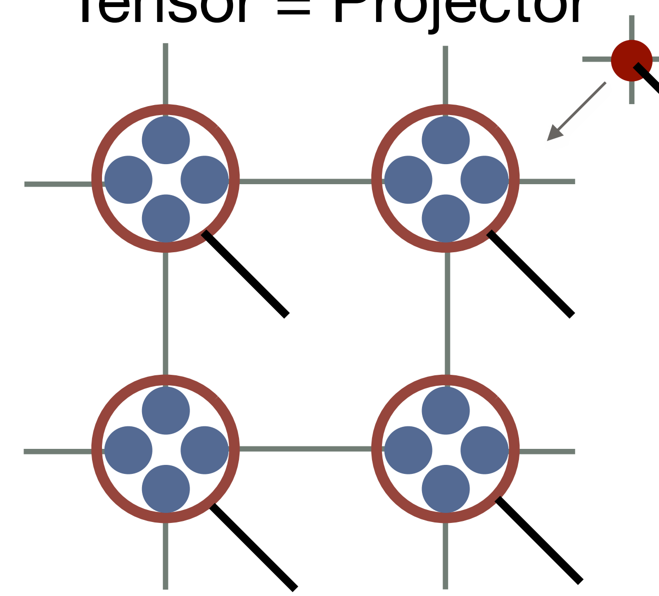
$m_i = 1, 2, \ldots m$ $m =$ dimension of the local Hilbert space

*D can be larger than m. "Virtual state "

TPS on square lattice

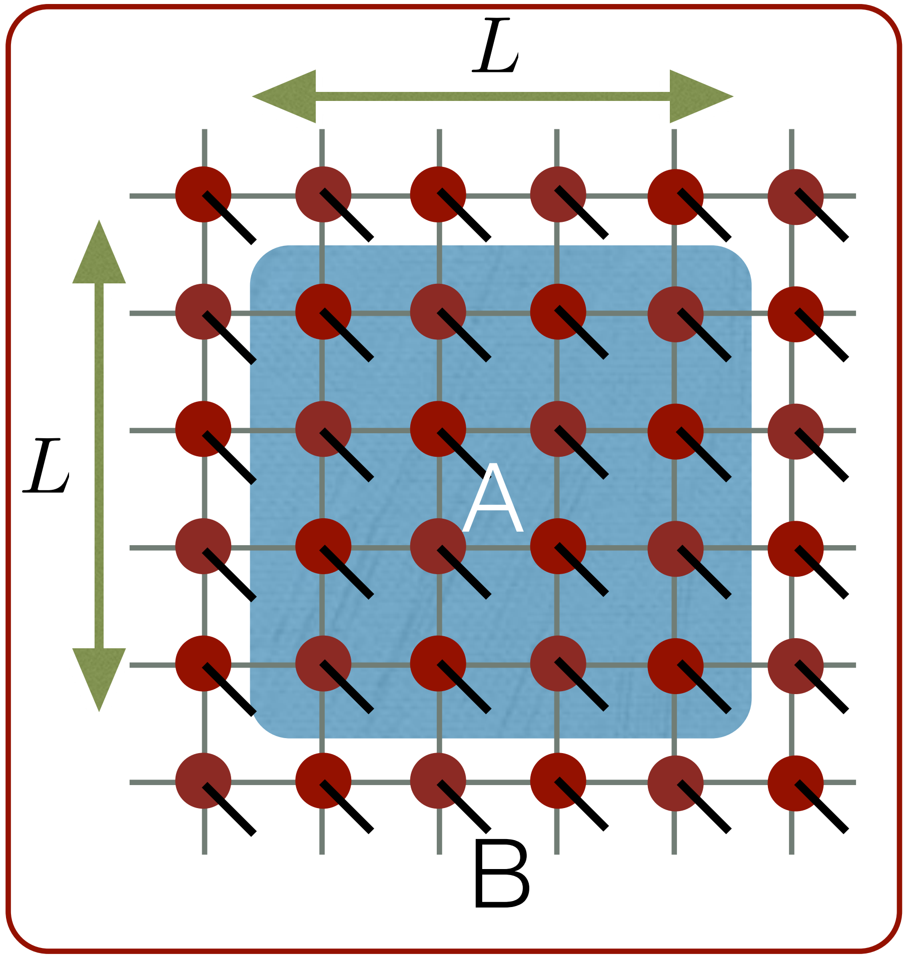


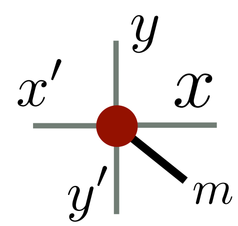
Tensor = Projector



Maximally entangled state between D-state spins

Entanglement entropy of TPS (PEPS)





Bond dimension = D

# of bonds connecting regions A and B

$N_c(L) = 4L$ (square lattice)

$N_c(L) = 2dL^{d-1}$ (d-dimensional hyper cubic lattice)

$$\mathrm{rank}\ \rho_A \leq D^{N_c(L)} \sim D^{2dL^{d-1}}$$

$$S_A = -\mathrm{Tr}\ \rho_A \log \rho_A \leq 2dL^{d-1} \log D$$

TPS can satisfies the area law even for d >1.

We can efficiently approximate vectors in higher dimensional space by TPS.

* Similar to the MPS in 1d, TPS can approximate infinite system!

Difference between MPS and TPS

Cost of tensor network contraction:

d-dimensional cubic lattice $N = L^d$

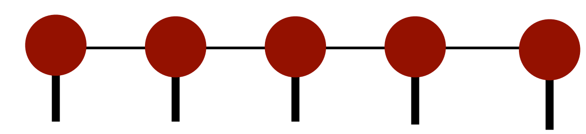
MPS: $O(N)$

TPS: $O(e^{L^{d-1}})$

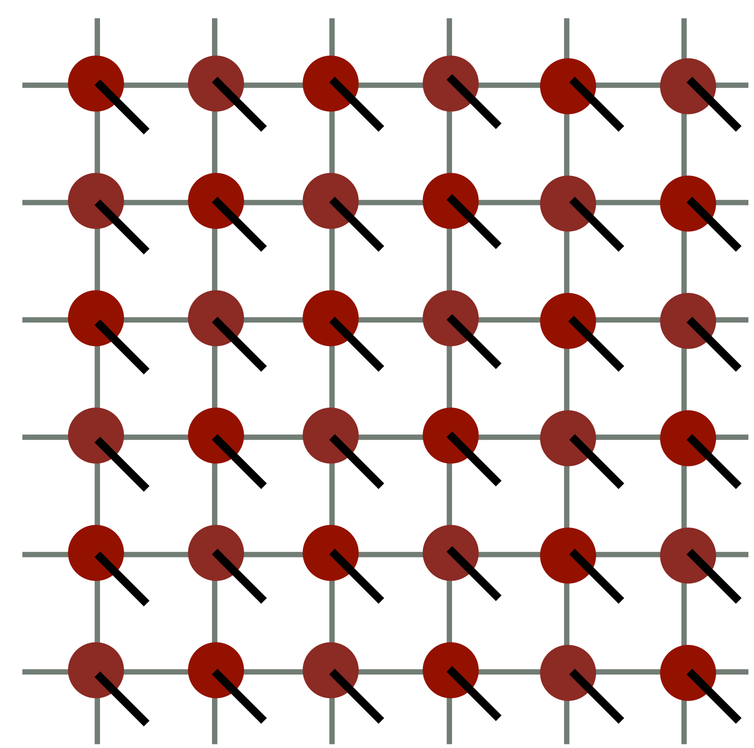
It is impossible to perform exact contraction even if we know local tensors in the case of TPS.

In the case of TPS, usually we approximately calculate the contraction.

MPS



TPS (PEPS)



Example of approximate contraction: CTM method

For (infinite) open boundary system

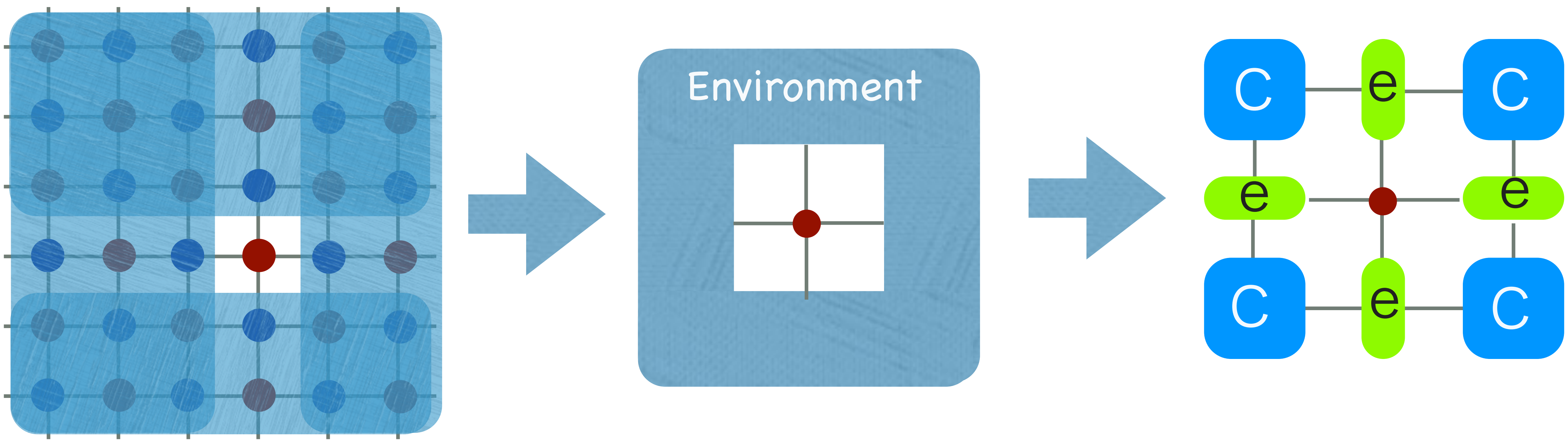
(T. Nishino and K. Okunishi, JPSJ **65**, 891 (1996))
(R. Orus *et al*, Phys. Rev. B **80**, 094403 (2009))

Infinite PEPS

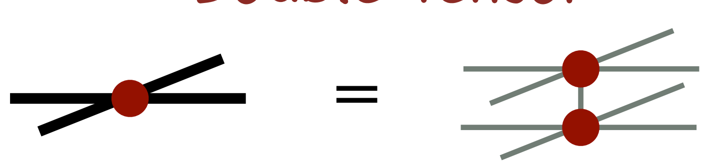
(with a translational invariance)

Environment

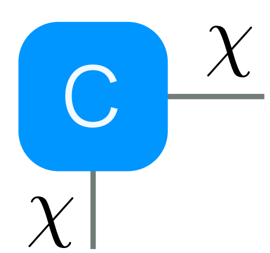
Corner transfer matrix Representation



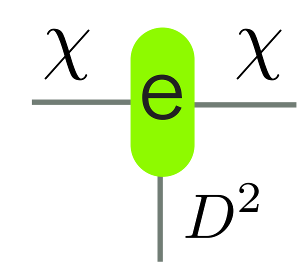
Double tensor



Corner transfer matrix



Edge tensor



Mapping to a "classical" system

χ =bond dimension $\chi \sim D^2$

Application of TPS to eigenvalue problem

For calculation of minimum eigenvalues and its eigenvector, we can use similar techniques to those in MPS

Variational method:

(P. Corboz, Phys. Rev. B **94**, 035133 (2016))
(L. Vanderstraeten, Phys. Rev. B **94**, 155123 (2016))
(H.-J. Liao *et a*l, Phys. Rev. X **9**, 031041(2019))

minimize cost function: $F = \dfrac{\vec{\psi}^\dagger(\mathcal{H}\vec{\psi})}{\vec{\psi}^\dagger\vec{\psi}}$

Imaginary time evolution:

Simulate imaginary time evolution: $|\Psi_{\mathrm{GS}}\rangle \propto \lim_{\beta\to\infty} e^{-\beta\mathcal{H}}|\Psi_0\rangle$
（虚時間発展）

For a initial state $\langle\Psi_{\mathrm{GS}}|\Psi_0\rangle \neq 0$

(H. G. Jiang *et al*, Phys. Rev. Lett. **101**, 090603 (2008))
(J. Jordan *et al*, Phys. Rev. Lett. **101**, 250602 (2008))

Example of application: Honeycomb lattice Kitaev Model

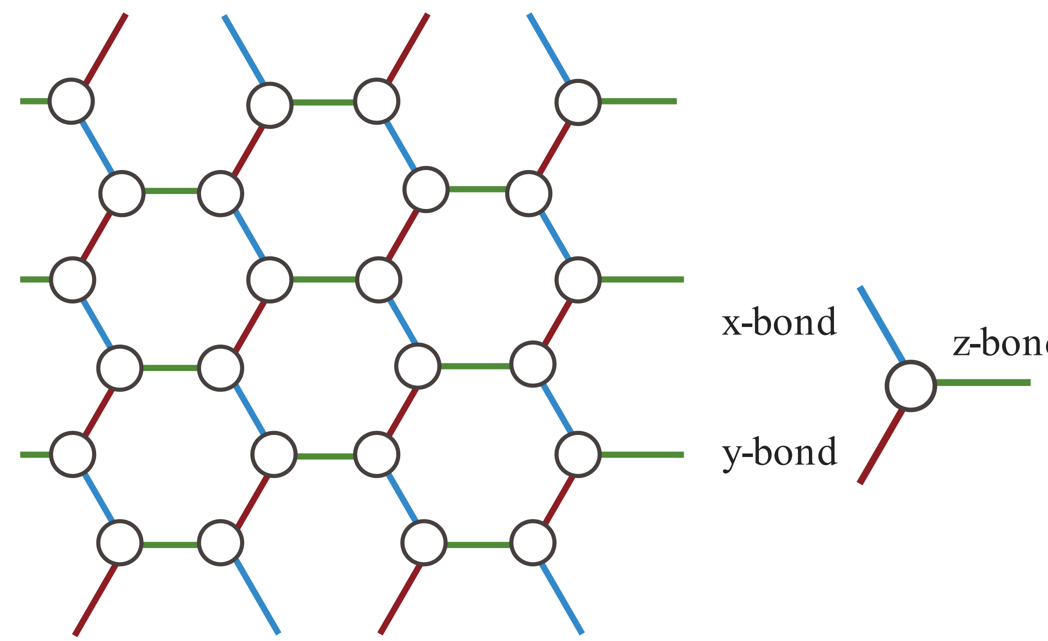
A. Kitaev, Annals of Physics 321, 2 (2006)

Kitaev model

$$\mathcal{H} = -\sum_{\gamma,\langle i,j\rangle_\gamma} J_\gamma S_i^\gamma S_j^\gamma$$

γ:bond direction

Depending on the bond direction, only specific spin components interact.



