ETH zürich

Statistical physics through the lens of real-space mutual information

Maciej Koch-Janusz

University of Zurich UZH

MARIE CURIE ACTIONS

THE UNIVERSITY OF CHICAGO

האוניברסיטה העברית בירושלים
THE HEBREW UNIVERSITY OF JERUSALEM

Zohar Ringel

Amit Gordon

Aditya Banerjee

- MKJ and Zohar Ringel, *Nature Physics* ***14****, 578-582 (2018)*
- A. Gordon, A. Banerjee, MKJ and Z. Ringel, "Relevance in RG and in Information Theory" *arXiv:2012.01447*

ETH zürich

Doruk Efe Gokmen

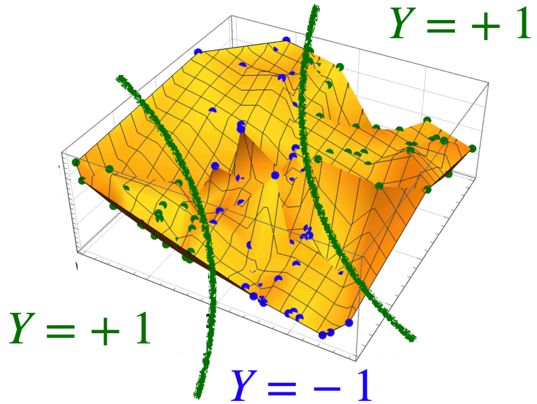
Patrick Lenggenhager

Sebastian D. Huber

- P. Lenngenhager, D.E. Gokmen, Z. Ringel, S.D. Huber and MKJ, *Phys. Rev. X* ***10****, 011037 (2020)*
- D.E. Gokmen, Z. Ringel, S.D. Huber and MKJ, "Statistical physics through the lens of real-space mutual information" *arXiv:2101.11633*

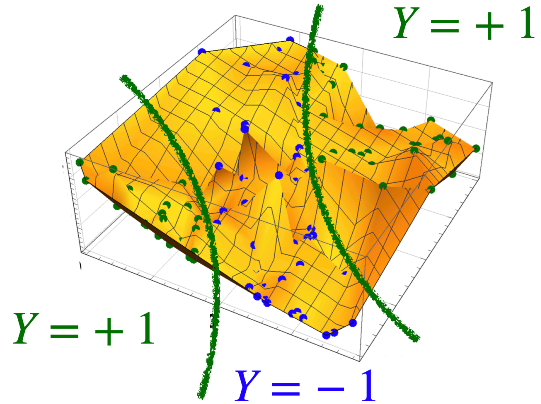
ML and physics

ML and physics

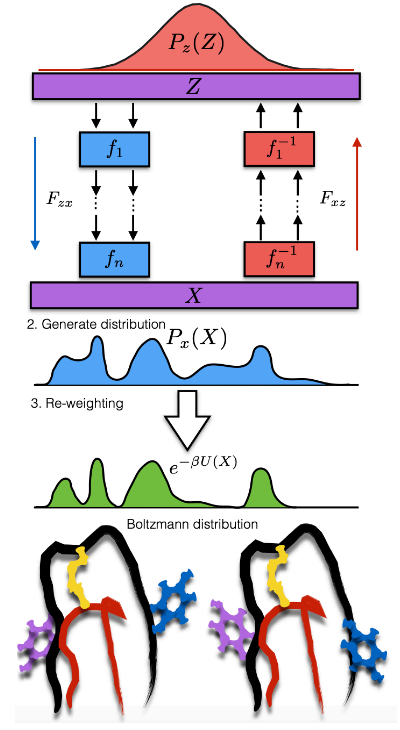


Goldt et al. PRX 10, 4 (2020)

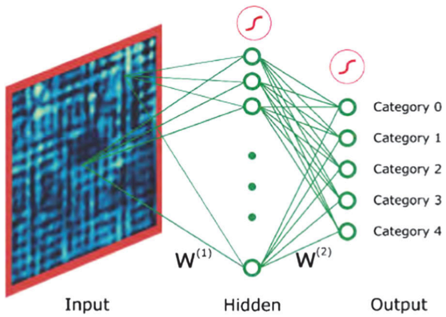
ML and physics

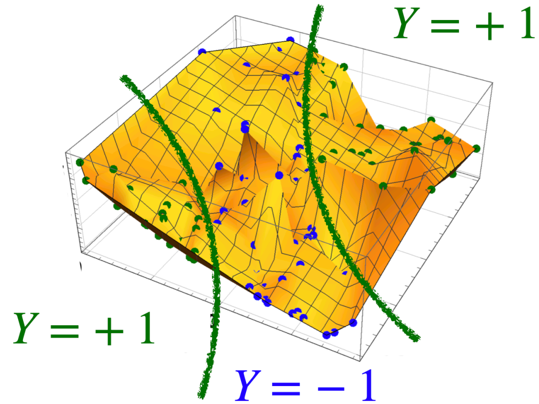


Goldt et al. PRX 10, 4 (2020)

