ETH zürich

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

Patrick Lenggenhager

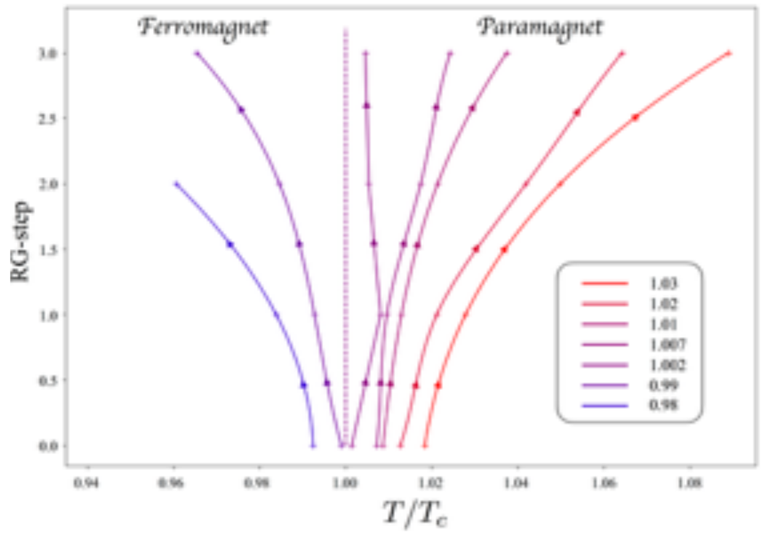
Outline

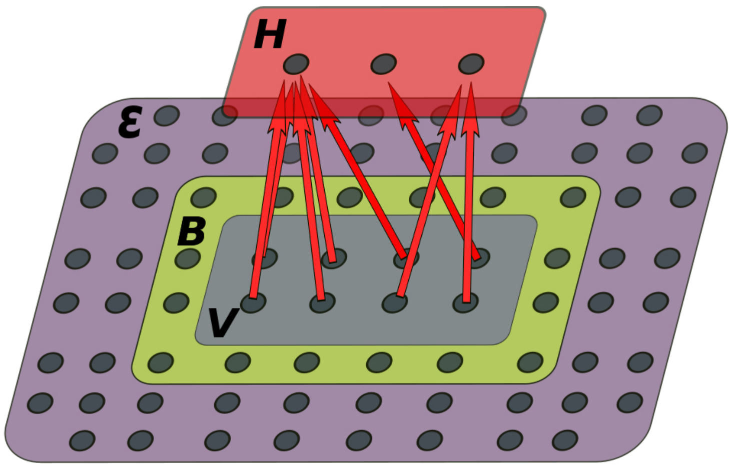
Outline

- Machine learning in condensed matter
- RBMs 101
- Information-theoretic approach to real-space RG
  - The Real Space Mutual Information algorithm
  - Results
  - “Optimality” of Mutual Information

The punchline

An information theoretic approach and an *unsupervised* machine learning algorithm performing real-space RG of classical statistical physical systems: degrees of freedom relevant for large length scales, RG flow, critical exponents.





ETH zürich

Phase transitions and classification

Phase transitions and classification

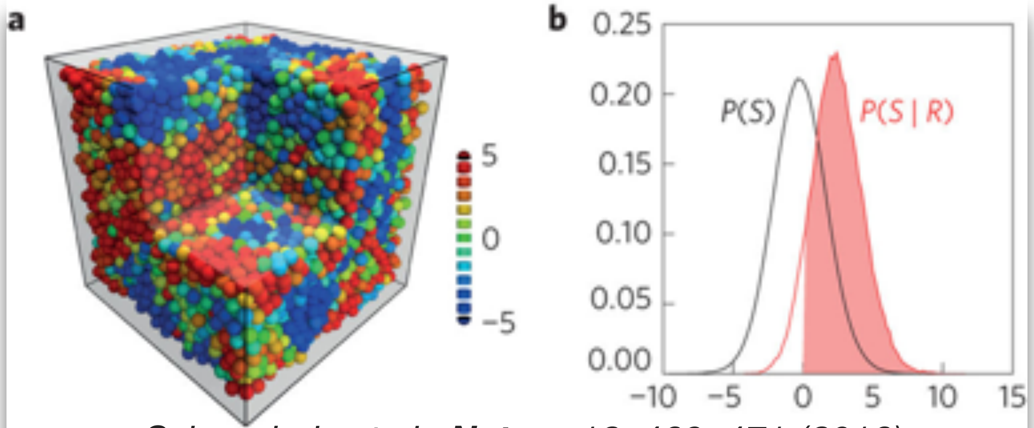
Lei Wang,
Phys. Rev. B **94,** 195105 *(2016)*

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)*

Phase transitions and classification

Phase transitions and classification



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

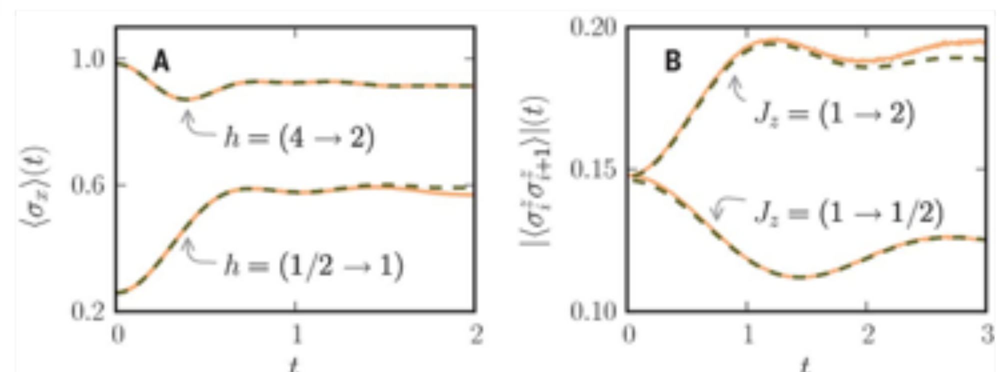
Phase transitions and classification

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)

Phase transitions and classification

State compression and representation

ETH zürich

Phase transitions and classification

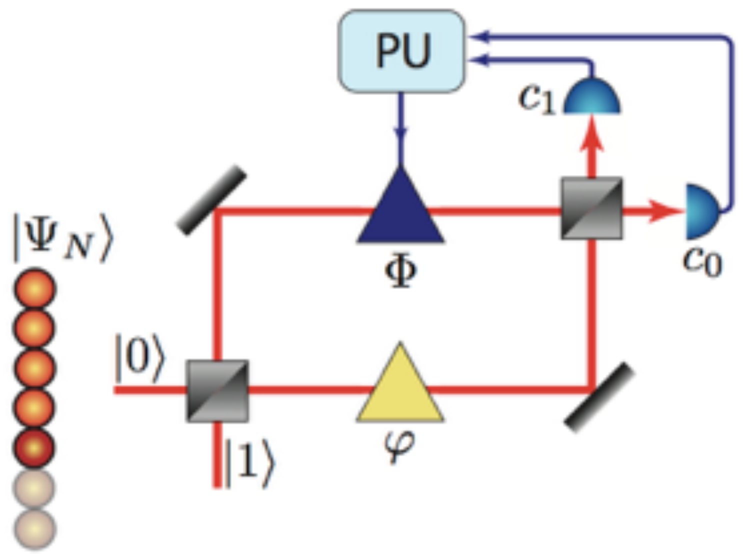
State compression and representation

Experimental / numerical protocols

Phase transitions and classification

State compression and representation

Experimental / numerical protocols



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

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

Phase transitions and classification

State compression and representation

Experimental / numerical protocols

Phase transitions and classification

State compression and representation

Experimental / numerical protocols

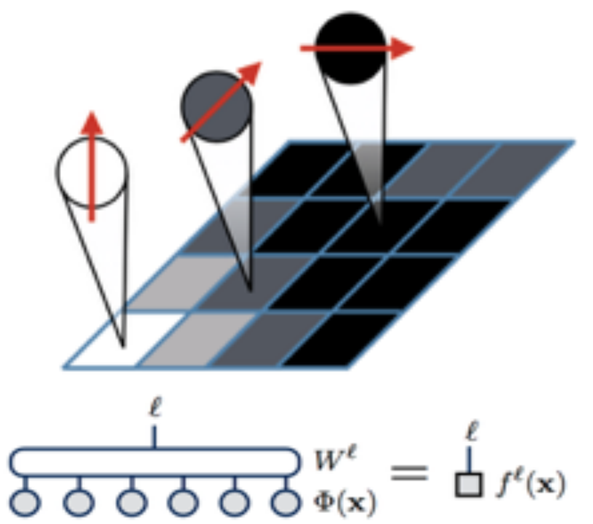
physics -> ML

Phase transitions and classification

State compression and representation

Experimental / numerical protocols

physics -> ML



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

Machine Learning

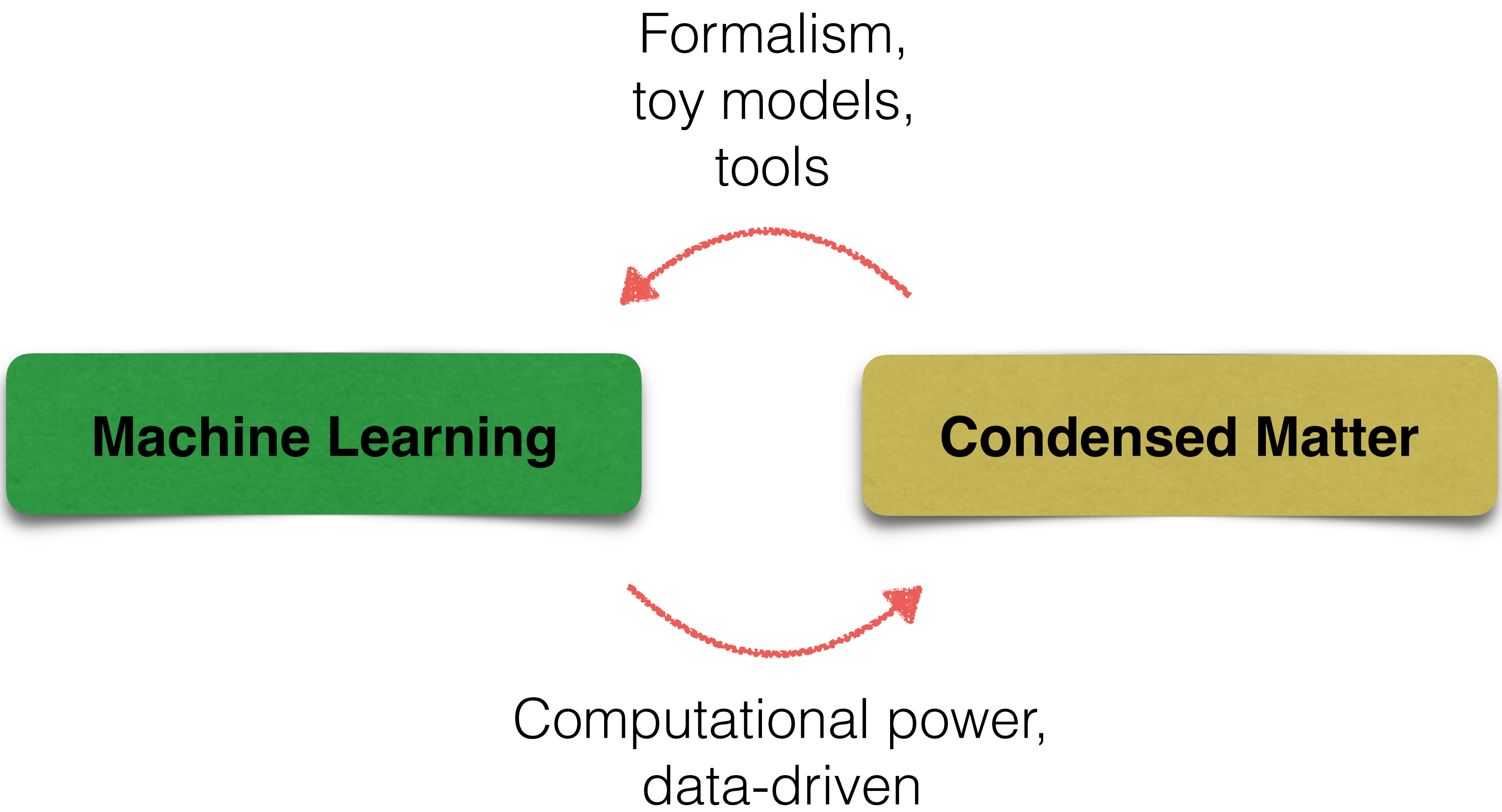
Machine Learning

Condensed Matter

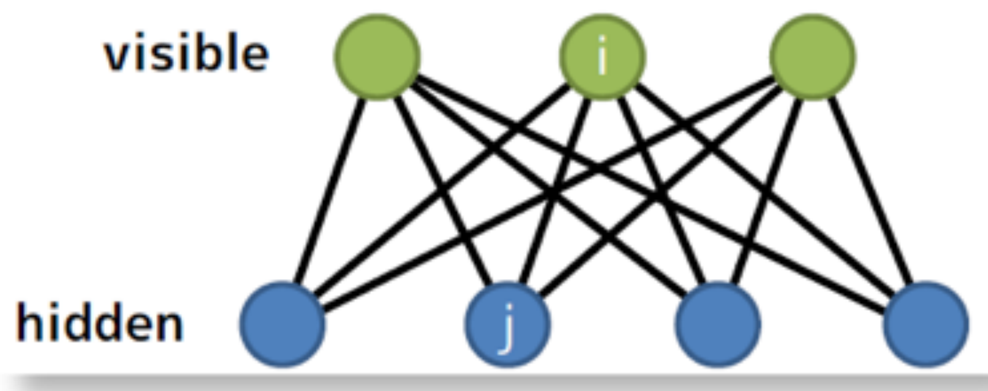
Machine Learning

Condensed Matter

Computational power,
data-driven

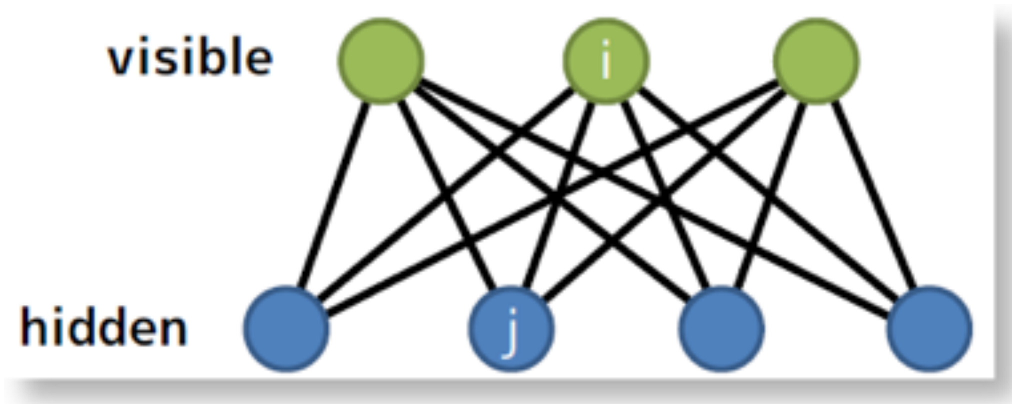

Formalism,
toy models,
tools
Machine Learning
Condensed Matter
Computational power,
data-driven