Exactly solvable by introducing Majorana fermion

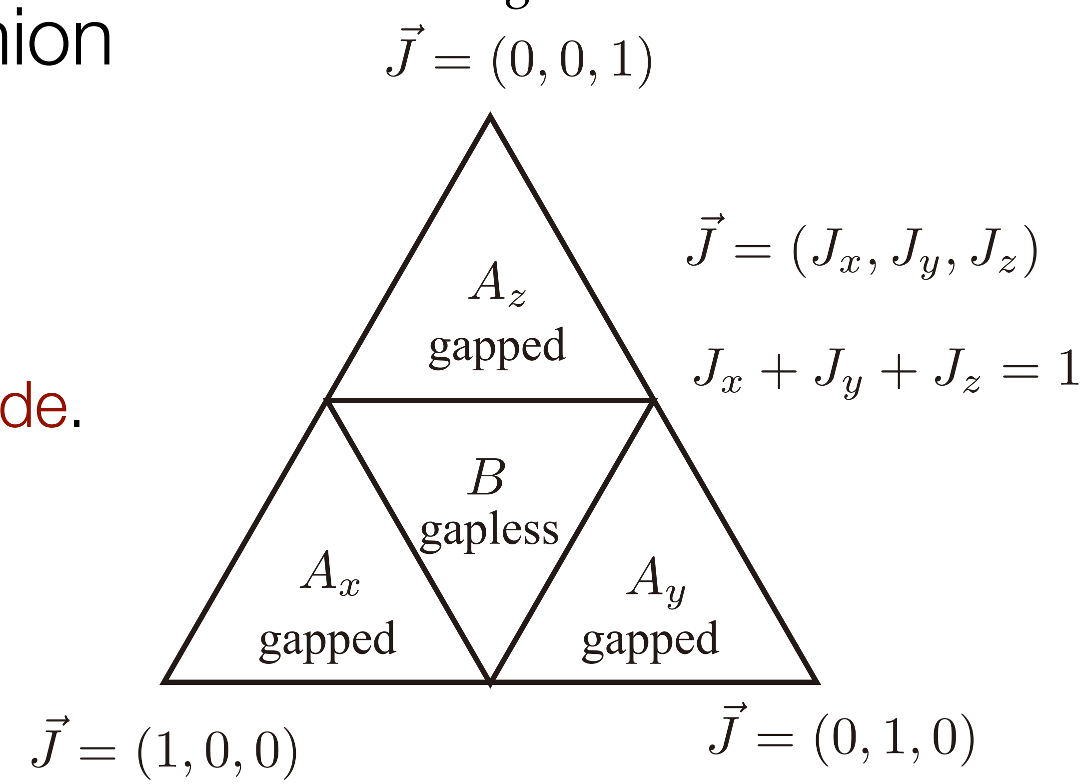
Isotropic region (B) : gapless spin liquid

Anisotropic region (A) : gapped spin liquid

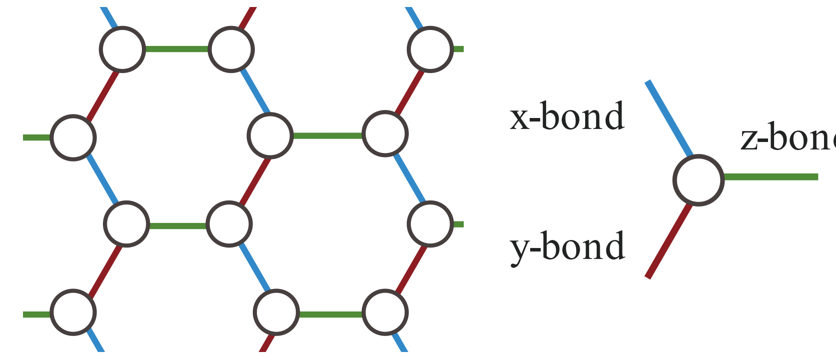
Cf. The anisotropic limit corresponds to the Toric code.

*Recently, researchers have realized that this type of models might appear in real materials. Hot topic!

Phase diagram



Application : Kitaev spin liquid



Honeycomb lattice Kitaev model

$$\mathcal{H} = -\sum_{\gamma,\langle i,j\rangle_\gamma} J_\gamma S_i^\gamma S_j^\gamma$$
$$(\gamma = x, y, z)$$

At $J_x = J_y = J_z$, the ground state is

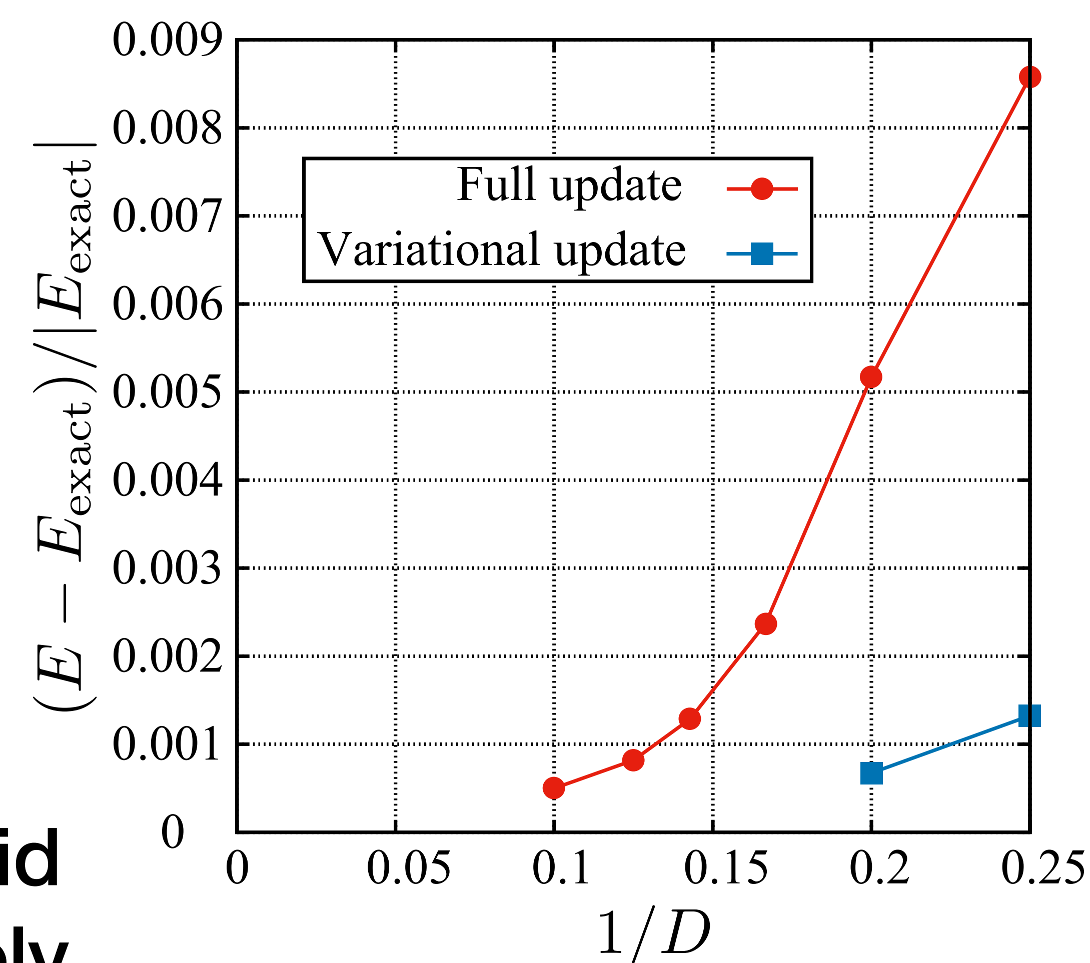
a gapless spin liquid.

In the present (super)computers,
we can access around D=10 (maybe 16)
by using massively parallel code.

The error of the ground state
energy is **less than 10^{-3}**
for infinite system!

Energy error obtained by iTPS

(T. okubo et al, unpublished)



iTPS can represent Kitaev spin liquid
in the thermodynamic limit accurately.

Interesting topics related to TPS

- Application to itinerant electron system, which may break the area law

  (P. Corboz et al, Phys. Rev. B. **81**, 165104 (2010))

  (P. Corboz, Phys. Rev. B. **93**, 045116 (2016))

- Characterization of topologies in wave function
  - Symmetric tensor network and modular matrix

    (J.-W. Mei et al, Phys. Rev. B. **95**, 235107 (2017))
- Application to three dimensions
  - So far, there is no practical calculations for non-trivial models.
  - Mainly, due to the scaling: $O(D^{18})$?
- Calculation of thermal states
  - Tensor network decomposition of the density matrix.

    (A. Kshetrimayum et al, PRL **122**, 070502 (2019))

    (Czarnik et al, PRB **99**, 035115 (2019))

References

- R. Orús, *A practical introduction to tensor networks: Matrix product states and projected entangled pair states*, Annals. of Physics **349**, 117 (2014).
- R. Orús, *Tensor networks for complex quantum systems*, Nature Review Physics **1**, 538 (2019).
- 西野友年、大久保毅 テンソルネットワーク形式の進展と応用, 日本物理学会誌 **72**, 702 (2017).
- 大久保毅 テンソルネットワークによる情報圧縮とフラストレート磁性体への応用, 物性研究・電子版 **7**, 072209 (2018) .

Additional information for tensor networks

You can find lecture videos (and slides) related to tensor networks at

https://www.issp.u-tokyo.ac.jp/public/caqmp2019/program.html

They are the lectures given in the recent workshop CAQMP2019.

Examples:

- **Tao Xiang**, *Overview on the Tensor Network Renormalization of (2+1)D Quantum Lattice Model*
- **Natalia Chepiga**, *Matrix Product States for frustrated spin chains, lattices with an extended Hilbert space and constrained models in 1D*
- **Tomotoshi Nishino**, *Optimization Schemes in TNS*
- **Laurens Vanderstraeten**, *Variational methods for tensor networks*
- **Tsuyoshi Okubo**, *Tensor network approach to two-dimensional frustrated spin systems*

数理科学 2022年2月号

特集"テンソルネットワークの進展"

多彩な表現形式が物理をつなぐ

『統計力学系や量子多体系などをテンソルの縮約によって記述するテンソルネットワークは，これまでに数値計算の手法開発など様々な研究を経て発展し，今日では物理学をはじめとする数理科学諸分野から注目を集めています．本特集では，そういったテンソルネットワークの基礎から，統計力学，物性，素粒子，量子情報といった物理分野との接点や，近年著しい発展をみせている機械学習との関連についてを取り上げます．また，それら諸分野との交流によるテンソルネットワーク自身の発展の展望についても迫っていきます．』

- 西野友年：微視から巨視へのネットワーク
- 原田健自：テンソルネットワーク入門 — テンソル縮約，統計力学・量子力学との関わり
- 森田悟史：テンソルネットワークと統計力学
- 柴田尚和 ：テンソルネットワークと物性物理
- 藏増嘉伸 ：テンソルネットワークと場の理論
- 吉田紅：テンソルネットワークと量子重力 — ホログラフィック量子誤り訂正符号
- 藤井啓祐 ：テンソルネットワークと測定型量子計算
- 御手洗光祐 ：テンソルネットワークと機械学習
- 上田宏：テンソルネットワークと数値計算 — 省計算資源による大規模計算から超並列計算技術まで
- 大久保毅：テンソルネットワークの将来

Open source code: TeNeS (and pTNS)

TeNeS: Tensor Network Solver

https://github.com/issp-center-dev/TeNeS

It can calculate the ground state of two-dimensional (spin) models by using iTPS method.

Y. Motoyama, T. Okubo, et.al., arXiv:2112.13184

pTNS:

https://github.com/TsuyoshiOkubo/pTNS

- It contains several tensor network codes written in python and c++
  - MPS, iTPS and TRG.
- c++ version of iTPS code was used as α version of TeNeS.
- So far, no documentation.

Topics treated in this lecture

1st: Huge data in modern physics (Today)

2nd: Information compression in modern physics
(+review of linear algebra)

3rd: Review of linear algebra (+ singular value decomposition)

4th: Singular value decomposition and low rank approximation

5th: Basics of sparse modeling

6th: Basics of Krylov subspace methods

7th: Information compression in materials science

8th: Accelerating data analysis: Application of sparse modeling

9th: Data compression: Application of Krylov subspace method

10th: Entanglement of information and matrix product states

11th: Matrix product states + Application of MPS to eigenvalue problems

12th: Application of MPS to time evolution and data science

13th: Other tensor network representations
+ (Appendix: Information compression by tensor network renormalization)

Evaluation

Evaluation will be done based on 2 reports:

1. Low rank approximation and MPS approximation (by Okubo)
2. Partial SVD (by Yamaji)

Deadline: **Jan. 27th**

- Please include your name and student id in your report.
- Please submit it through **ITC-LMS**
  If you have any troubles, please send us email:
  t-okubo@phys.s.u-tokyo.ac.jp
  YAMAJI.Youhei@nims.go.jp

Notice!

> The grade will be evaluated based on
> the sum of scores of two reports.
> (So, if you will miss one of them, it will be big disadvantage.)

Slides: The lecture materials and recordings are uploaded on

- ITC-LMS
- github (only slides and codes)
  - https://github.com/compsci-alliance/information-compression

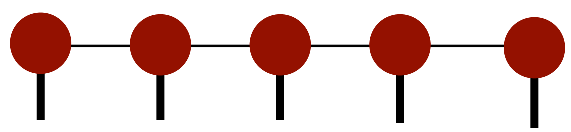
Appendixes

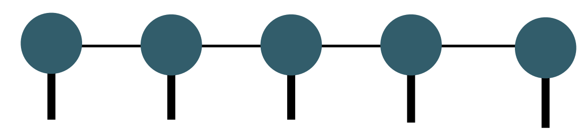
Tensor network representation of a scalar

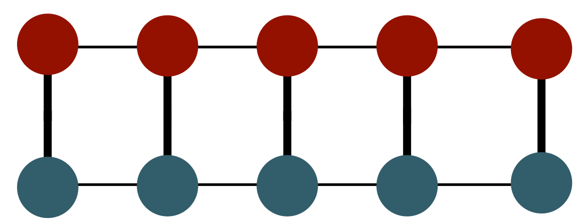
Tensor network representation of a scalar

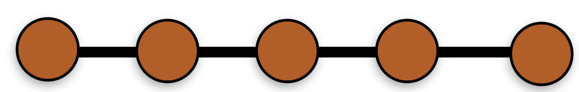
Example: inner product of two TNSs

MPS

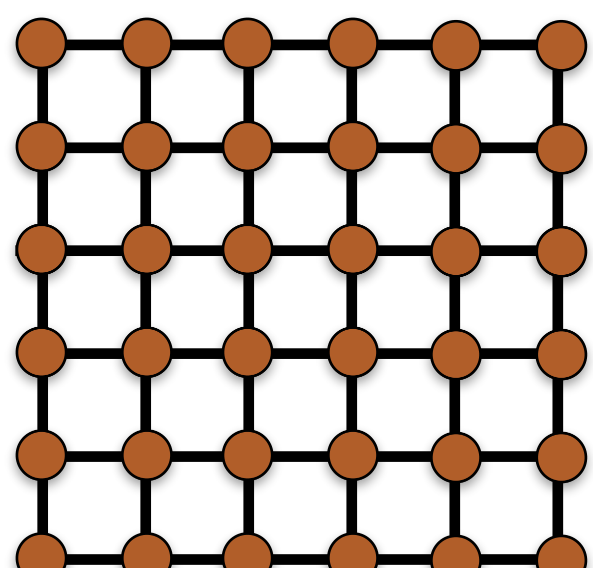
$\vec{v} =$ 

$\vec{u} =$ 

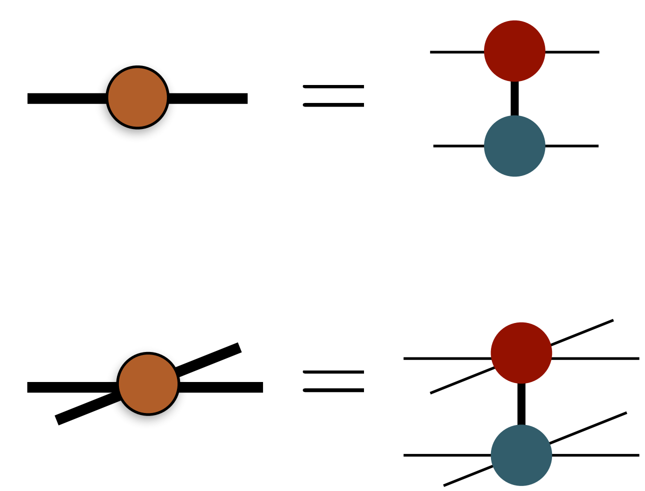
$\vec{v} \cdot \vec{u}^* =$ 

$=$ 

TPS (in two dimension)