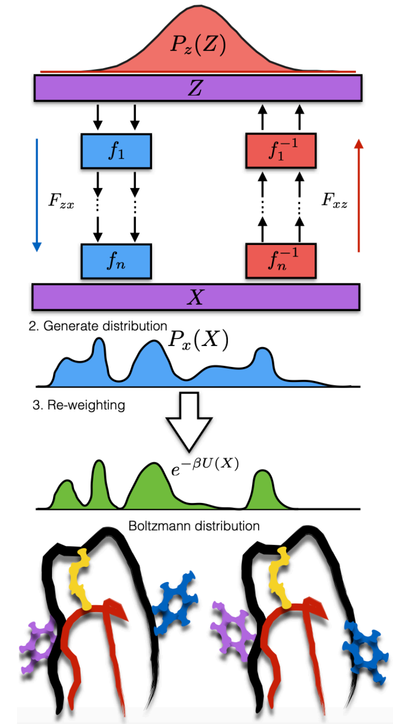


ML and physics

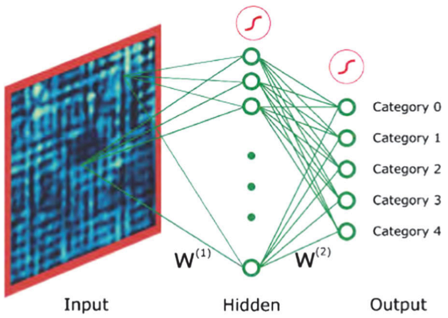


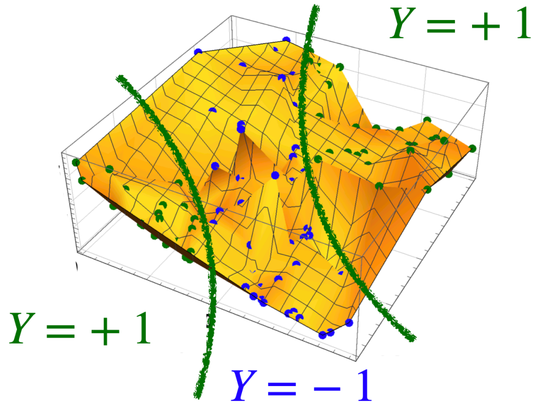


Goldt et al. PRX 10, 4 (2020)

