ETH zürich

Information Theory, Machine Learning and the Real-Space Renormalization Group

Maciej Koch-Janusz

האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

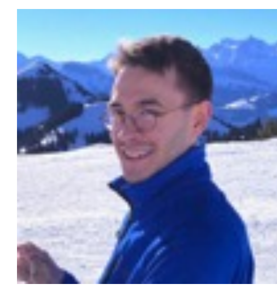
Zohar Ringel

``Mutual Information, Neural Networks and the Renormalization Group''
MKJ and Zohar Ringel, *Nature Physics* ***14****, 578-582 (2018)*

ETH zürich

Sebastian D. Huber

Patrick Lenggenhager



Doruk Efe Gokmen

``Optimal Renormalization Group Transformation from Information Theory''
P. Lenngenhager, Z. Ringel, S.D. Huber and MKJ, *arXiv:1809.09632*

Outline

Outline

- Machine learning in condensed matter
- Information-theoretic approach to real-space RG:
  - The Real Space Mutual Information algorithm
  - Optimality of the RSMI algorithm
  - Disordered systems
- Conclusions and Outlook

Why machine learning?

ETH zürich

Why machine learning?

- Machine learning has a lot of practical applications already

Why machine learning?

- Machine learning has a lot of practical applications already

Why machine learning?

- Machine learning has a lot of practical applications already

Why machine learning?

- Machine learning has a lot of practical applications already

Why machine learning?

- Machine learning has a lot of practical applications already

- Can science benefit in a more fundamental fashion?

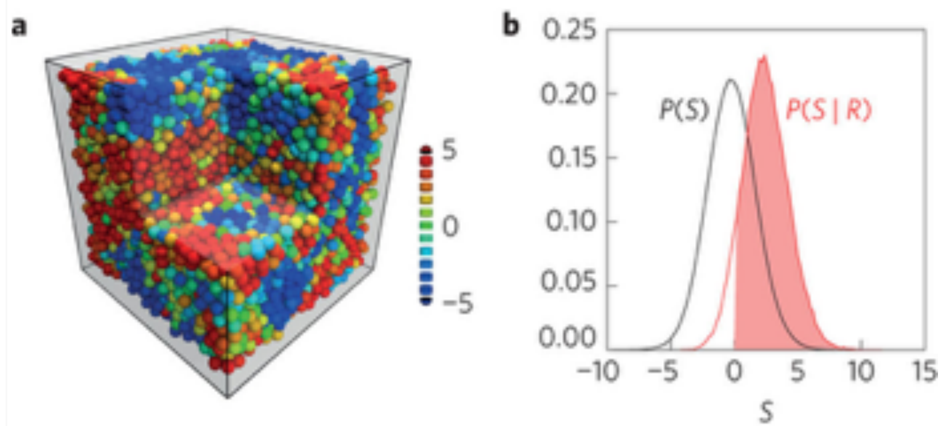
Phase transitions and classification

Phase transitions and classification

Lei Wang,
Phys. Rev. B **94,** 195105 *(2016)*

J. Carrasquilla and R. Melko
Nature Physics **13,** *431–434 (2017)*

E.P. van Nieuwenburg, Y. Liu, S. Huber
Nature Physics **13,** *435–439 (2017)*



*Schoenholz et al., **Nature** 12, 469–471 (2016)*

ETH zürich

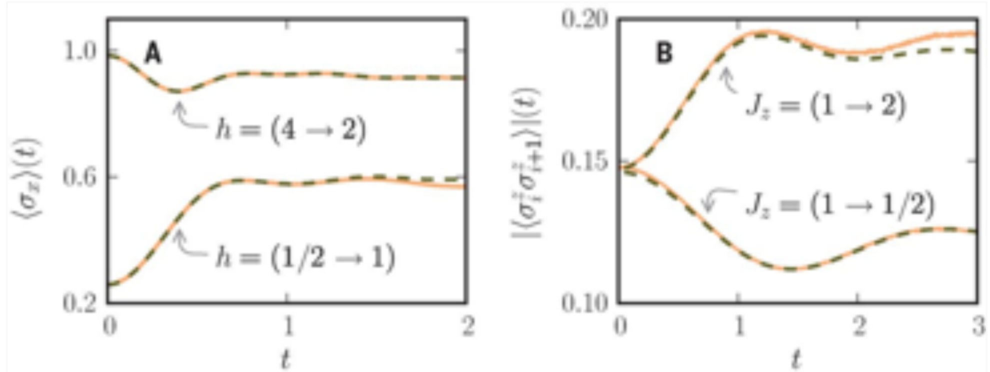
Phase transitions and classification

Phase transitions and classification

State compression and representation

Phase transitions and classification

State compression and representation



G. Carleo, M. Troyer, *Science* **355**, 602–606 (2017)

G. Torlai et al., *Nature Physics* **314**, 447–450 (2018

J. Carrasquilla et al., *arXiv:1810.10584*, (2018)

Phase transitions and classification

State compression and representation

Phase transitions and classification

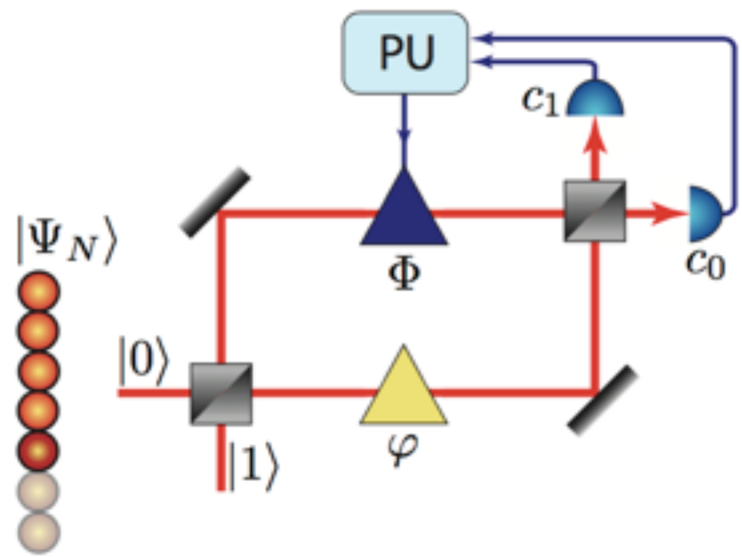
State compression and representation

Experimental / numerical protocols / error correction

Phase transitions and classification

State compression and representation

Experimental / numerical protocols / error correction



A. Hentschel, B.Sanders, *PRL* **104**, 063603 (2010)

J. Wang et al., *Nature Physics* **13**, 551–555 (2017)

T. Foesel et al., *Phys. Rev. X* **8**, 031084 (2018)

R. Sweke et al., *arXiv:* 1810.07207 (2018)

Phase transitions and classification

State compression and representation

Experimental / numerical protocols / error correction

Phase transitions and classification

State compression and representation

Experimental / numerical protocols / error correction

physics -> ML

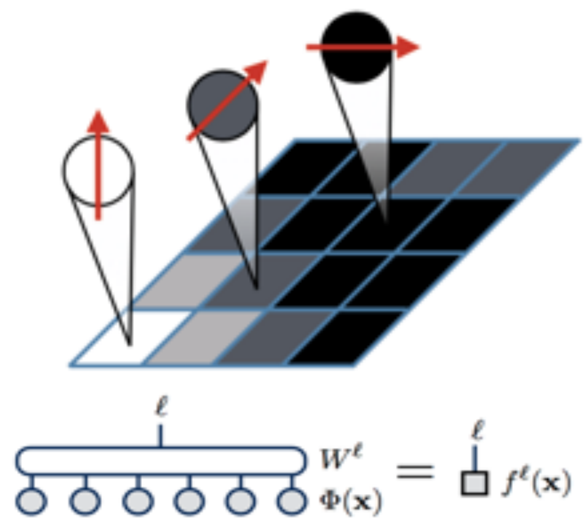
ETH zürich

Phase transitions and classification

State compression and representation

Experimental / numerical protocols / error correction

physics -> ML



M. Stoudenmire, D. Schwab,
Advances in Neural Information Processing Systems 29, 4799 (2016)

T. Chao, P. Mazaheri, B. Sun, N, Weingartner, Z. Nussinov,
arXiv: 1708.05715 (2017)

Works of M. Mezard, G. Biroli, L. Zdeborova,….

Machine Learning

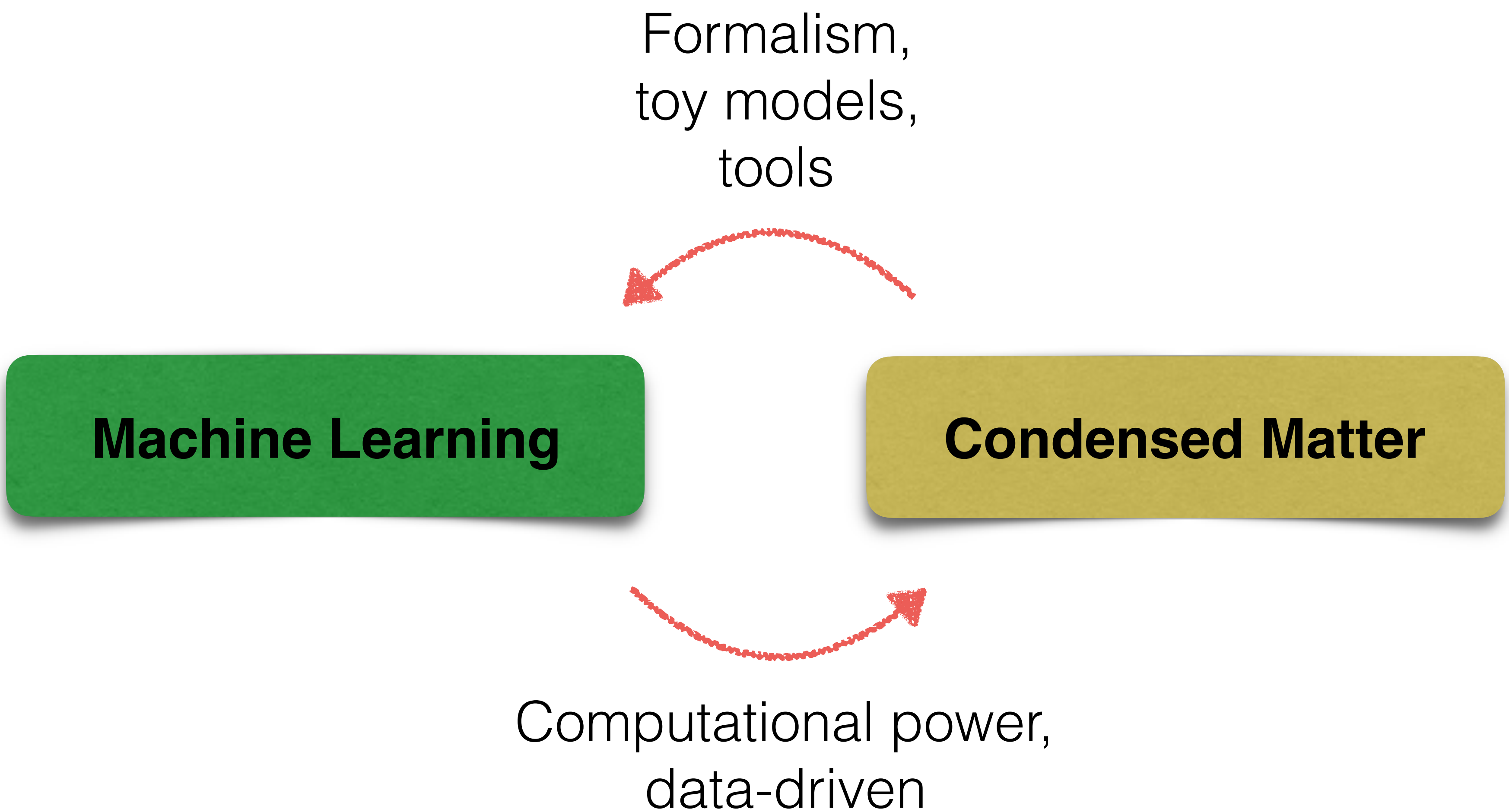
Machine Learning

Condensed Matter

Machine Learning

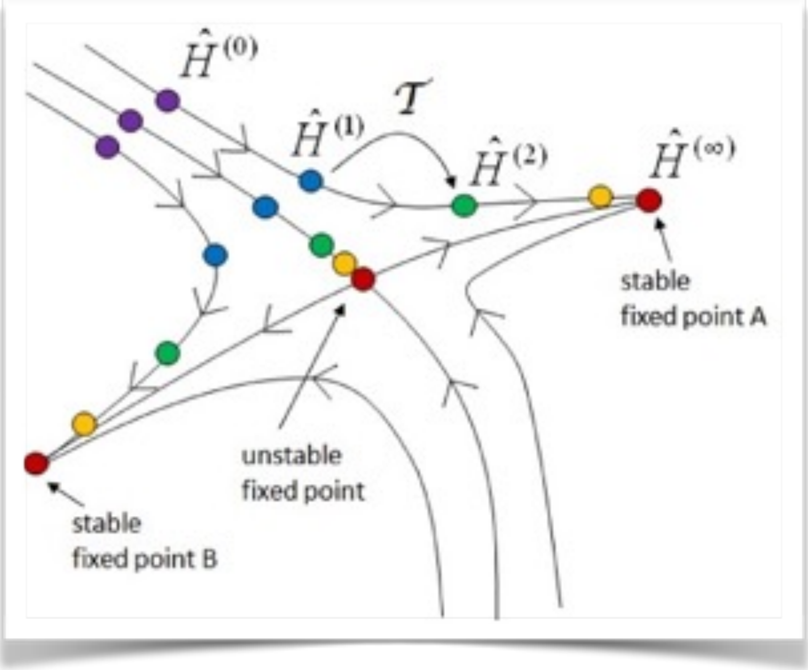
Condensed Matter

Computational power,
data-driven


Formalism,
toy models,
tools
Machine Learning
Condensed Matter
Computational power,
data-driven