$\vec{v} \cdot \vec{u}^* =$ 

Double layer tensor



Statistical mechanics and canonical ensemble

Canonical ensemble:
（カノニカル分布）

$$P(\Gamma) \propto e^{-\beta\mathcal{H}(\Gamma)}$$

Γ ：State (e.g. {S1, S2, ... SL})

$P(\Gamma)$: Probability to appear state Γ

$\beta = \dfrac{1}{k_B T}$: Inverse temperature

$\mathcal{H}$ ：Hamiltonian

Partition function（分配関数）

＝Normalization factor of the canonical ensemble

$$Z = \sum_{\Gamma} e^{-\beta\mathcal{H}(\Gamma)}$$

Relation to the free energy in thermodynamics

$$F = -k_B T \ln Z$$

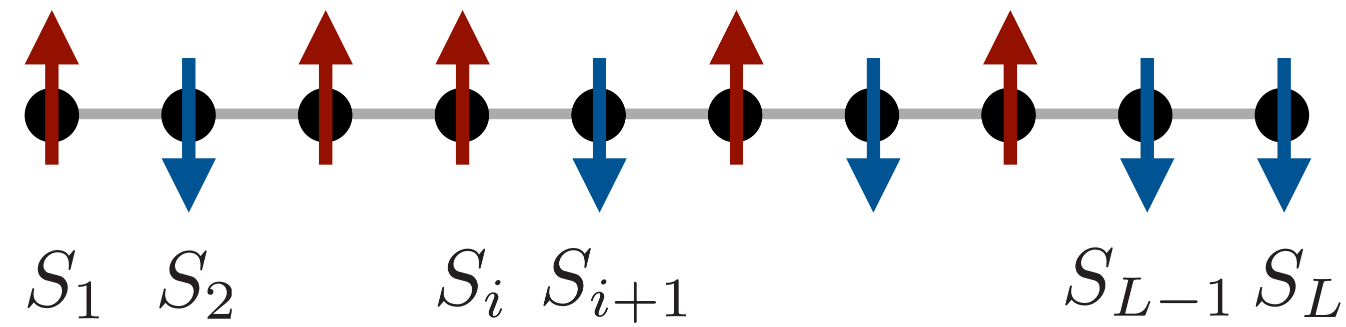
log of the partition function＝Free energy

Tensor network representation of partition function

Classical Ising model on a chain

$$\mathcal{H} = -J \sum_{i=1}^{L-1} S_i S_{i+1}$$

$$S_i = 1, -1$$



$1 = \uparrow \qquad -1 = \downarrow$

Partition function:

$$Z = \sum_{\{S_i = \pm 1\}} e^{\beta J \sum_i S_i S_{i+1}}$$

$$= \sum_{\{S_i = \pm 1\}} \prod_{i=1}^{L-1} e^{\beta J S_i S_{i+1}}$$

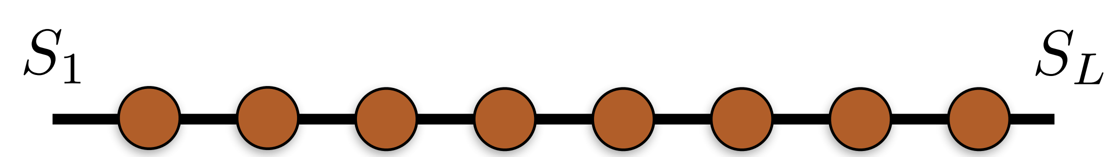
$$= \sum_{S_1 = \pm 1, S_L = \pm 1} \left(T^{L-1}\right)_{S_1, S_L}$$

Transfer matrix (転送行列)

$$T_{S_i, S_{i+1}} = e^{\beta J S_i S_{i+1}}$$

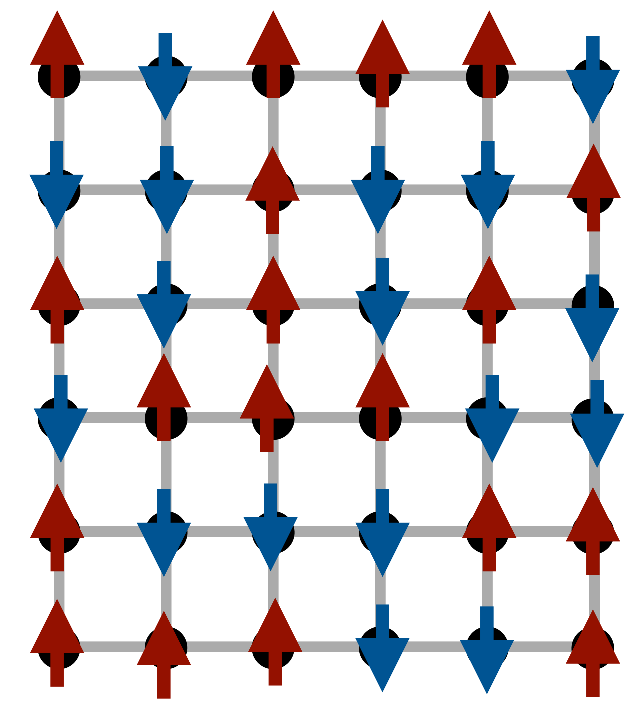
$$\begin{array}{cc} & \begin{array}{cc} +1 & -1 \end{array} \\ T = & \begin{pmatrix} e^{\beta J} & e^{-\beta J} \\ e^{-\beta J} & e^{\beta J} \end{pmatrix} \begin{array}{c} +1 \\ -1 \end{array} \end{array}$$

$$\sum_{S_1 = \pm 1, S_L = \pm 1}$$



Tensor network representation in two dimension

Classical Ising model on the square lattice



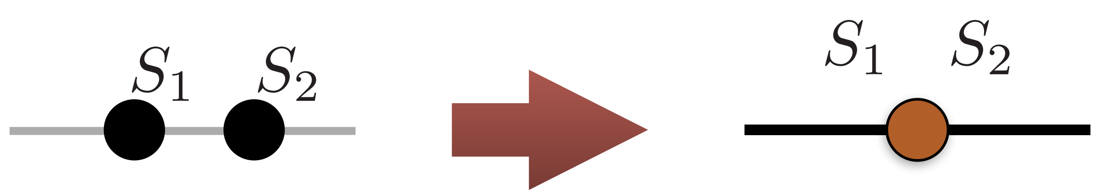
$$\mathcal{H} = -J\sum_{\langle i,j\rangle} S_i S_j \quad (S_i = \pm 1 = \uparrow, \downarrow)$$

$$Z = \sum_{\{S_i = \pm 1\}} e^{\beta J \sum_{\langle i,j\rangle} S_i S_j}$$

We can use a tensor instead of the transfer matrix.

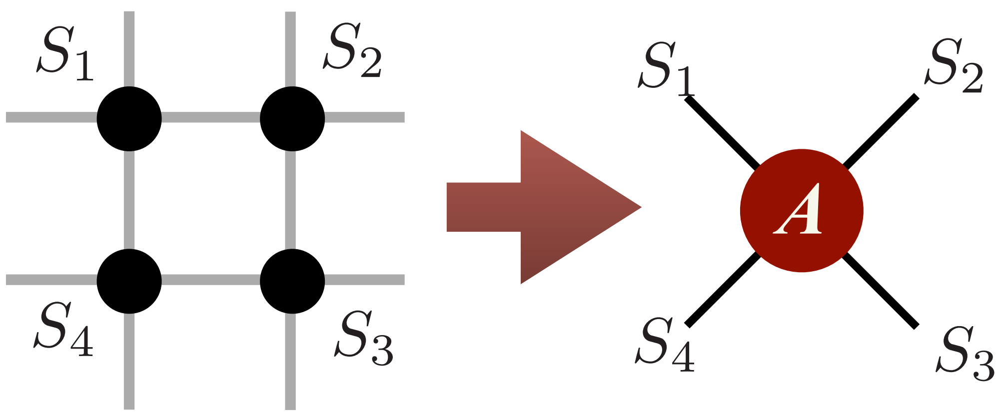
$$e^{\beta J S_1 S_2} = T_{S_1 S_2}$$

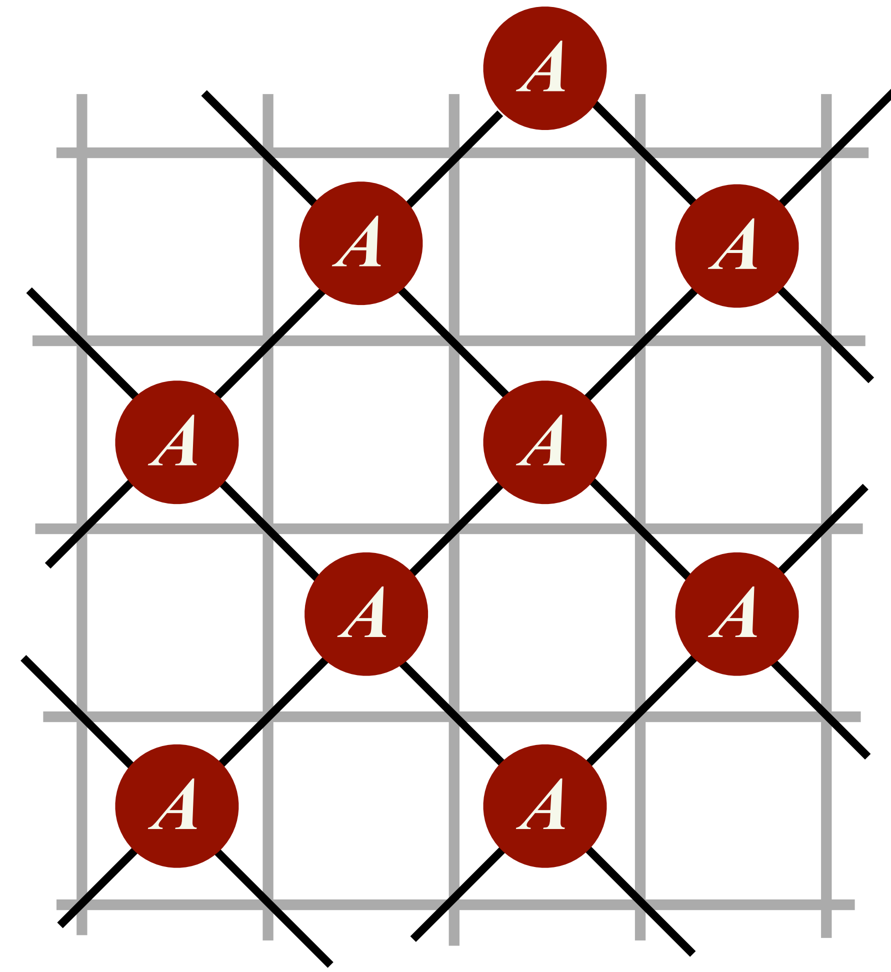
Tensor?



Tensor network representation in two dimension

$$e^{\beta J(S_1S_2+S_2S_3+S_3S_4+S_4S_1)} = A_{S_1S_2S_3S_4}$$



$Z =$ 

Partition function = Tensor network of tensor A

Square lattice Ising model→Square lattice tensor network rotating 45 degrees.

*We can construct a tensor network where tensors are on the nodes of original lattice.

Calculation cost of "classical" tensor network

Cost of tensor network contraction:

d-dimensional cubic lattice $N = L^d$

Chain: $O(ND^2)$ (Open)

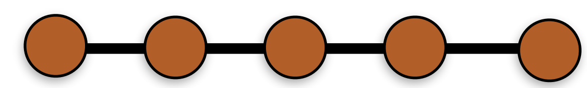
$O(ND^3)$ (Periodic)

Square: $O(D^L)$ (Open)

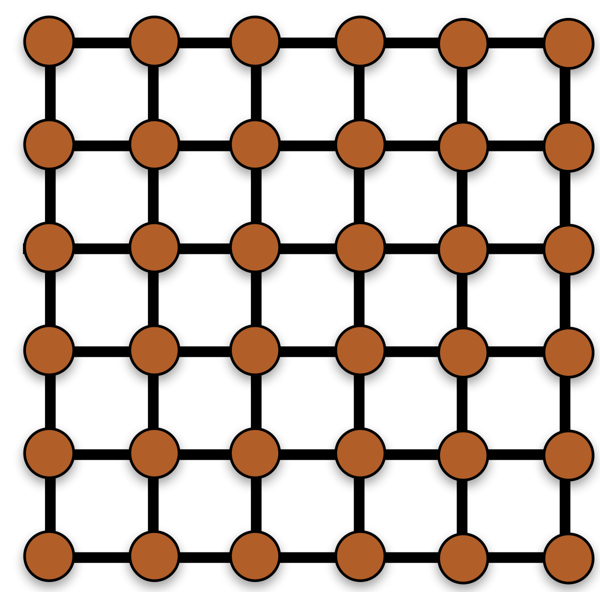
$O(D^{2L})$ (Periodic)

d-dimensional cubic: $O(D^{L^{d-1}})$

Chain



Square lattice



It is impossible to perform exact contraction.

We need efficient approximations for the contraction.

Tensor network renormalization

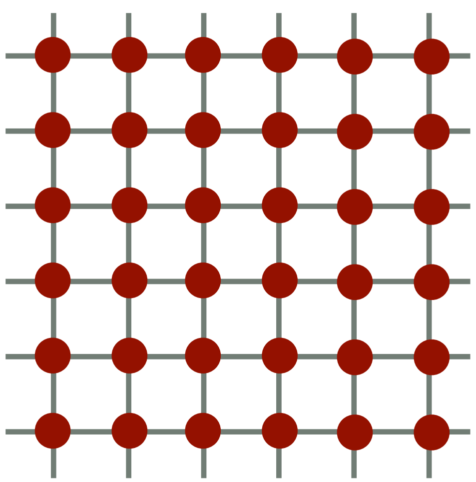
Tensor network renormalization
（テンソルネットワーク繰り込み）

- Approximate calculation of a tensor network contraction by using "coarse graining" （粗視化） of the network
    - Coarse graining $\longleftrightarrow$ Real space renormalization
    - （粗視化） $\longleftrightarrow$ （実空間繰り込み）
- It can be applicable to (basically) any lattices, and the idea (algorithm) is independent on "models" represented by tensor networks.
    - Potential application to wide range of the science.

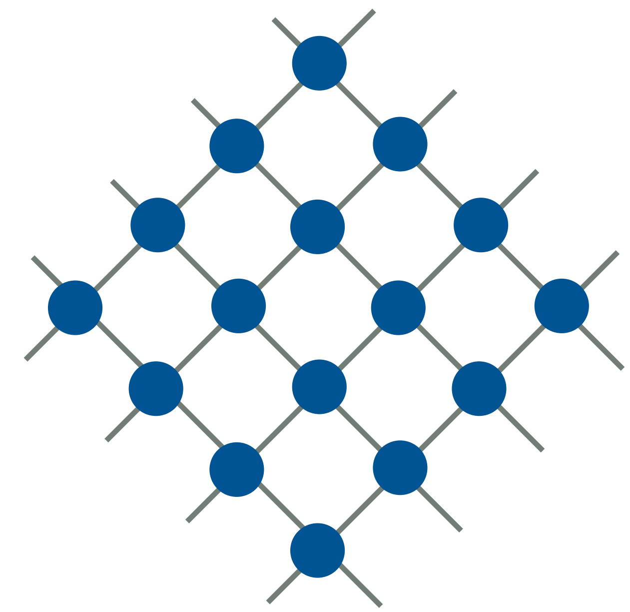
Outline of tensor network renormalization

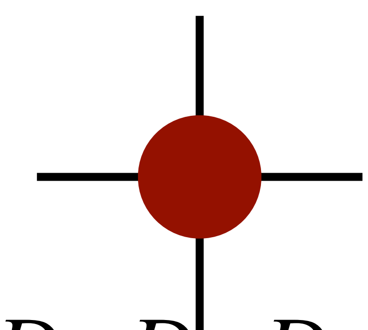
Scaler represented by $L \times L$ tensors

$(L \times L)/2$ tensors



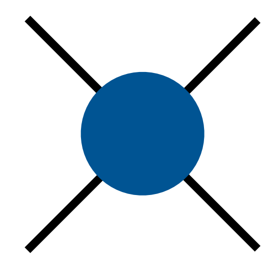
Corse graining (Renormalization) into $\sqrt{2}$ times longer scale.