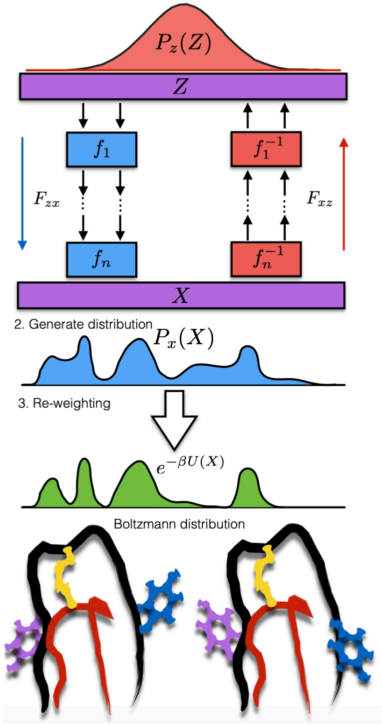


ML and physics



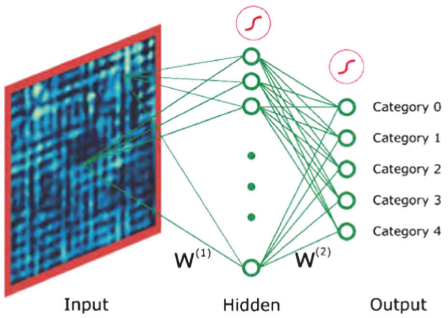


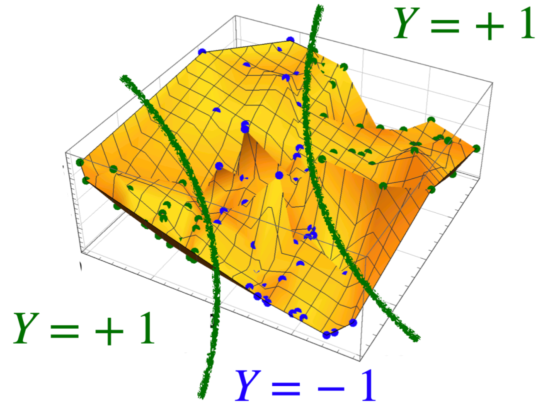
Goldt et al. PRX 10, 4 (2020)



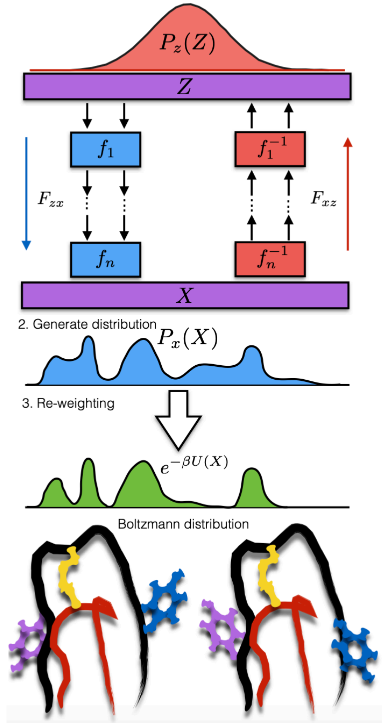
- Carleo et al. "Machine Learning and the physical sciences" *Rev. Mod. Phys. 91, 045002 (2019)*
- Juan Carrasquilla "Machine Learning for Quantum matter" *Advances in Physics: X, 5:1 (2020)*

ML and physics





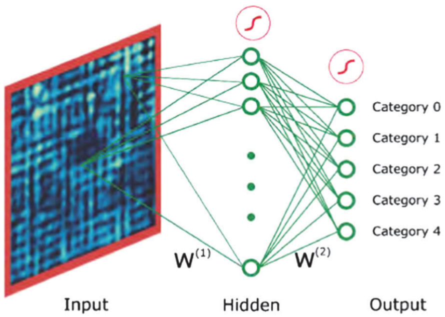
Goldt et al. PRX 10, 4 (2020)

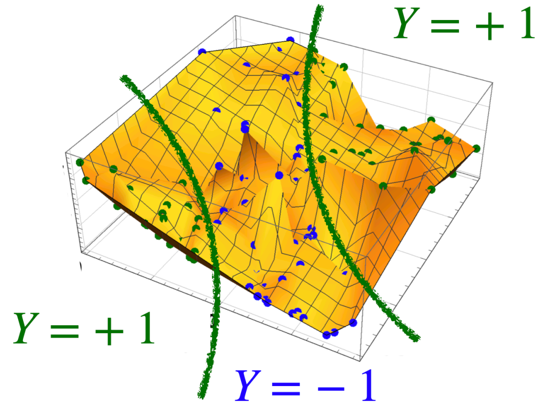


- Carleo et al. "Machine Learning and the physical sciences" *Rev. Mod. Phys. 91, 045002 (2019)*
- Juan Carrasquilla "Machine Learning for Quantum matter" *Advances in Physics: X, 5:1 (2020)*

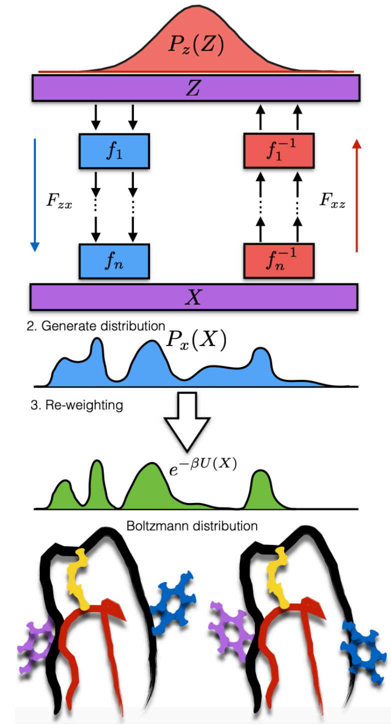
- Challenges: reproducibility, interpretability, re-usability