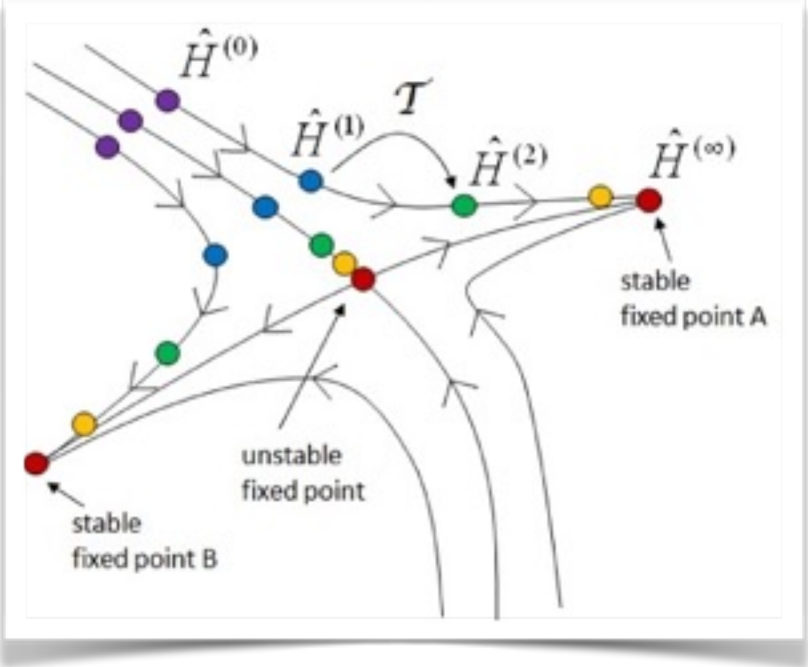
Renormalization Group

Renormalization Group



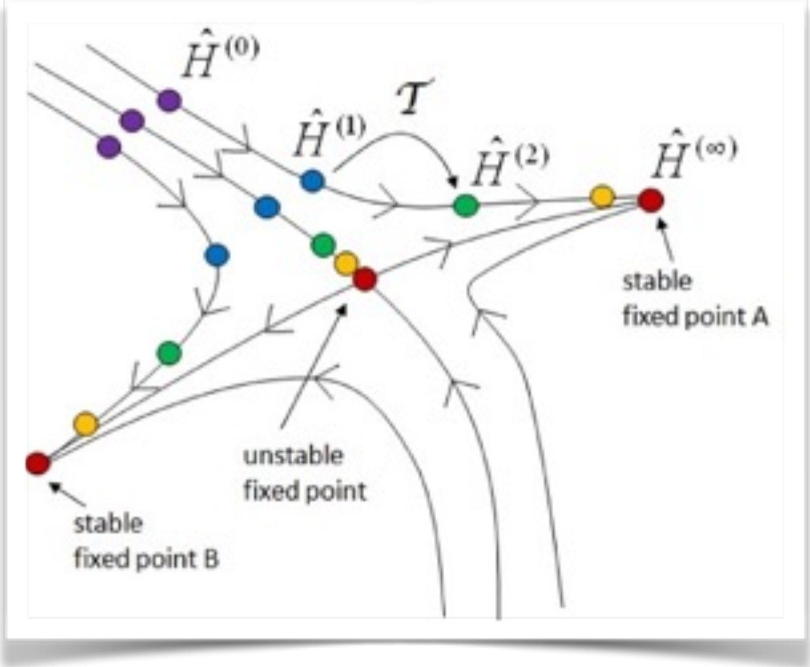
Renormalization Group

- Conceptually important: formalizes the notion of universality



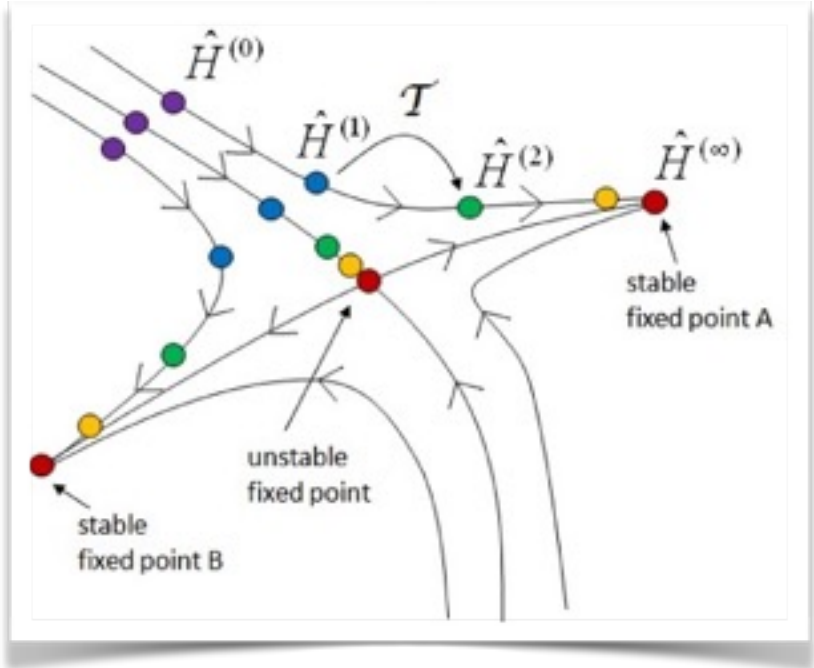
Renormalization Group

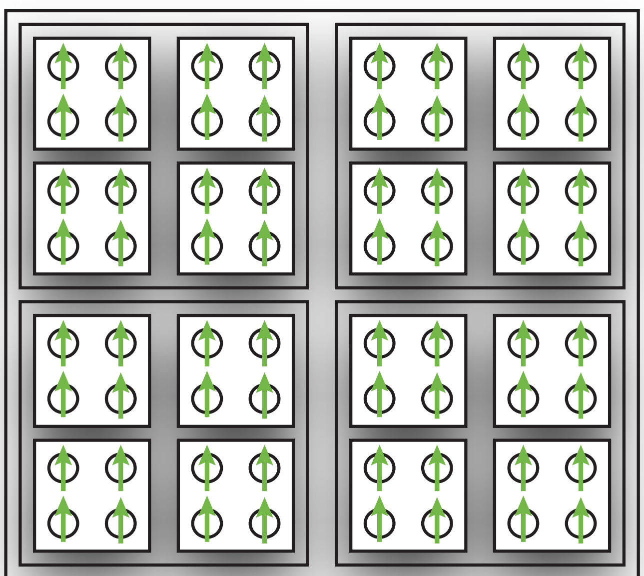
- Conceptually important: formalizes the notion of universality
- Many flavors exist: Wilsonian, DMRG, real-space, ….



Renormalization Group

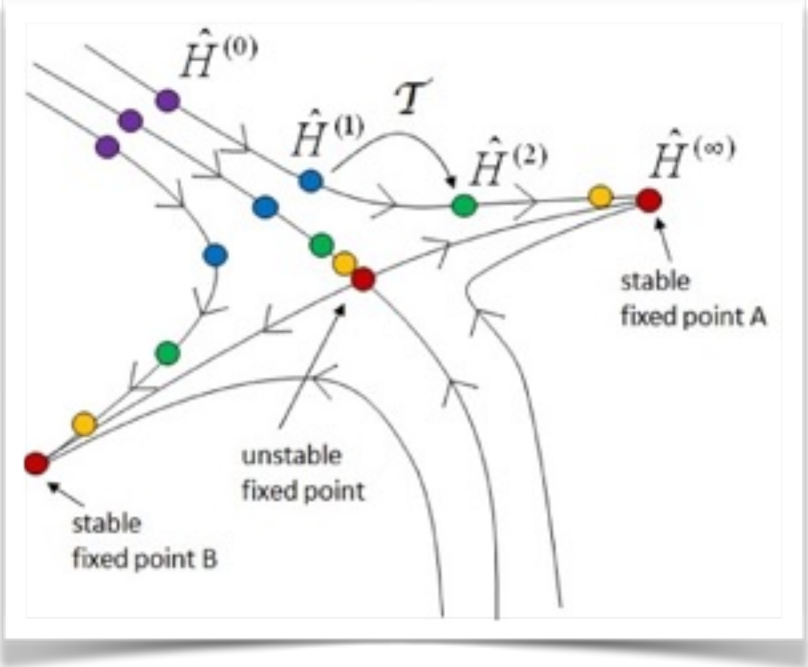
- Conceptually important: formalizes the notion of universality
- Many flavors exist: Wilsonian, DMRG, real-space, ….

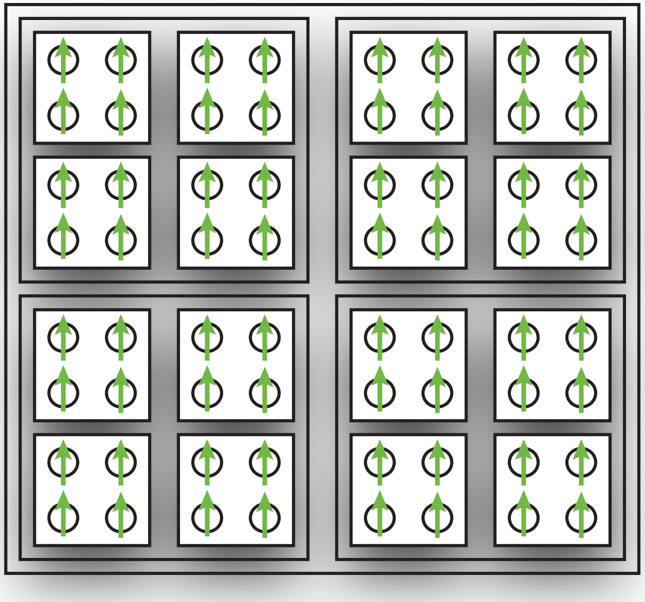




Renormalization Group

- Conceptually important: formalizes the notion of universality
- Many flavors exist: Wilsonian, DMRG, real-space, ….





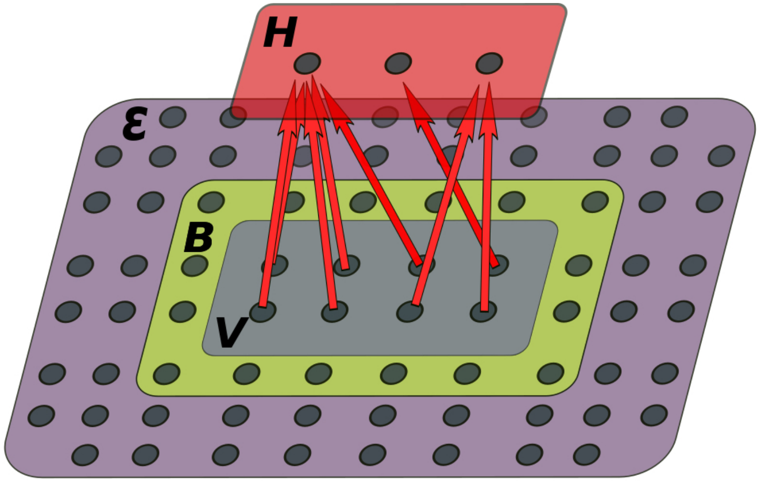
Leitmotiv: integrate out some ‘fast’ degrees of freedom to obtain effective theory of the ‘slow’ ones

In a sense the relation of RG to information theory is obvious as averaging loses information

In a sense the relation of RG to information theory is obvious as averaging loses information

- Can it be formalized?
- Is it useful in practice?

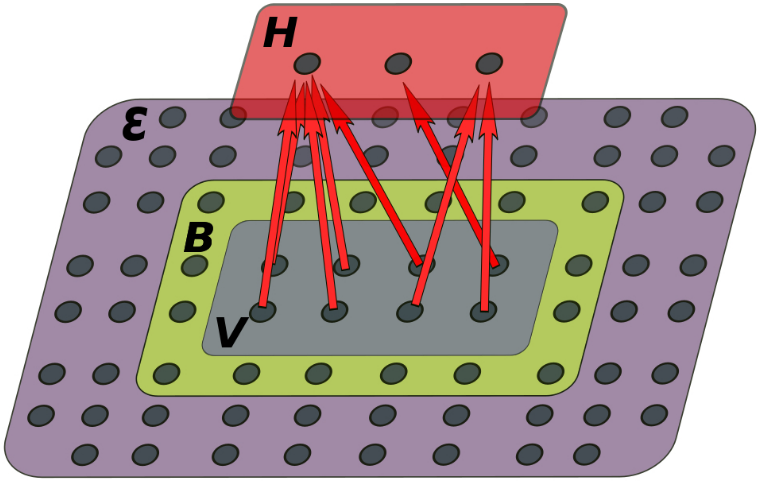
Real-space RG from Information Theory perspective



$$P(\mathcal{X}) = \frac{1}{Z} e^{\mathcal{K}(\mathcal{X})}$$

Real-space RG from Information Theory perspective

$$e^{\mathcal{K}'(\mathcal{X}')} = \sum_{\mathcal{X}} e^{\mathcal{K}(\mathcal{X})} P_{\Lambda}(\mathcal{X}'|\mathcal{X})$$



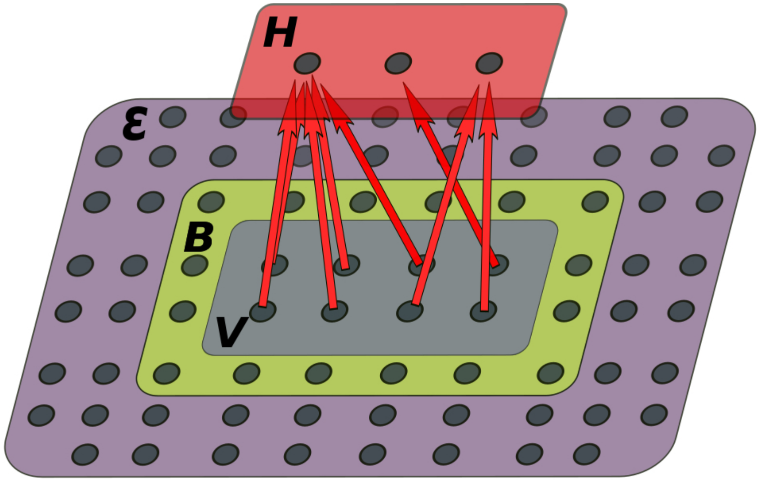
$$P(\mathcal{X}) = \frac{1}{Z} e^{\mathcal{K}(\mathcal{X})}$$

ETH zürich

Real-space RG from Information Theory perspective

$$e^{\mathcal{K}'(\mathcal{X}')} = \sum_{\mathcal{X}} e^{\mathcal{K}(\mathcal{X})} P_\Lambda(\mathcal{X}'|\mathcal{X})$$

Task: Learn $P_\Lambda(\mathcal{H}\,|\,\mathcal{V})$ such that $\mathcal{H}$ tracks the *slow degrees of freedom* within region $\mathcal{V}$



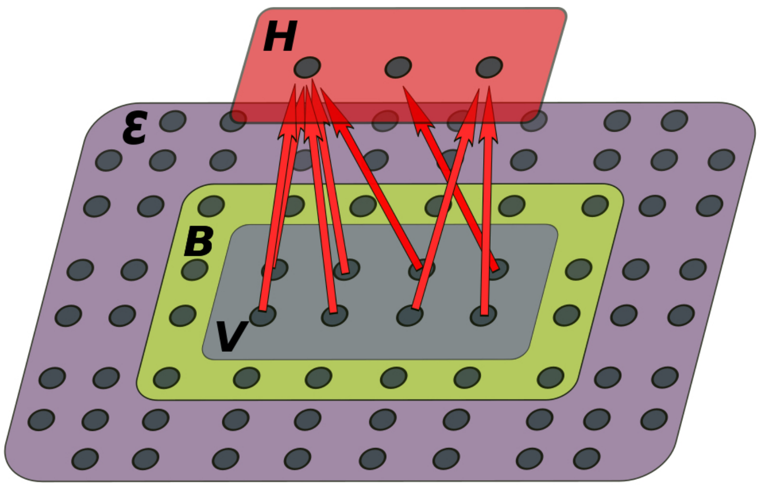
$$P(\mathcal{X}) = \frac{1}{Z} e^{\mathcal{K}(\mathcal{X})}$$

Real-space RG from Information Theory perspective

$$e^{\mathcal{K}'(\mathcal{X}')} = \sum_{\mathcal{X}} e^{\mathcal{K}(\mathcal{X})} P_\Lambda(\mathcal{X}'|\mathcal{X})$$

Task: Learn $P_\Lambda(\mathcal{H}|\mathcal{V})$ such that $\mathcal{H}$ tracks the *slow degrees of freedom* within region $\mathcal{V}$

$$\boxed{\Lambda = (a_i, b_j, \lambda_i^j)}$$



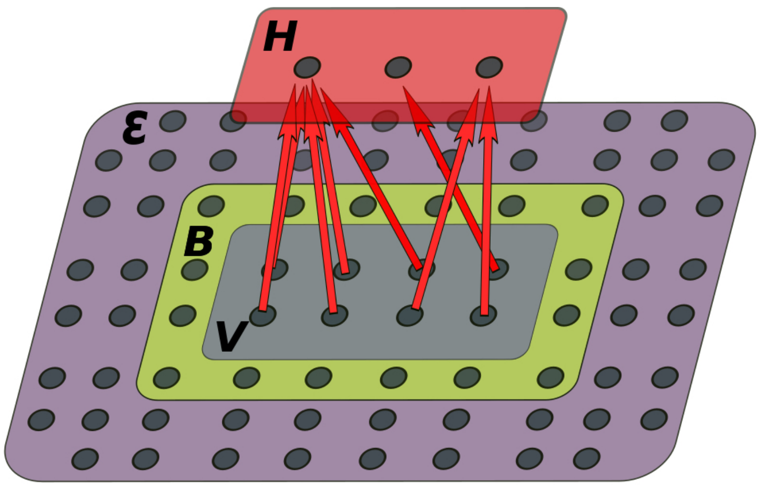
$$P(\mathcal{X}) = \frac{1}{Z} e^{\mathcal{K}(\mathcal{X})}$$

Real-space RG from Information Theory perspective

$$e^{\mathcal{K}'(\mathcal{X}')} = \sum_{\mathcal{X}} e^{\mathcal{K}(\mathcal{X})} P_\Lambda(\mathcal{X}'|\mathcal{X})$$

Task: Learn $P_\Lambda(\mathcal{H}|\mathcal{V})$ such that $\mathcal{H}$ tracks the *slow degrees of freedom* within region $\mathcal{V}$

$$\boxed{\Lambda = (a_i, b_j, \lambda_i^j)}$$

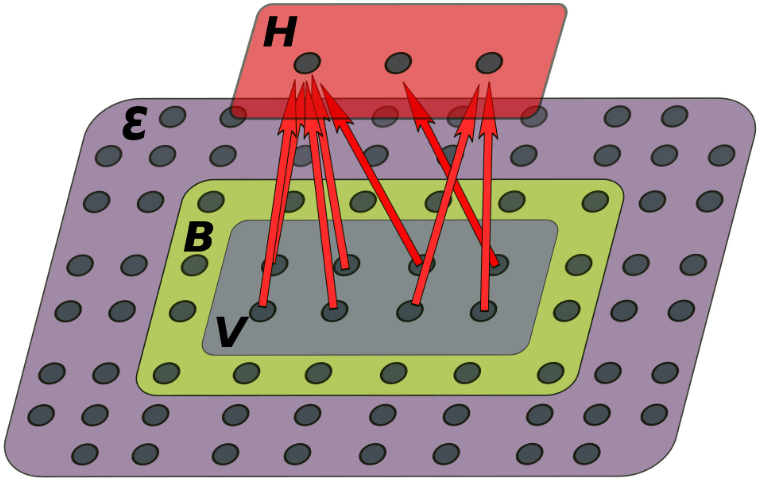


$$P(\mathcal{X}) = \frac{1}{Z} e^{\mathcal{K}(\mathcal{X})}$$

Method: Require that *slow degrees of freedom* maximize spatial mutual information

Real-space RG from Information Theory perspective

$$\boxed{e^{\mathcal{K}'(\mathcal{X}')} = \sum_{\mathcal{X}} e^{\mathcal{K}(\mathcal{X})} P_\Lambda(\mathcal{X}'|\mathcal{X})}$$



Task: Learn $P_\Lambda(\mathcal{H}|\mathcal{V})$ such that $\mathcal{H}$ tracks the *slow degrees of freedom* within region $\mathcal{V}$

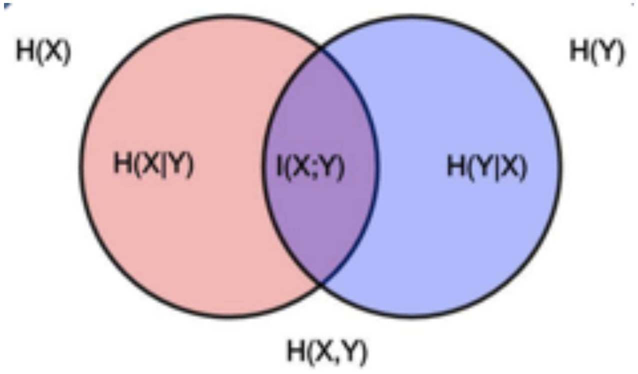
$$\boxed{\Lambda = (a_i, b_j, \lambda_i^j)}$$

$$P(\mathcal{X}) = \frac{1}{Z} e^{\mathcal{K}(\mathcal{X})}$$

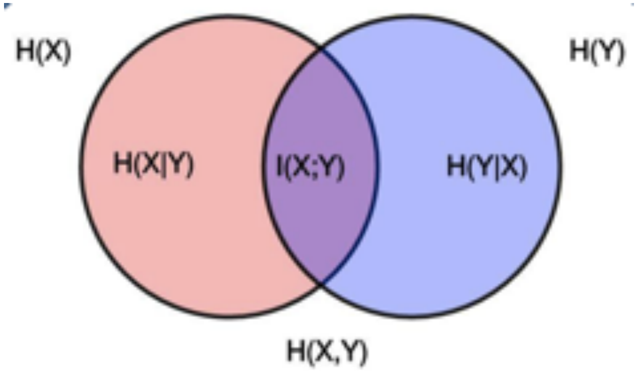
Method: Require that *slow degrees of freedom* maximize spatial mutual information