(Restricted) Boltzmann Machines



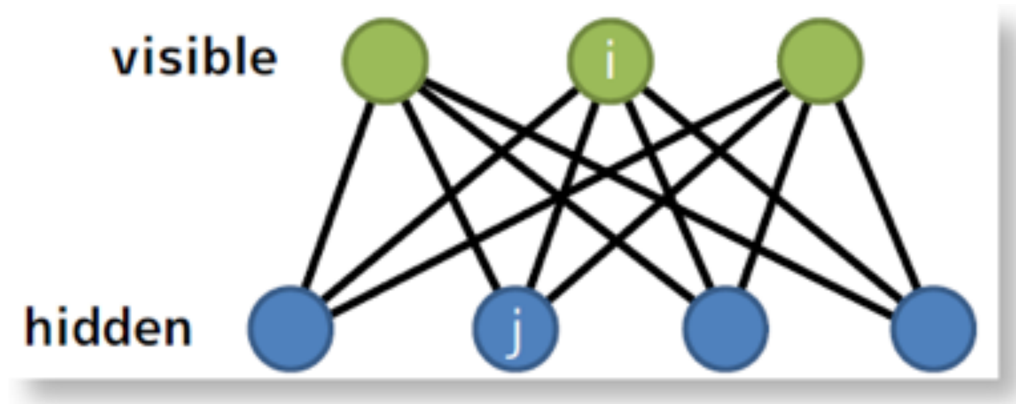
(Restricted) Boltzmann Machines

- Stochastic networks



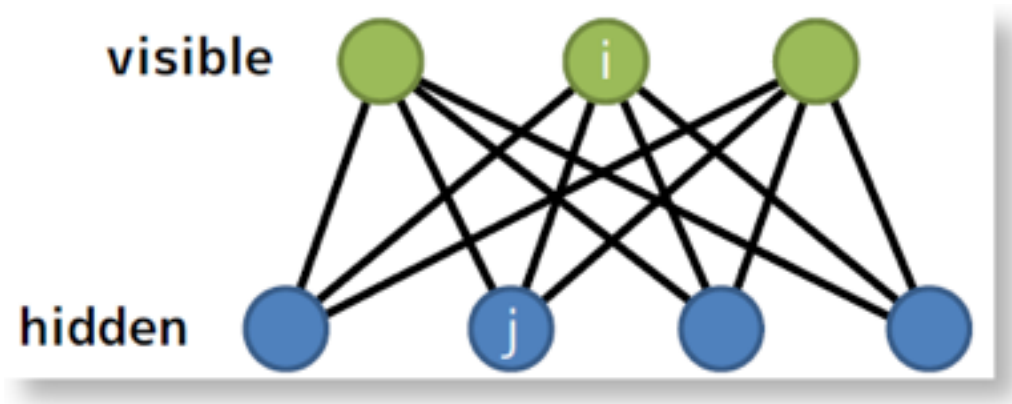
(Restricted) Boltzmann Machines

- Stochastic networks
- Model probability distributions



(Restricted) Boltzmann Machines

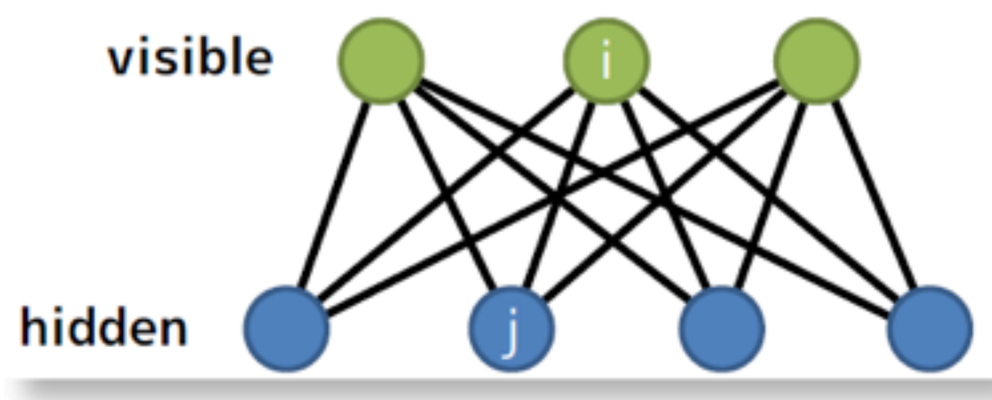
- Stochastic networks
- Model probability distributions



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

(Restricted) Boltzmann Machines

- Stochastic networks
- Model probability distributions

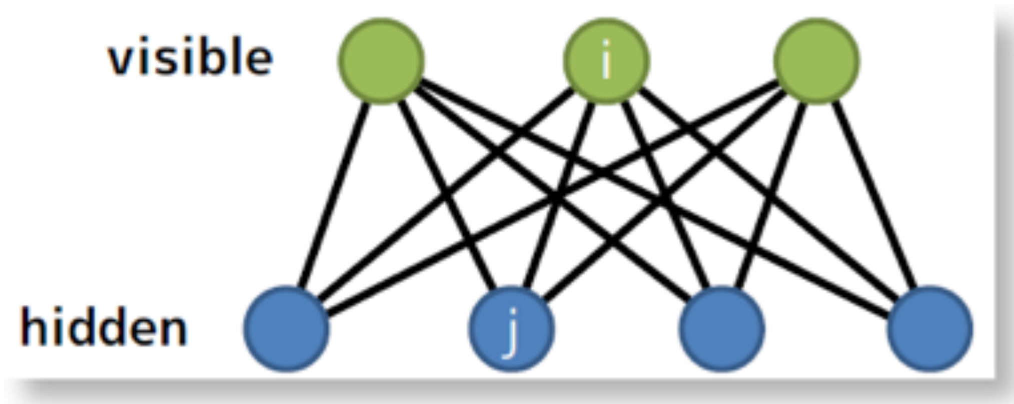


$$E_\Theta \equiv E_{a,b,\theta}(\mathcal{V},\mathcal{H}) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_{ij} v_i \theta_{ij} h_j$$

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

(Restricted) Boltzmann Machines

- Stochastic networks
- Model probability distributions



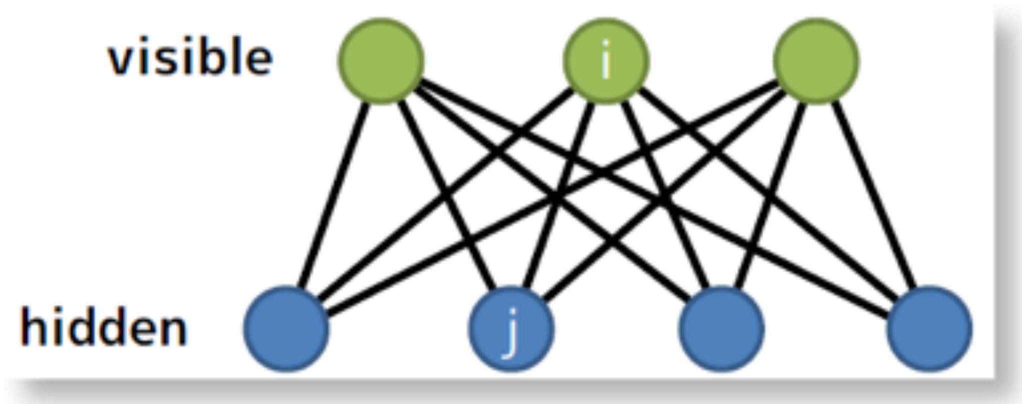
$$E_\Theta \equiv E_{a,b,\theta}(\mathcal{V},\mathcal{H}) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_{ij} v_i \theta_{ij} h_j$$

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

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

(Restricted) Boltzmann Machines

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



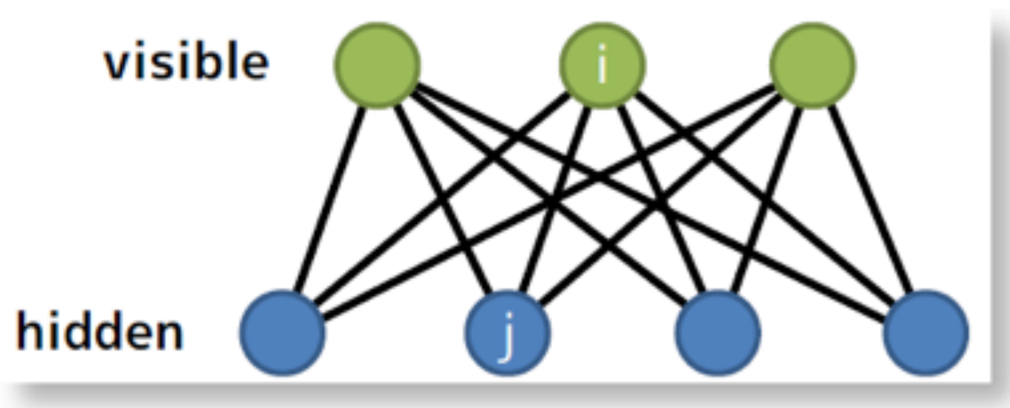
$$E_\Theta \equiv E_{a,b,\theta}(\mathcal{V},\mathcal{H}) = -\sum_i a_i v_i - \sum_j b_j h_j - \sum_{ij} v_i \theta_{ij} h_j$$

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

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

(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

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

(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)

(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)

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

(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)

(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

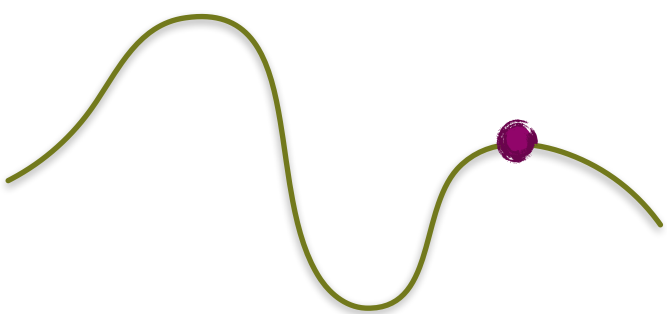
$$P_\Theta(\mathcal{V},\mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}$$

- 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

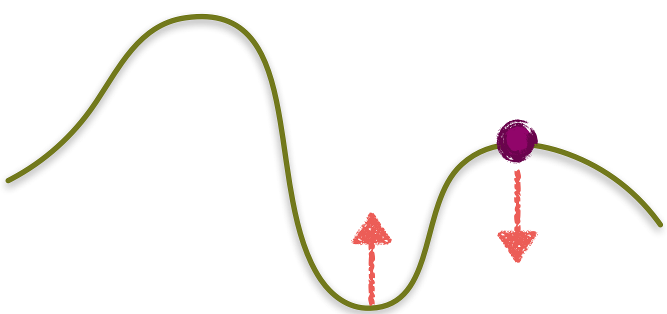
$$P_\Theta(\mathcal{V}, \mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}$$

- 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

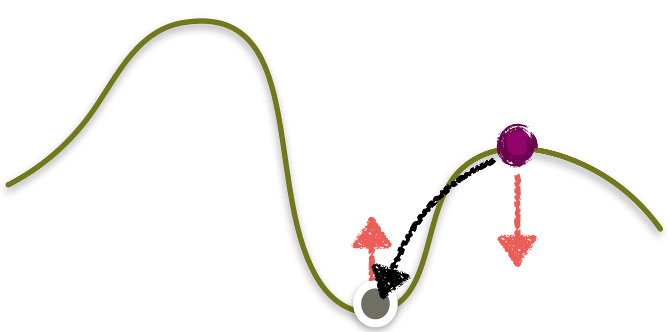
$$P_\Theta(\mathcal{V},\mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}$$

- 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

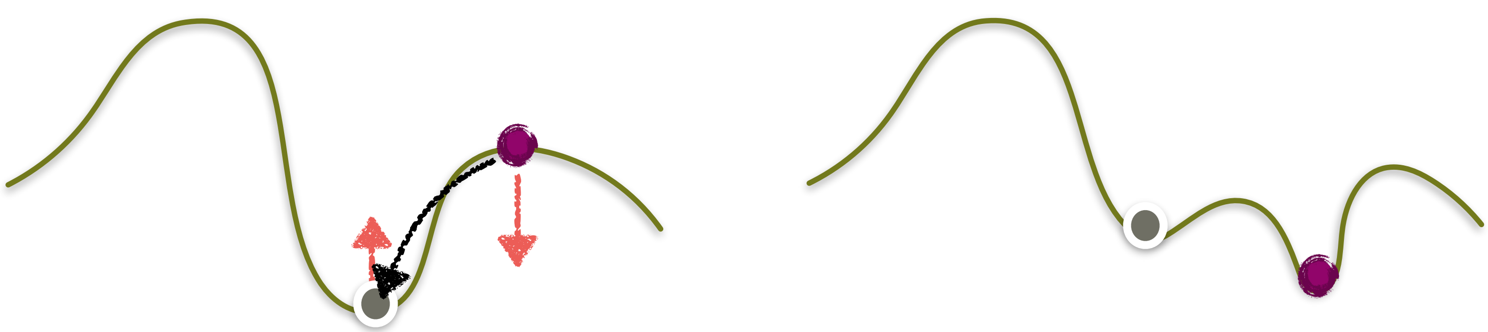
$$P_\Theta(\mathcal{V},\mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}$$

- First attempt: Maximal Likelihood (ML)

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

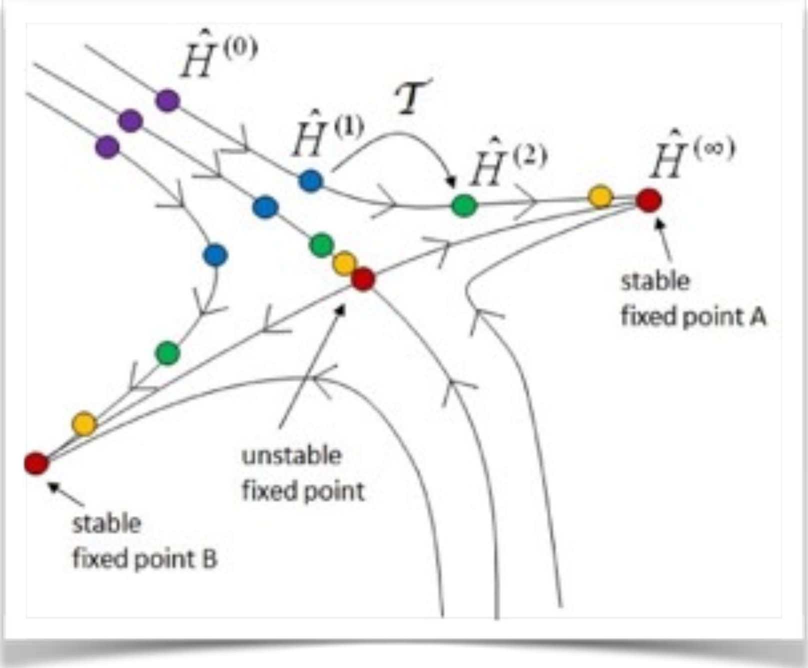
- Better solution: Contrastive Divergence (CD)