ML and physics





Goldt et al. PRX 10, 4 (2020)

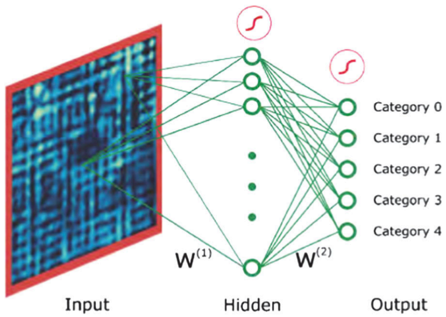


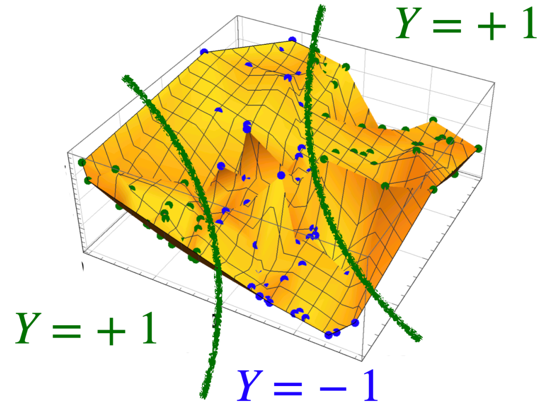
- Carleo et al. "Machine Learning and the physical sciences" *Rev. Mod. Phys. 91, 045002 (2019)*
- Juan Carrasquilla "Machine Learning for Quantum matter" *Advances in Physics: X, 5:1 (2020)*

- Challenges: reproducibility, interpretability, re-usability

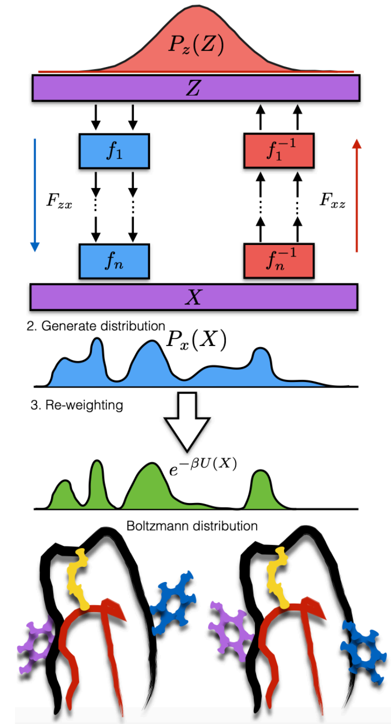
Automating theory-building

ML and physics





Goldt et al. PRX 10, 4 (2020)