A: $D \times D \times D \times D$

Approximation

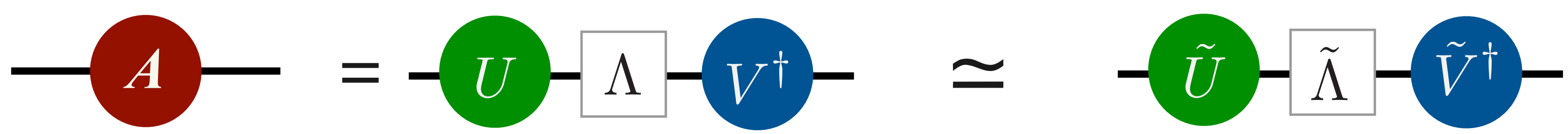


$\tilde{A} : D \times D \times D \times D$

Reduce the number of tensors keeping their size constant

Key technique: low rank approximation by SVD

$\boxed{\text{Best low-rank approximation of a matrix = SVD}}$

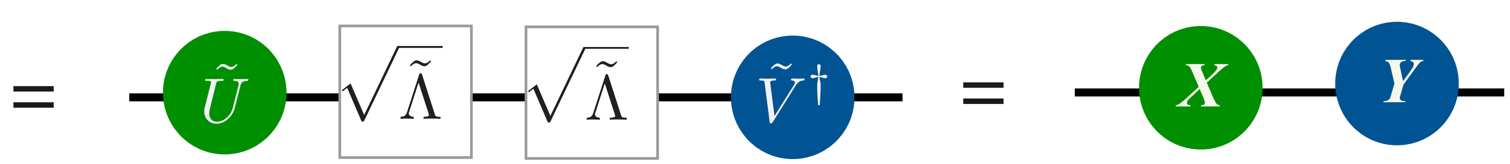


$A: M \times N$
(M ≦ N)

$\Lambda: M \times M$
(Diagonal matrix)
$U, V: (\mathrm{M,N}) \times M$

$\tilde{\Lambda}: R \times R$
(Keeping the R largest singular values)
$\tilde{U}, \tilde{V}: (\mathrm{M,N}) \times R$

In addition,



$\sqrt{\tilde{\Lambda}}$:Diagonal matrix
those elements are $\sqrt{\lambda}$

$X = \tilde{U}\sqrt{\tilde{\Lambda}}$ $\;:\mathrm{M} \times R$

$Y = \sqrt{\tilde{\Lambda}}\tilde{V}^{\dagger}$ $\;:\mathrm{R} \times M$

By SVD, we can decompose a matrix into a product of "small" matrices.

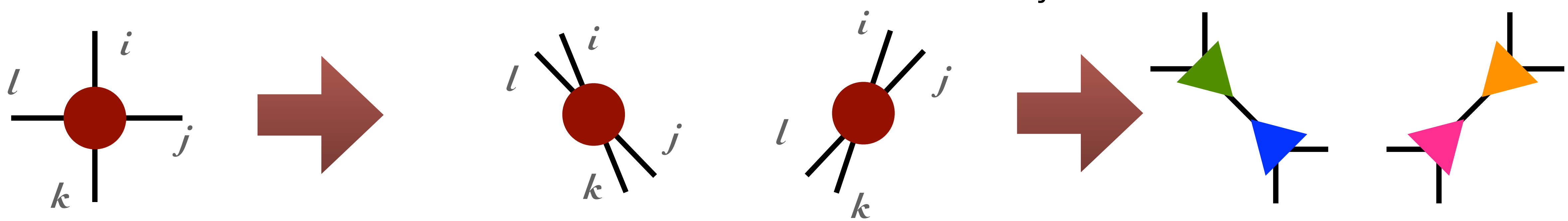
Recipe of Tensor Renormalization Group (TRG)

M. Levin and C. P. Nave, Phys. Rev. Lett. **99**, 120601 (2007)
Z.-C. Gu, M. Levin and X.-G. Wen, Phys. Rev. B **78**, 205116 (2008)

1. Decomposition

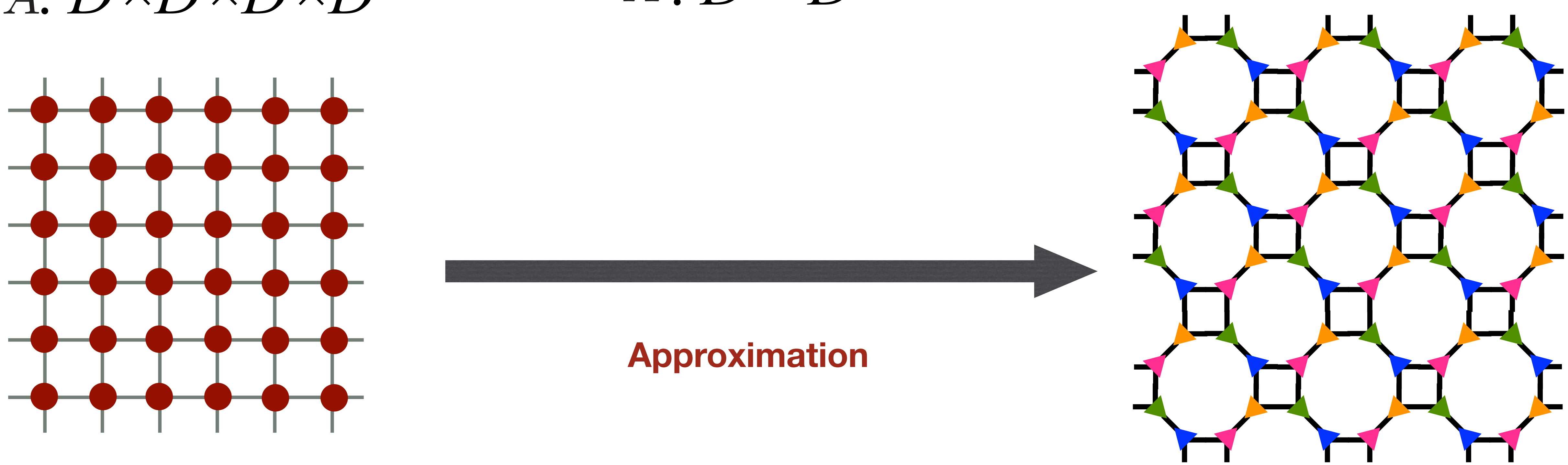
Regard a tensor as a matrix

D-rank approximation
by SVD



$A_{i,j,k,l}$ $\quad A_{(i,l),(j,k)}$ $\quad A_{(i,j),(k,l)}$

A: $D \times D \times D \times D$ $\qquad A : D^2 \times D^2$



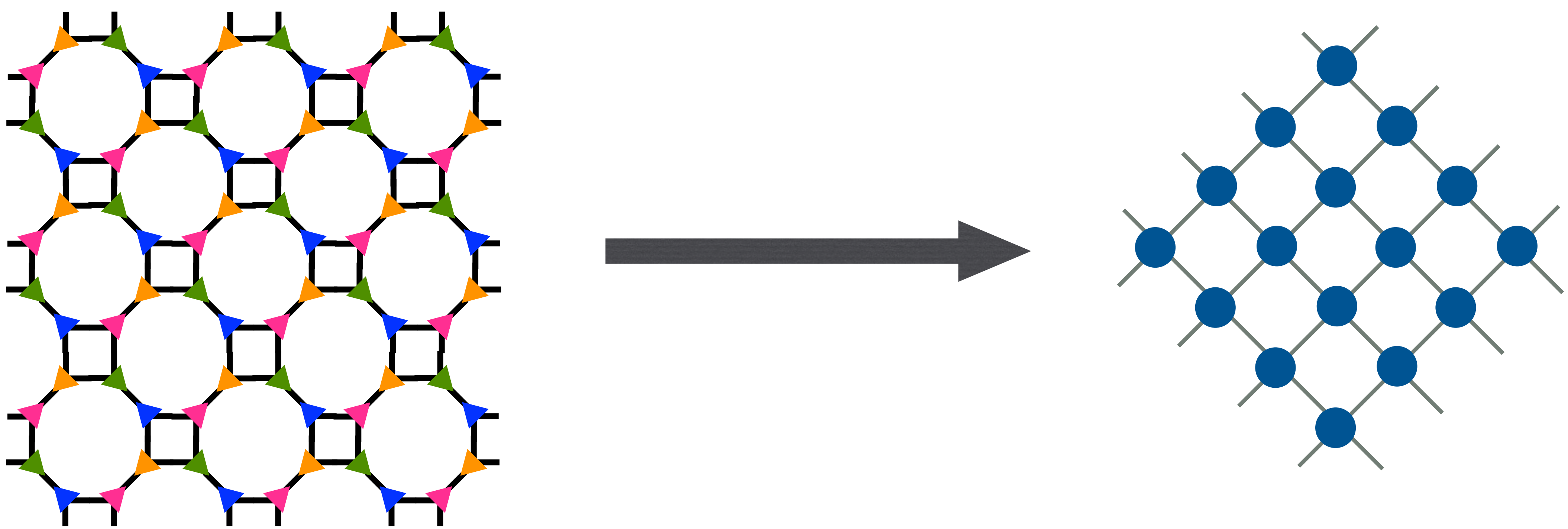
Recipe of Tensor Renormalization Group (TRG)

M. Levin and C. P. Nave, Phys. Rev. Lett. **99**, 120601 (2007)
Z.-C. Gu, M. Levin and X.-G. Wen, Phys. Rev. B **78**, 205116 (2008)

2. Coarse graining

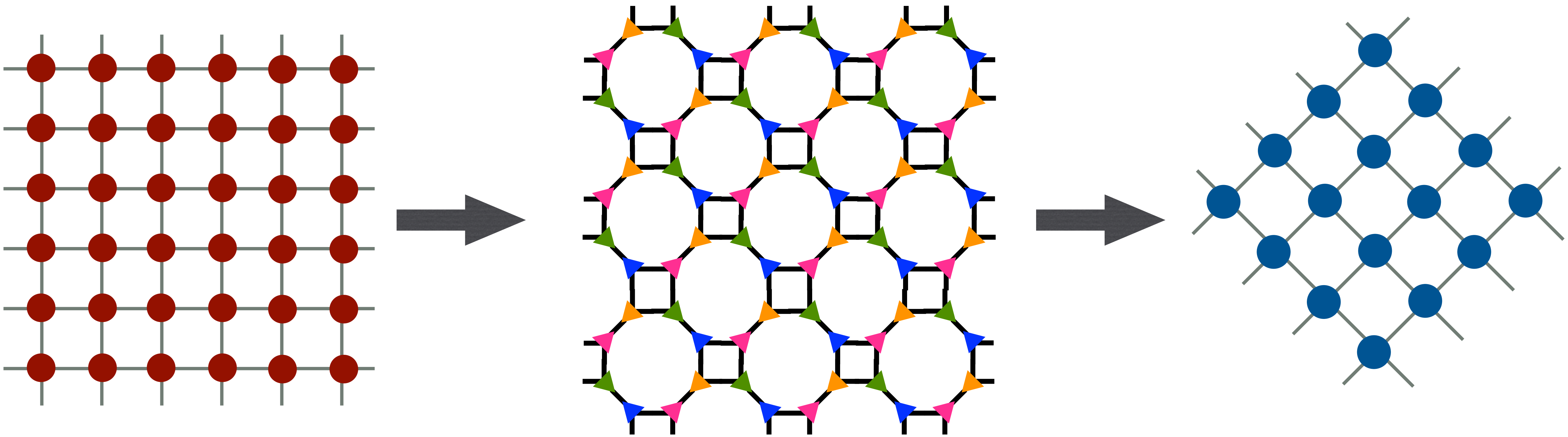
$$\tilde{A} : D \times D \times D \times D$$

In total, two original tensors are coarse grained into a new tensor.



Recipe of Tensor Renormalization Group (TRG)

M. Levin and C. P. Nave, Phys. Rev. Lett. **99**, 120601 (2007)
Z.-C. Gu, M. Levin and X.-G. Wen, Phys. Rev. B **78**, 205116 (2008)



Calculation cost: SVD= $O(D^6)$

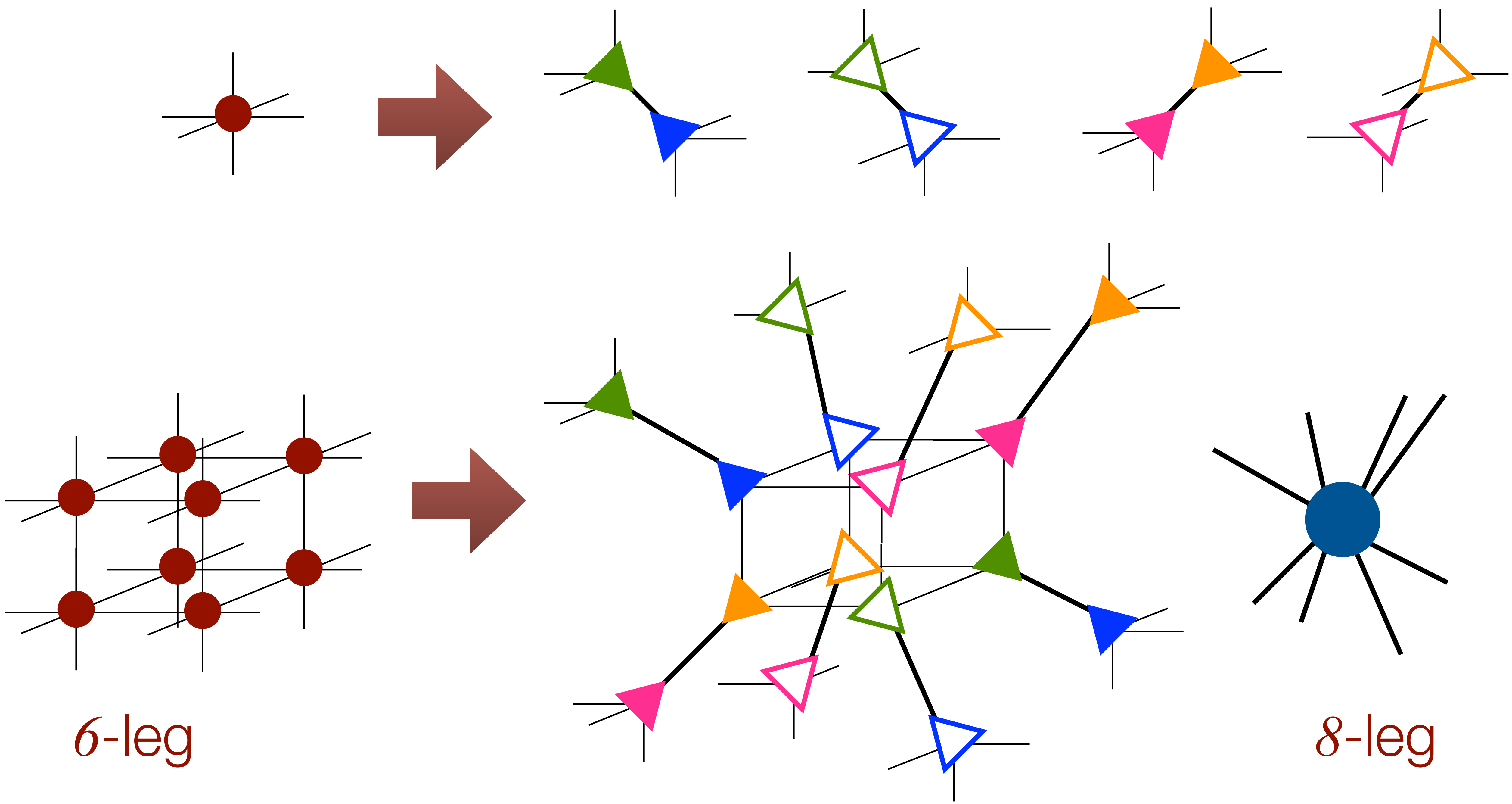
Contraction= $O(D^6)$ (per tensor)

*By one TRG step, # of tensors is reduced by 1/2.