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



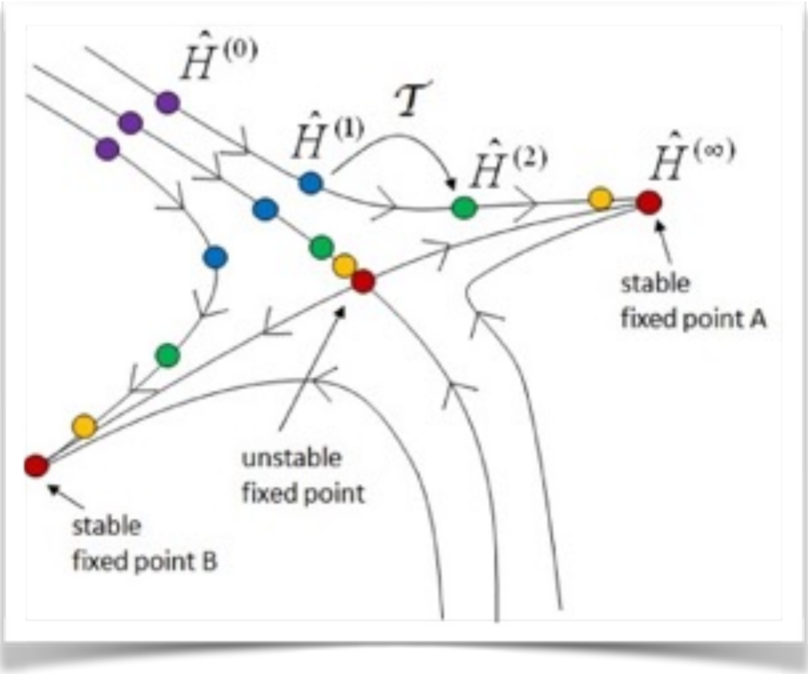
Renormalization Group

Renormalization Group



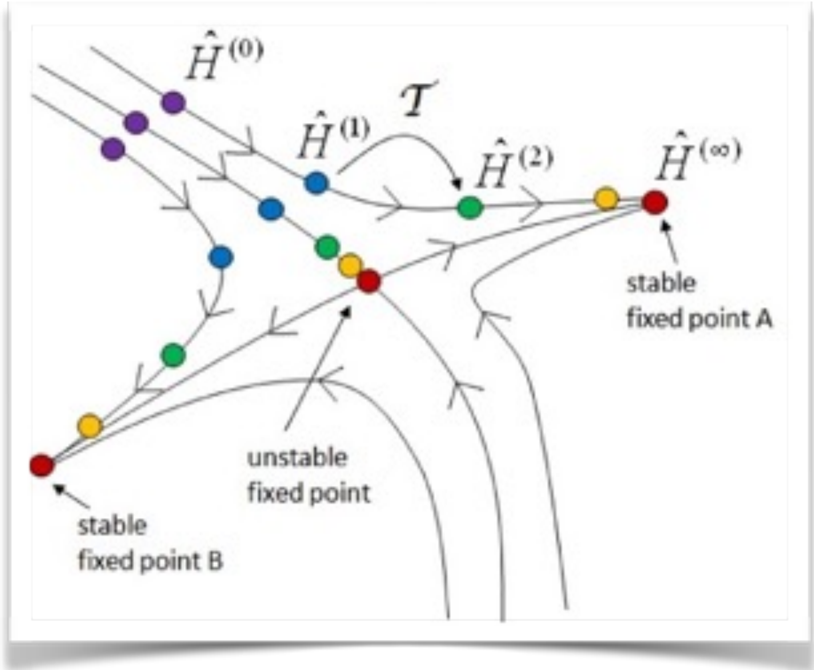
Renormalization Group

- Conceptually important: formalizes the notion of separation of scales



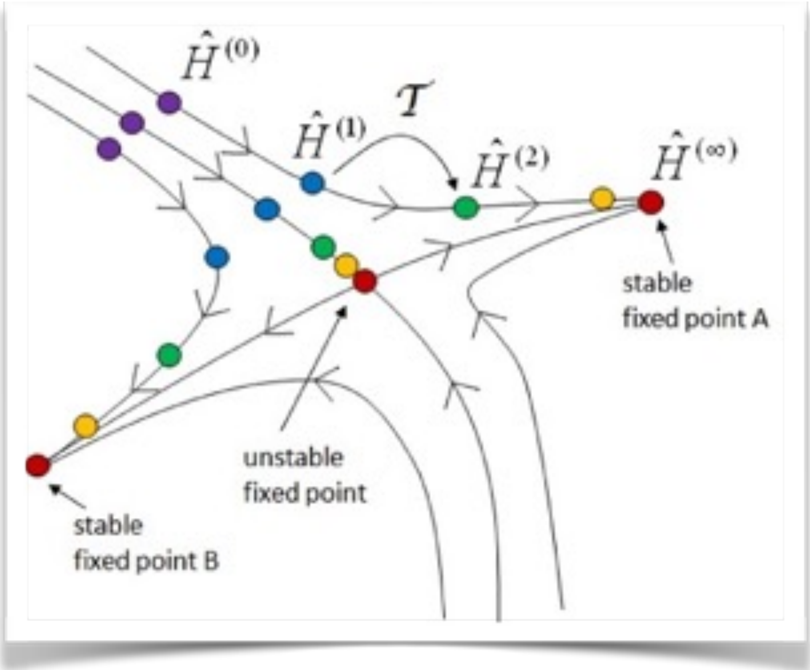
Renormalization Group

- Conceptually important: formalizes the notion of separation of scales
- Universality



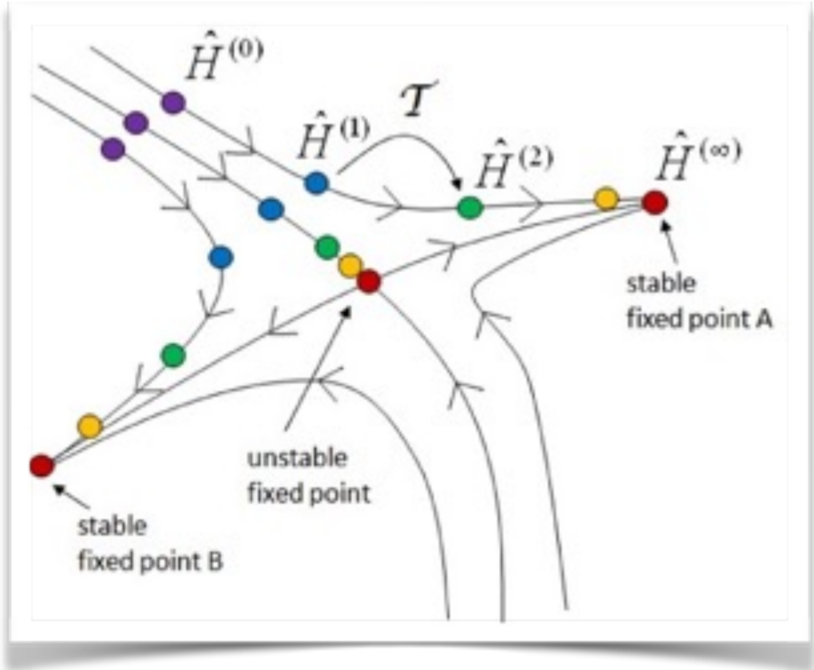
Renormalization Group

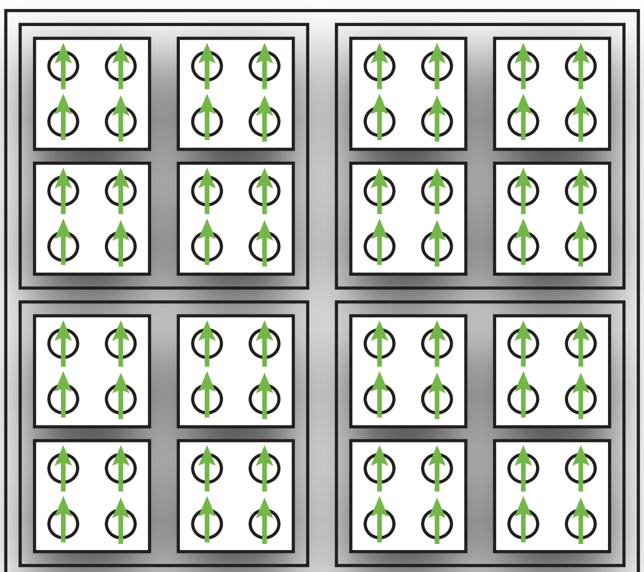
- Conceptually important: formalizes the notion of separation of scales
- Universality
- Many flavors exist: Wilsonian, DMRG, real-space, ….



Renormalization Group

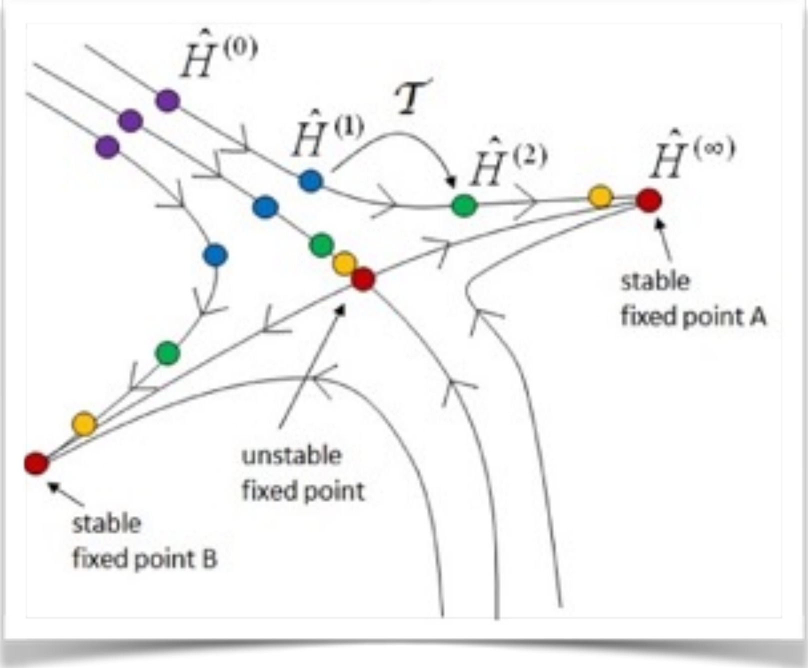
- Conceptually important: formalizes the notion of separation of scales
- Universality
- Many flavors exist: Wilsonian, DMRG, real-space, ….

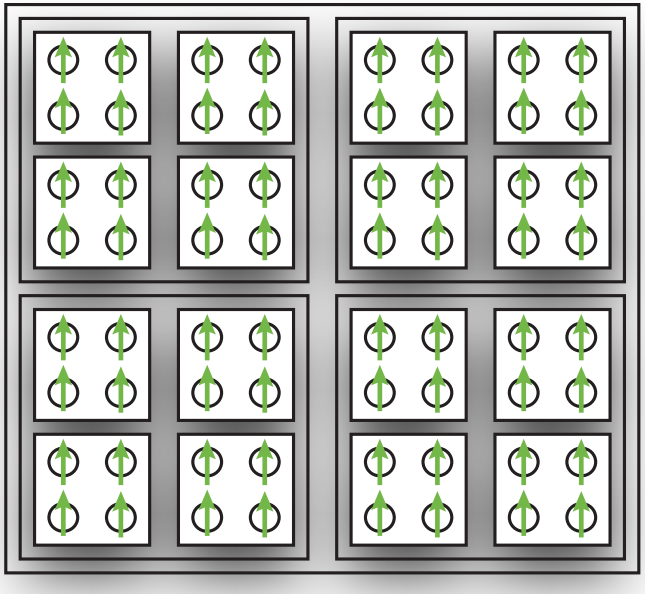




Renormalization Group

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





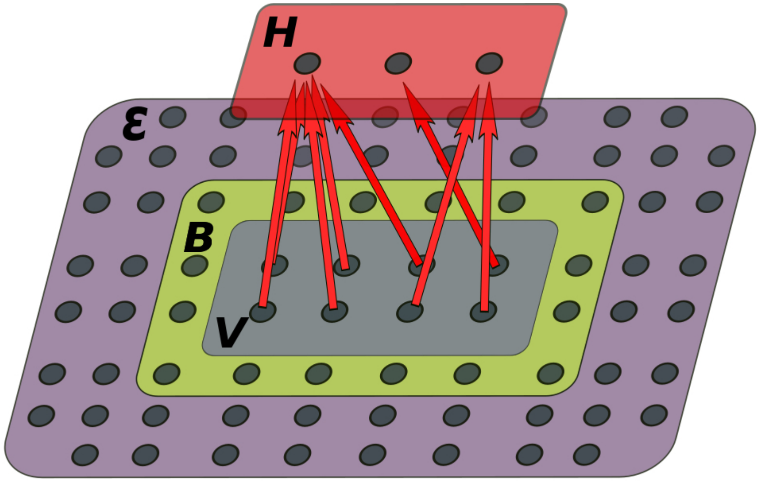
Leitmotiv: integrate out some 'fast' degrees of freedom to obtain effective theory of the 'slow' ones

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?

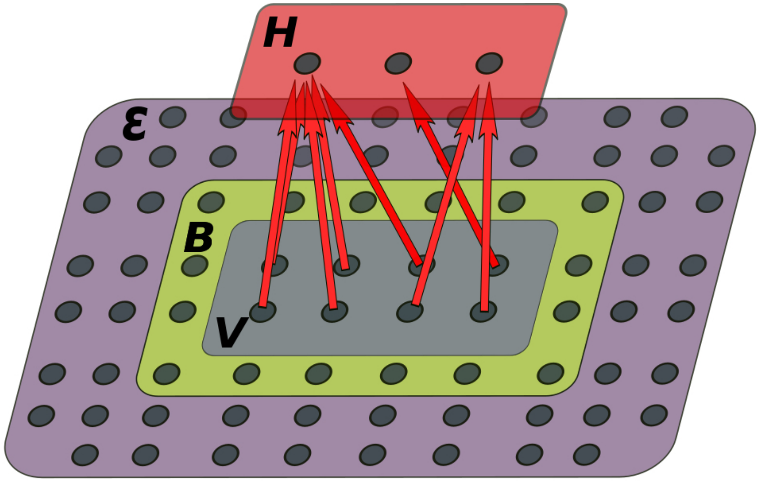
Real-space RG from Information Theory perspective



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

Real-space RG from Information Theory perspective

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

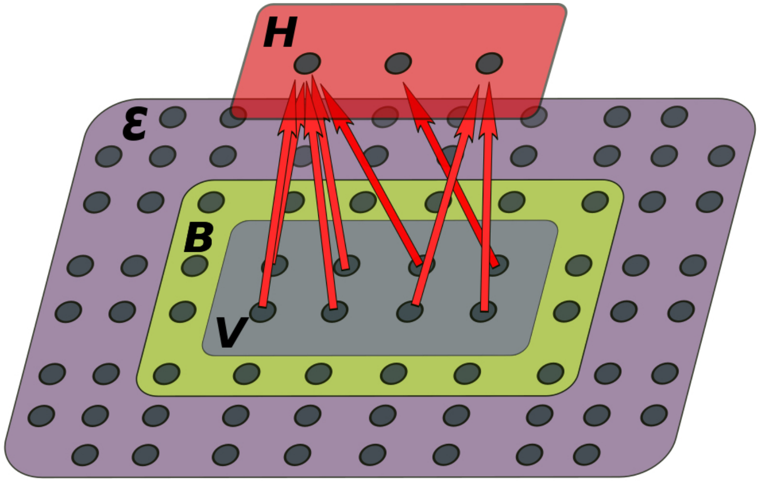


$$P(\mathcal{X}) = \frac{1}{Z} e^{\mathcal{K}(\mathcal{X})}$$

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 $\mathcal{V}$



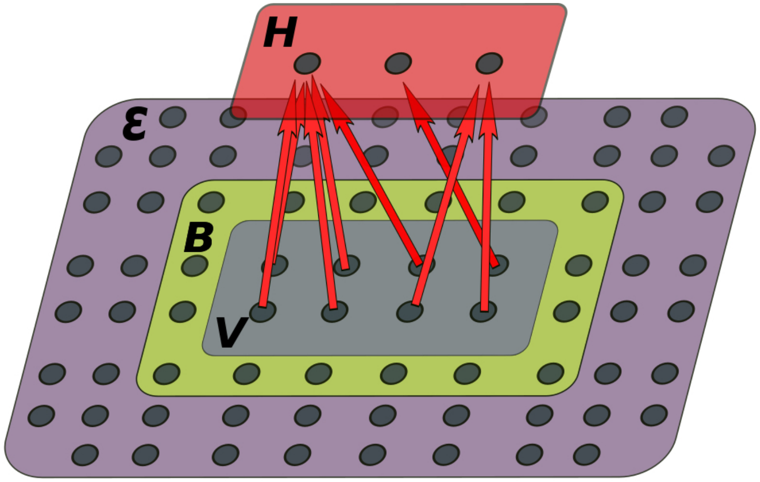
$$P(\mathcal{X}) = \frac{1}{Z} e^{\mathcal{K}(\mathcal{X})}$$

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 $\mathcal{V}$

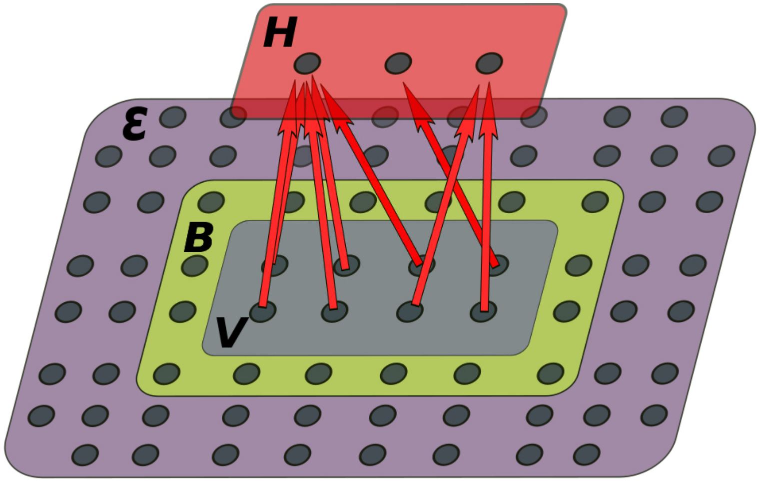
$$\boxed{\Lambda = (a_i, b_j, \lambda_i^j)}$$