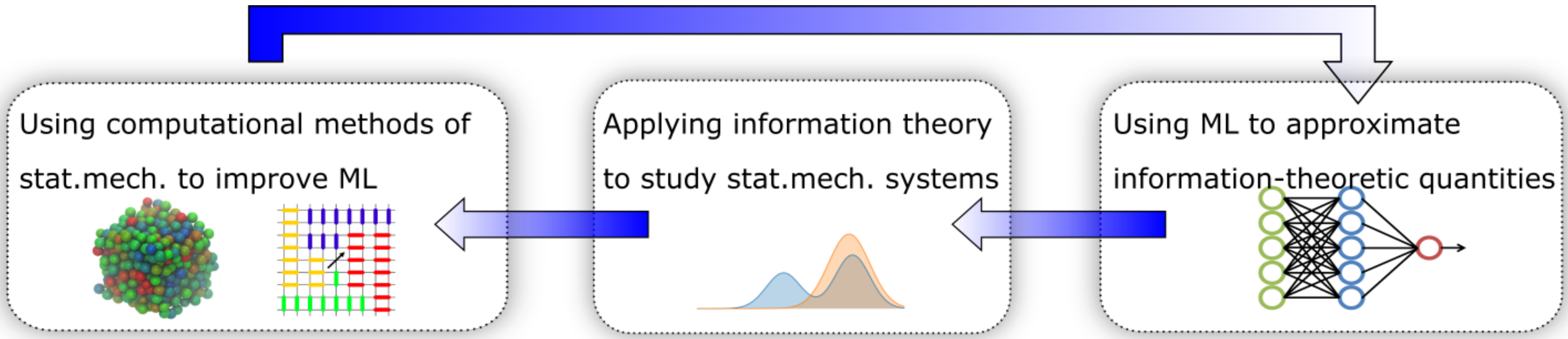


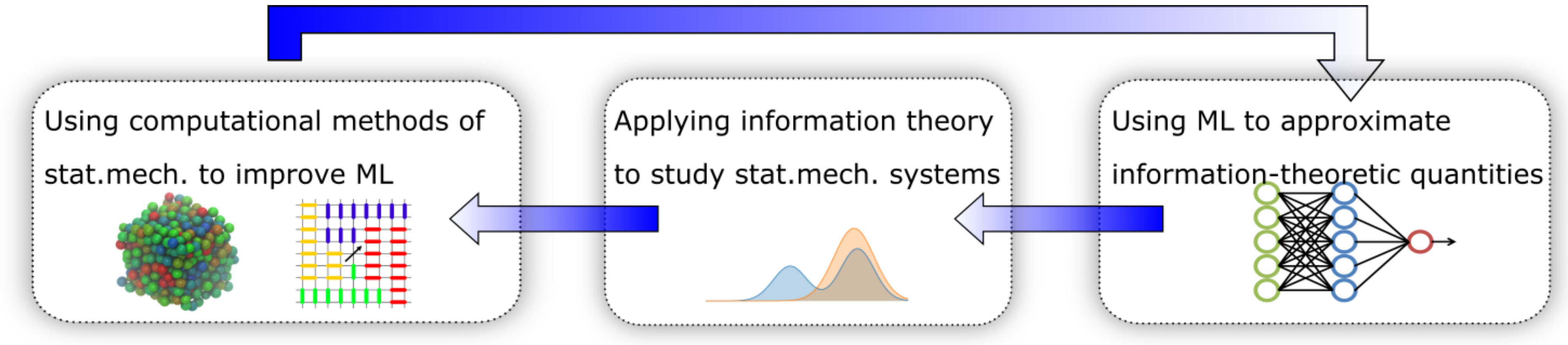
- Carleo et al. “Machine Learning and the physical sciences” *Rev. Mod. Phys. 91, 045002 (2019)*
- Juan Carrasquilla “Machine Learning for Quantum matter” *Advances in Physics: X, 5:1 (2020)*

- Challenges: reproducibility, interpretability, re-usability

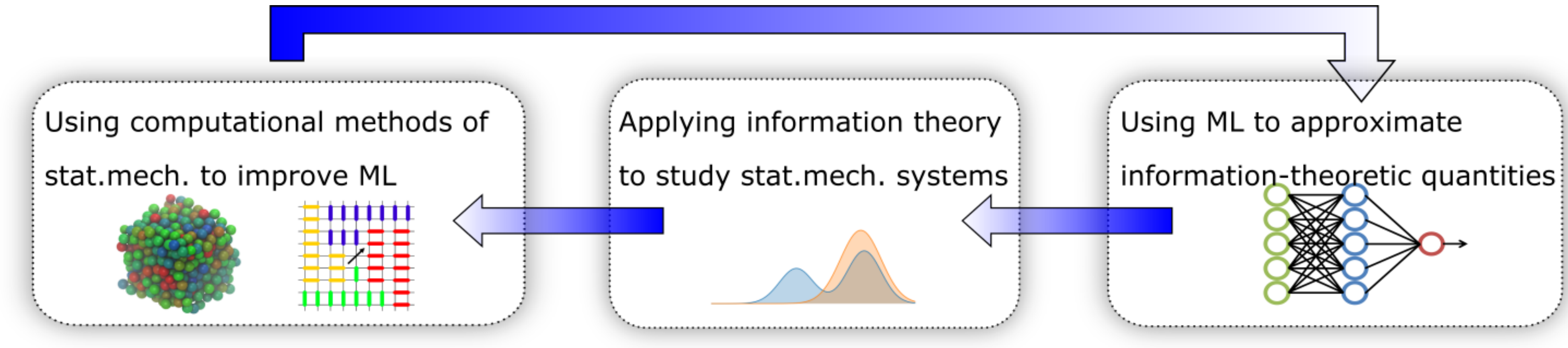
Automating theory-building

- **Formal** interpretability without prior knowledge


Using computational methods of
stat.mech. to improve ML
Applying information theory
to study stat.mech. systems
Using ML to approximate
information-theoretic quantities