Formally: find $\max[I_\Lambda(\mathcal{H}:\mathcal{E})]$ over parameters Λ

Mutual Information

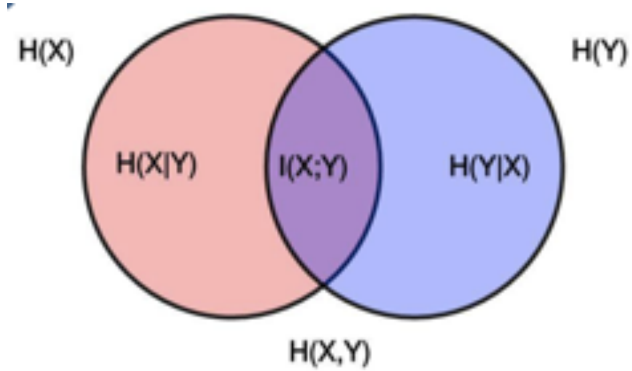


Mutual Information



- Vanishes for independent variables
- Bounded from above by information entropy
- More general than correlation functions

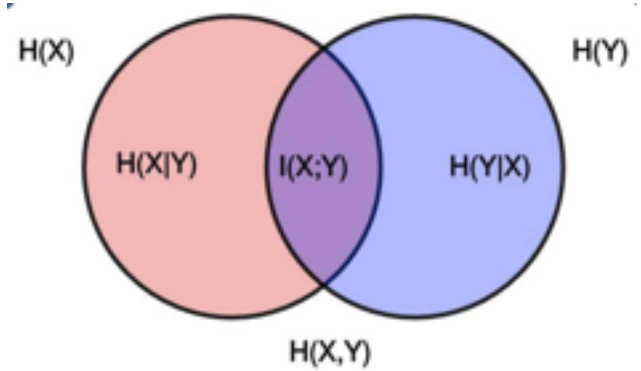
Mutual Information



$$I_\Lambda(\mathcal{H}:\mathcal{E}) = \sum_{\mathcal{H},\mathcal{E}} P_\Lambda(\mathcal{E},\mathcal{H}) \log\left(\frac{P_\Lambda(\mathcal{E},\mathcal{H})}{P_\Lambda(\mathcal{H})P(\mathcal{E})}\right)$$

- Vanishes for independent variables
- Bounded from above by information entropy
- More general than correlation functions

Mutual Information

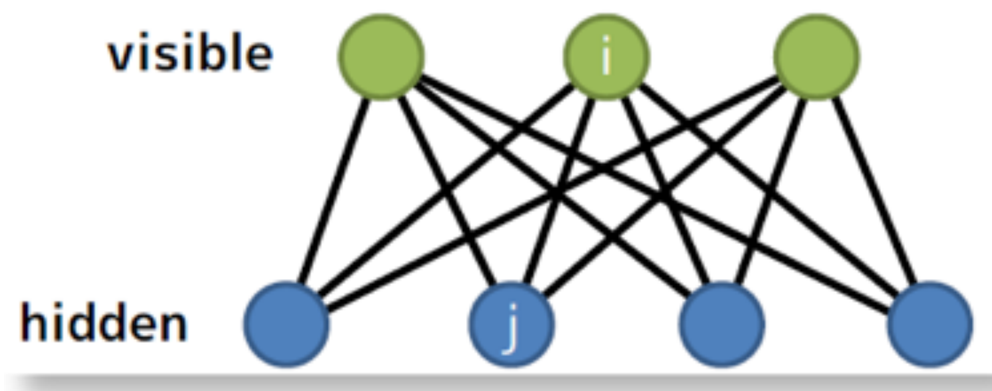


$$I_\Lambda(\mathcal{H}:\mathcal{E}) = \sum_{\mathcal{H},\mathcal{E}} P_\Lambda(\mathcal{E},\mathcal{H}) \log\left(\frac{P_\Lambda(\mathcal{E},\mathcal{H})}{P_\Lambda(\mathcal{H})P(\mathcal{E})}\right)$$

- Vanishes for independent variables
- Bounded from above by information entropy
- More general than correlation functions

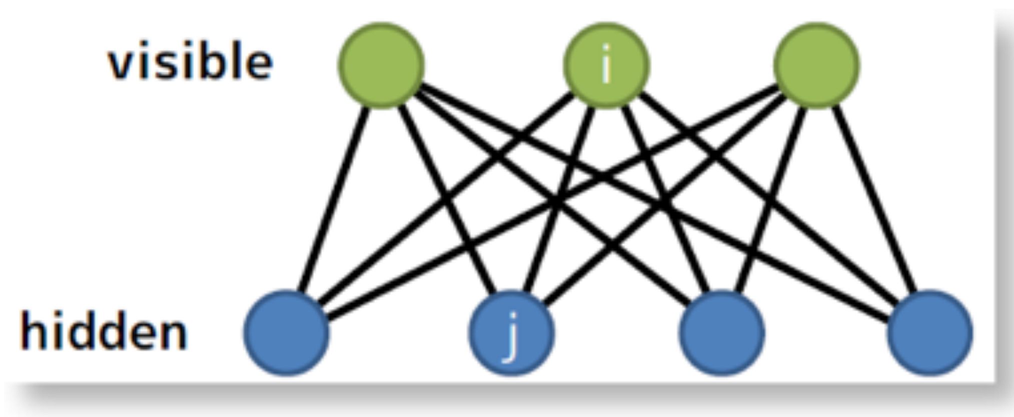
$$I_\Lambda(\mathcal{H}:\mathcal{E}) = H(\mathcal{H}) - H(\mathcal{H}|\mathcal{E})$$

(Restricted) Boltzmann Machines



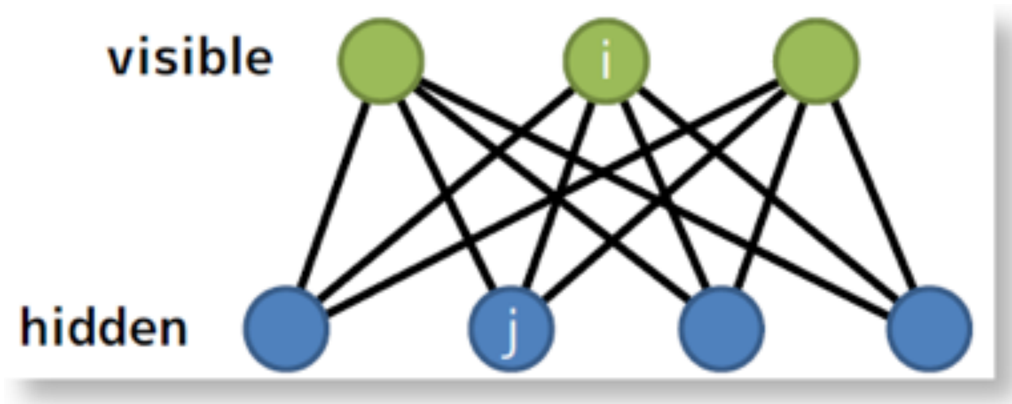
(Restricted) Boltzmann Machines

- Stochastic networks



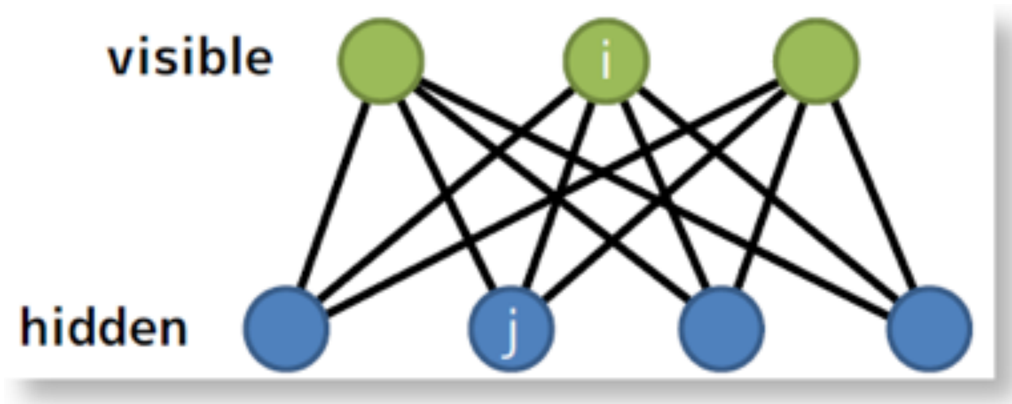
(Restricted) Boltzmann Machines

- Stochastic networks
- Model probability distributions



(Restricted) Boltzmann Machines

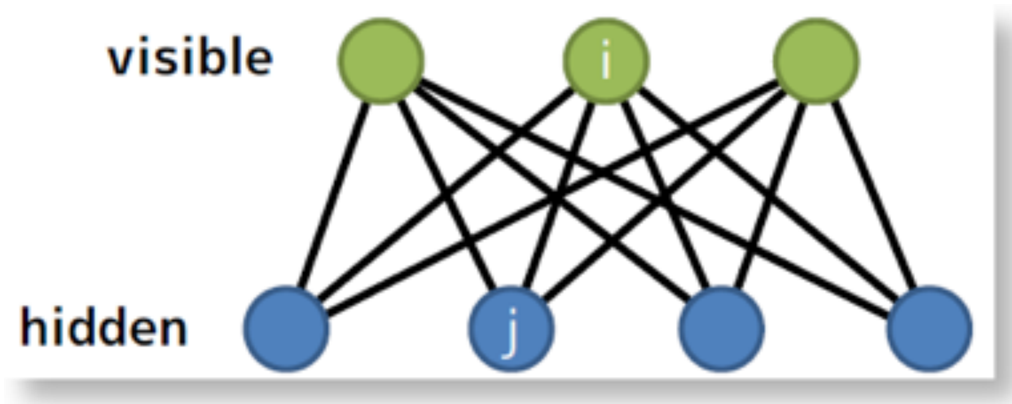
- Stochastic networks
- Model probability distributions



$$E_\Theta \equiv E_{a,b,\theta}(\mathcal{V},\mathcal{H}) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_{ij} v_i \theta_{ij} h_j$$

(Restricted) Boltzmann Machines

- Stochastic networks
- Model probability distributions

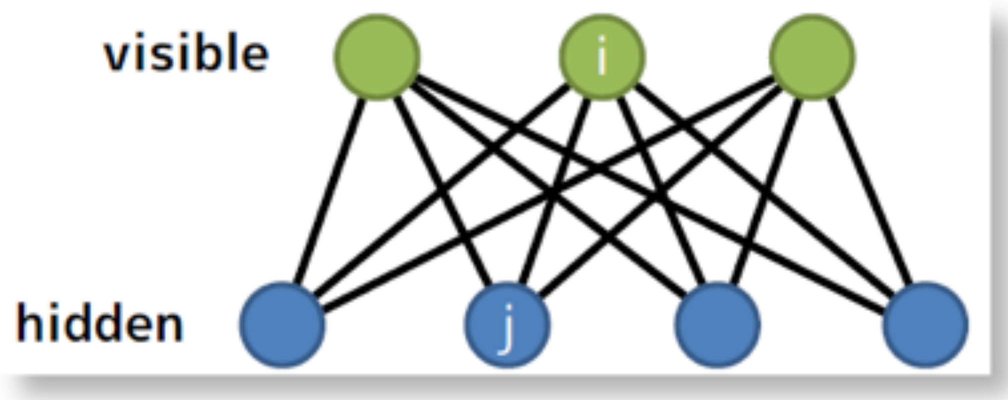


$$E_\Theta \equiv E_{a,b,\theta}(\mathcal{V},\mathcal{H}) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_{ij} v_i \theta_{ij} h_j$$

$$\boxed{P_\Theta(\mathcal{V},\mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}}$$

(Restricted) Boltzmann Machines

- Stochastic networks
- Model probability distributions



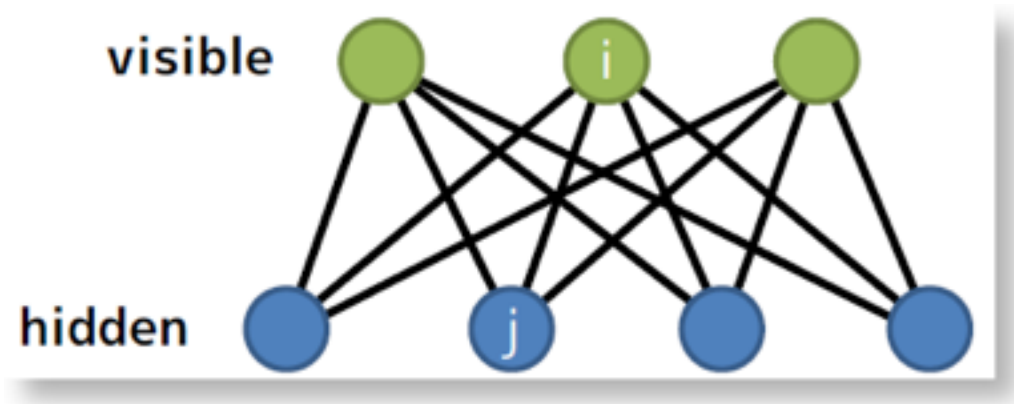
$$E_\Theta \equiv E_{a,b,\theta}(\mathcal{V},\mathcal{H}) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_{ij} v_i \theta_{ij} h_j$$

$$\boxed{P_\Theta(\mathcal{V},\mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}}$$

$$P_\Theta(\mathcal{V}) \qquad\qquad P_\Theta(\mathcal{H}|\mathcal{V})$$

(Restricted) Boltzmann Machines

- Stochastic networks
- Model probability distributions
- Can be used for efficient sampling



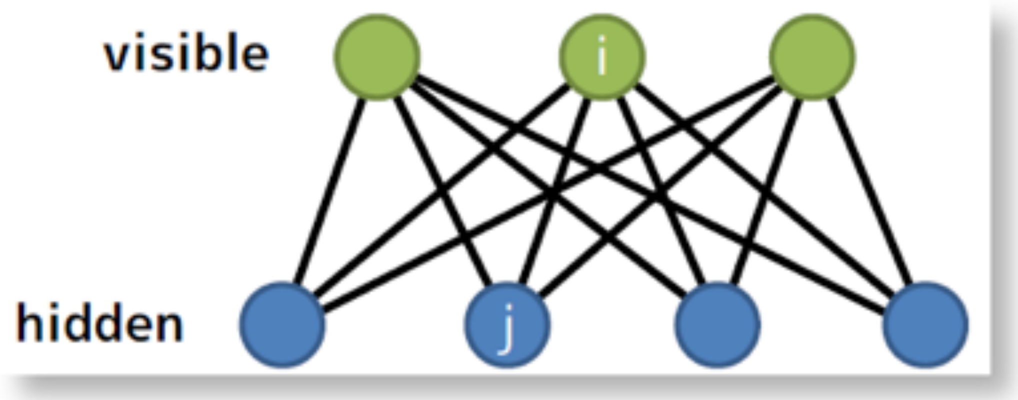
$$E_\Theta \equiv E_{a,b,\theta}(\mathcal{V},\mathcal{H}) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_{ij} v_i \theta_{ij} h_j$$

$$\boxed{P_\Theta(\mathcal{V},\mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}}$$

$$P_\Theta(\mathcal{V}) \qquad P_\Theta(\mathcal{H}|\mathcal{V})$$

(Restricted) Boltzmann Machines

- Stochastic networks
- Model probability distributions
- Can be used for efficient sampling



$$E_\Theta \equiv E_{a,b,\theta}(\mathcal{V},\mathcal{H}) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_{ij} v_i \theta_{ij} h_j$$

$$\boxed{P_\Theta(\mathcal{V},\mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}}$$

$$P_\Theta(\mathcal{V}) \qquad P_\Theta(\mathcal{H}|\mathcal{V})$$

- How to chose the parameters?