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

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 $\mathcal{V}$

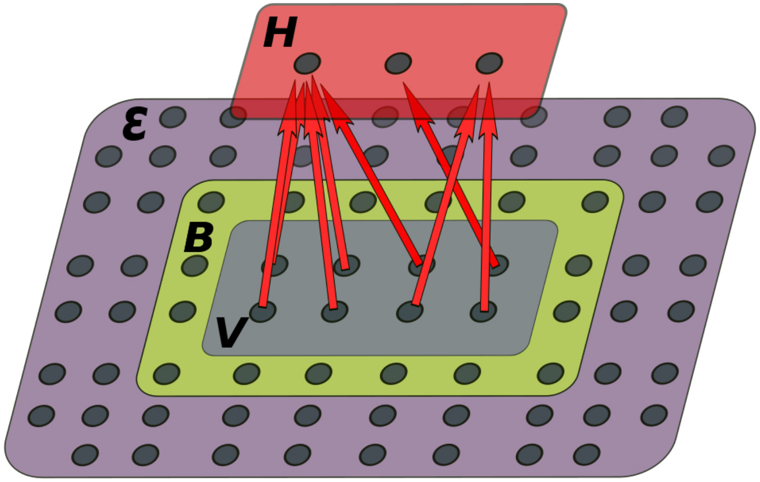
$$\Lambda = (a_i, b_j, \lambda_i^j)$$

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

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

Real-space RG from Information Theory perspective

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



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

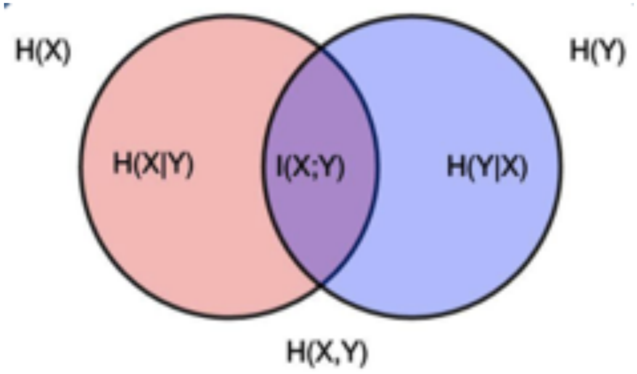
Task: Learn $P_\Lambda(\mathcal{H}|\mathcal{V})$ such that $\mathcal{H}$ tracks the *slow degrees of freedom* within region $\mathcal{V}$

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

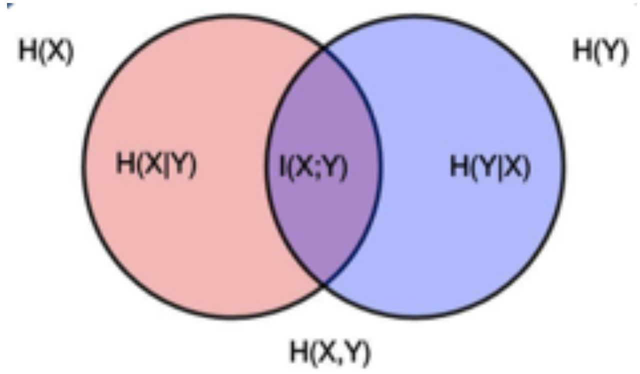
Method: Require that *slow degrees of freedom* maximize spatial mutual information

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

Mutual Information

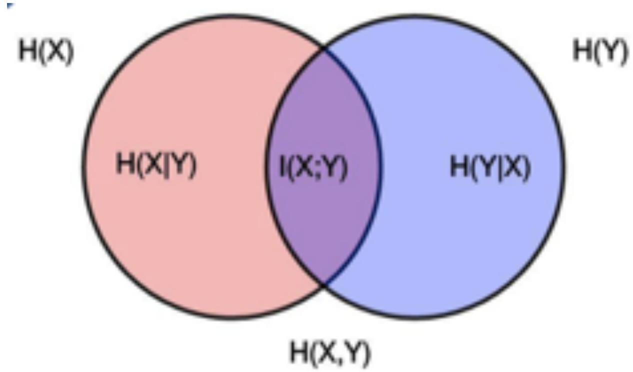


Mutual Information



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

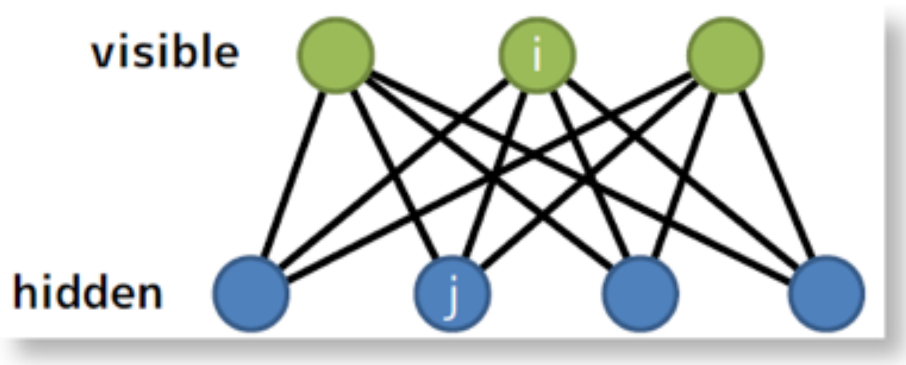
Mutual Information



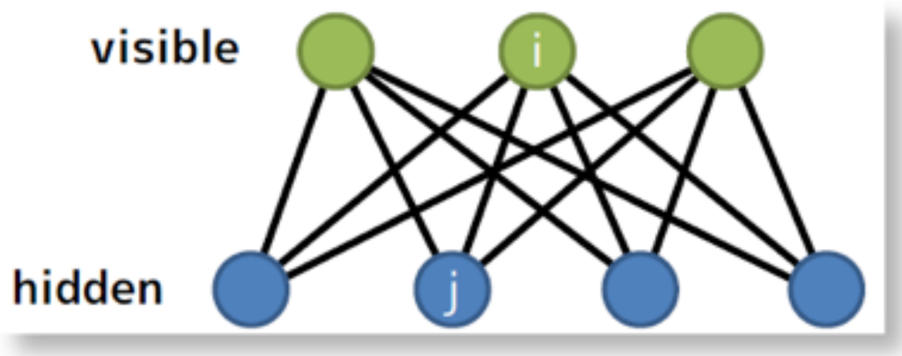
$$I_\Lambda(\mathcal{H}:\mathcal{E}) = \sum_{\mathcal{H},\mathcal{E}} P_\Lambda(\mathcal{E},\mathcal{H}) \log\left(\frac{P_\Lambda(\mathcal{E},\mathcal{H})}{P_\Lambda(\mathcal{H})P(\mathcal{E})}\right)$$

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

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

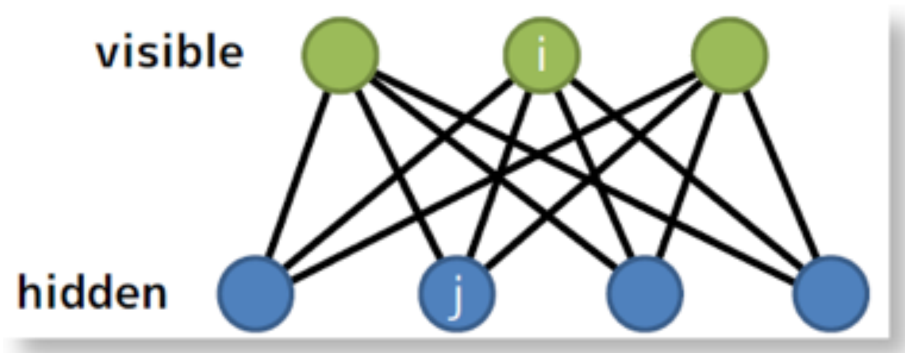


$$P_\Theta(\mathcal{V},\mathcal{H}) = \frac{1}{\mathcal{Z}} e^{-E_{a,b,\theta}(\mathcal{V},\mathcal{H})}$$



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

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



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

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

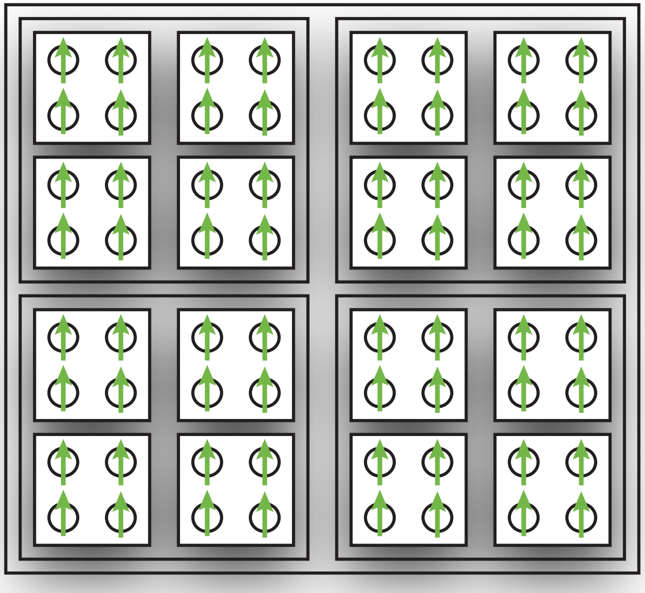
Test case 1: the 2D Ising model

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

Test case 1: the 2D Ising model

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

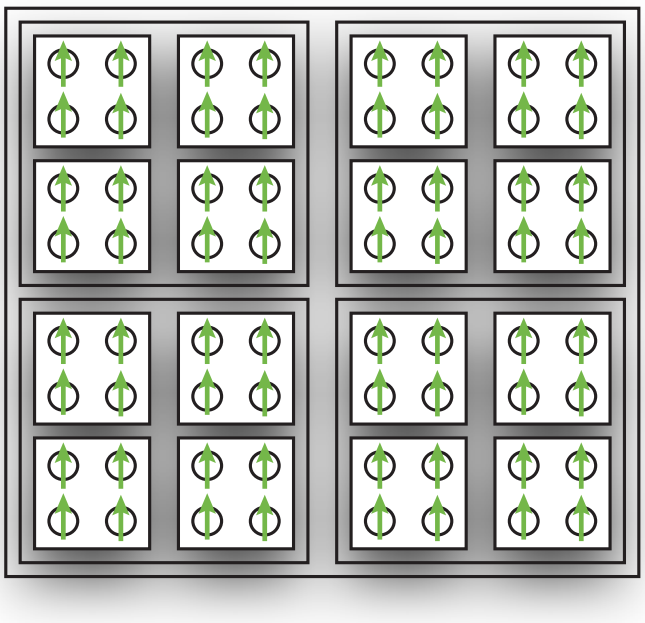
Migdal-Kadanoff block-spins:

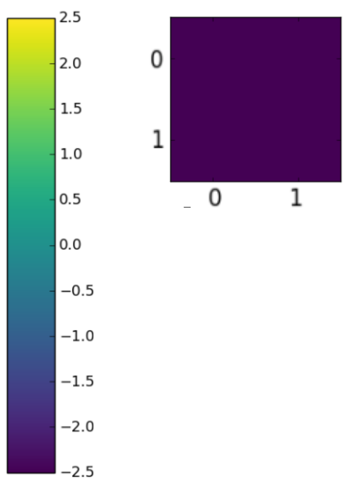


Test case 1: the 2D Ising model

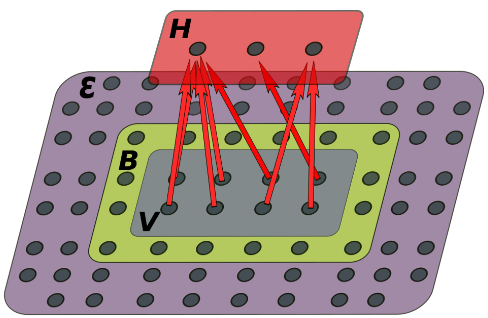
$$H_I = -\sum_{<i,j>} s_i s_j$$

Migdal-Kadanoff block-spins:





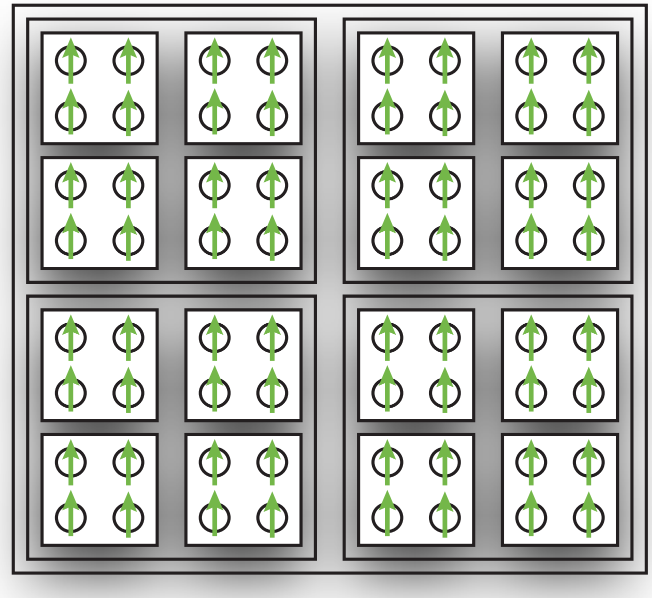
N=1

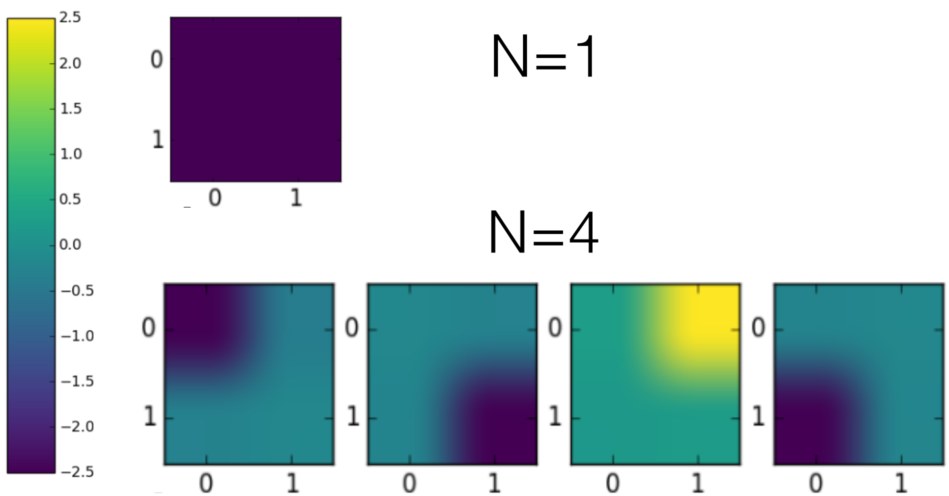


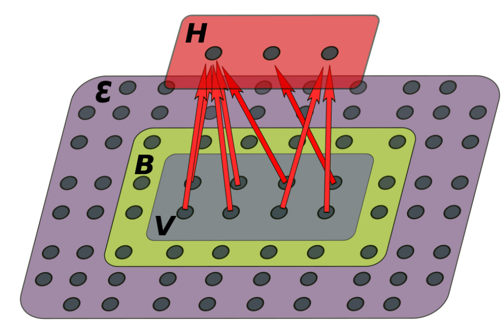
Test case 1: the 2D Ising model

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

Migdal-Kadanoff block-spins:

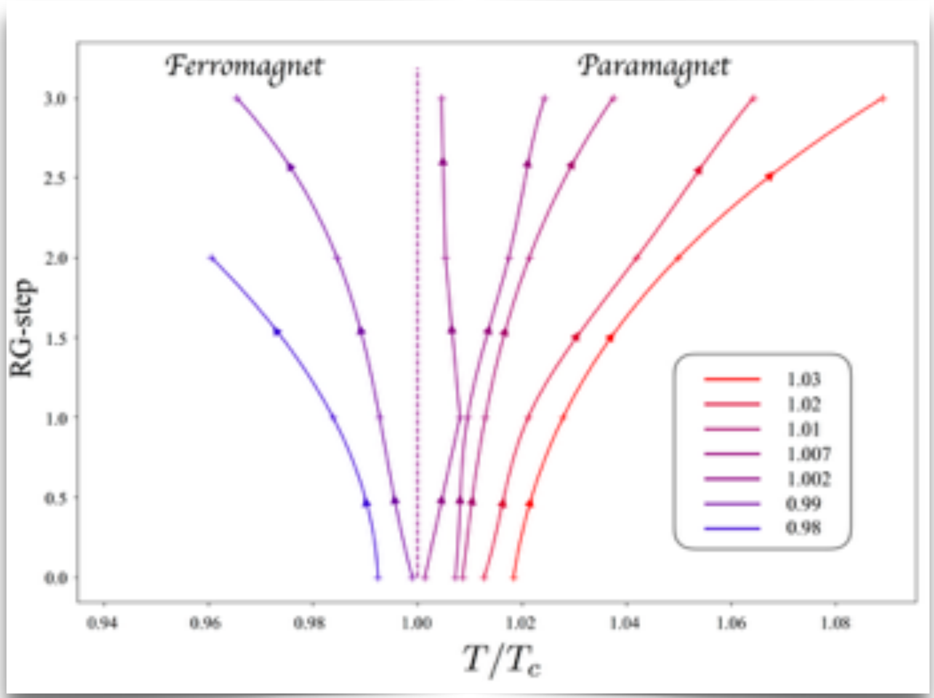






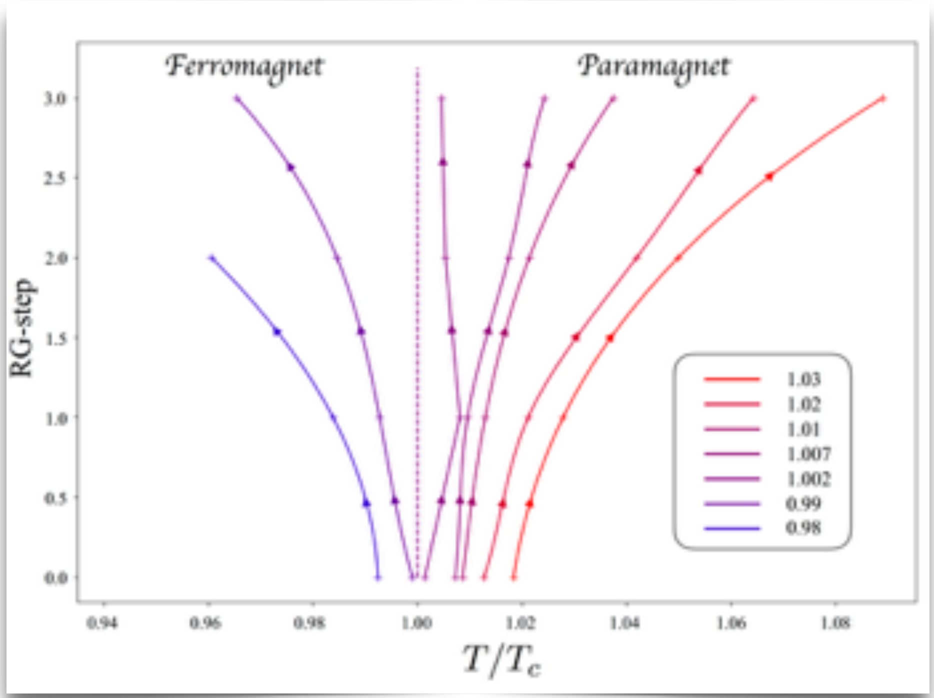
RG flow and critical exponents

- RG flow reconstruction



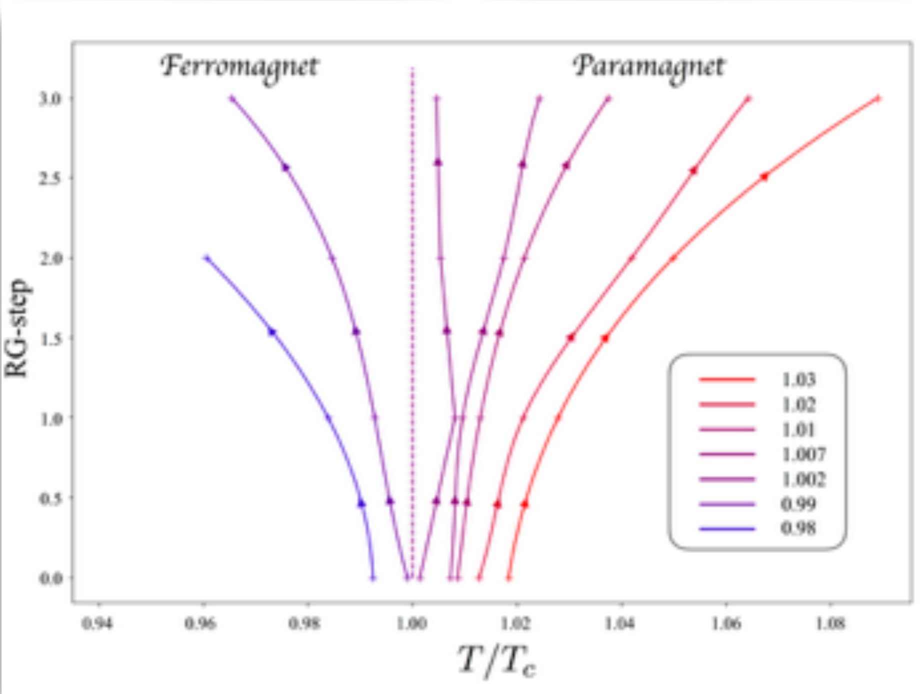
RG flow and critical exponents

- RG flow reconstruction
- Position and type of critical points

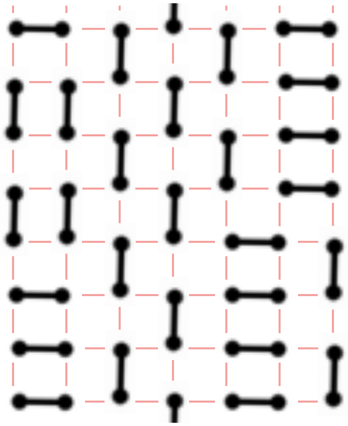


RG flow and critical exponents

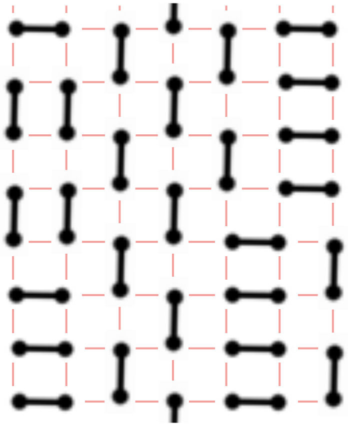
- RG flow reconstruction
- Position and type of critical points
- Critical exponents



Test case 2: the dimer model

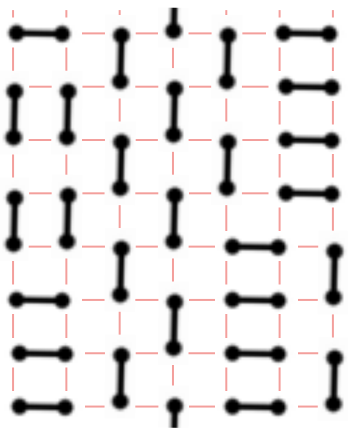


Test case 2: the dimer model



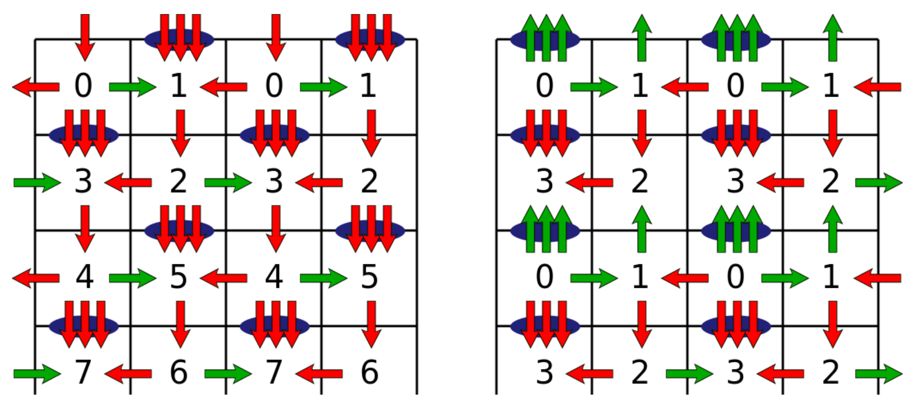
- Defined by local constraints
- Partition function counts configurations

Test case 2: the dimer model

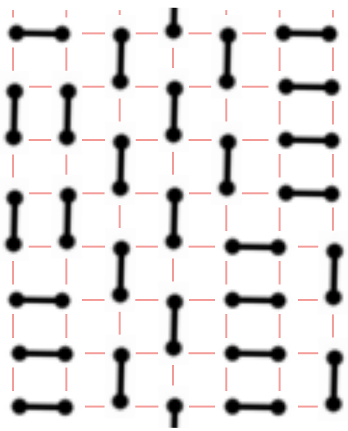


- Defined by local constraints
- Partition function counts configurations

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



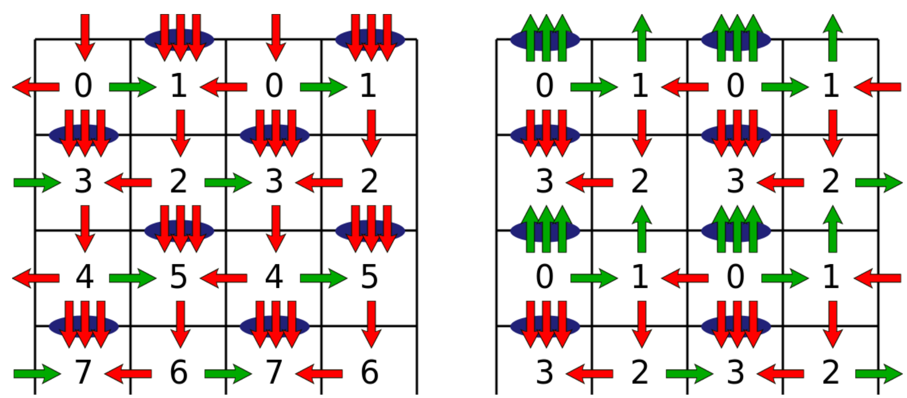
Test case 2: the dimer model



- Defined by local constraints
- Partition function counts configurations

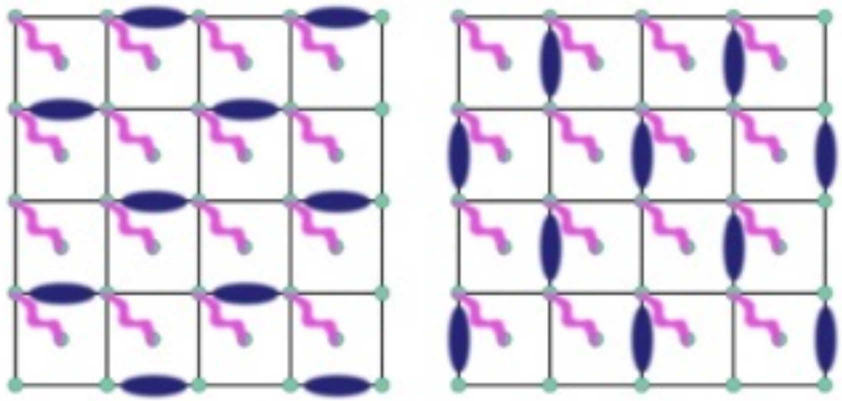
RG of dimer model:
mapping to height field h(x)

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

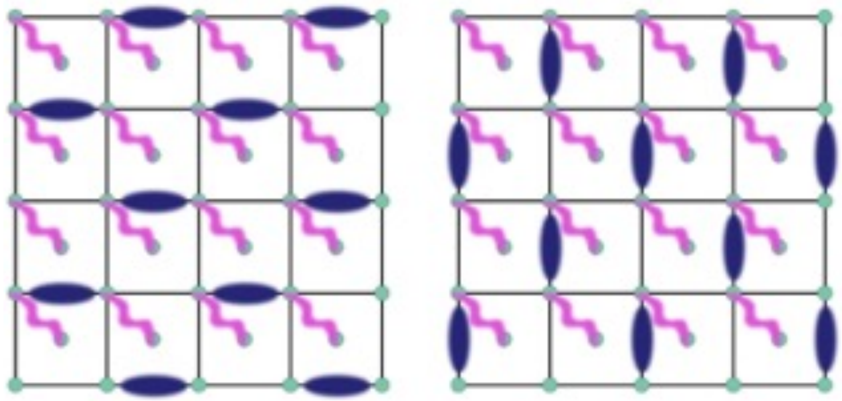


- Let’s add noise!

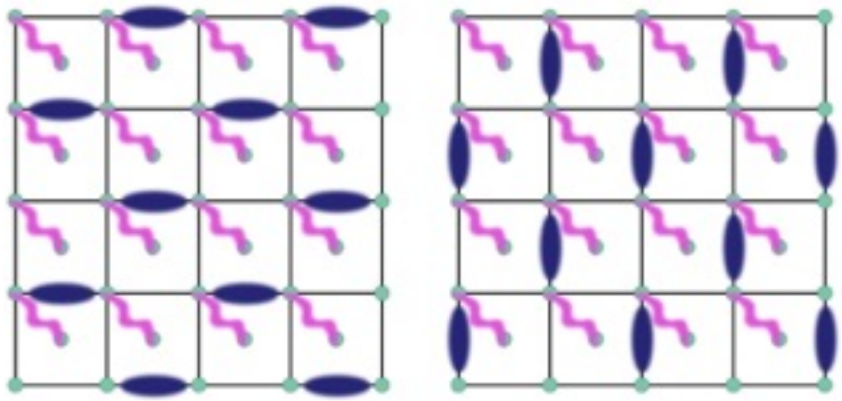
- Let's add noise!



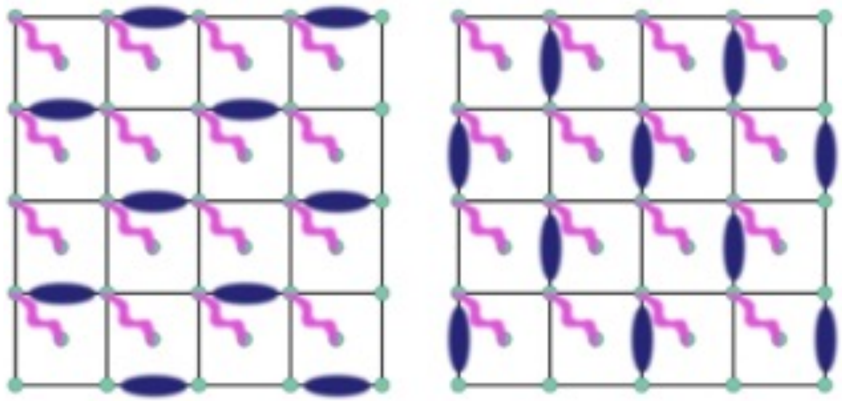
- Let's add noise!
- Physically irrelevant, but strong pattern



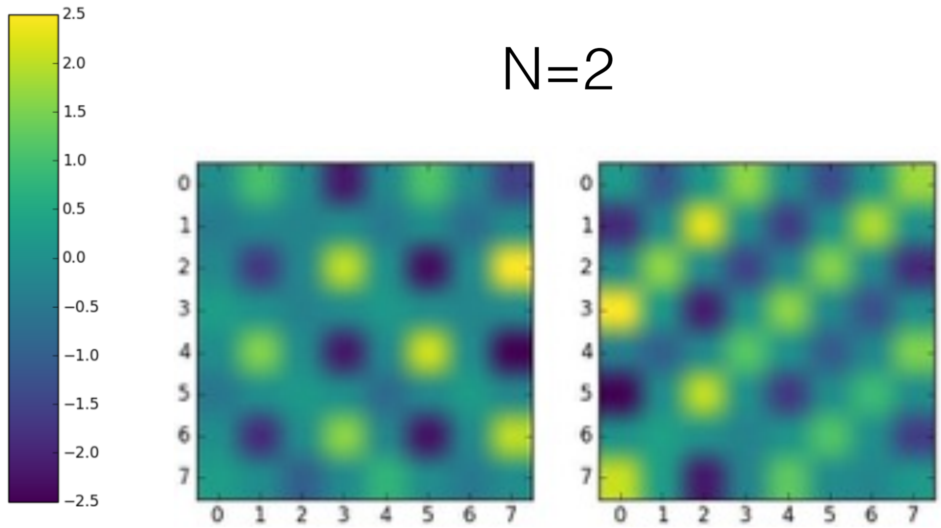
- Let's add noise!
- Physically irrelevant, but strong pattern



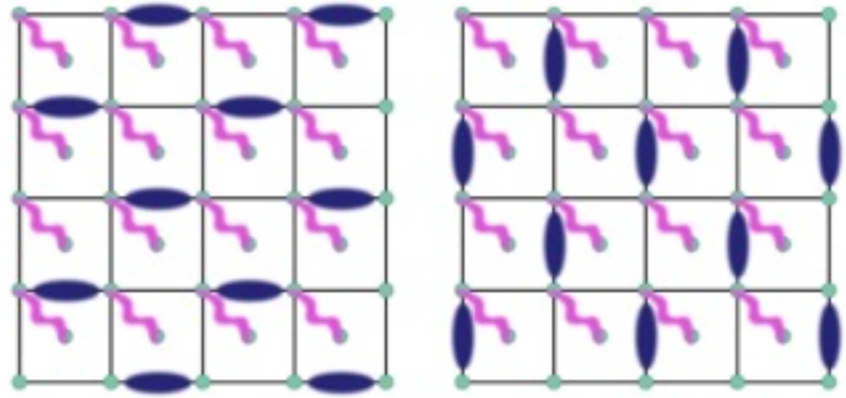
- Let's add noise!
- Physically irrelevant, but strong pattern