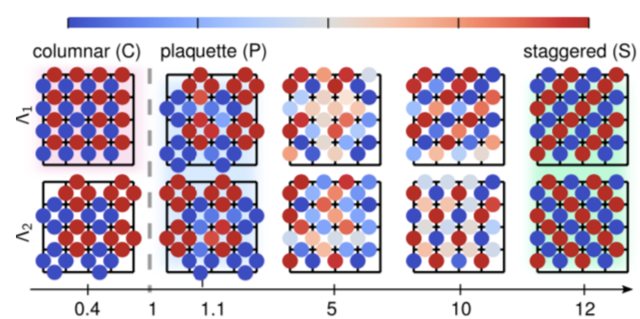


- Optimal RG transformation as a variational problem in information theory
- https://github.com/RSMI-NE/RSMI-NE
- Comprehensive view of long-distance physics
- Constructing the relevant operators
- Formal connection between compression theory and field theory formalism
- Possible extensions to non-equilibrium



- Optimal RG transformation as a variational problem in information theory
- https://github.com/RSMI-NE/RSMI-NE
- Comprehensive view of long-distance physics
- Constructing the relevant operators
- Formal connection between compression theory and field theory formalism
- Possible extensions to non-equilibrium

Nature Physics 14, 578-582 (2018)
PRX 10, 011037 (2020)
arXiv: 2012.01447 (2020)
arXiv: 2101.11633 (2021)