We can calculate the contraction in polynomial cost!

Tensor renormalization group for higher dimensions

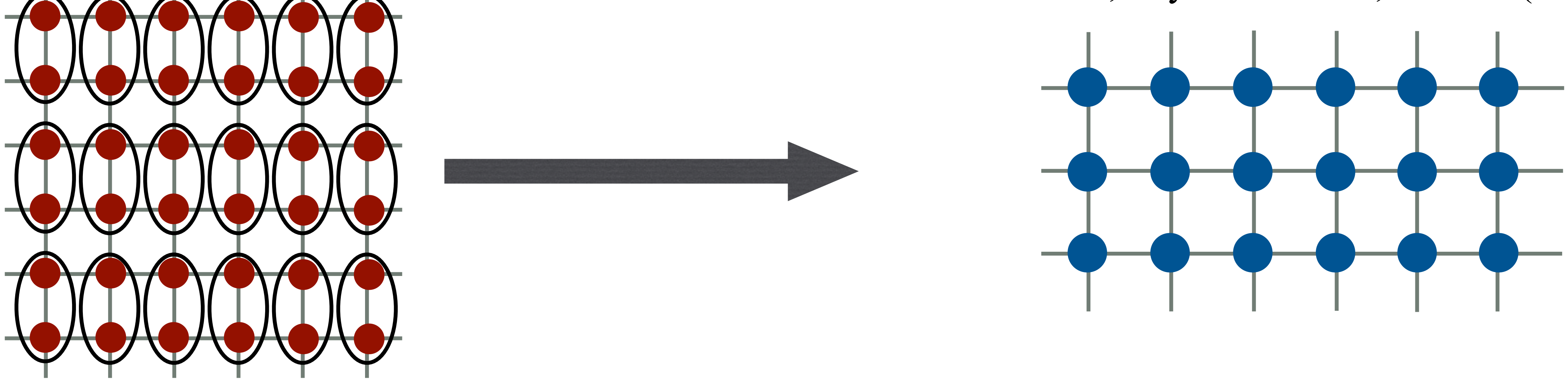
Simple generalization of TRG to cubic lattice (three dimension)



Tensor renormalization group by using HOSVD

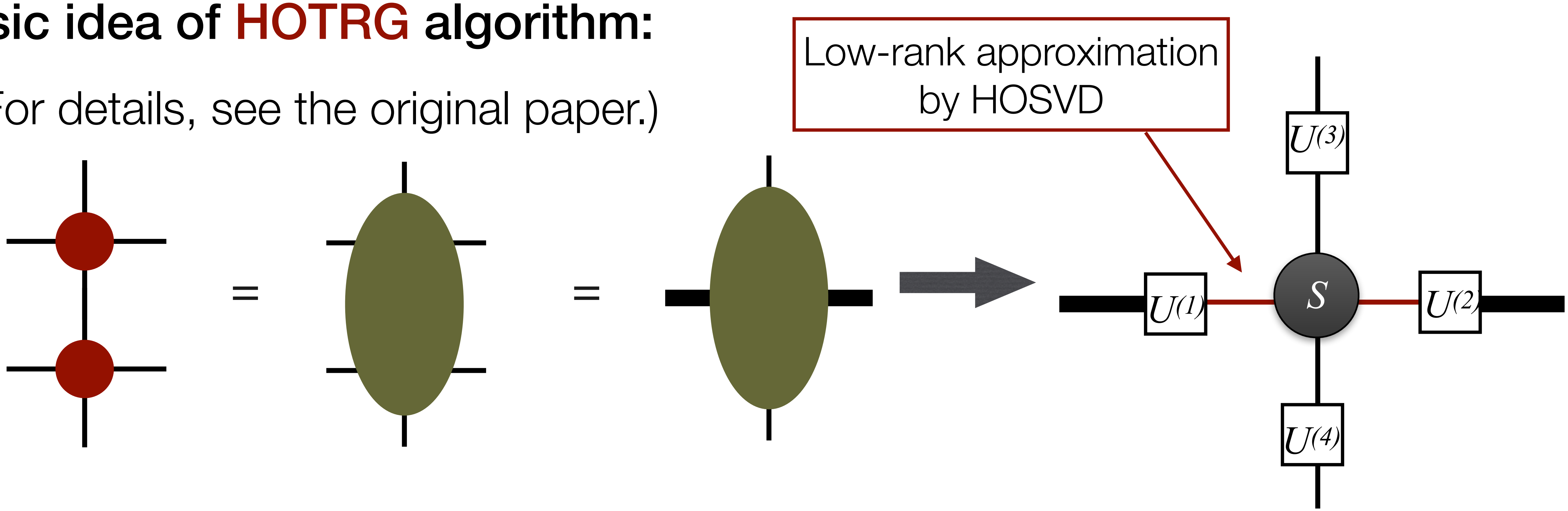
Anisotropic coarse graining by using HOSVD instead of SVD

Z. Y. Xie *et a*l, Phys. Rev. B **86**, 045139 (2012)



Basic idea of HOTRG algorithm:

(For details, see the original paper.)



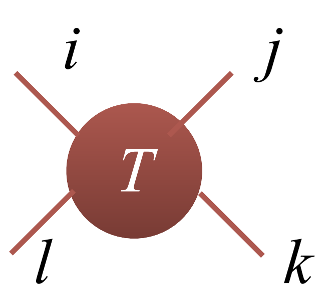
(Reference of HOSVD)

Tucker decomposition: generalization of SVD

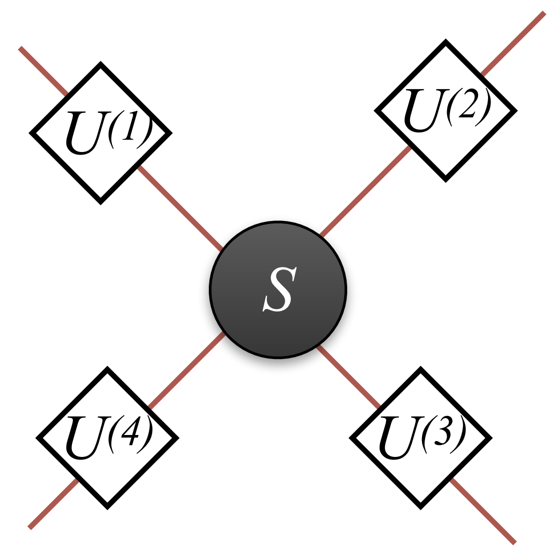
Review: T. G. Kolda et al, SIAM Review **51**, 455 (2009)

Tucker decomposition:

(Tucker (1963))



$=$



$U^{(i)}$: Factor matrix (usually unitary)

S :Core tensor

$$T_{ijkl} = \sum_{i'=1}^{I}\sum_{j'=1}^{J}\sum_{k'=1}^{K}\sum_{l'=1}^{L} S_{i'j'k'l'} U^{(1)}_{ii'} U^{(2)}_{jj'} U^{(3)}_{kk'} U^{(4)}_{ll'}$$

*If S is "diagonal", Tucker decomposition becomes CP decomposition.

Low "rank" approximation

$$T_{ijkl} = \sum_{i'=1}^{I'}\sum_{j'=1}^{J'}\sum_{k'=1}^{K'}\sum_{l'=1}^{L'} \tilde{S}_{i'j'k'l'} \tilde{U}^{(1)}_{ii'} \tilde{U}^{(2)}_{jj'} \tilde{U}^{(3)}_{kk'} \tilde{U}^{(4)}_{ll'}$$

$I' < I, \quad J' < J, \quad K' < K, \quad L' < L$

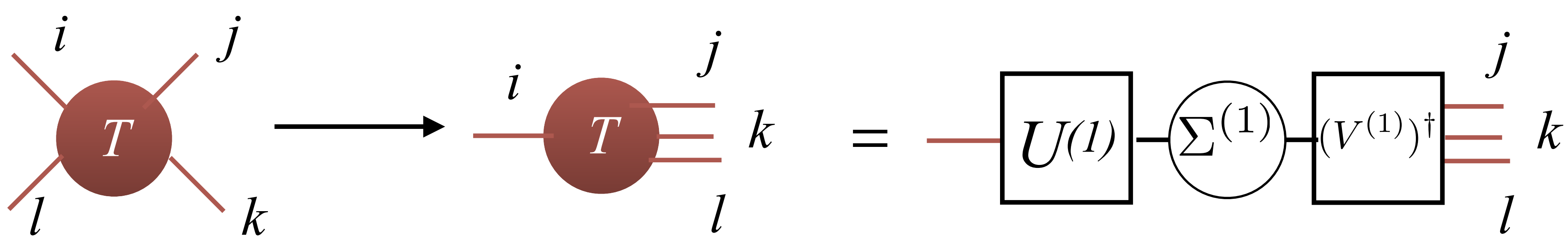
rank-(I',J', K', L') approximation

(Reference of HOSVD)

Higher order SVD (HOSVD)

L. De Lathauwer et al, SIAM J. Matrix Anal. & Appl., **21**, 1253 (2000)

Define a factor matrix from matrix SVD:



Core tensor is calculated as

$$S_{i'j'k'l'} \equiv \sum_{ijkl} T_{ijkl} (U^{(1)})^{\dagger}_{i'i} (U^{(2)})^{\dagger}_{j'j} (U^{(3)})^{\dagger}_{k'k} (U^{(4)})^{\dagger}_{l'l}$$

Properties of the core tensor

$$S^{*}_{:,i_n=\alpha,:,:} \cdot S_{:,i_n=\beta,:,:} = \begin{cases} 0 & (\alpha \neq \beta) \\ (\sigma^{(n)}_{\alpha})^2 & (\alpha = \beta) \end{cases}$$

Dot product

$$A \cdot B \equiv \sum_{i,j,k,l} A_{ijkl} B_{ijkl}$$

Generalization of the diagonal matrix Σ in matrix SVD.

* Low-rank approximation based on HOSVD is not optimal.

Power of the HOTRG

Advantage:

- HOTRG does not change the network structure.
  - We can easily generalize it to higher dimensions.
- Low-rank approximation is based on the cluster of two tensors.
  - At the approximation, we take into account more information.
  - More efficient than TRG where SVD is done for a single tensor.

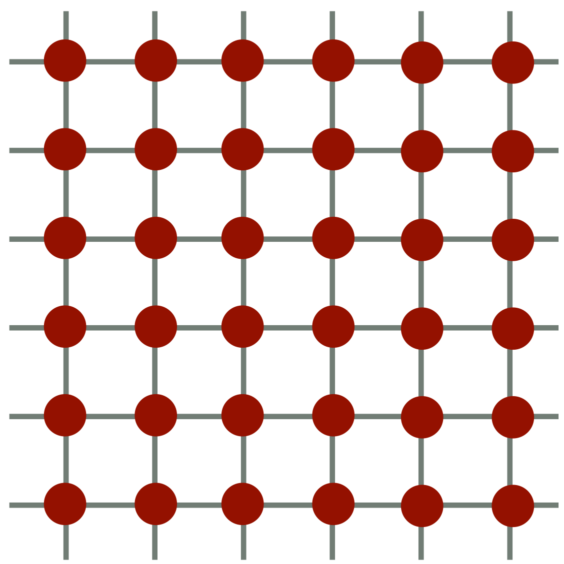
Disadvantage:

- HOTRG needs higher cost than TRG.
  - $\mathrm{O}(D^7)$ in HOTRG $\longleftrightarrow$ $\mathrm{O}(D^6)$ in TRG

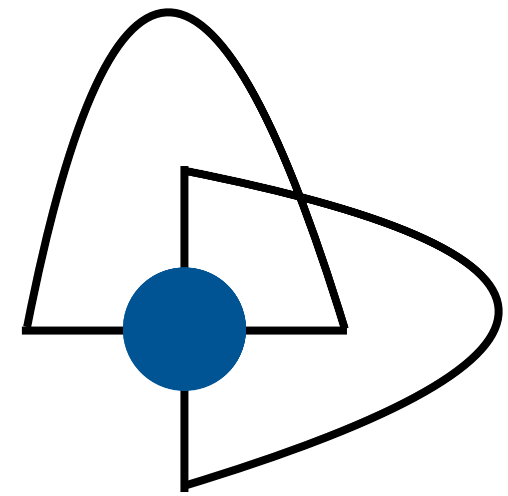
Application to a classical partition function

Partition function

(Periodic boundary condition)

$Z =$ 

Repeat TRG step until only a few tensors remain.



We can easily calculate physical quantities from Z.

Free energy: $F = -k_B T \ln Z$

Energy: $E = -\dfrac{\partial \ln Z}{\partial \beta}$ (Use difference approximation)

Specific heat: $C = \dfrac{1}{k_B T^2} \dfrac{\partial^2 \ln Z}{\partial \beta^2}$ (Use difference approximation)