(R)BM training

$$\boxed{P_\Theta(\mathcal{V},\mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}}$$

(R)BM training

$$\boxed{P_{\Theta}(\mathcal{V},\mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}}$$

- First attempt: Maximal Likelihood (ML)

(R)BM training

$$P_\Theta(\mathcal{V},\mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}$$

- First attempt: Maximal Likelihood (ML)

$$\sum_{\mathcal{V}\in\mathrm{data}} \log P_\Theta(\mathcal{V}) = \sum_{\mathrm{all}\ \mathcal{V}} P_{\mathrm{data}}(\mathcal{V}) \log P_\Theta(\mathcal{V})$$

(R)BM training

$$\boxed{P_{\Theta}(\mathcal{V},\mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}}$$

- First attempt: Maximal Likelihood (ML)

(R)BM training

$$P_\Theta(\mathcal{V},\mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}$$

- First attempt: Maximal Likelihood (ML)

$$\sum_{\text{all } \mathcal{V}} P_{data}(\mathcal{V}) \log P_{data}(\mathcal{V}) - \sum_{\text{all } \mathcal{V}} P_{data}(\mathcal{V}) \log P_\Theta(\mathcal{V}) = \sum_{\text{all } \mathcal{V}} P_{data}(\mathcal{V}) \log \left(\frac{P_{data}(\mathcal{V})}{P_\Theta(\mathcal{V})} \right)$$

(R)BM training

$$P_\Theta(\mathcal{V}, \mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}$$

- First attempt: Maximal Likelihood (ML)

$$\sum_{\text{all } \mathcal{V}} P_{data}(\mathcal{V}) \log P_{data}(\mathcal{V}) - \sum_{\text{all } \mathcal{V}} P_{data}(\mathcal{V}) \log P_\Theta(\mathcal{V}) = \sum_{\text{all } \mathcal{V}} P_{data}(\mathcal{V}) \log \left(\frac{P_{data}(\mathcal{V})}{P_\Theta(\mathcal{V})} \right)$$

- Better solution: Contrastive Divergence (CD)

(R)BM training

$$P_\Theta(\mathcal{V},\mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}$$

- First attempt: Maximal Likelihood (ML)

$$\sum_{\text{all } \mathcal{V}} P_{data}(\mathcal{V}) \log P_{data}(\mathcal{V}) - \sum_{\text{all } \mathcal{V}} P_{data}(\mathcal{V}) \log P_\Theta(\mathcal{V}) = \sum_{\text{all } \mathcal{V}} P_{data}(\mathcal{V}) \log\left(\frac{P_{data}(\mathcal{V})}{P_\Theta(\mathcal{V})}\right)$$

- Better solution: Contrastive Divergence (CD)

$$KL(P_{data}||P_\Theta) - KL(P_n||P_\Theta)$$

(R)BM training

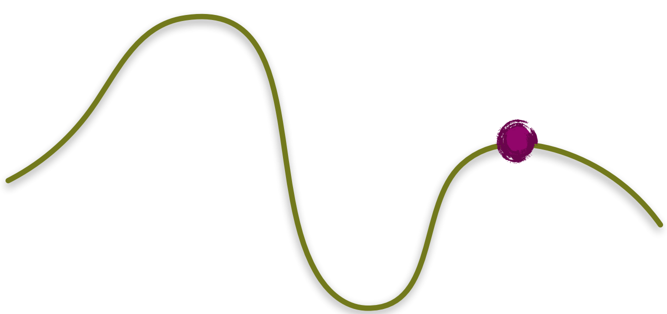
$$P_\Theta(\mathcal{V},\mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}$$

- First attempt: Maximal Likelihood (ML)

$$\sum_{\text{all } \mathcal{V}} P_{data}(\mathcal{V}) \log P_{data}(\mathcal{V}) - \sum_{\text{all } \mathcal{V}} P_{data}(\mathcal{V}) \log P_\Theta(\mathcal{V}) = \sum_{\text{all } \mathcal{V}} P_{data}(\mathcal{V}) \log \left(\frac{P_{data}(\mathcal{V})}{P_\Theta(\mathcal{V})} \right)$$

- Better solution: Contrastive Divergence (CD)

$$KL(P_{data}||P_\Theta) - KL(P_n||P_\Theta)$$



(R)BM training

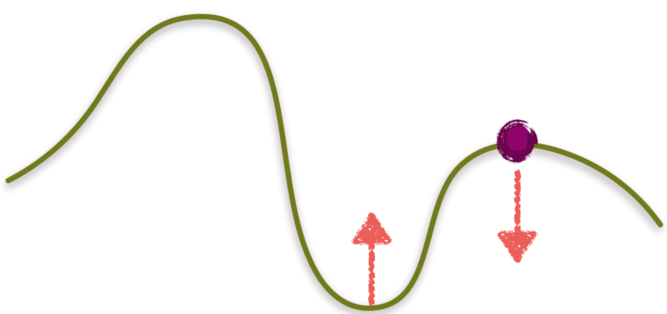
$$P_\Theta(\mathcal{V},\mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}$$

- First attempt: Maximal Likelihood (ML)

$$\sum_{\text{all } \mathcal{V}} P_{data}(\mathcal{V}) \log P_{data}(\mathcal{V}) - \sum_{\text{all } \mathcal{V}} P_{data}(\mathcal{V}) \log P_\Theta(\mathcal{V}) = \sum_{\text{all } \mathcal{V}} P_{data}(\mathcal{V}) \log \left(\frac{P_{data}(\mathcal{V})}{P_\Theta(\mathcal{V})} \right)$$

- Better solution: Contrastive Divergence (CD)

$$KL(P_{data}||P_\Theta) - KL(P_n||P_\Theta)$$



(R)BM training

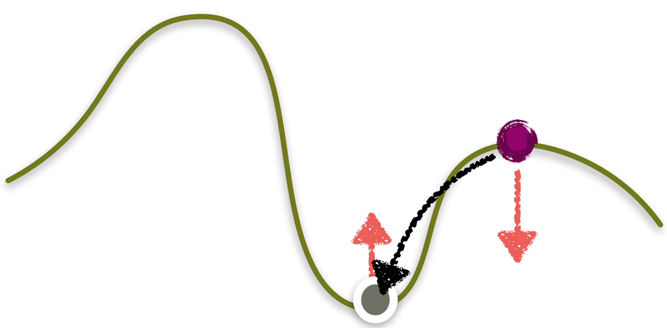
$$P_\Theta(\mathcal{V},\mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}$$

- First attempt: Maximal Likelihood (ML)

$$\sum_{\text{all } \mathcal{V}} P_{data}(\mathcal{V}) \log P_{data}(\mathcal{V}) - \sum_{\text{all } \mathcal{V}} P_{data}(\mathcal{V}) \log P_\Theta(\mathcal{V}) = \sum_{\text{all } \mathcal{V}} P_{data}(\mathcal{V}) \log \left(\frac{P_{data}(\mathcal{V})}{P_\Theta(\mathcal{V})} \right)$$

- Better solution: Contrastive Divergence (CD)

$$KL(P_{data}||P_\Theta) - KL(P_n||P_\Theta)$$



(R)BM training

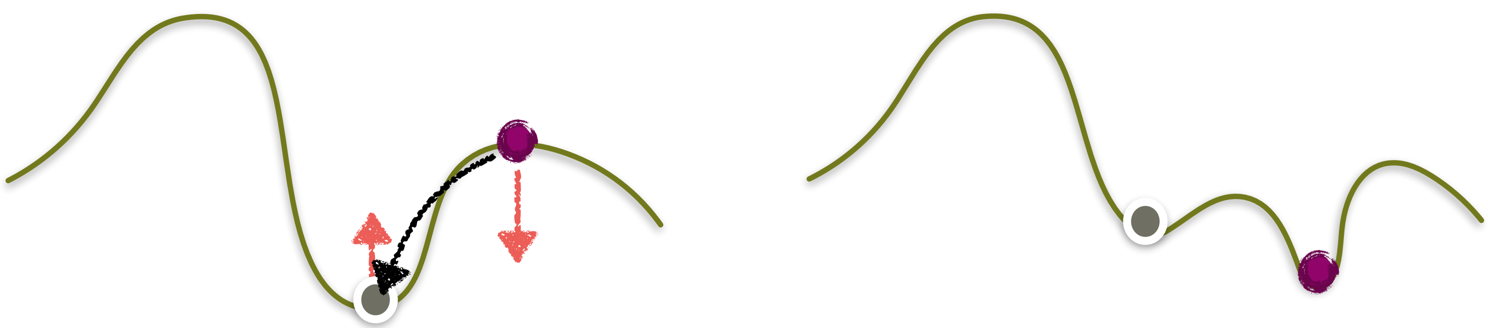
$$P_\Theta(\mathcal{V},\mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}$$

- First attempt: Maximal Likelihood (ML)

$$\sum_{\text{all } \mathcal{V}} P_{data}(\mathcal{V}) \log P_{data}(\mathcal{V}) - \sum_{\text{all } \mathcal{V}} P_{data}(\mathcal{V}) \log P_\Theta(\mathcal{V}) = \sum_{\text{all } \mathcal{V}} P_{data}(\mathcal{V}) \log \left(\frac{P_{data}(\mathcal{V})}{P_\Theta(\mathcal{V})}\right)$$

- Better solution: Contrastive Divergence (CD)

$$KL(P_{data}||P_\Theta) - KL(P_n||P_\Theta)$$



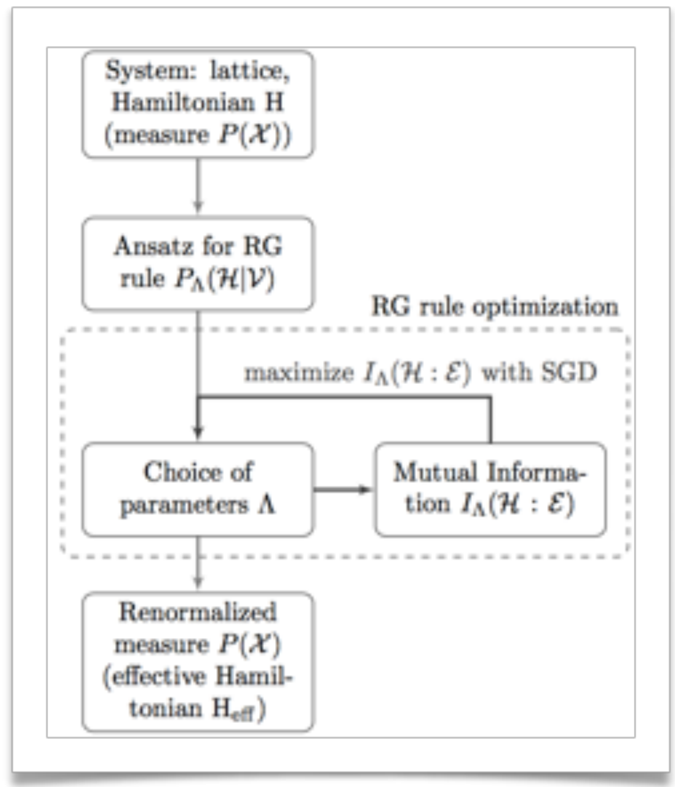
Stage I. - Train RBMs to reproduce P(V,E) and P(V) via contrastive divergence

$$\boxed{P_{\Theta}(\mathcal{V},\mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}}$$

Stage I. - Train RBMs to reproduce P(V,E) and P(V) via contrastive divergence

$$\boxed{P_\Theta(\mathcal{V},\mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}}$$

Stage II. - Model $P_\lambda(H \mid V)$ as an RBM, obtain $P_\lambda(H,E)$, do SGD to maximize I(H:E) (involves Monte Carlo)



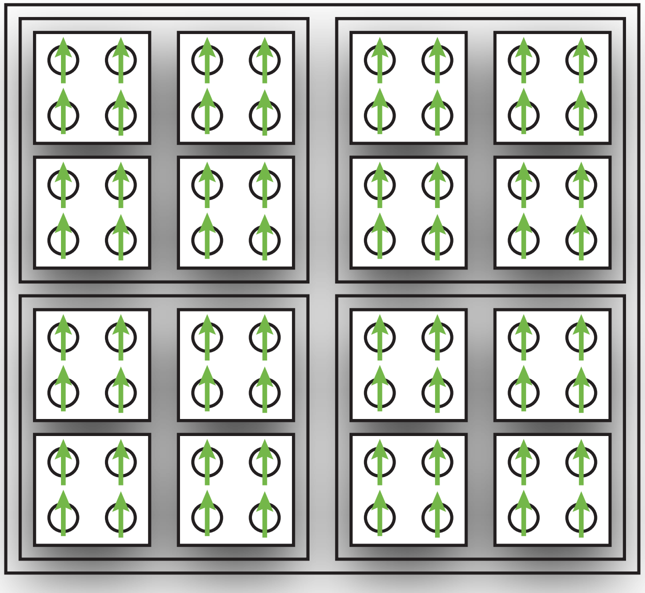
Example 1: the 2D Ising model

$$\boxed{H_I = -\sum_{<i,j>} s_i s_j}$$

Example 1: the 2D Ising model

$$\boxed{H_I = -\sum_{<i,j>} s_i s_j}$$

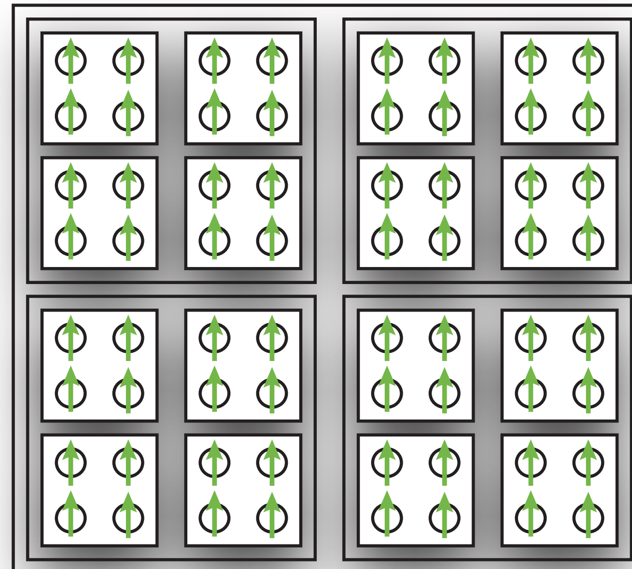
Migdal-Kadanoff block-spins:

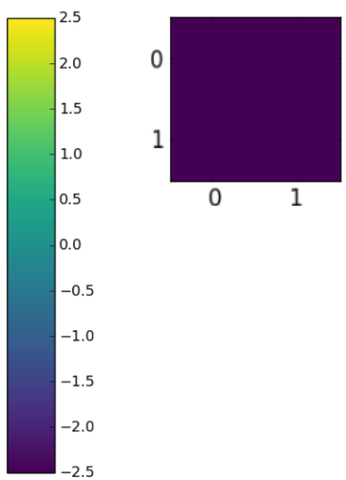


Example 1: the 2D Ising model

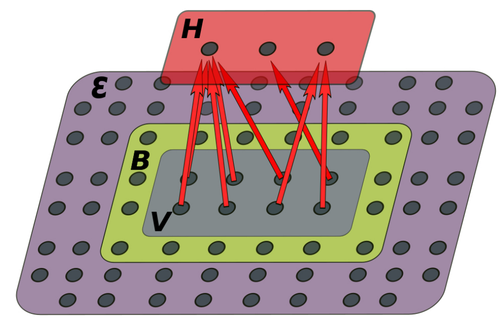
$$\boxed{H_I = -\sum_{<i,j>} s_i s_j}$$

Migdal-Kadanoff block-spins:





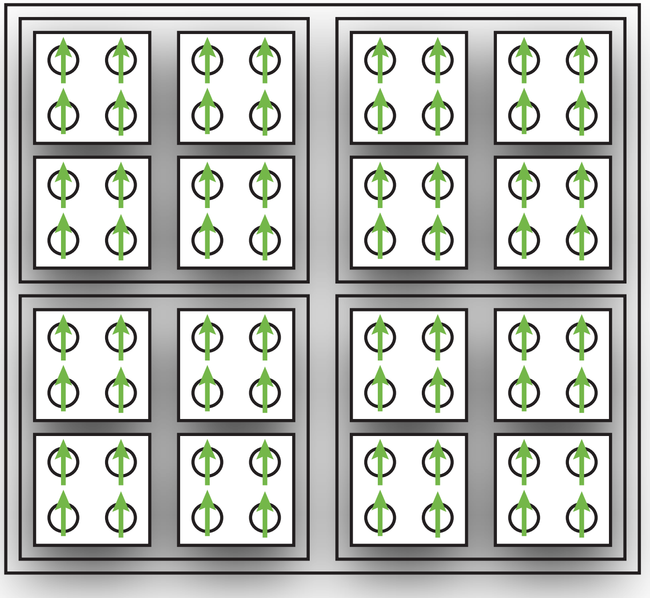
N=1

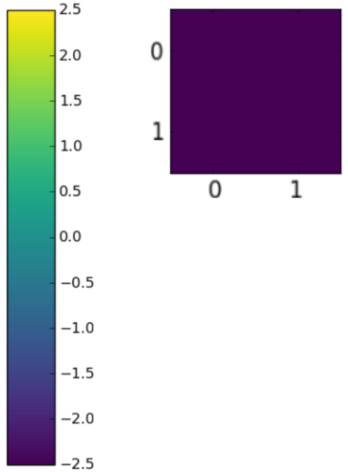


Example 1: the 2D Ising model

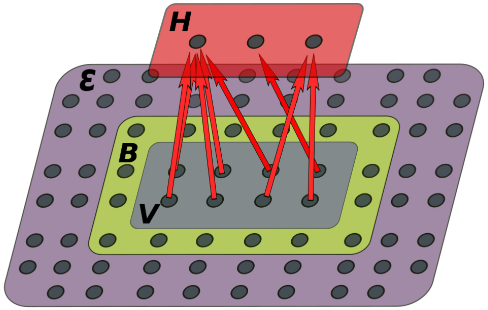
$$\boxed{H_I = -\sum_{<i,j>} s_i s_j}$$

Migdal-Kadanoff block-spins:

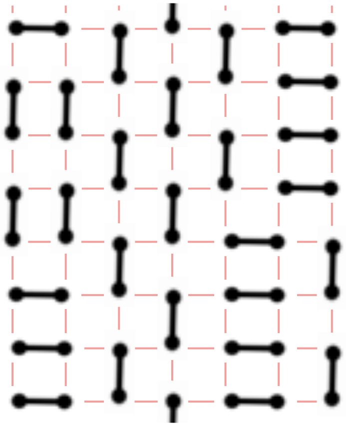




N=1

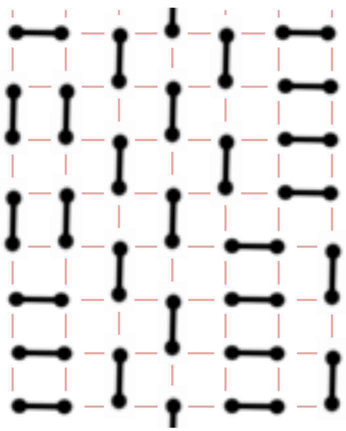


Example 2: the dimer model



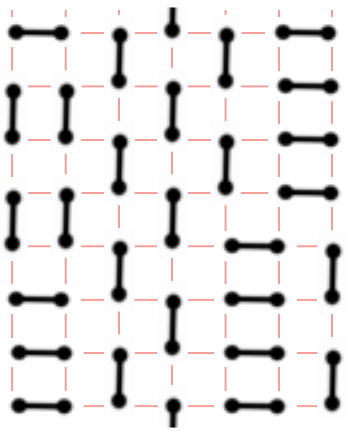
ETH zürich

Example 2: the dimer model



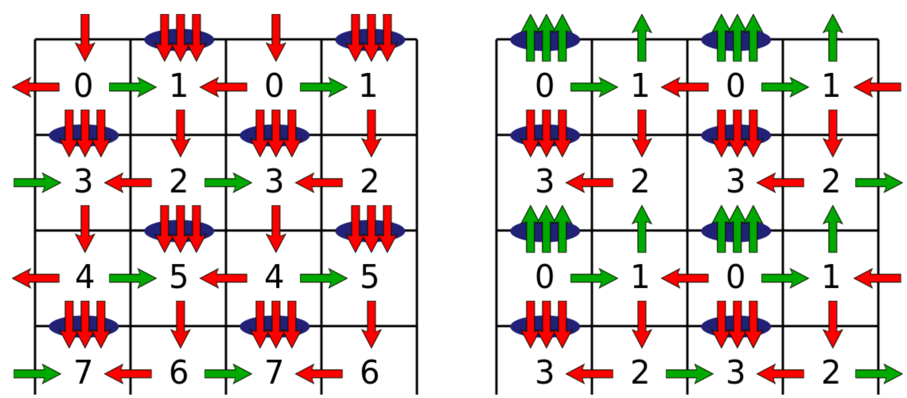
- Defined by local constraints
- Partition function counts configurations

Example 2: the dimer model

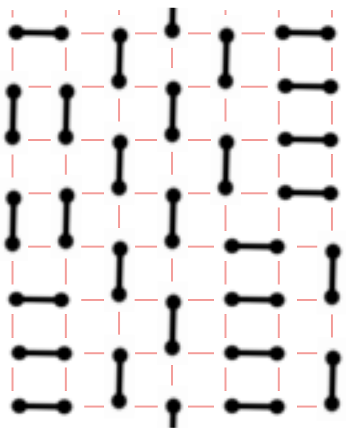


- Defined by local constraints
- Partition function counts configurations

RG of dimer model:
mapping to height field h(x)



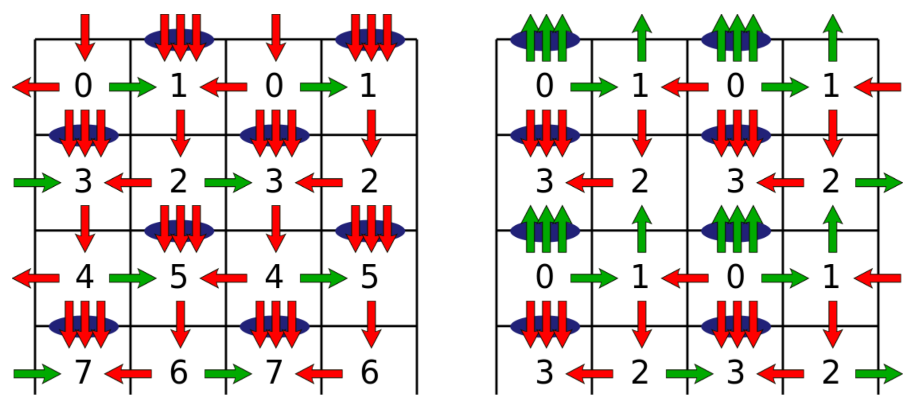
Example 2: the dimer model



- Defined by local constraints
- Partition function counts configurations

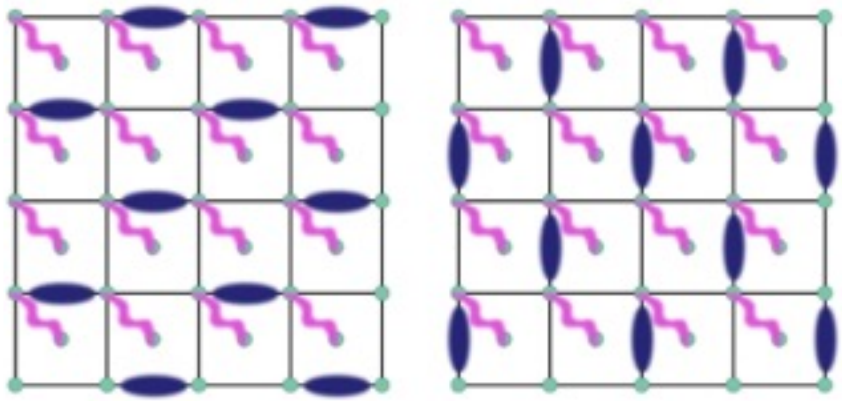
RG of dimer model: mapping to height field h(x)

$$S_{dim}[h] = \int d^2x \; (\nabla h(\vec{x}))^2 \equiv \int d^2x \; \vec{E}^2(\vec{x})$$

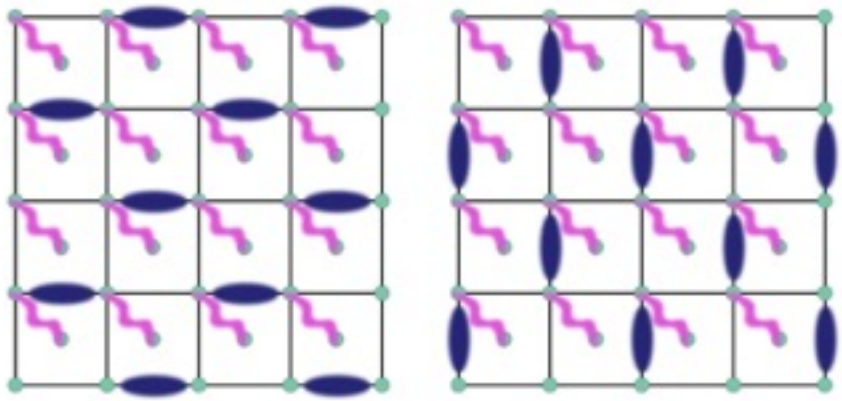


- Let’s add noise!

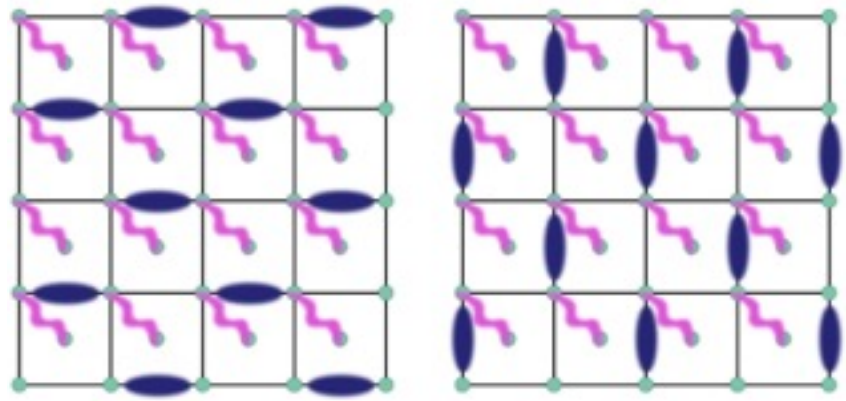
- Let’s add noise!



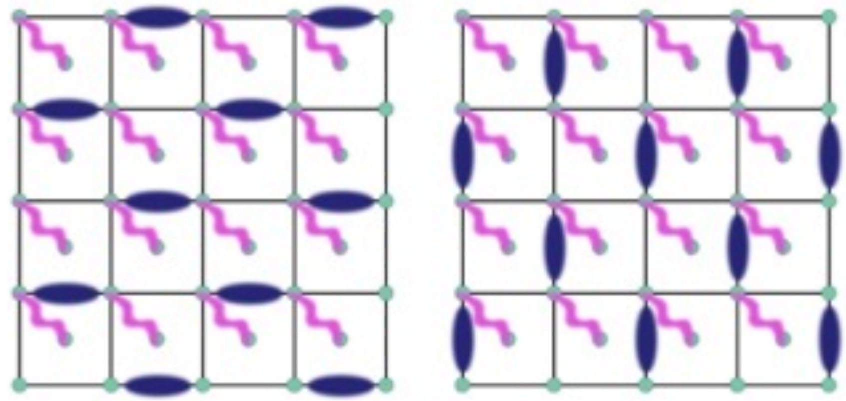
- Let's add noise!
- Physically irrelevant, but strong pattern



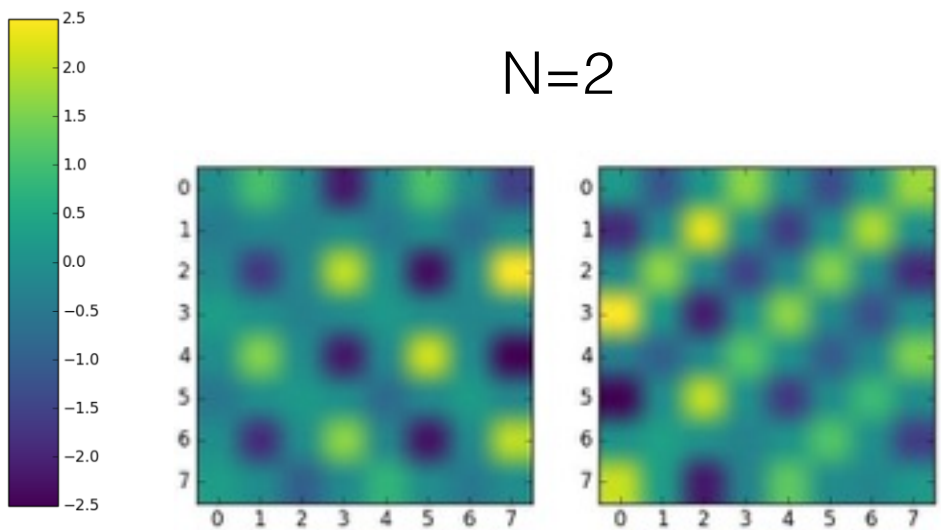
- Let’s add noise!
- Physically irrelevant, but strong pattern



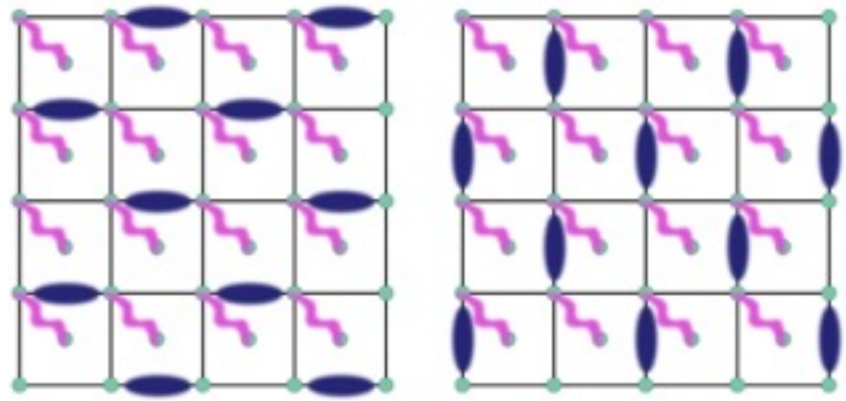
- Let's add noise!
- Physically irrelevant, but strong pattern



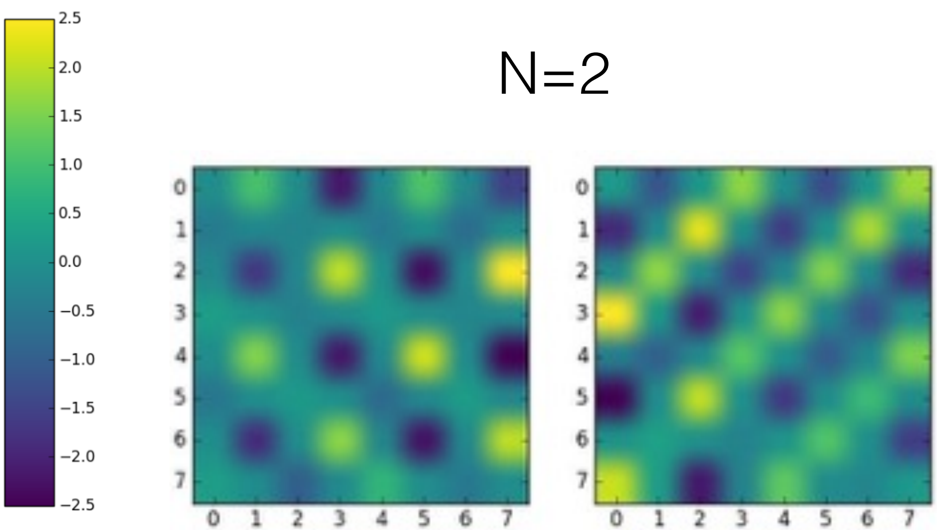
N=2

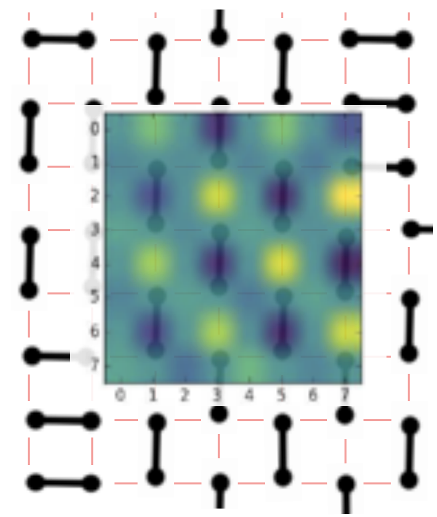


- Let’s add noise!
- Physically irrelevant, but strong pattern

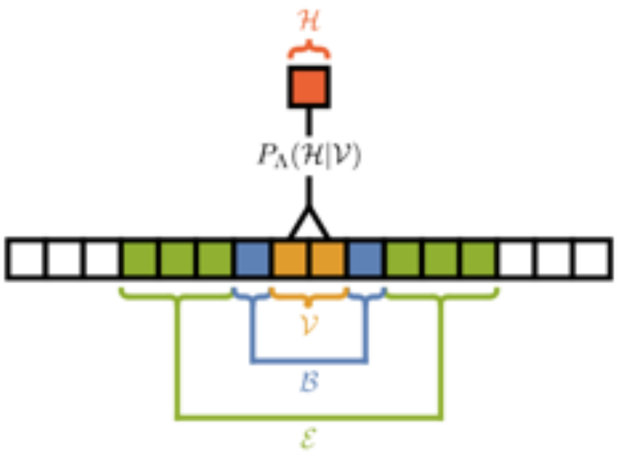


N=2



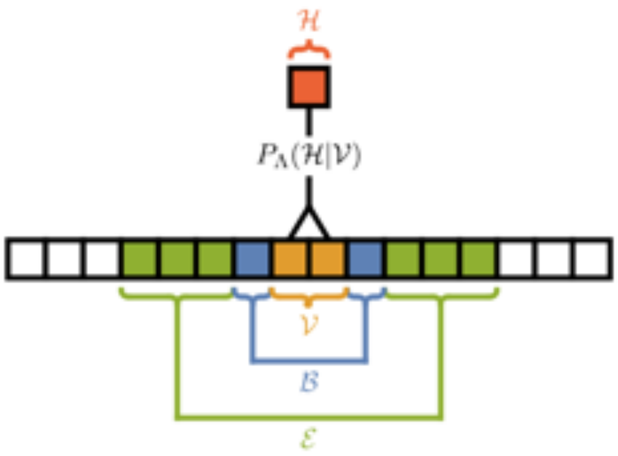


Optimality of the RSMI approach



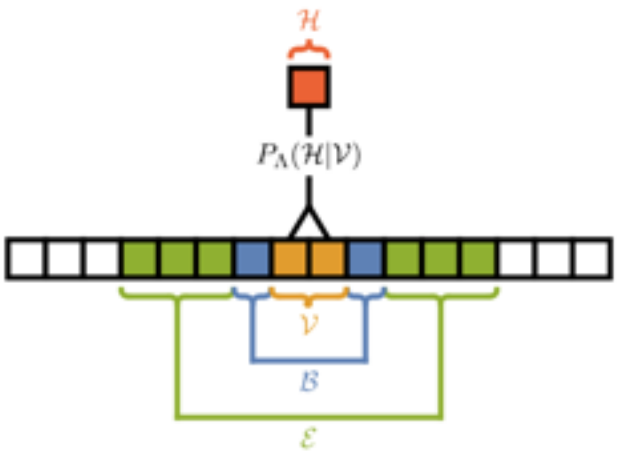
Optimality of the RSMI approach

$$\boxed{I_\Lambda(\mathcal{H}:\mathcal{E}) \leq I(\mathcal{V}:\mathcal{E})}$$



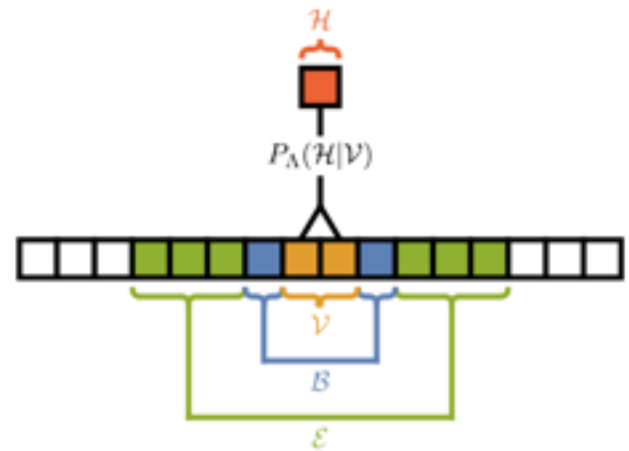
Optimality of the RSMI approach

$$I_\Lambda(\mathcal{H} : \mathcal{E}) \leq I(\mathcal{V} : \mathcal{E})$$



- Assume first perfect RSMI capture

Optimality of the RSMI approach

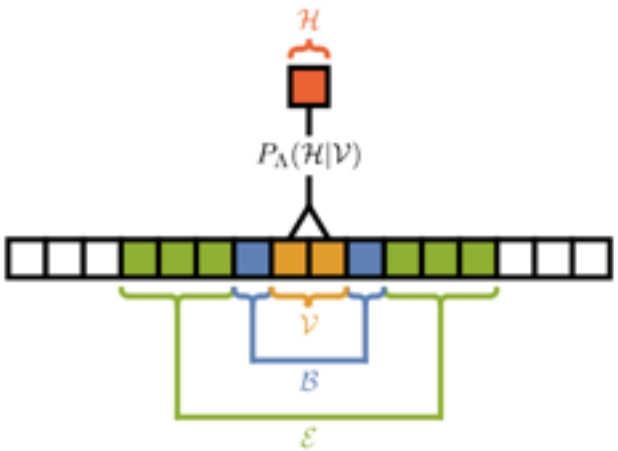


$$I_\Lambda(\mathcal{H}:\mathcal{E}) \leq I(\mathcal{V}:\mathcal{E})$$

- Assume first perfect RSMI capture

$$P(\mathcal{H}_{j\leq 2}, \mathcal{H}_{j\geq 2}|\mathcal{H}_0) = P(\mathcal{H}_{j\leq -2}|\mathcal{H}_0) \cdot P(\mathcal{H}_{j\geq 2}|\mathcal{H}_0)$$