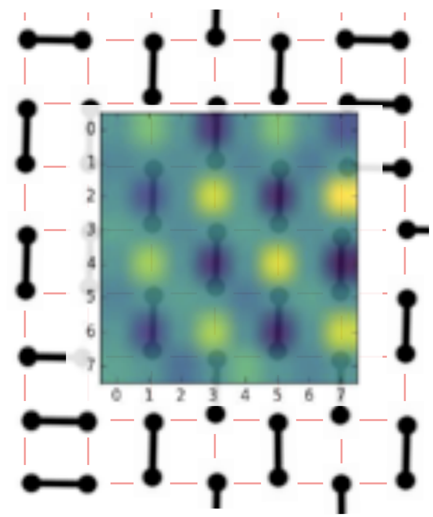


N=2

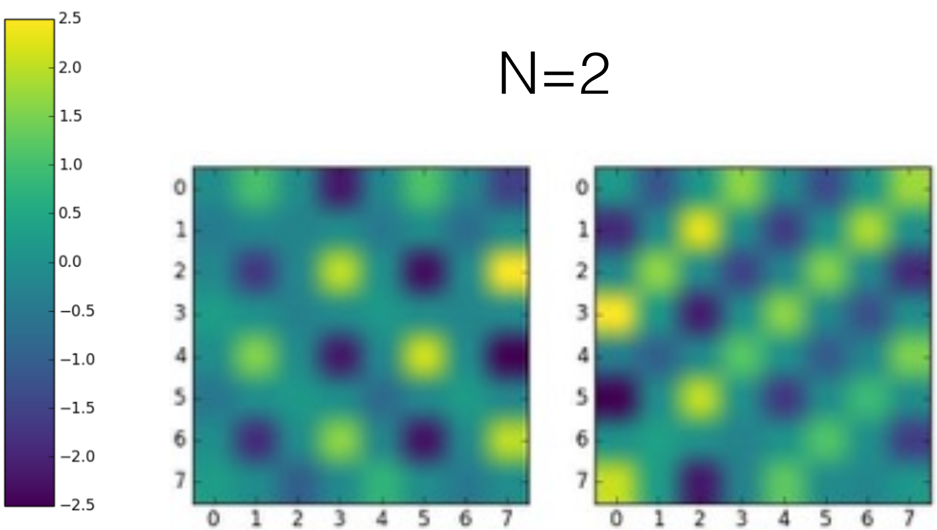


- Let's add noise!
- Physically irrelevant, but strong pattern





N=2



ETH zürich

"Optimality" of mutual information

“Optimality” of mutual information

- Effective Hamiltonian from cumulant expansion

"Optimality" of mutual information

- Effective Hamiltonian from cumulant expansion

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

ETH zürich

"Optimality" of mutual information

- Effective Hamiltonian from cumulant expansion

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

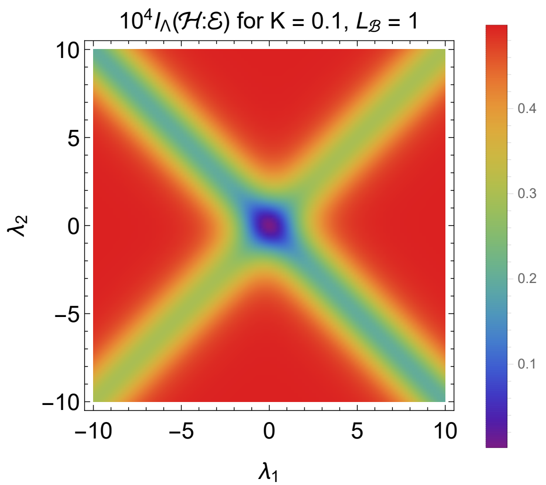
- The "rangeness" and "n-body-ness"

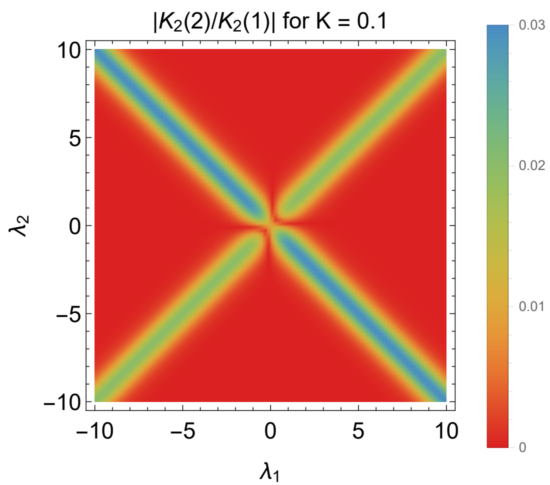
“Optimality” of mutual information

- Effective Hamiltonian from cumulant expansion

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

- The “rangeness” and “n-body-ness”





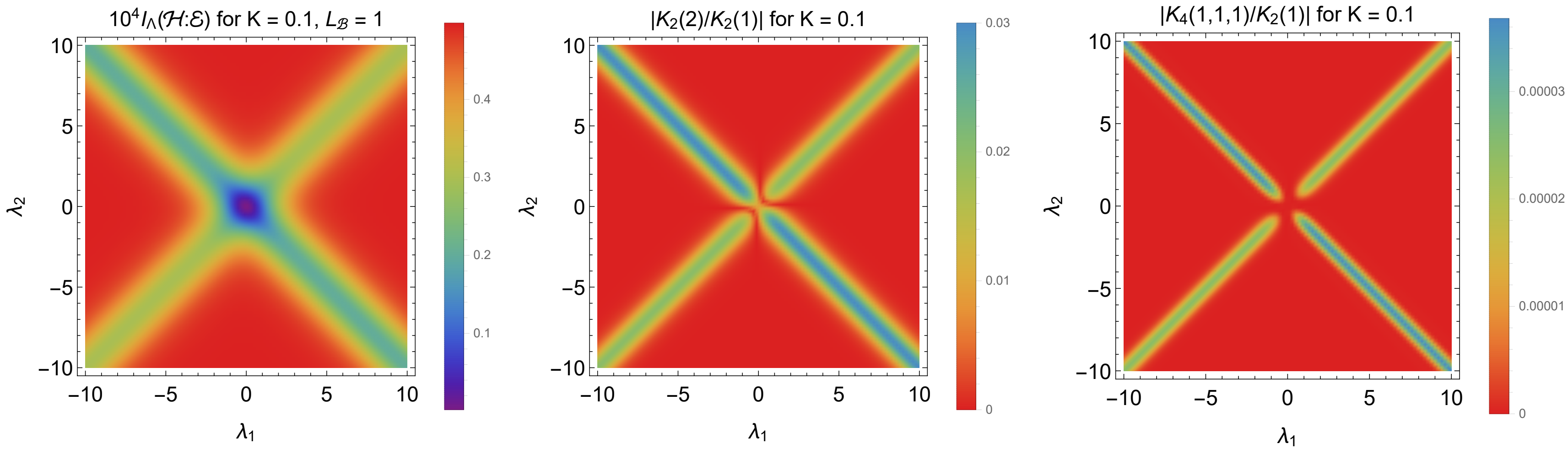
ETH zürich

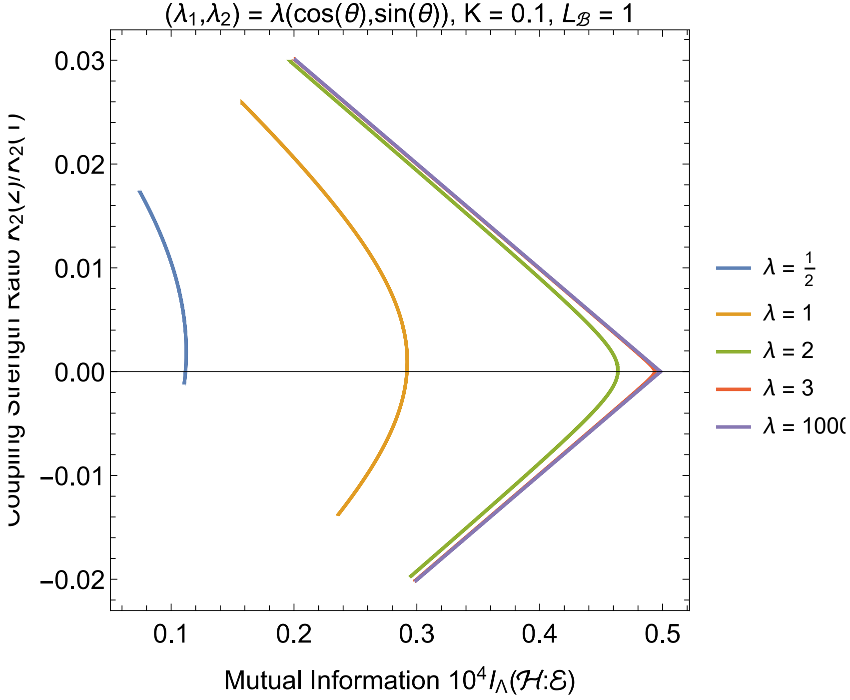
"Optimality" of mutual information

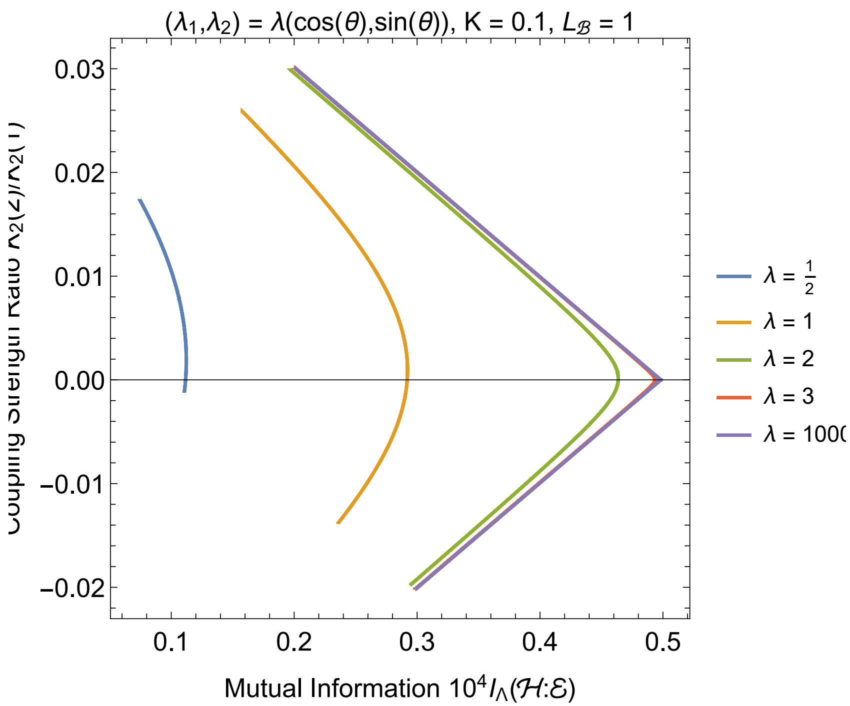
- Effective Hamiltonian from cumulant expansion

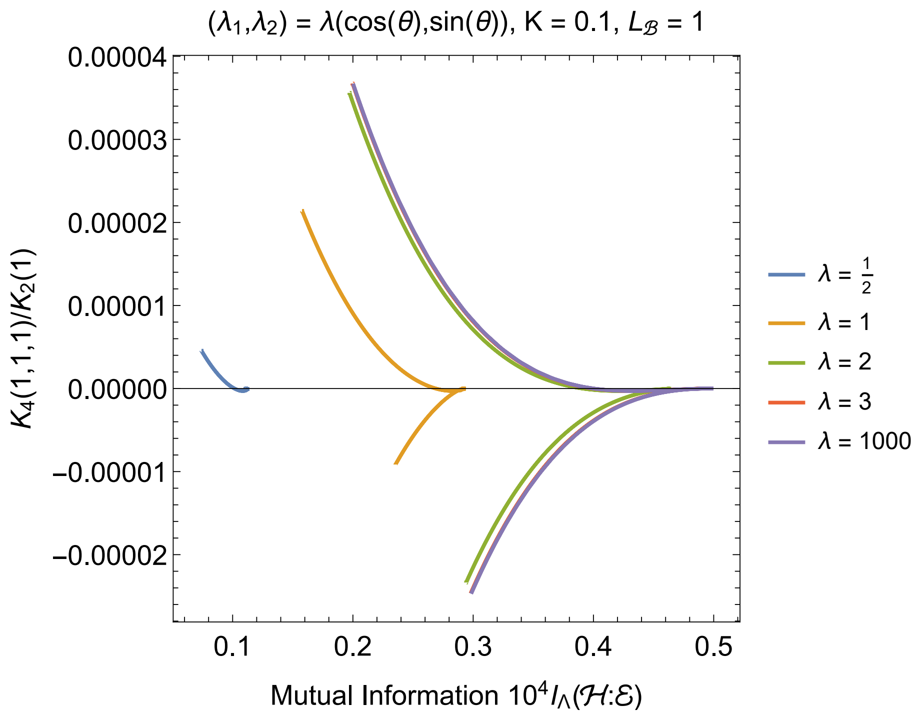
$$\mathcal{K}'[\mathcal{X}'] = \log(Z_{\Lambda,0}[\mathcal{X}']) + \sum_{k=0}^{\infty} \frac{1}{k!} C_k[\mathcal{X}']$$

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

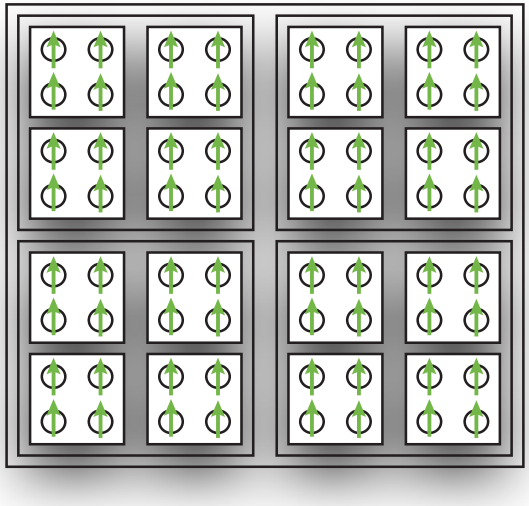



(λ1,λ2) = λ(cos(θ),sin(θ)), K = 0.1, L_B = 1
0.03
0.02
0.01
0.00
−0.01
−0.02
0.1
0.2
0.3
0.4
0.5
Mutual Information 10^4 I_Λ(H:E)
λ = 1/2
λ = 1
λ = 2
λ = 3
λ = 100


(λ1,λ2) = λ(cos(θ),sin(θ)), K = 0.1, L_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_Λ(H:E)
λ = 1/2
λ = 1
λ = 2
λ = 3
λ = 100(

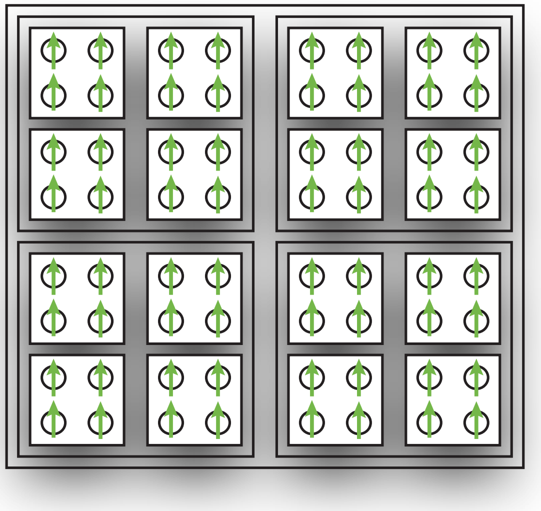

(λ1,λ2) = λ(cos(θ),sin(θ)), K = 0.1, L_B = 1
0.00004
0.00003
0.00002
0.00001
0.00000
−0.00001
−0.00002
K4(1,1,1)/K2(1)
0.1
0.2
0.3
0.4
0.5
Mutual Information 10^4 I_Λ(H:E)
λ = 1/2
λ = 1
λ = 2
λ = 3
λ = 1000