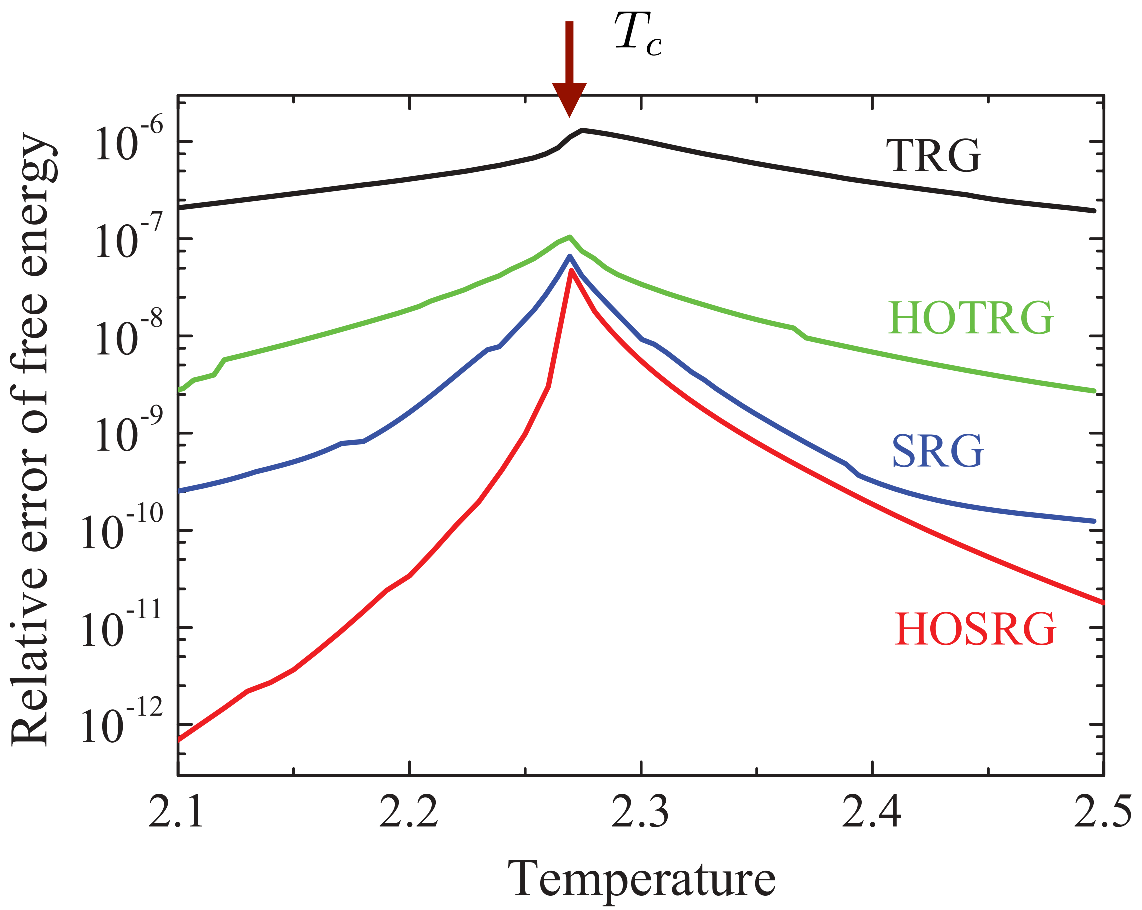
Example of calculation

Ising model in infinite size

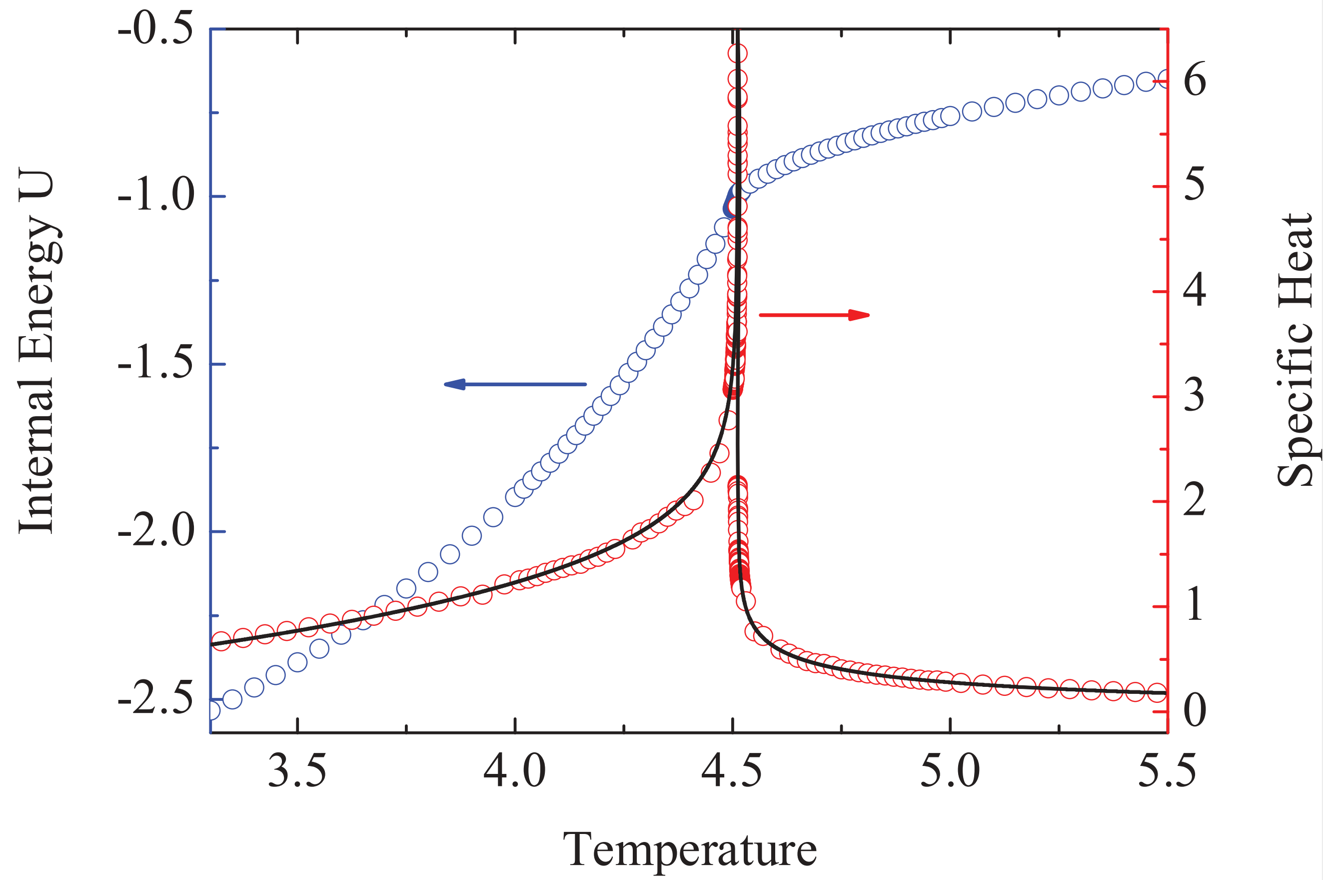
$$\mathcal{H} = -J \sum_{\langle i,j \rangle} S_i S_j$$

Z. Y. Xie *et a*l, Phys. Rev. B **86**, 045139 (2012)

Error of free energy for 2D Ising model



Energy and specific heat of 3D Ising model



$$T_c/J = \frac{2}{\ln(1+\sqrt{2})} \simeq 2.269$$

Interesting topics in tensor network renormalization

- Try to find efficient algorithm to remove "short range" entanglement
  - TNR, Loop-TNR, GILT, Gauge fixing

TNR: G. Evenbly and G. Vidal, Phys. Rev. Lett. **115**, 180405 (2015)

Loop-TNR: S. Yang, Z.-C. Gu and , X.-G. Wen, Phys. Rev. Lett. **118**, 110504 (2017)

GILT: M. Hauru, C. Delcamp. S. Mizera Phys. Rev. B **97**, 045111 (2018)

Gauge fixing: G. Evenbly, Phys. Rev. B **98**, 085155 (2018)

- Application to lattice QCD
  - TRG with Grassmann algebra

Z.-C. Gu, F. Verstraete, and X.-G. Wen, arXiv:1004.2563

S. Takeda, and Y. Yoshimura PTEP **2015**, 043B1 (2015).

- Property at the criticality
  - Relation between TNR and MERA
  - Relation to Conformal invariance

G. Evenbly and G. Vidal,Phys. Rev. Lett. **115**, 200401 (2015)

G. Evenbly, Phys. Rev. B **95**, 045117 (2017)

Tensor network renormalization at the critical point
— When the accuracy of TRG becomes worse?
(Will be skipped)

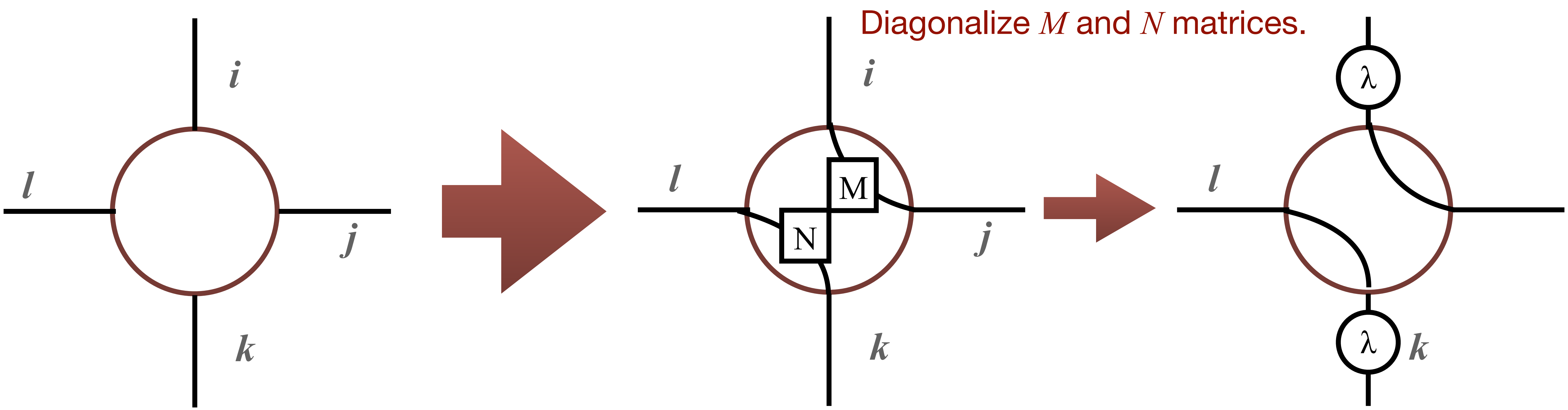
Correlation (entanglement) within a tensor

General tensor

Eg. *Correlation* in (i,j) and (k,j)

$$A_{ijkl} = M_{ij} N_{kl} \Rightarrow A_{ijkl} = \lambda_i^{(M)} \lambda_k^{(N)} \delta_{ij} \delta_{kl}$$

$$A_{ijkl}$$



New rule for representation of the correlation:

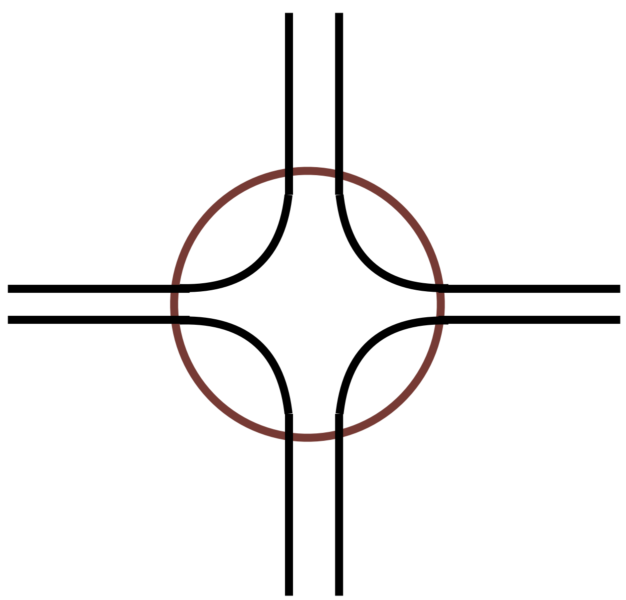
$$i \;\text{——}\; j = \delta_{ij}$$

(+ we neglect eigenvalues in the graph.)

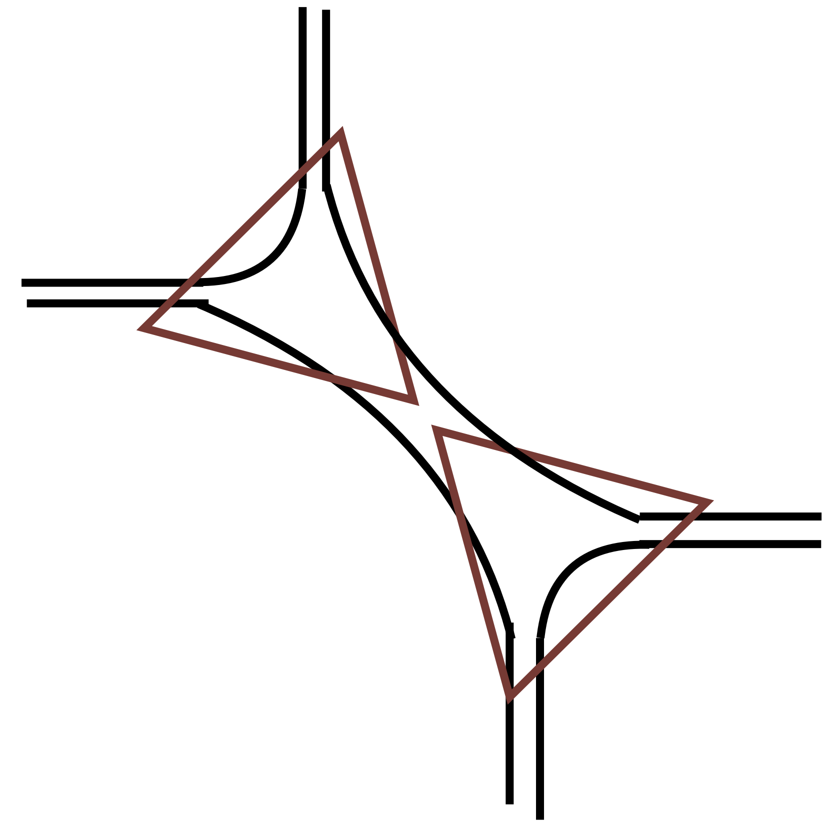
Fixed point of TRG: Corner Double Line tensor (固定点)

Corner Double Line (CDL) tensor:

There are correlations among the nearest legs.



SVD



Degree of freedoms connecting two tensors.

Two lines = D

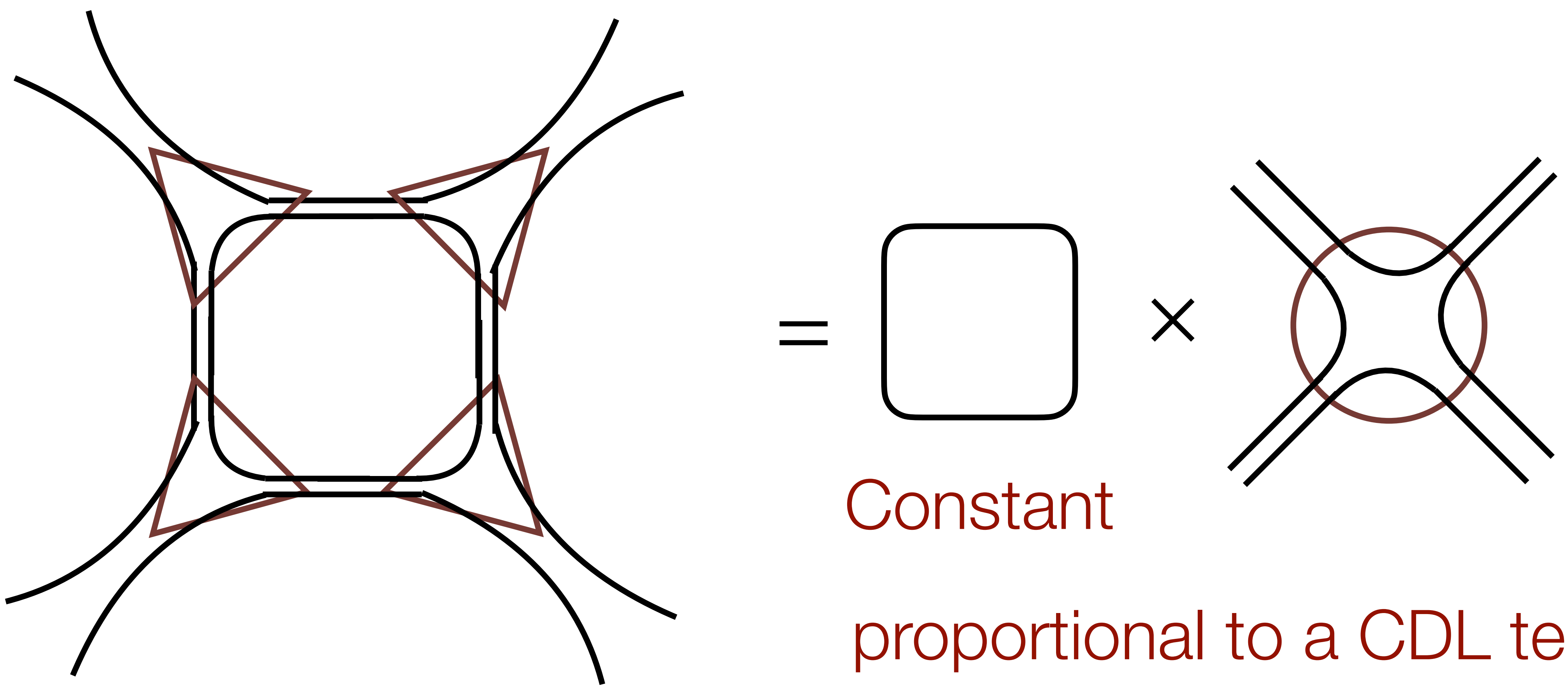
Original bond dimension = D

→ Single line: bond dimension $\sqrt{D}$

→ No truncation error at SVD (Original rank = D)