Optimality of the RSMI approach



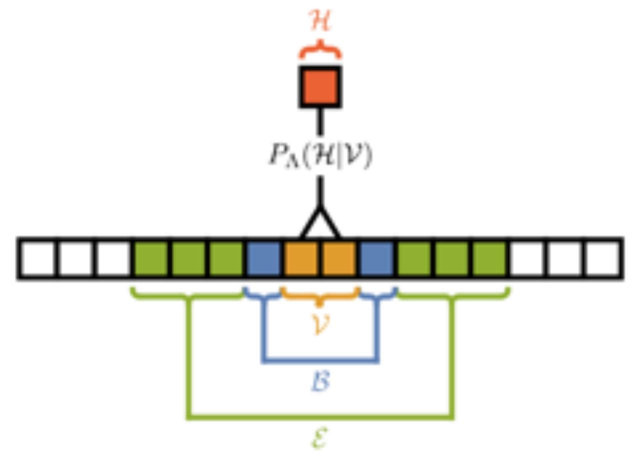
$$I_\Lambda(\mathcal{H}:\mathcal{E}) \leq I(\mathcal{V}:\mathcal{E})$$

- Assume first perfect RSMI capture

$$P(\mathcal{H}_{j\leq 2}, \mathcal{H}_{j\geq 2}|\mathcal{H}_0) = P(\mathcal{H}_{j\leq -2}|\mathcal{H}_0) \cdot P(\mathcal{H}_{j\geq 2}|\mathcal{H}_0)$$

The range of the Hamiltonian does not increase!

Optimality of the RSMI approach



$$I_\Lambda(\mathcal{H}:\mathcal{E}) \leq I(\mathcal{V}:\mathcal{E})$$

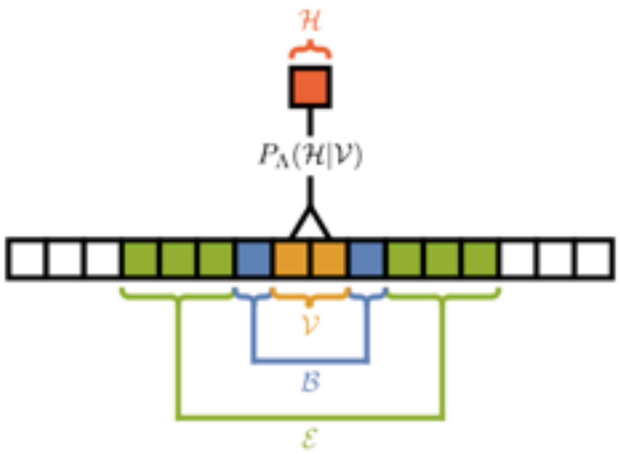
- Assume first perfect RSMI capture

$$P(\mathcal{H}_{j\leq 2}, \mathcal{H}_{j\geq 2}|\mathcal{H}_0) = P(\mathcal{H}_{j\leq -2}|\mathcal{H}_0)\cdot P(\mathcal{H}_{j\geq 2}|\mathcal{H}_0)$$

The range of the Hamiltonian does not increase!

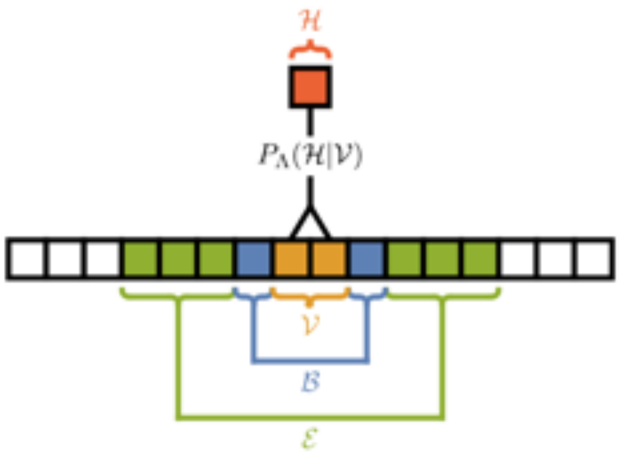
BUT: Realistically, this assumption is rarely true

Optimality: 1D Ising model



Optimality: 1D Ising model

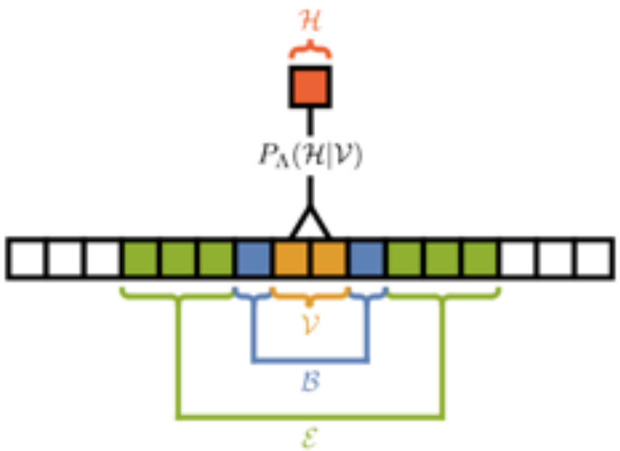
- Effective Hamiltonian:



Optimality: 1D Ising model

- Effective Hamiltonian:

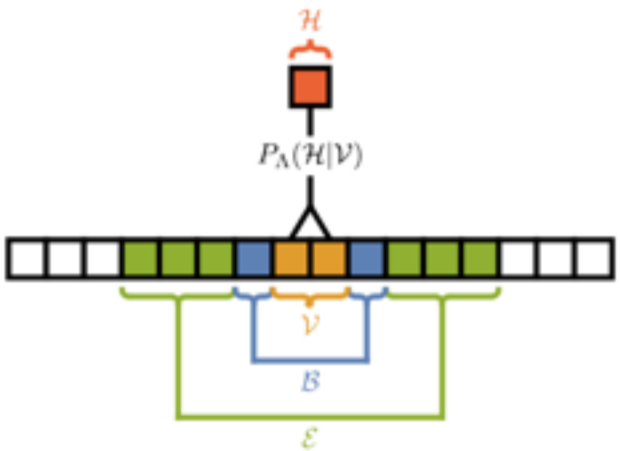
$$\mathcal{K}'[\mathcal{X}'] = \log(Z_{\Lambda,0}[\mathcal{X}']) + \sum_{k=0}^{\infty} \frac{1}{k!} C_k[\mathcal{X}']$$



Optimality: 1D Ising model

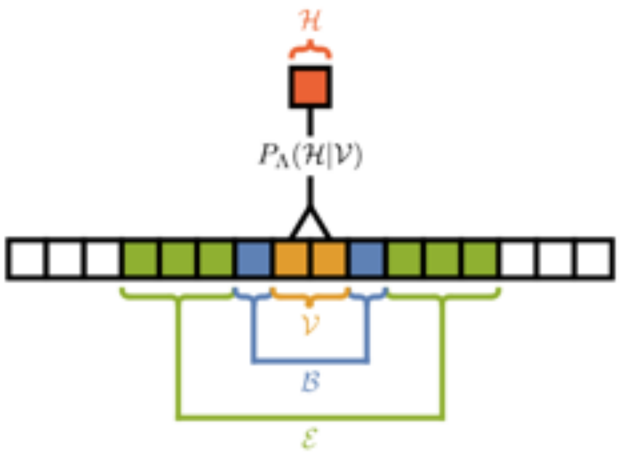
- Effective Hamiltonian:

$$\mathcal{K}'[\mathcal{X}'] = \log(Z_{\Lambda,0}[\mathcal{X}']) + \sum_{k=0}^{\infty} \frac{1}{k!} C_k[\mathcal{X}']$$



- The "rangeness" and "n-body-ness"

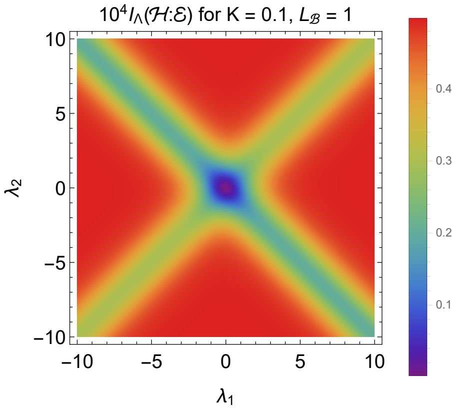
Optimality: 1D Ising model

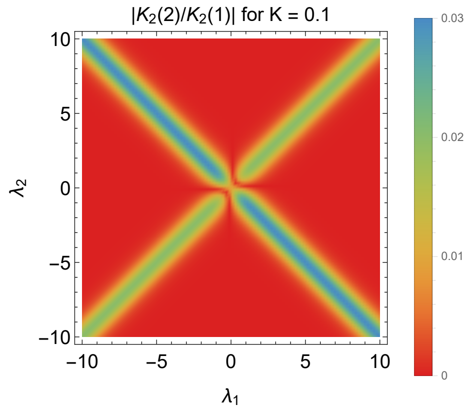


- Effective Hamiltonian:

$$\boxed{\mathcal{K}'[\mathcal{X}'] = \log(Z_{\Lambda,0}[\mathcal{X}']) + \sum_{k=0}^{\infty} \frac{1}{k!} C_k[\mathcal{X}']}$$

- The “rangeness” and “n-body-ness”

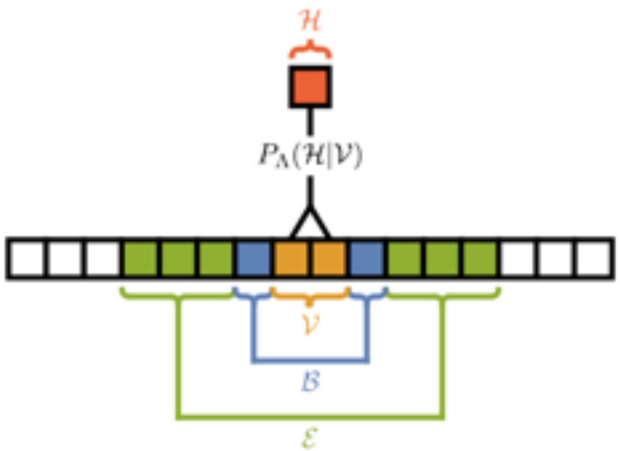




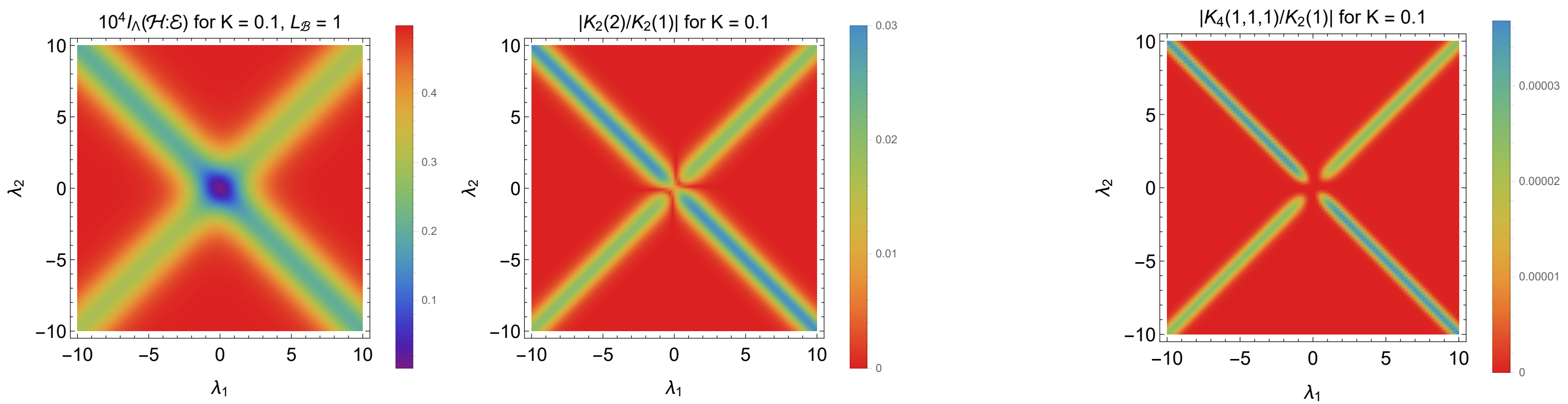
Optimality: 1D Ising model

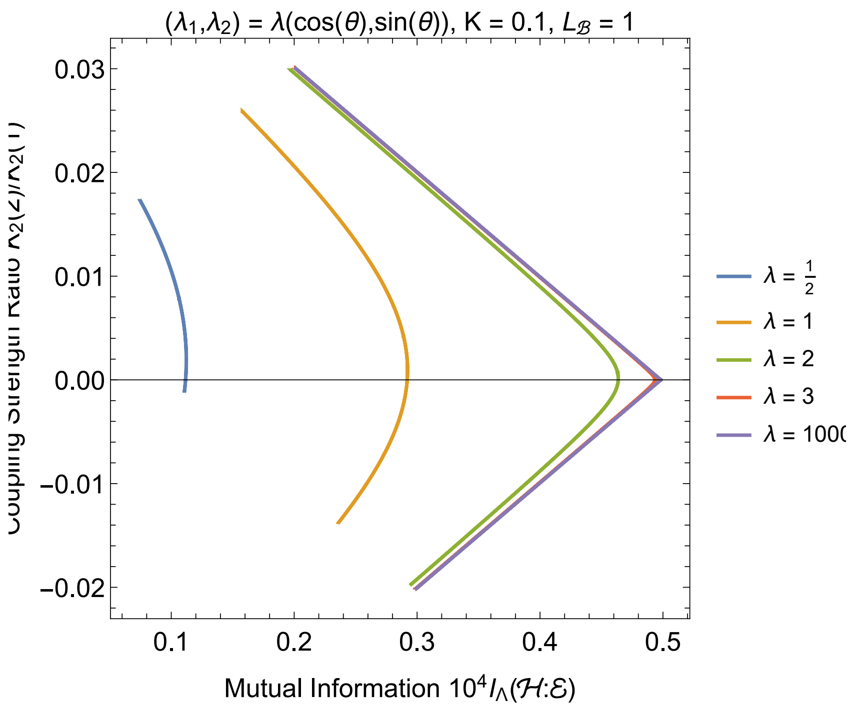
- Effective Hamiltonian:

$$\mathcal{K}'[\mathcal{X}'] = \log(Z_{\Lambda,0}[\mathcal{X}']) + \sum_{k=0}^{\infty} \frac{1}{k!} C_k[\mathcal{X}']$$



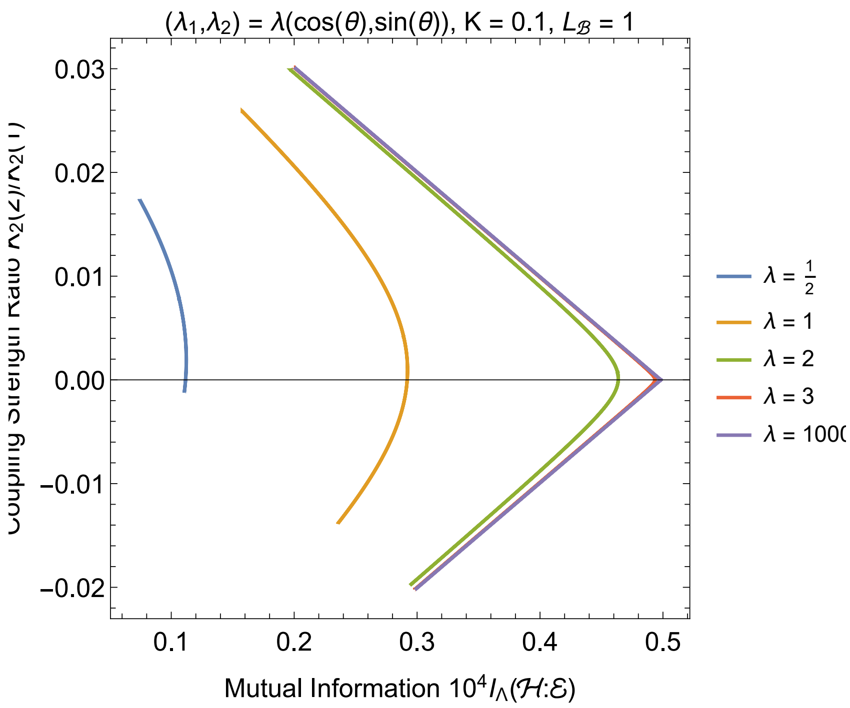
- The "rangeness" and "n-body-ness"