Effective hamiltonian



Effective hamiltonian

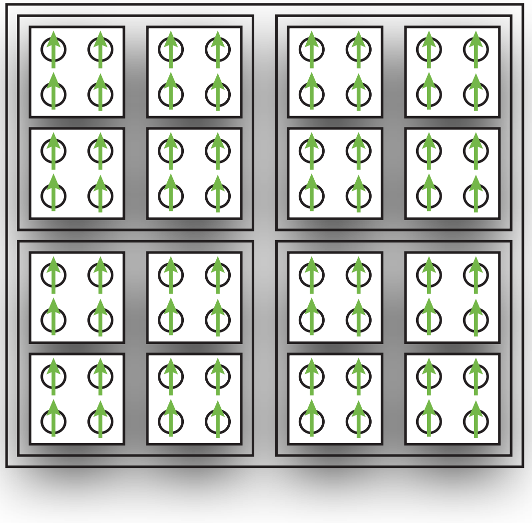
$$\mathcal{K}[\mathcal{X}] = \mathcal{K}_0[\mathcal{X}] + \mathcal{K}_1[\mathcal{X}]$$



Effective hamiltonian

$$\mathcal{K}[\mathcal{X}] = \mathcal{K}_0[\mathcal{X}] + \mathcal{K}_1[\mathcal{X}]$$

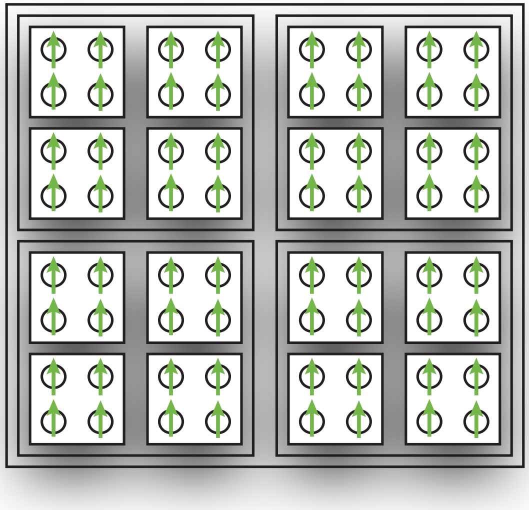
$$\mathcal{K}_0[\mathcal{X}] = \sum_{j=1}^{n} \mathcal{K}_{\mathrm{b}}[\mathcal{V}_j]$$



Effective hamiltonian

$$\mathcal{K}[\mathcal{X}] = \mathcal{K}_0[\mathcal{X}] + \mathcal{K}_1[\mathcal{X}] \qquad \mathcal{K}_0[\mathcal{X}] = \sum_{j=1}^{n} \mathcal{K}_\mathrm{b}[\mathcal{V}_j]$$

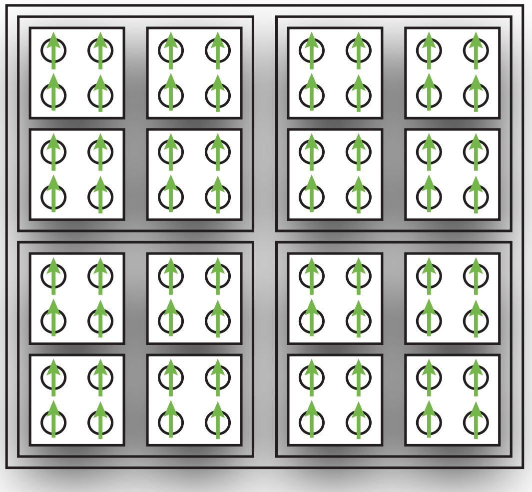
$$Z_0 = \sum_{\mathcal{X}} e^{\mathcal{K}_0[\mathcal{X}]} = \prod_{j=1}^{n} Z_\mathrm{b}, \qquad Z_\mathrm{b} = \sum_{\mathcal{V}} e^{\mathcal{K}_\mathrm{b}[\mathcal{V}]}$$



Effective hamiltonian

$$\mathcal{K}[\mathcal{X}] = \mathcal{K}_0[\mathcal{X}] + \mathcal{K}_1[\mathcal{X}] \qquad \mathcal{K}_0[\mathcal{X}] = \sum_{j=1}^{n} \mathcal{K}_\mathrm{b}[\mathcal{V}_j]$$

$$Z_0 = \sum_{\mathcal{X}} e^{\mathcal{K}_0[\mathcal{X}]} = \prod_{j=1}^{n} Z_\mathrm{b}, \qquad Z_\mathrm{b} = \sum_{\mathcal{V}} e^{\mathcal{K}_\mathrm{b}[\mathcal{V}]}$$

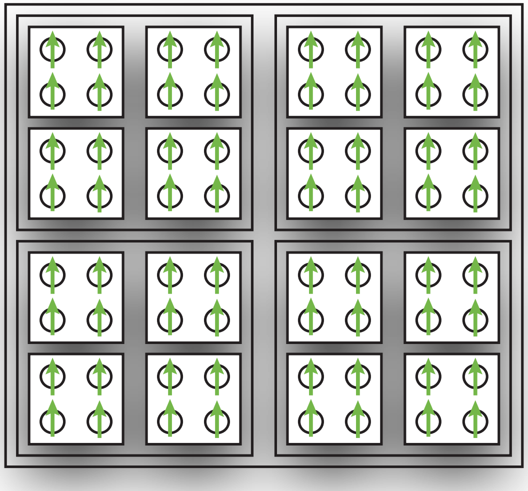


$$\begin{aligned}
e^{\mathcal{K}'[\mathcal{X}']} &= Z_0 \sum_{\mathcal{X}} e^{\mathcal{K}_1[\mathcal{X}]} \prod_{j=1}^{n} \underbrace{\frac{e^{\mathcal{K}_\mathrm{b}[\mathcal{V}_j]}}{Z_\mathrm{b}} P_\Lambda(\mathcal{H}_j|\mathcal{V}_j)}_{=:P_{\Lambda,\mathrm{b}}(\mathcal{H}_j,\mathcal{V}_j)} \\
&= Z_0 \sum_{\mathcal{X}} e^{\mathcal{K}_1[\mathcal{X}]} \prod_{j=1}^{n} P_{\Lambda,\mathrm{b}}(\mathcal{V}_j|\mathcal{H}_j) P_{\Lambda,\mathrm{b}}(\mathcal{H}_j) \\
&= Z_0 \underbrace{\prod_{j=1}^{n} P_{\Lambda,\mathrm{b}}(\mathcal{H}_j)}_{P_{\Lambda,0}(\mathcal{X}')} \sum_{\mathcal{X}} e^{\mathcal{K}_1[\mathcal{X}]} \underbrace{\prod_{j=1}^{n} P_{\Lambda,\mathrm{b}}(\mathcal{V}_j|\mathcal{H}_j)}_{=:P_{\Lambda,0}(\mathcal{X}|\mathcal{X}')} \\
&= Z_0 P_{\Lambda,0}(\mathcal{X}') \left\langle e^{\mathcal{K}_1[\mathcal{X}]} \right\rangle_{\Lambda,0} [\mathcal{X}'],
\end{aligned}$$

Effective hamiltonian

$$\mathcal{K}[\mathcal{X}] = \mathcal{K}_0[\mathcal{X}] + \mathcal{K}_1[\mathcal{X}] \qquad \mathcal{K}_0[\mathcal{X}] = \sum_{j=1}^{n} \mathcal{K}_\mathrm{b}[\mathcal{V}_j]$$

$$Z_0 = \sum_{\mathcal{X}} e^{\mathcal{K}_0[\mathcal{X}]} = \prod_{j=1}^{n} Z_\mathrm{b}, \qquad Z_\mathrm{b} = \sum_{\mathcal{V}} e^{\mathcal{K}_\mathrm{b}[\mathcal{V}]}$$



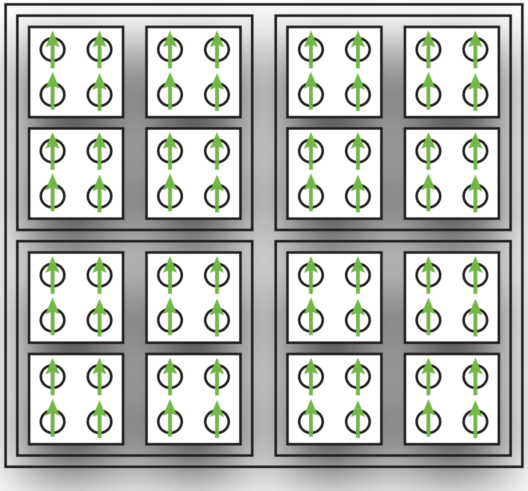
$$\begin{aligned} e^{\mathcal{K}'[\mathcal{X}']} &= Z_0 \sum_{\mathcal{X}} e^{\mathcal{K}_1[\mathcal{X}]} \prod_{j=1}^{n} \underbrace{\frac{e^{\mathcal{K}_\mathrm{b}[\mathcal{V}_j]}}{Z_\mathrm{b}} P_\Lambda(\mathcal{H}_j|\mathcal{V}_j)}_{=:P_{\Lambda,\mathrm{b}}(\mathcal{H}_j,\mathcal{V}_j)} \\ &= Z_0 \sum_{\mathcal{X}} e^{\mathcal{K}_1[\mathcal{X}]} \prod_{j=1}^{n} P_{\Lambda,\mathrm{b}}(\mathcal{V}_j|\mathcal{H}_j) P_{\Lambda,\mathrm{b}}(\mathcal{H}_j) \\ &= Z_0 \underbrace{\prod_{j=1}^{n} P_{\Lambda,\mathrm{b}}(\mathcal{H}_j)}_{P_{\Lambda,0}(\mathcal{X}')} \sum_{\mathcal{X}} e^{\mathcal{K}_1[\mathcal{X}]} \underbrace{\prod_{j=1}^{n} P_{\Lambda,\mathrm{b}}(\mathcal{V}_j|\mathcal{H}_j)}_{=:P_{\Lambda,0}(\mathcal{X}|\mathcal{X}')} \\ &= Z_0 P_{\Lambda,0}(\mathcal{X}') \left\langle e^{\mathcal{K}_1[\mathcal{X}]} \right\rangle_{\Lambda,0} [\mathcal{X}'], \end{aligned}$$

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

Effective hamiltonian

$$\mathcal{K}[\mathcal{X}] = \mathcal{K}_0[\mathcal{X}] + \mathcal{K}_1[\mathcal{X}] \qquad \mathcal{K}_0[\mathcal{X}] = \sum_{j=1}^{n} \mathcal{K}_{\mathrm{b}}[\mathcal{V}_j]$$

$$Z_0 = \sum_{\mathcal{X}} e^{\mathcal{K}_0[\mathcal{X}]} = \prod_{j=1}^{n} Z_{\mathrm{b}}, \qquad Z_{\mathrm{b}} = \sum_{\mathcal{V}} e^{\mathcal{K}_{\mathrm{b}}[\mathcal{V}]}$$



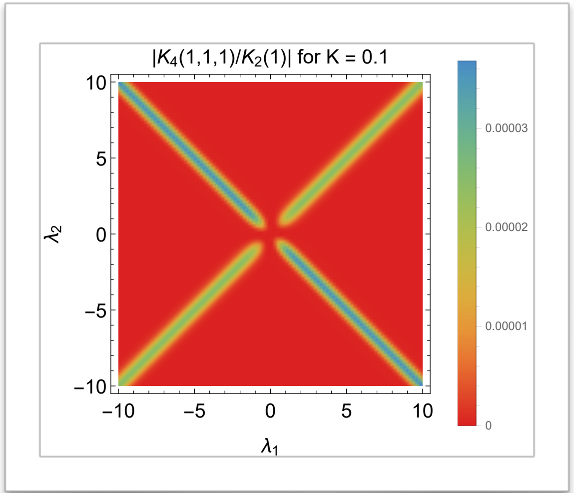
$$\begin{aligned} e^{\mathcal{K}'[\mathcal{X}']} &= Z_0 \sum_{\mathcal{X}} e^{\mathcal{K}_1[\mathcal{X}]} \prod_{j=1}^{n} \underbrace{\frac{e^{\mathcal{K}_{\mathrm{b}}[\mathcal{V}_j]}}{Z_{\mathrm{b}}} P_\Lambda(\mathcal{H}_j|\mathcal{V}_j)}_{=:P_{\Lambda,\mathrm{b}}(\mathcal{H}_j,\mathcal{V}_j)} \\ &= Z_0 \sum_{\mathcal{X}} e^{\mathcal{K}_1[\mathcal{X}]} \prod_{j=1}^{n} P_{\Lambda,\mathrm{b}}(\mathcal{V}_j|\mathcal{H}_j) P_{\Lambda,\mathrm{b}}(\mathcal{H}_j) \\ &= Z_0 \underbrace{\prod_{j=1}^{n} P_{\Lambda,\mathrm{b}}(\mathcal{H}_j)}_{P_{\Lambda,0}(\mathcal{X}')} \sum_{\mathcal{X}} e^{\mathcal{K}_1[\mathcal{X}]} \underbrace{\prod_{j=1}^{n} P_{\Lambda,\mathrm{b}}(\mathcal{V}_j|\mathcal{H}_j)}_{=:P_{\Lambda,0}(\mathcal{X}|\mathcal{X}')} \\ &= Z_0 P_{\Lambda,0}(\mathcal{X}') \left\langle e^{\mathcal{K}_1[\mathcal{X}]} \right\rangle_{\Lambda,0} [\mathcal{X}'], \end{aligned}$$

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

$$\begin{aligned} C_1 &= \langle \mathcal{K}_1 \rangle_{\Lambda,0}\,, \\ C_2 &= \langle \mathcal{K}_1^2 \rangle_{\Lambda,0} - \langle \mathcal{K}_1 \rangle_{\Lambda,0}^2\,, \\ C_3 &= \langle \mathcal{K}_1^3 \rangle_{\Lambda,0} - 3 \langle \mathcal{K}_1^2 \rangle_{\Lambda,0} \langle \mathcal{K}_1 \rangle_{\Lambda,0} + 2 \langle \mathcal{K}_1 \rangle_{\Lambda,0}^3 \end{aligned}$$

Conclusions

- ML in condensed matter
- Information-theoretic view of RG
- The Real Space Mutual Information algorithm
- Optimality of MI



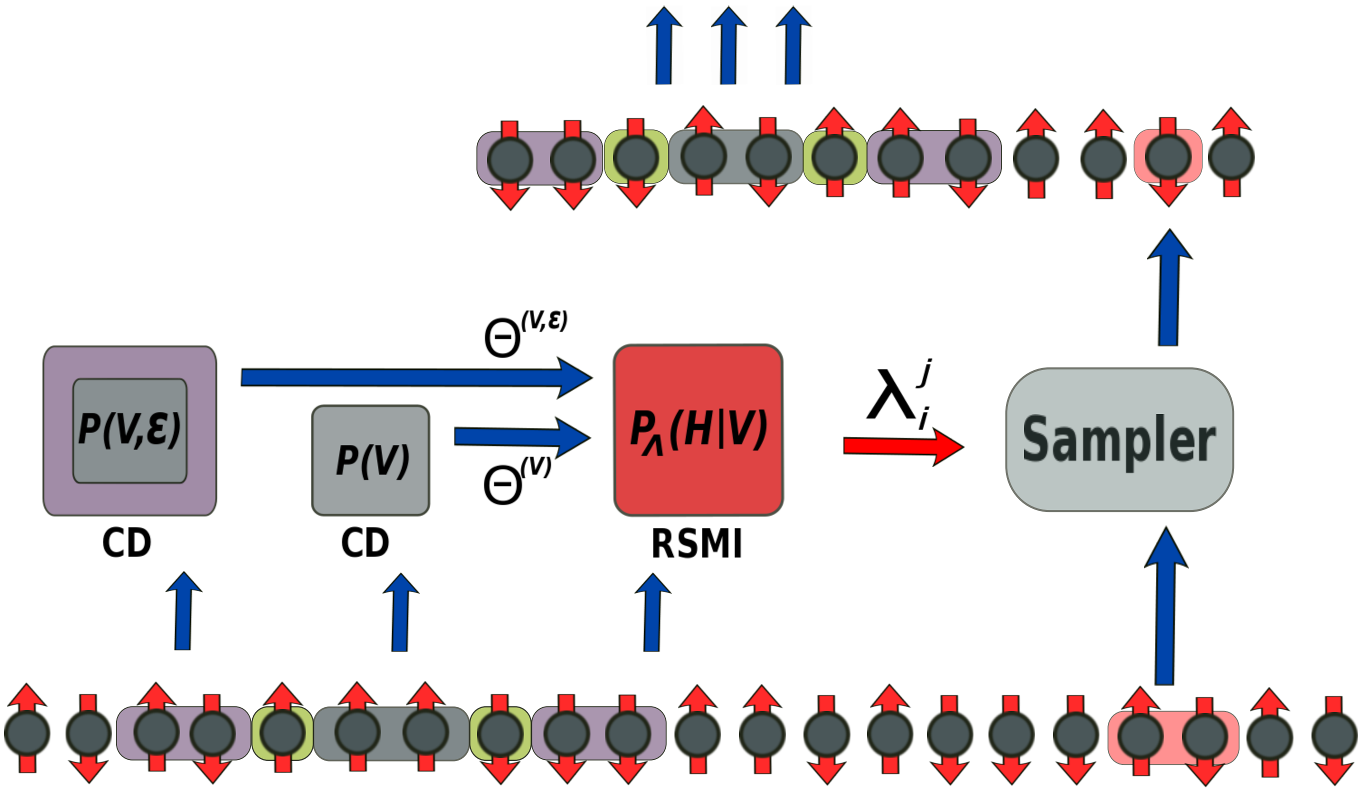
Lenggenhager, MKJ, Huber,
(*in preparation)*

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

Thank you!