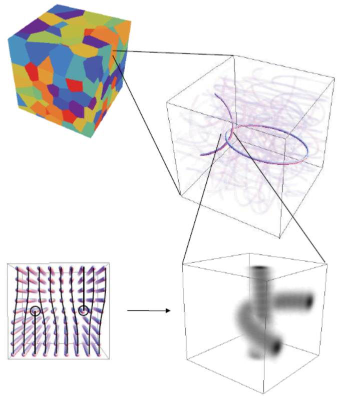


Renormalization Group

Renormalization Group

- Coarse-graining is important in theory/applications

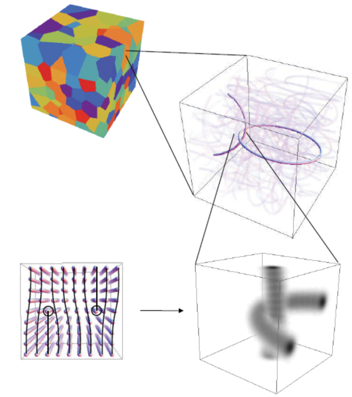
Renormalization Group

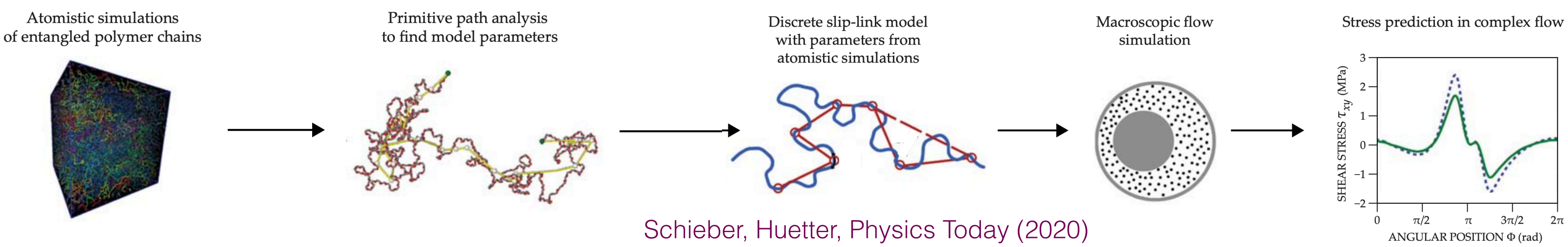


- Coarse-graining is important in theory/applications

Renormalization Group

- Coarse-graining is important in theory/applications

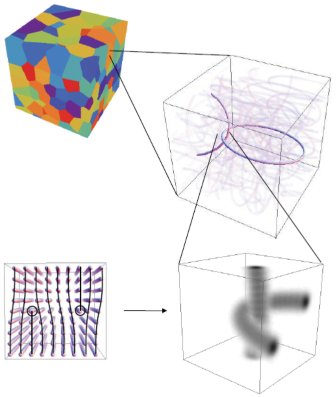


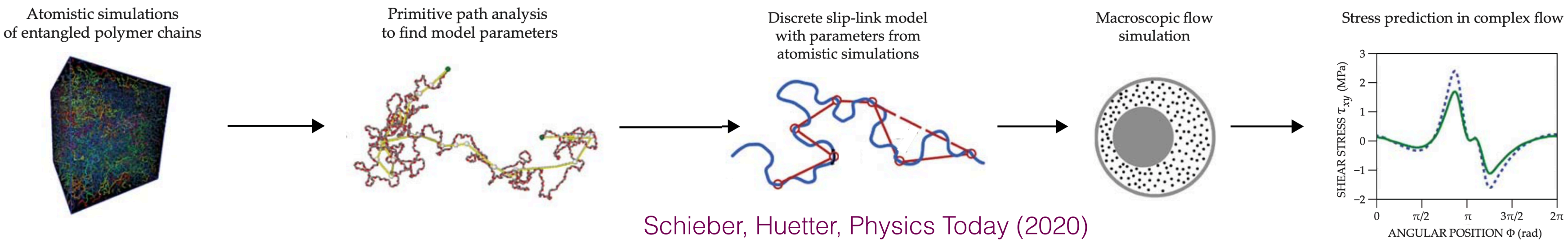


Schieber, Huetter, Physics Today (2020)

Renormalization Group

- Coarse-graining is important in theory/applications
- Deriving effective theories

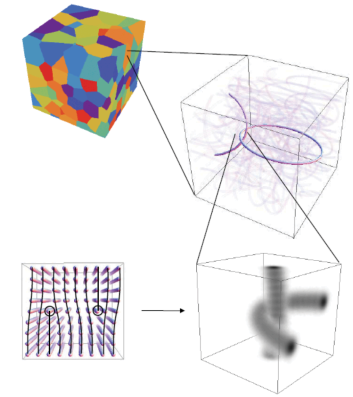


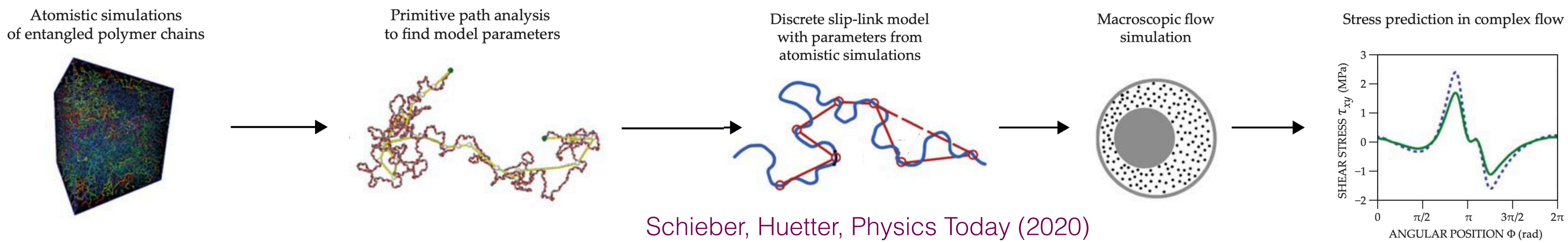


Schieber, Huetter, Physics Today (2020)

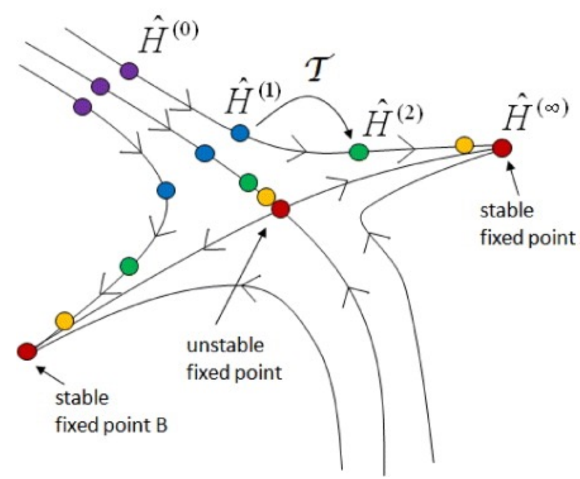
Renormalization Group

- Coarse-graining is important in theory/applications
- Deriving effective theories