Fixed point of TRG: Corner Double Line tensor

Contraction of four tensors in TRG:



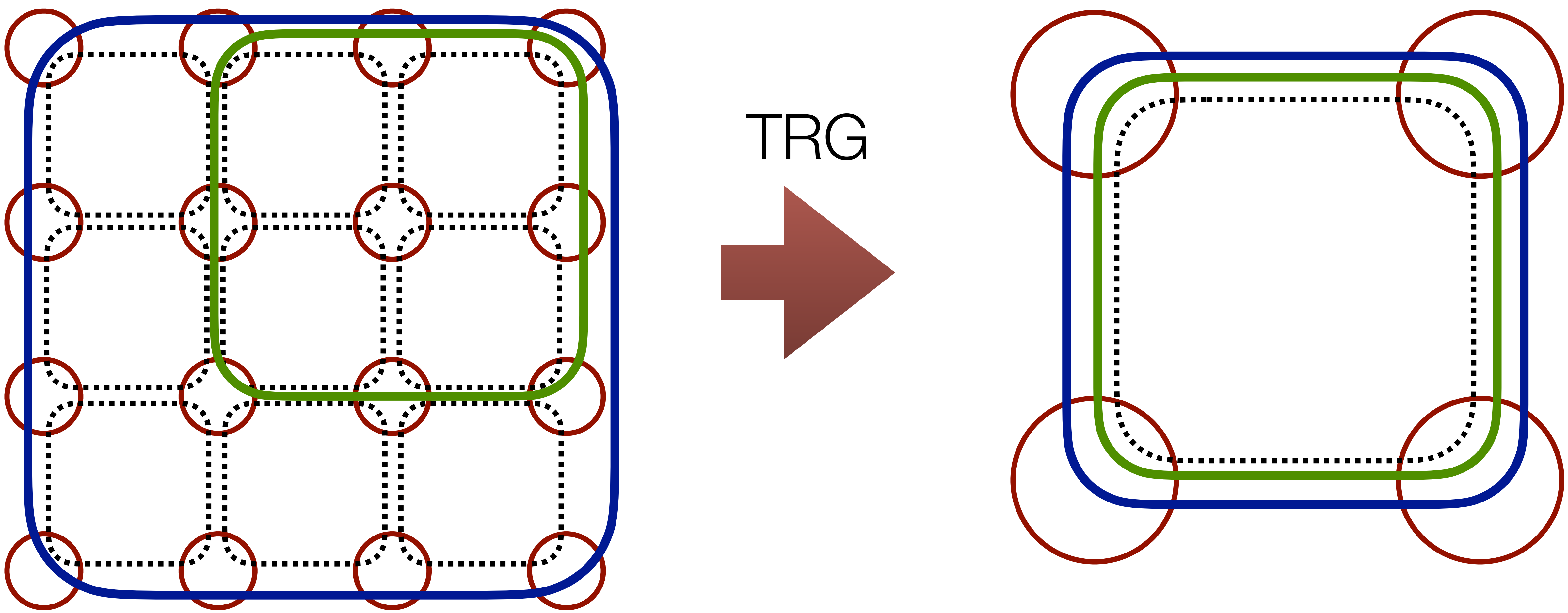
proportional to a CDL tensor!

CDL tensor is a fixed point of TRG (and also HOTRG).

CDL tensor remains as CDL tensor along TRG.

Problems in TRG: accumulation of correlations

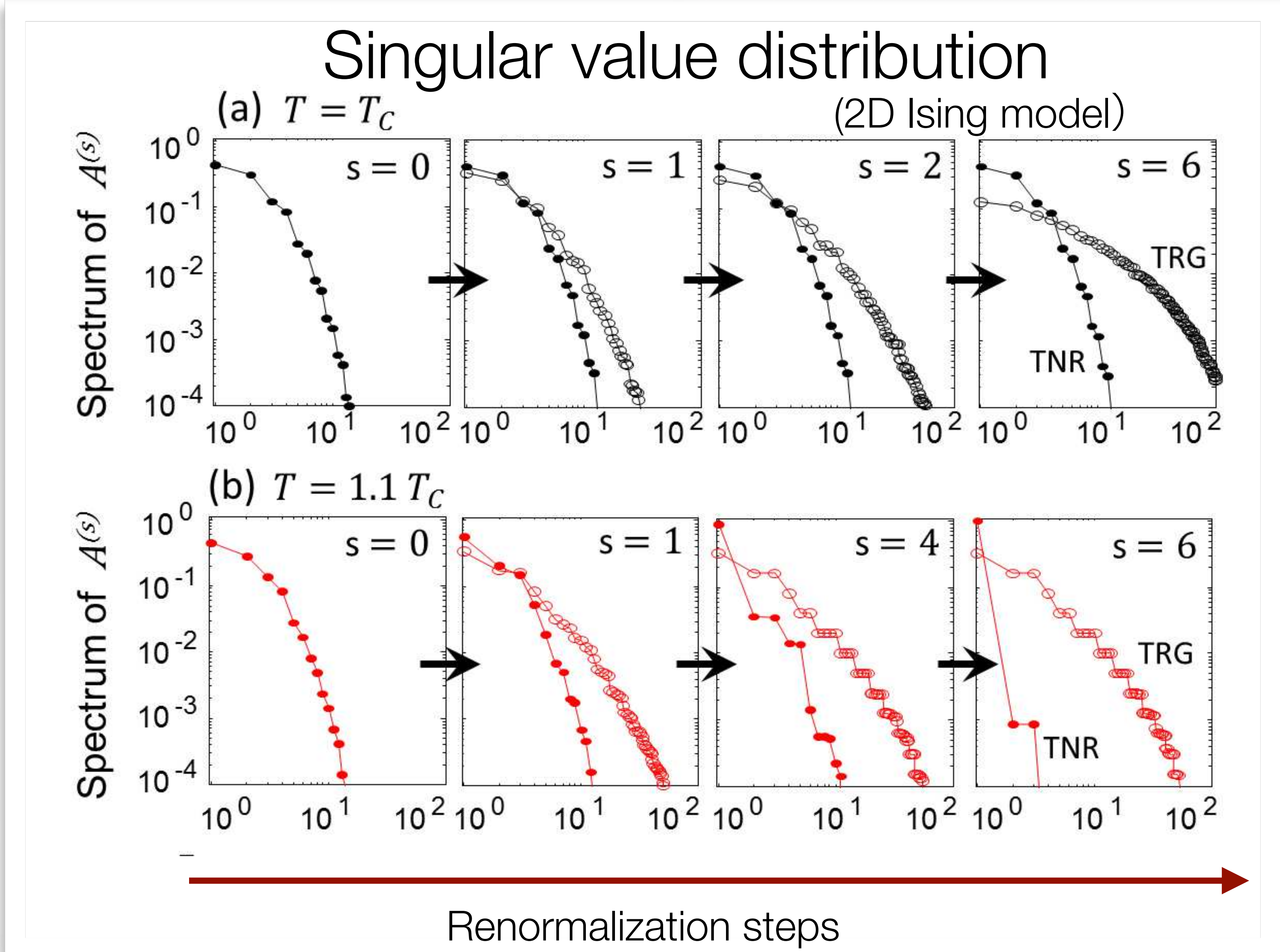
Correlation in several scales | **Correlations remains after TRG.**



Ideal renormalization: Correlation in shorter scales should be removed. Only the correlation in the present scale exists.

TRG : Correlations in all scales remain.

Problem in TRG: increase of truncation error



G. Evenbly and G. Vidal Phys. Rev. Lett. **115**, 180405 (2015)

In TRG, the width of the singular value distribution increases along renormalization.

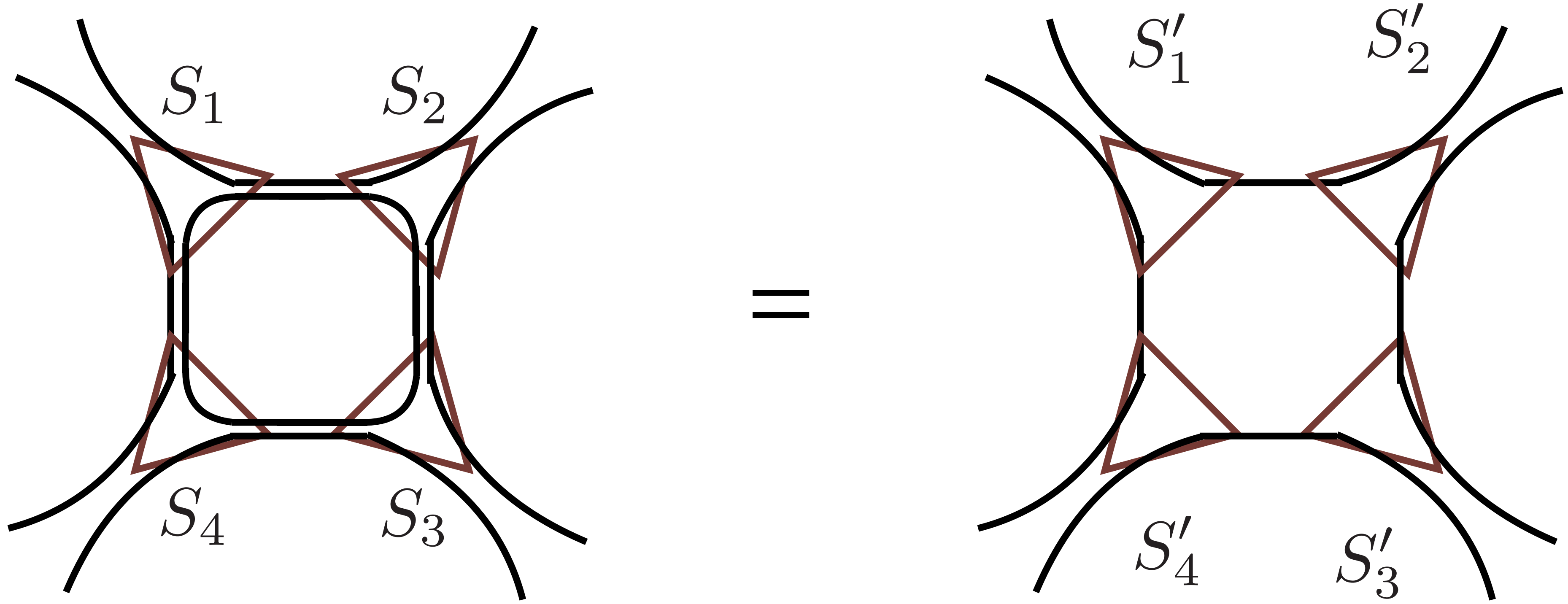
Increase of truncation error (decrease of accuracy)

Improvement of TRG：Entanglement Filtering (optional)

Try to remove CDL structure at renormalization steps.

Z.-C. Gu and X.-G Wen, Phys. Rev. B 80, 155131 (2009)

Idea:



S: $D \times D \times D$

S': $D \times D' \times D'$

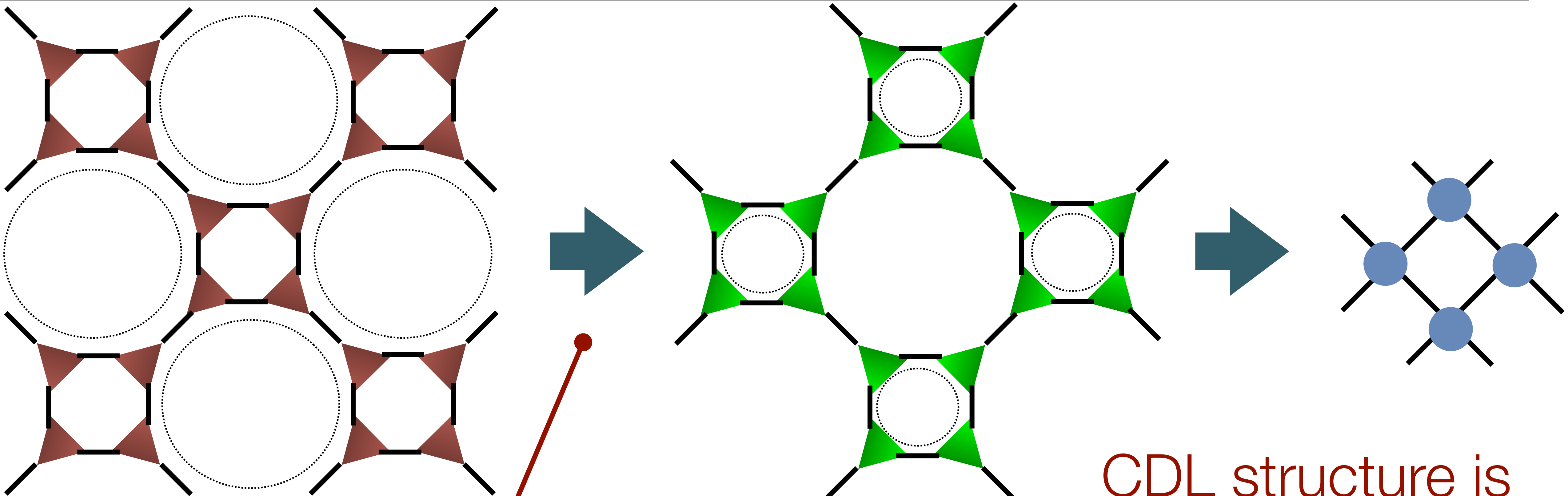
S' $= \square^{1/4}$ **S**

$D' \sim \sqrt{D}$

Insert this "approximation" into the TRG algorithm.

Tensor Entanglement Filtering Renormalization (optional)

Z.-C. Gu and X.-G Wen, Phys. Rev. B 80, 155131 (2009)

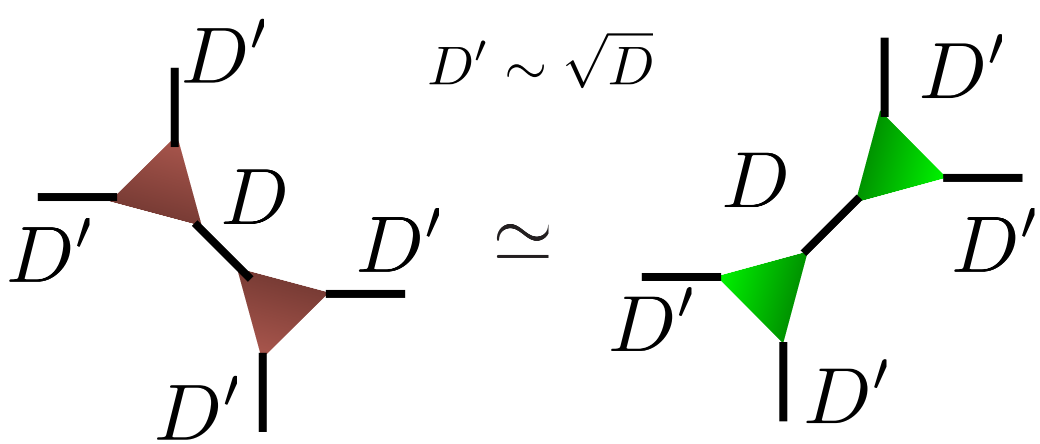


S' (after approximation)

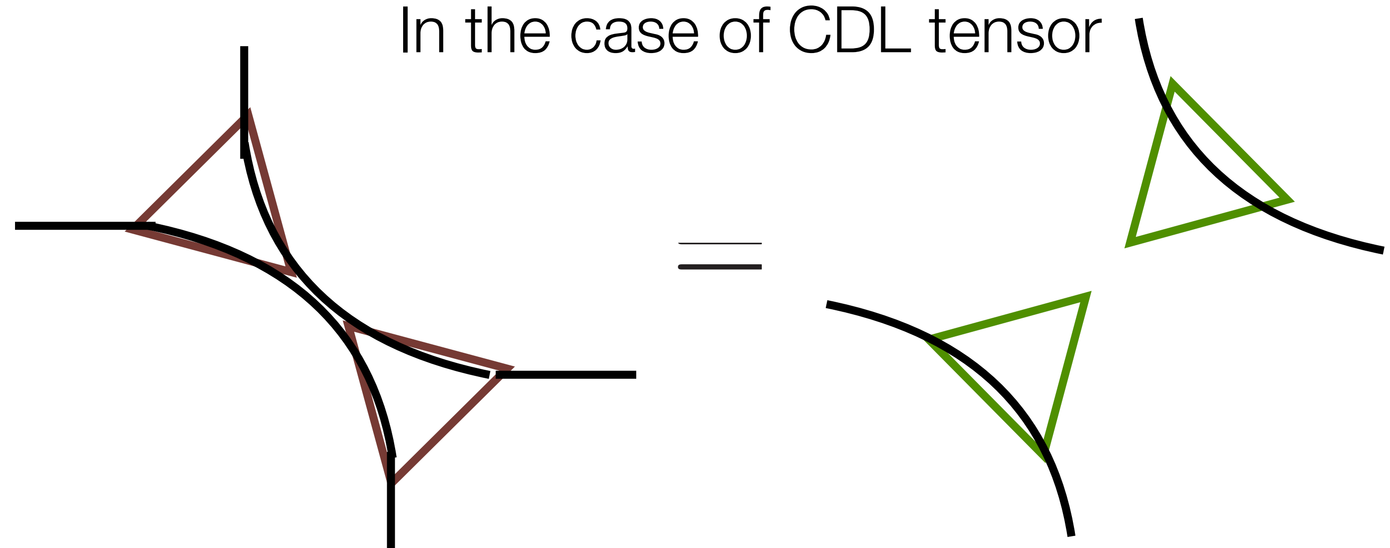
CDL structure is perfectly removed!