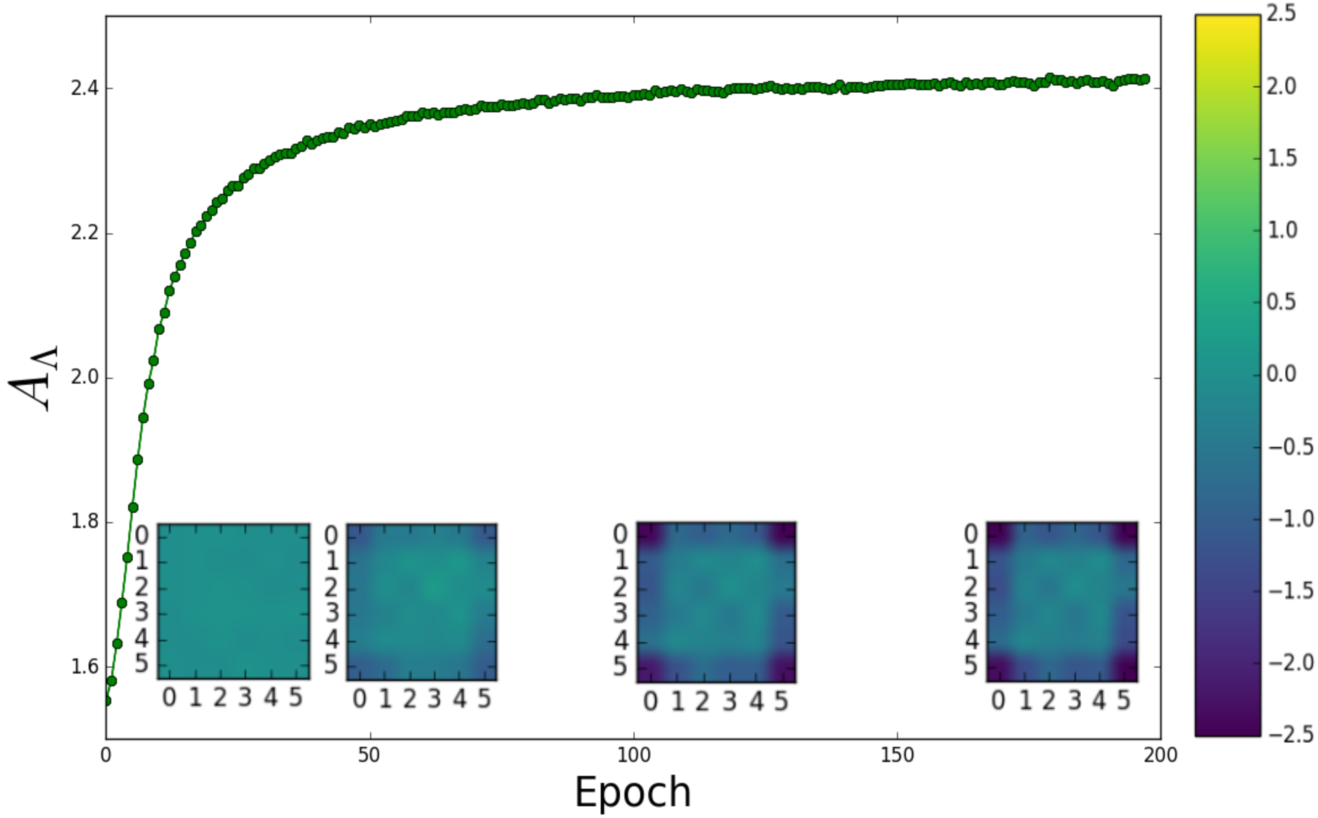
Multiple RG steps

Multiple RG steps



MI - Training

MI - Training

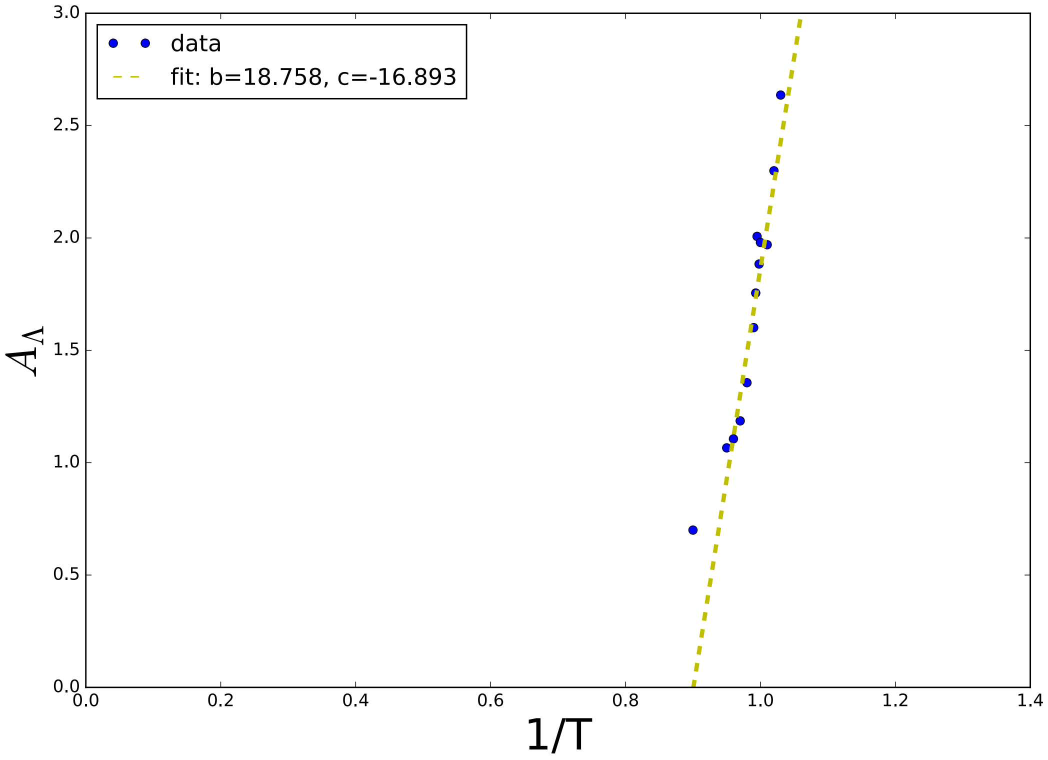


A_Λ
Epoch
2.4
2.2
2.0
1.8
1.6
0
50
100
150
200
0 1 2 3 4 5
2.5
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
−2.5

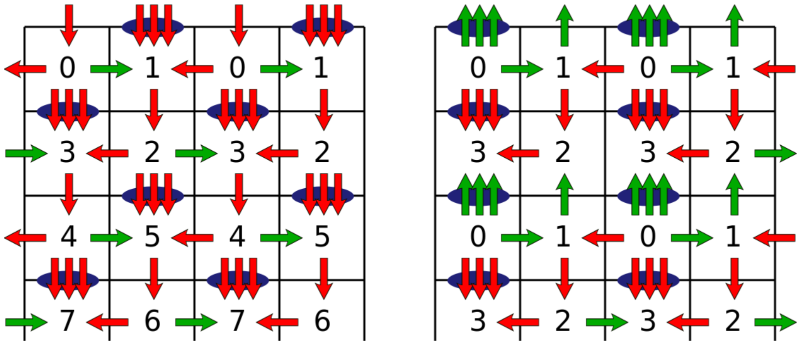
The MI “thermometer”

ETH zürich

The MI “thermometer”

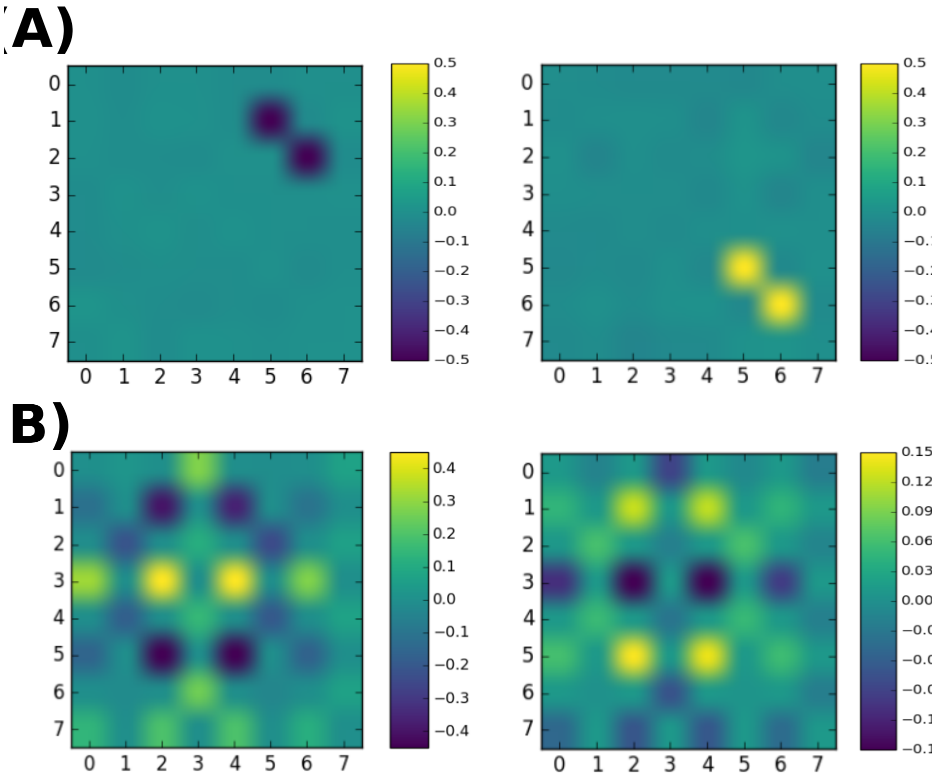


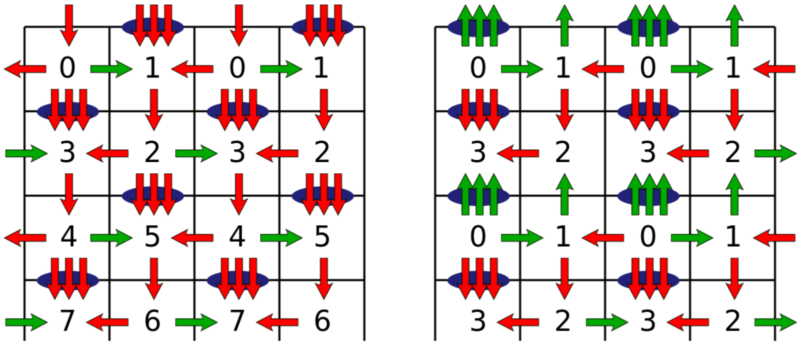
KL-training failure



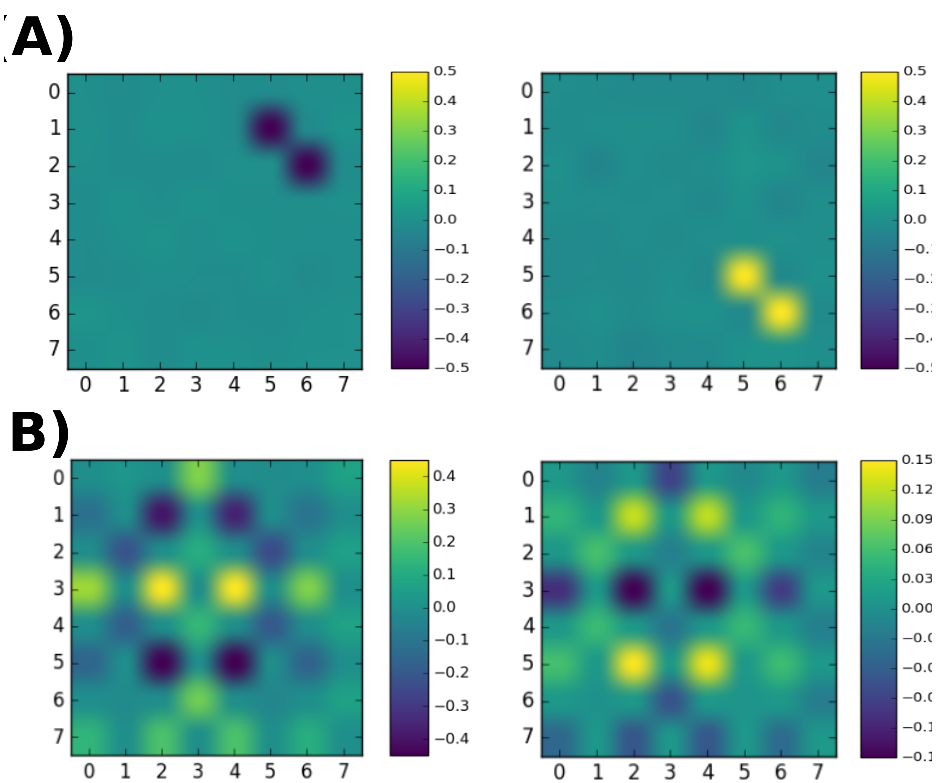
KL-training failure

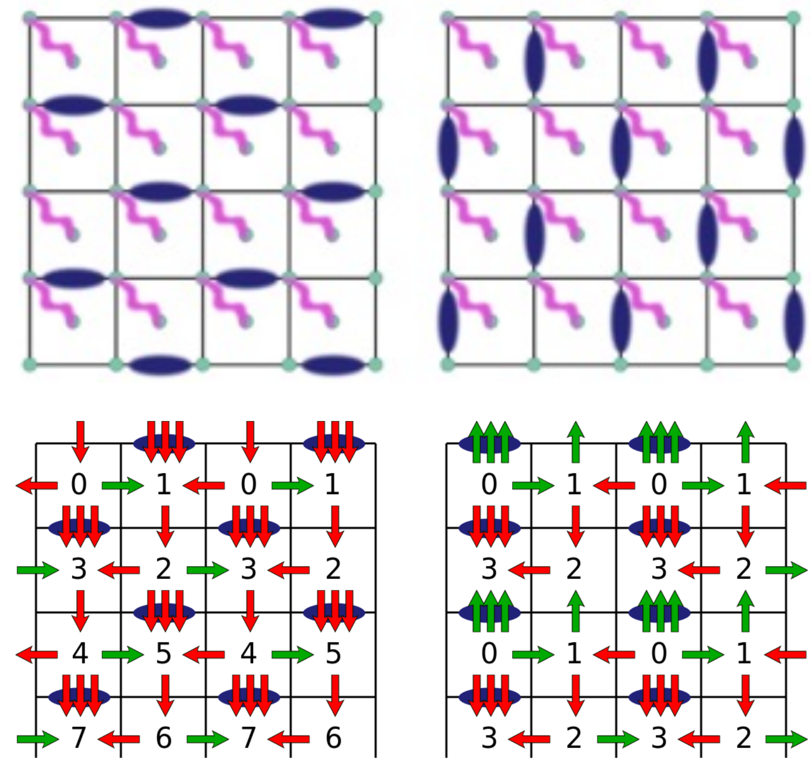
(A)





KL-training failure





Critical exponent

Critical exponent