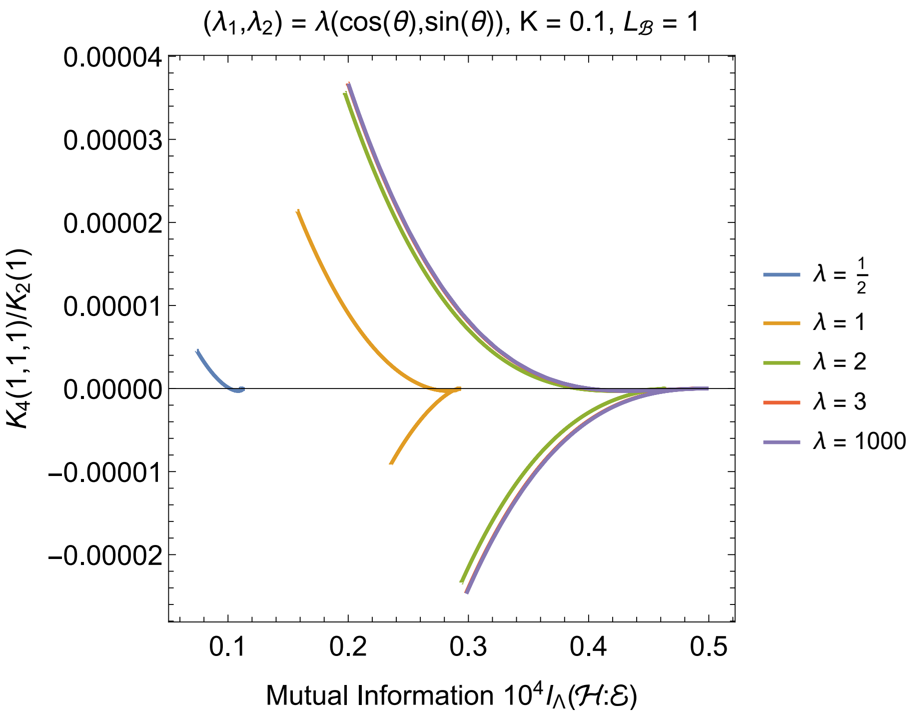


$(\lambda_1,\lambda_2) = \lambda(\cos(\theta),\sin(\theta))$, K = 0.1, $L_{\mathcal{B}} = 1$
0.03
0.02
0.01
0.00
−0.01
−0.02
0.1
0.2
0.3
0.4
0.5
Mutual Information $10^4 I_\Lambda(\mathcal{H}:\mathcal{E})$
$\lambda = \frac{1}{2}$
$\lambda = 1$
$\lambda = 2$
$\lambda = 3$
$\lambda = 1000$

ETH zürich



(λ1,λ2) = λ(cos(θ),sin(θ)), K = 0.1, L𝓑 = 1
0.03
0.02
0.01
0.00
−0.01
−0.02
0.1
0.2
0.3
0.4
0.5
Mutual Information 10^4 IΛ(𝓗:𝓔)
λ = 1/2
λ = 1
λ = 2
λ = 3
λ = 100(



(λ1,λ2) = λ(cos(θ),sin(θ)), K = 0.1, L𝓑 = 1
0.00004
0.00003
0.00002
0.00001
0.00000
−0.00001
−0.00002
K4(1,1,1)/K2(1)
0.1
0.2
0.3
0.4
0.5
Mutual Information 10^4 IΛ(𝓗:𝓔)
λ = 1/2
λ = 1
λ = 2
λ = 3
λ = 1000

Disordered systems

Disordered systems

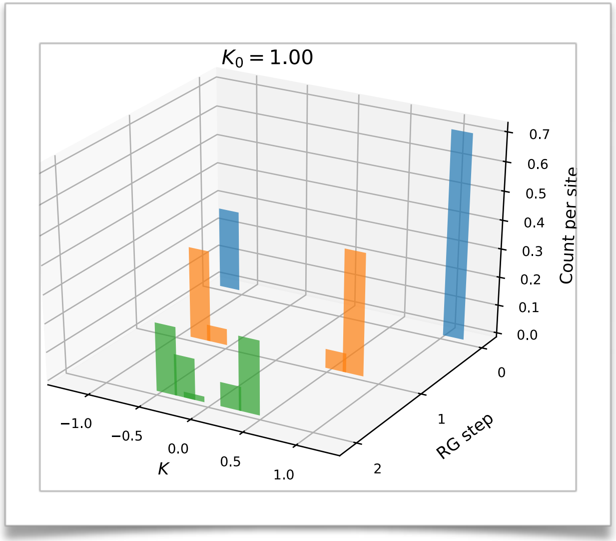
- RG acts on distributions of couplings:

Disordered systems

- RG acts on distributions
  of couplings: $\mathcal{P}(\mathbf{K}) \to \mathcal{P}'(\mathbf{K})$

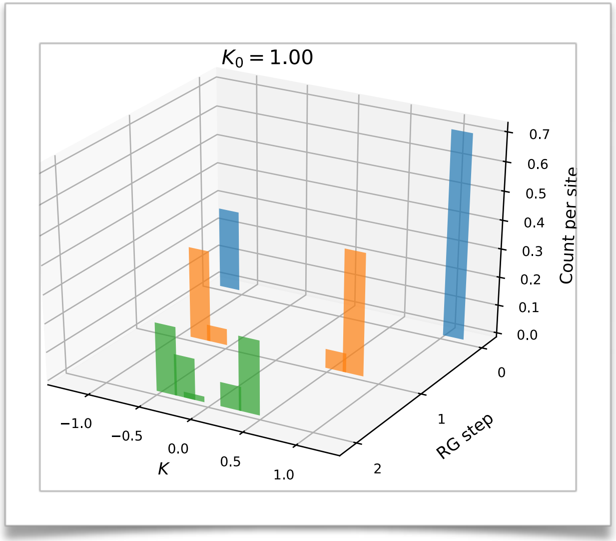
Disordered systems

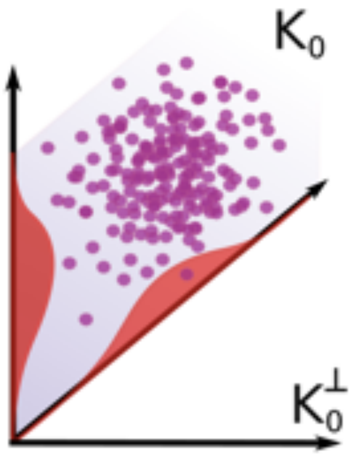
- RG acts on distributions of couplings: $\mathcal{P}(\mathbf{K}) \to \mathcal{P}'(\mathbf{K})$



Disordered systems

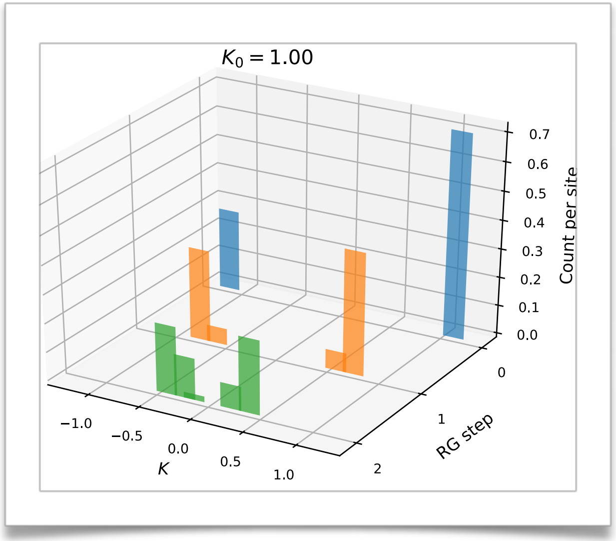
- RG acts on distributions of couplings: $\mathcal{P}(\mathbf{K}) \to \mathcal{P}'(\mathbf{K})$

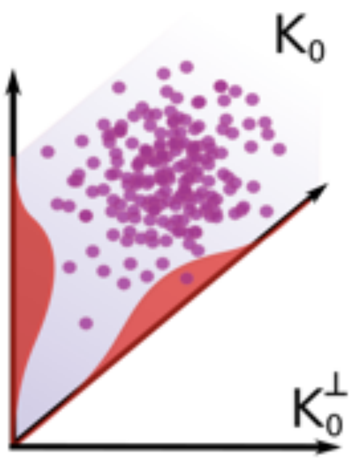




Disordered systems

- RG acts on distributions of couplings: $\mathcal{P}(\mathbf{K}) \to \mathcal{P}'(\mathbf{K})$

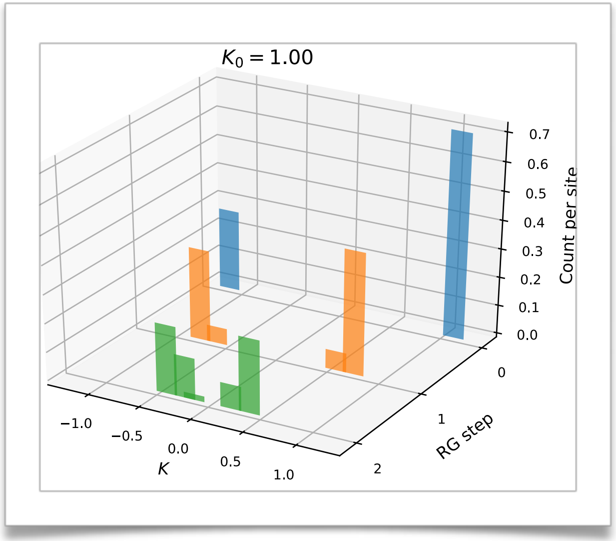


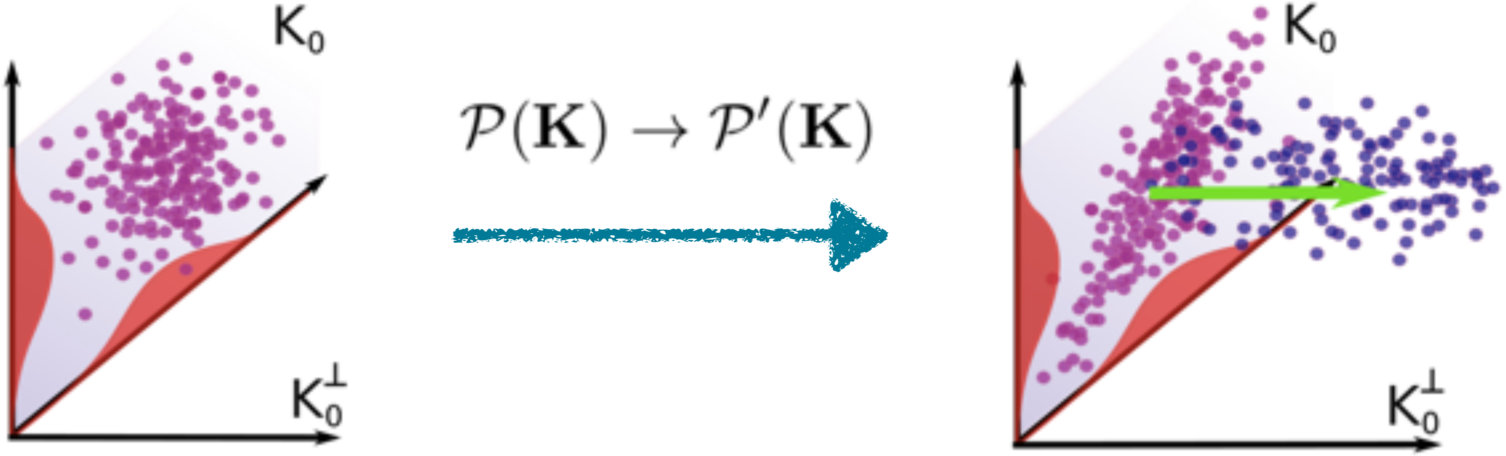


$\mathcal{P}(\mathbf{K}) \to \mathcal{P}'(\mathbf{K})$

Disordered systems

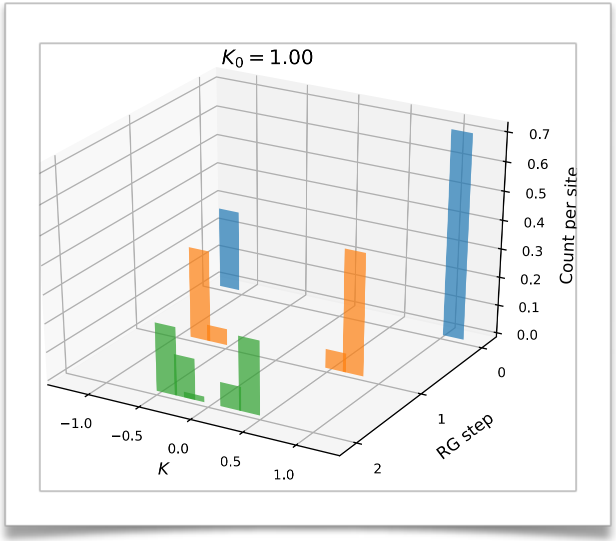
- RG acts on distributions of couplings: $\mathcal{P}(\mathbf{K}) \to \mathcal{P}'(\mathbf{K})$

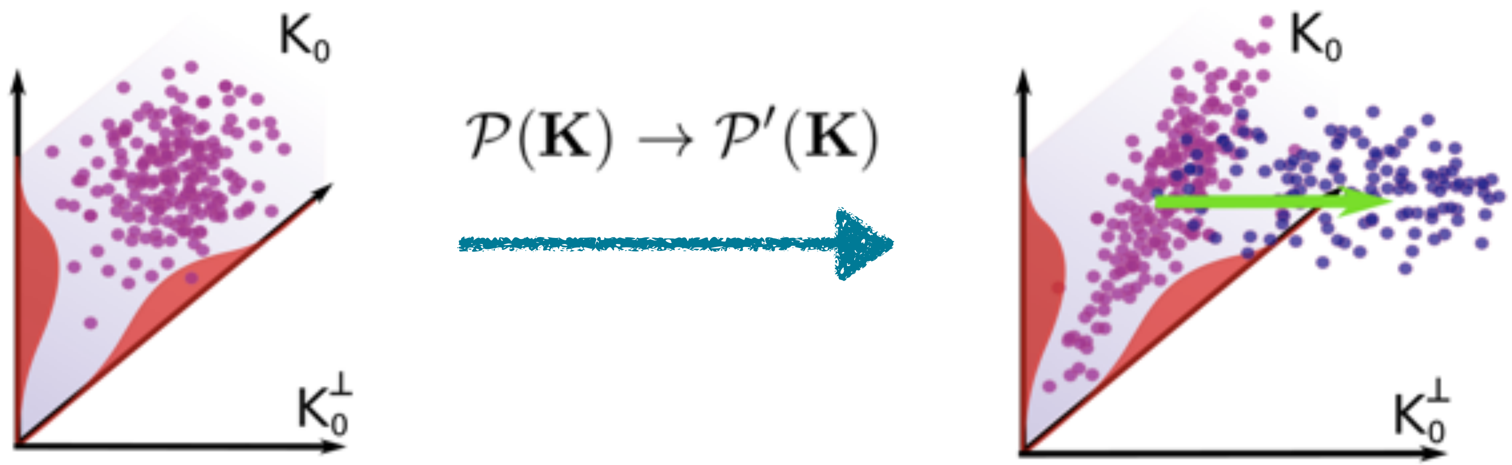




Disordered systems

- RG acts on distributions of couplings: $\mathcal{P}(\mathbf{K}) \to \mathcal{P}'(\mathbf{K})$
- Assume perfect RSMI capture

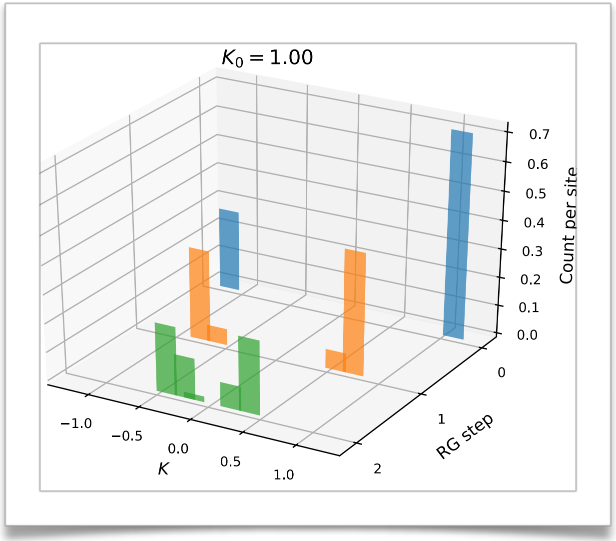


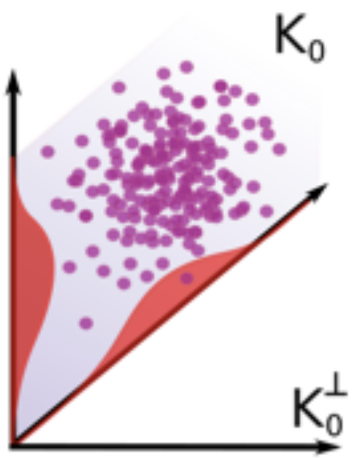


Disordered systems

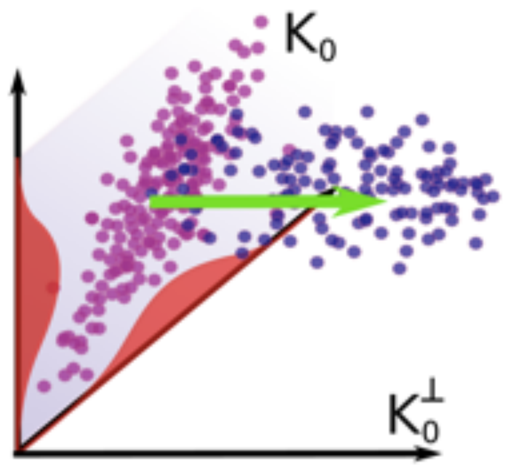
- RG acts on distributions of couplings: $\mathcal{P}(\mathbf{K}) \to \mathcal{P}'(\mathbf{K})$
- Assume perfect RSMI capture