Change of SVD: $D' < D$

$D' \sim \sqrt{D}$



In the case of CDL tensor

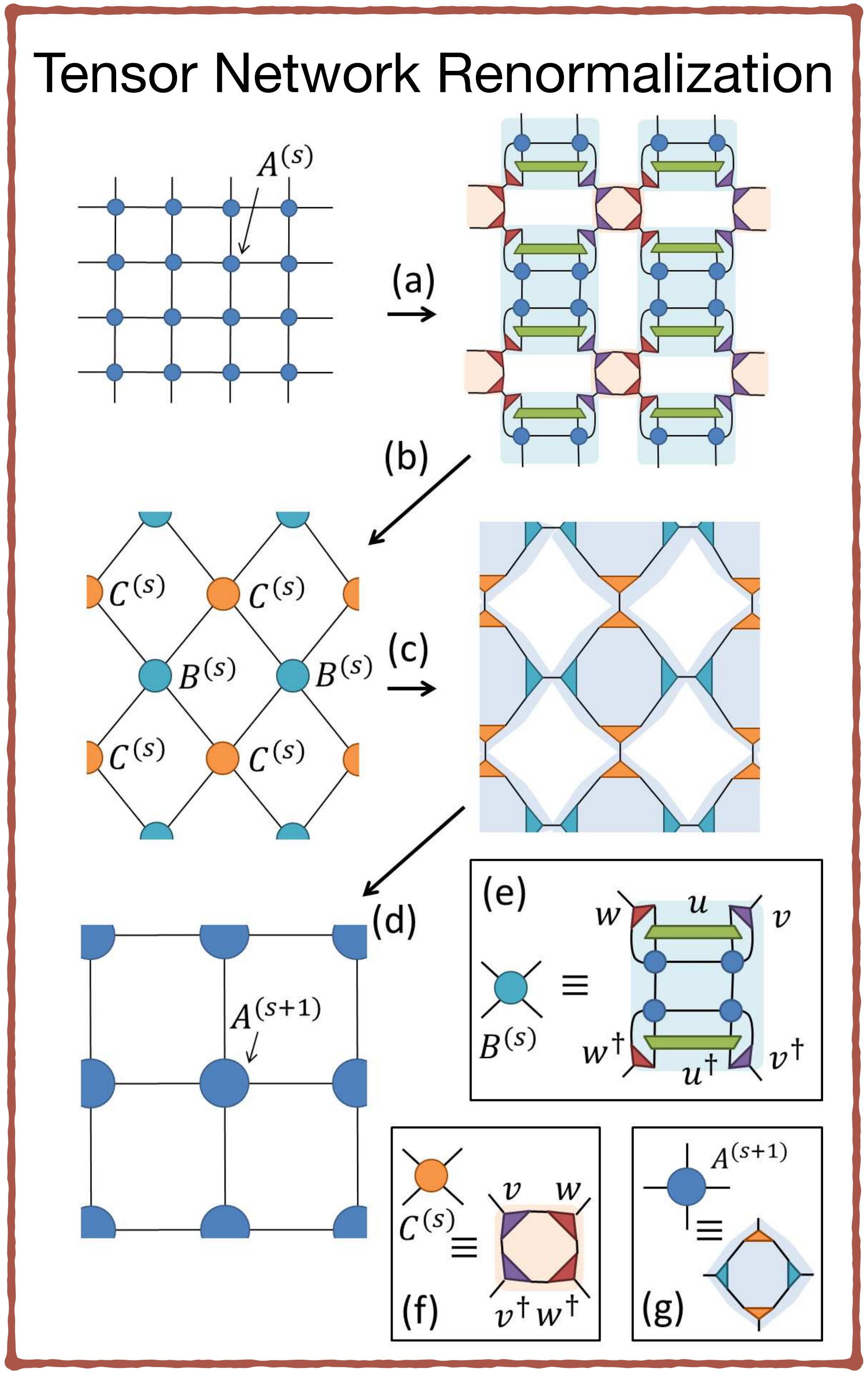


Remaining problem in TEFR (optional)

- TEFR works well far from the critical point.
  - Because it can remove CDL structure.
- In the vicinity of the critical point, the accuracy is still poor.
  - Because the actual entanglement is not necessarily perfect CDL structure.
- In order to improve further, we need to consider the entanglement structure beyond CDL tensor.

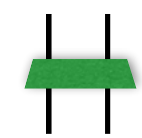
Recent progress: Tensor Network Renormalization

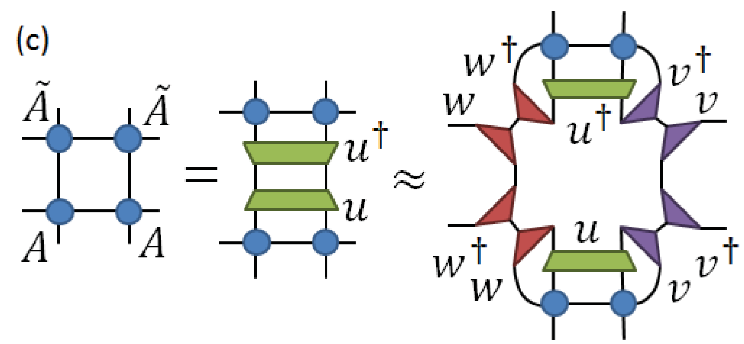
G. Evenbly and G. Vidal, Phys. Rev. Lett. **115**, 180405(2015).



Point of TNR

Use of a disentangler(Unitary tensor)

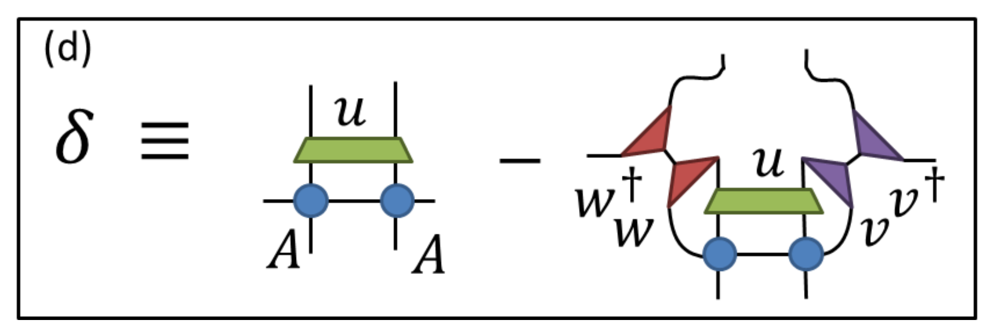




It can remove short range entanglement efficiently.

(Not only CDL)

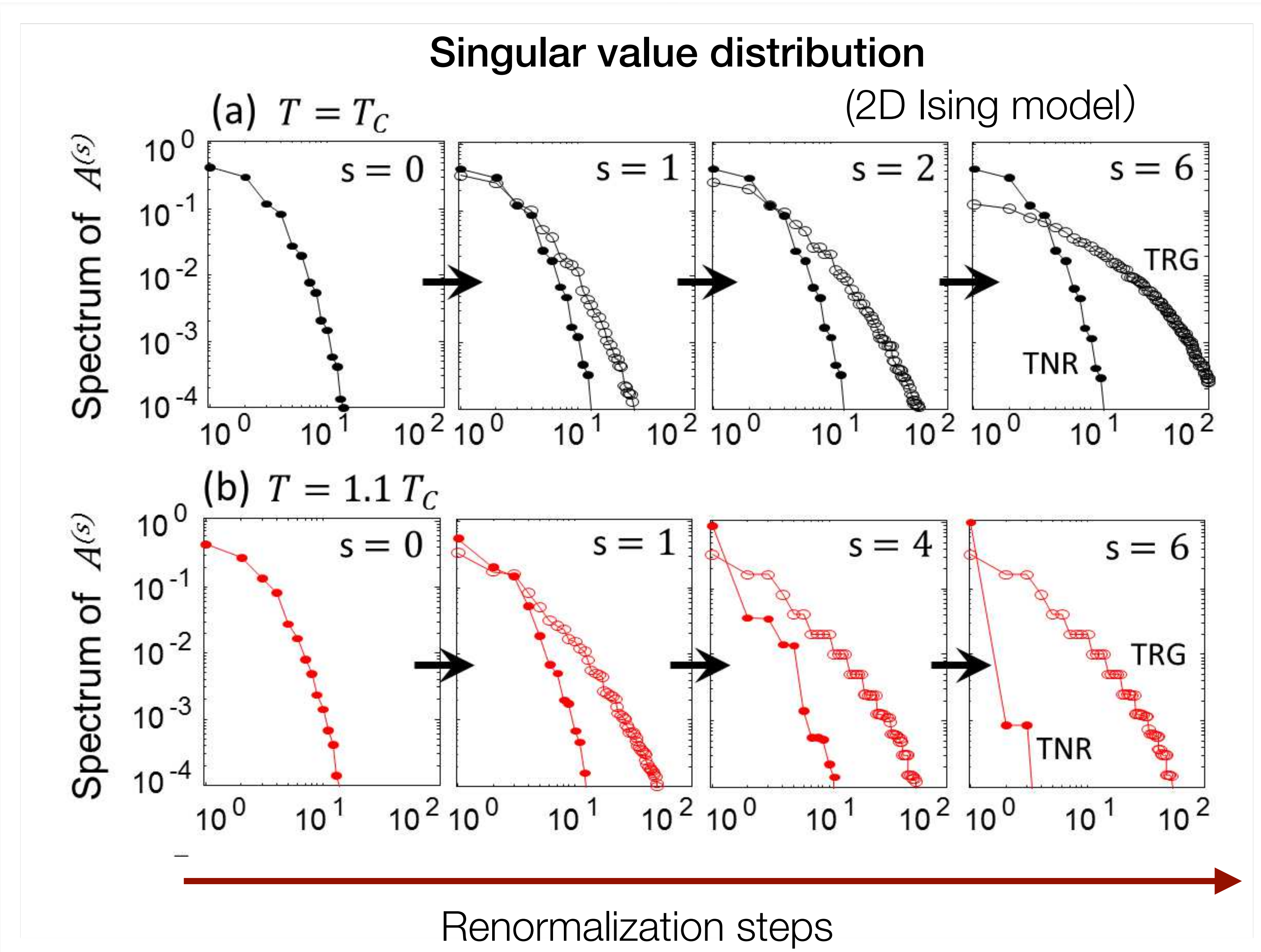
Approximation by using two-tensor cluster:

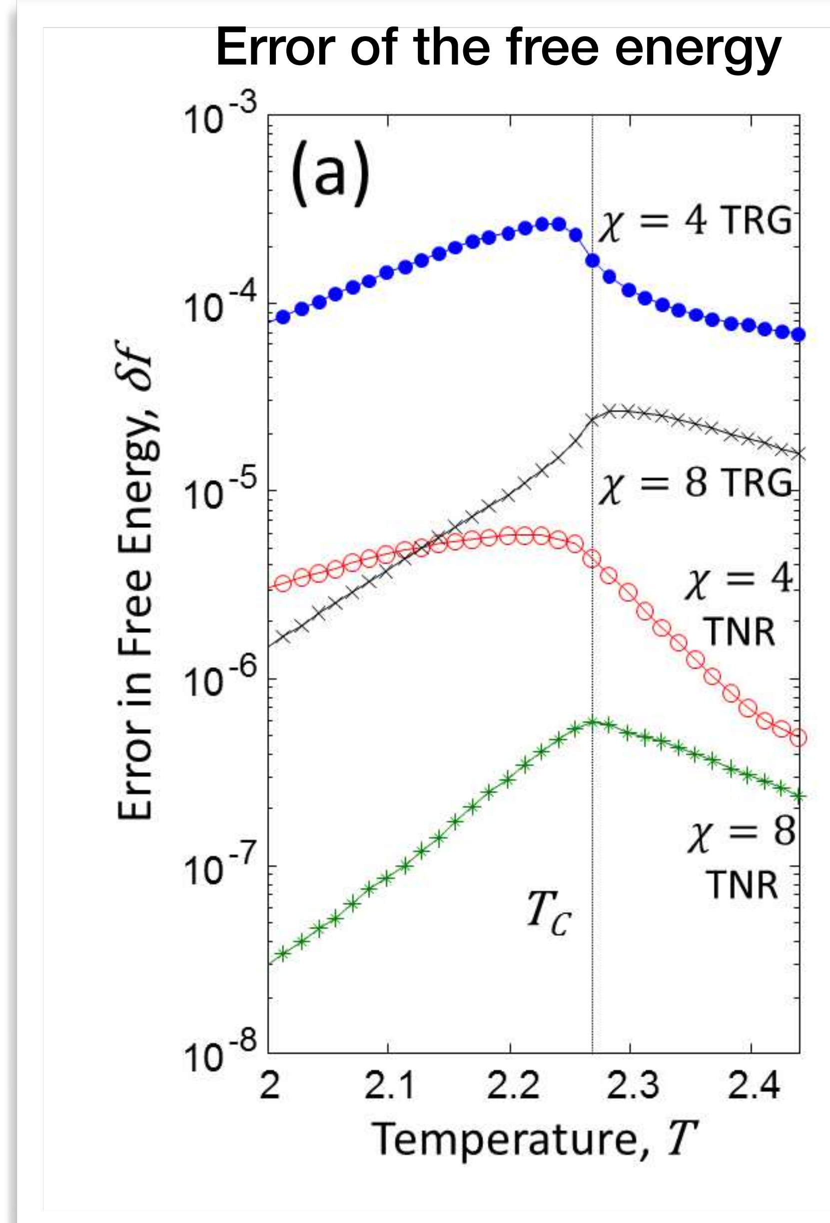


Better accuracy than the simple SVD of single tensor

Power of TNR

G. Evenbly and G. Vidal,Phys. Rev. Lett. **115**, 180405 (2015)
,arXiv: 1412.0732v2 (free energy).





In TNR:
- The singular value distribution is narrower than that of TRG.
- It is almost unchanged at Tc.
  - Indicating scale invariance of the critical system.