The factorization is stable to local disorder changes





$\mathcal{P}(\mathbf{K}) \to \mathcal{P}'(\mathbf{K})$



Example 3: disordered 1d Ising model

ETH zürich

Example 3: disordered 1d Ising model

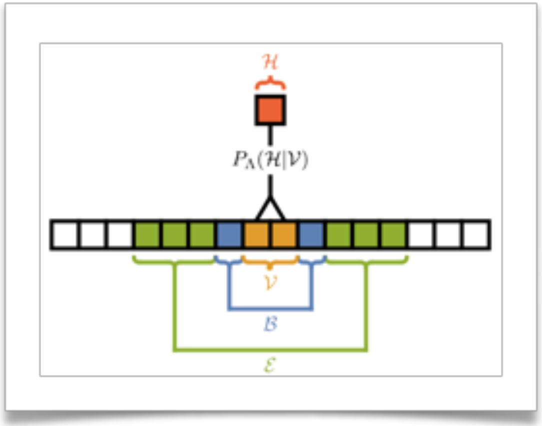
$$\mathcal{P}'(\mathbf{K}') = \int \delta(\mathbf{K}' - \mathbf{K}'(\mathbf{K}))\mathbf{d}\mathcal{P}(\mathbf{K})$$

Example 3: disordered 1d Ising model

$$\mathcal{P}'(\mathbf{K}') = \int \delta(\mathbf{K}' - \mathbf{K}'(\mathbf{K}))\mathbf{d}\mathcal{P}(\mathbf{K})$$

- Can be obtained by sampling

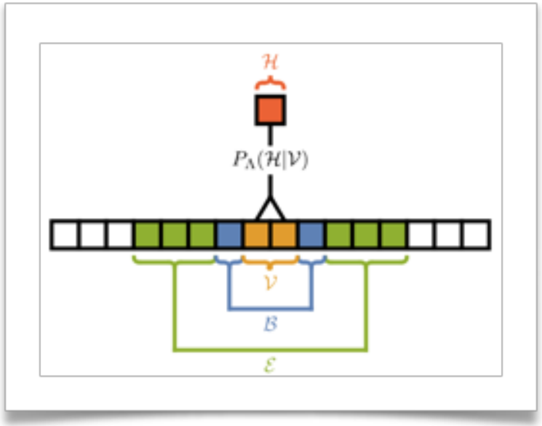
Example 3: disordered 1d Ising model



$$\mathcal{P}'(\mathbf{K}') = \int \delta(\mathbf{K}' - \mathbf{K}'(\mathbf{K}))\mathbf{d}\mathcal{P}(\mathbf{K})$$

- Can be obtained by sampling

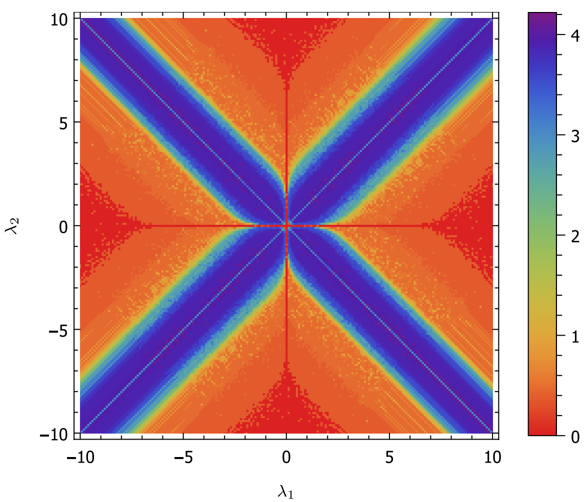
Example 3: disordered 1d Ising model



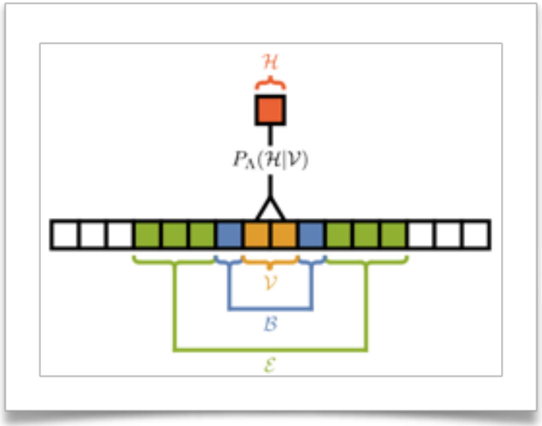
$$\mathcal{P}'(\mathbf{K}') = \int \delta(\mathbf{K}' - \mathbf{K}'(\mathbf{K}))\mathbf{d}\mathcal{P}(\mathbf{K})$$

- Can be obtained by sampling

KL-divergence



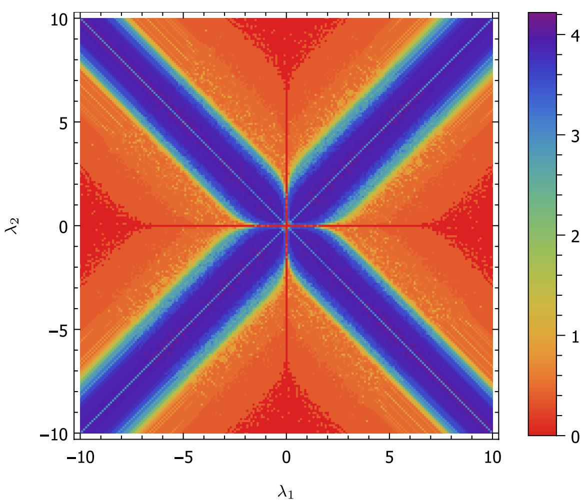
Example 3: disordered 1d Ising model



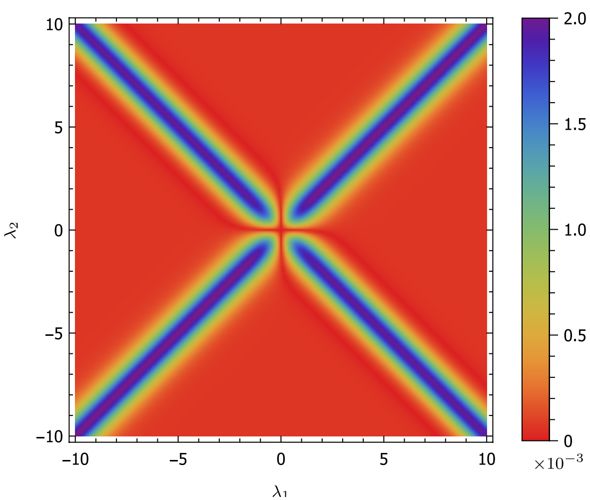
$$\mathcal{P}'(\mathbf{K}') = \int \delta(\mathbf{K}' - \mathbf{K}'(\mathbf{K}))\mathbf{d}\mathcal{P}(\mathbf{K})$$

- Can be obtained by sampling

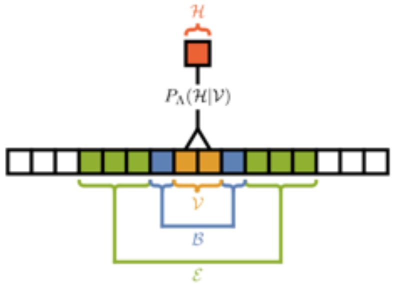
KL-divergence



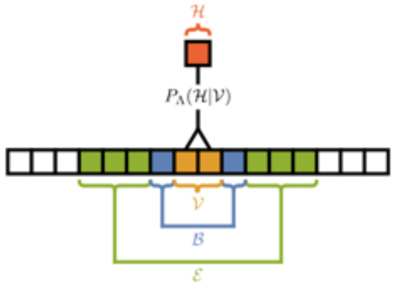
dCOM



The “shape” of coarse-grained variables



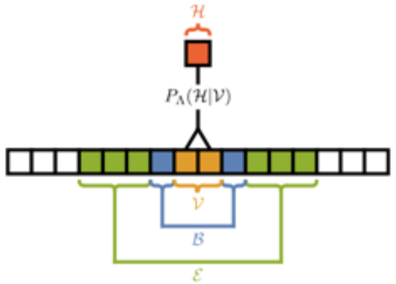
The "shape" of coarse-grained variables



- The hidden couples to:

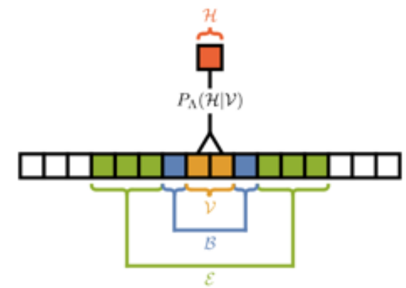
The “shape” of coarse-grained variables

- The hidden couples to: $\mathcal{V}_{\Lambda_i} = \frac{1}{||\Lambda_i||} \sum_j \lambda_{ij} v_j$



The “shape” of coarse-grained variables

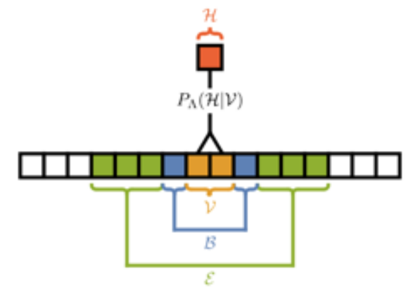
- The hidden couples to: $\mathcal{V}_{\Lambda_i} = \frac{1}{||\Lambda_i||} \sum_j \lambda_{ij} v_j$



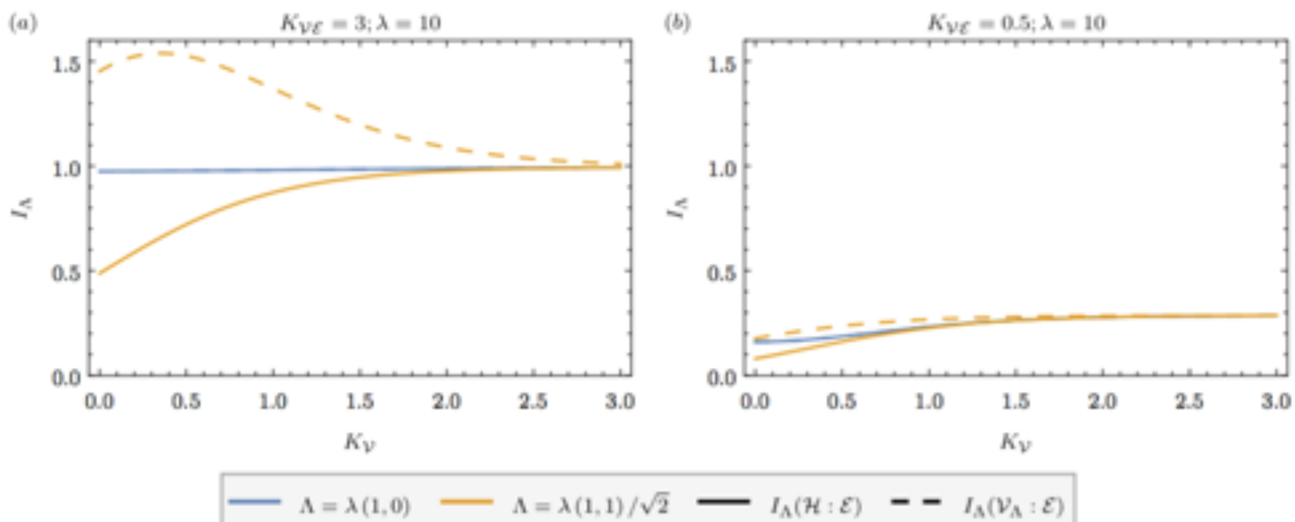
$$\boxed{I_\Lambda(\mathcal{H} : \mathcal{E}) \leq I(\mathcal{V}_\Lambda : \mathcal{E}) \leq I(\mathcal{V} : \mathcal{E})}$$

The “shape” of coarse-grained variables

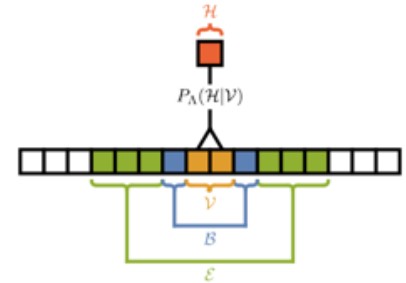
- The hidden couples to: $\mathcal{V}_{\Lambda_i} = \frac{1}{||\Lambda_i||}\sum_j \lambda_{ij} v_j$



$$\boxed{I_\Lambda(\mathcal{H}:\mathcal{E}) \leq I(\mathcal{V}_\Lambda:\mathcal{E}) \leq I(\mathcal{V}:\mathcal{E})}$$

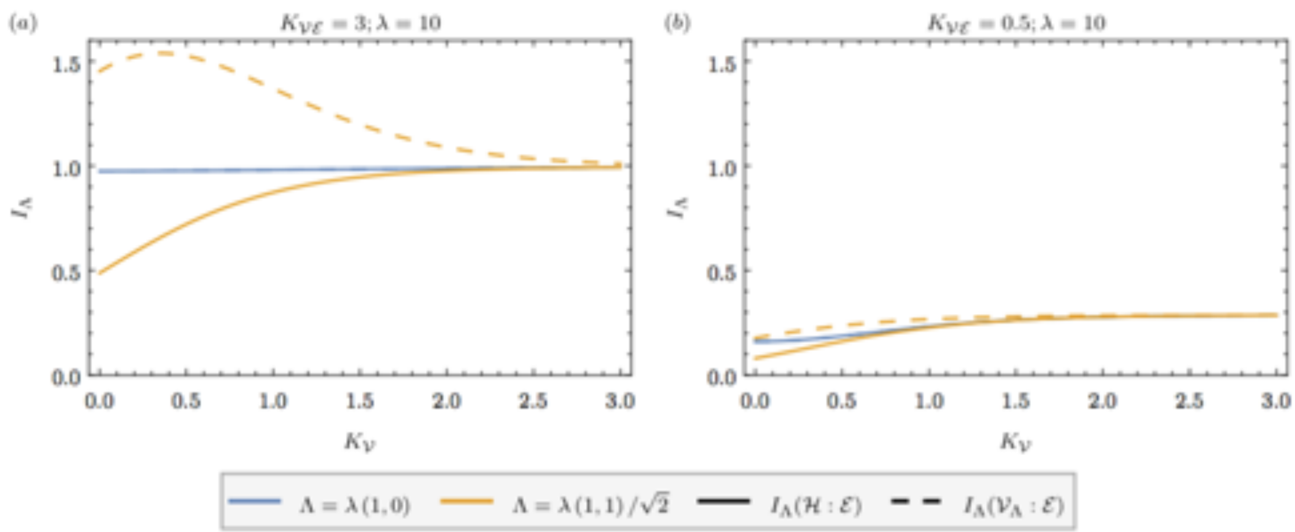


The “shape” of coarse-grained variables



- The hidden couples to: $\mathcal{V}_{\Lambda_i} = \frac{1}{||\Lambda_i||}\sum_j \lambda_{ij} v_j$

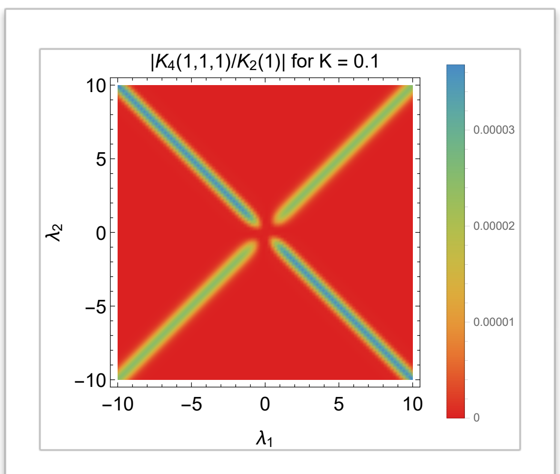
$$\boxed{I_\Lambda(\mathcal{H}:\mathcal{E}) \leq I(\mathcal{V}_\Lambda:\mathcal{E}) \leq I(\mathcal{V}:\mathcal{E})}$$



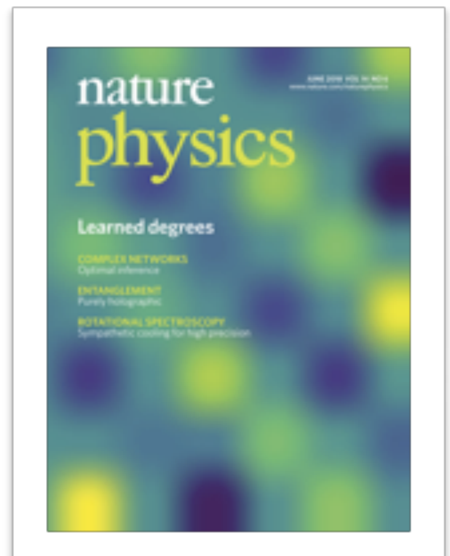
$$\boxed{I_\Lambda(\mathcal{H}:\mathcal{E}) = I(\mathcal{V}_\Lambda:\mathcal{E}) - I(\mathcal{V}_\Lambda:\mathcal{E}|\mathcal{H})}$$

Conclusions & Outlook

- Information-theoretic view of real-space RG
- Applications to correlated disorder, soft matter
- New numerical techniques GRISE, MINE, ….



Lenggenhager, Ringel, Huber, MKJ, *arXiv:1809.09632*



MKJ and Z. Ringel *Nature Physics* ***14****, 578-582 (2018)*

Thank you!