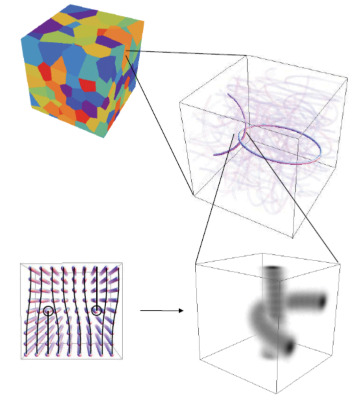




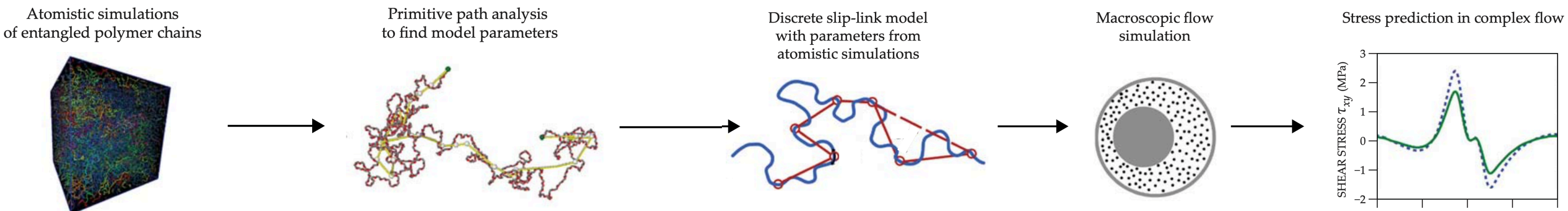
Schieber, Huetter, Physics Today (2020)



Renormalization Group

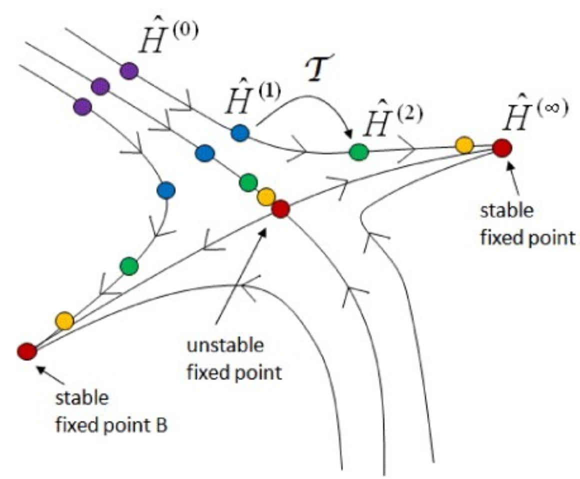


- Coarse-graining is important in theory/applications
- Deriving effective theories

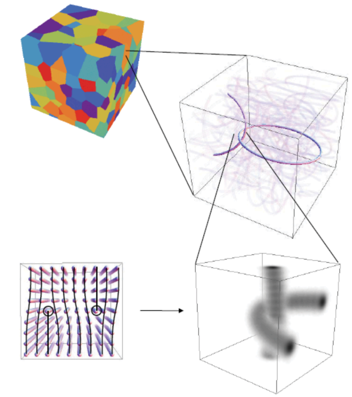


Schieber, Huetter, Physics Today (2020)

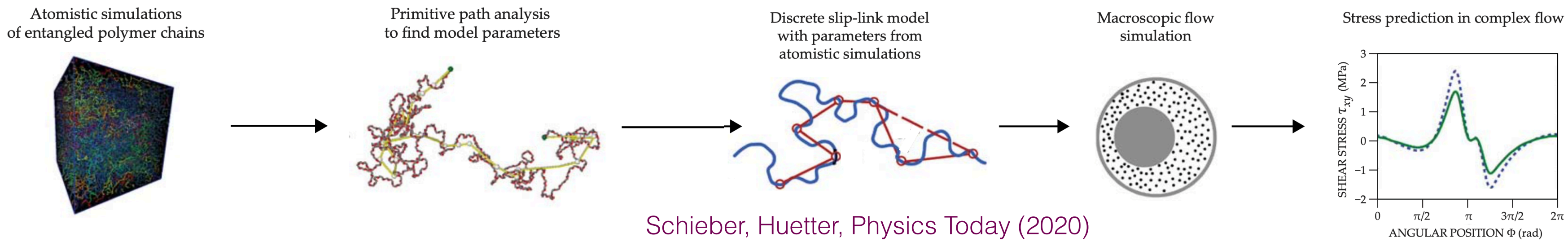
- Formalizes the notion of universality



Renormalization Group

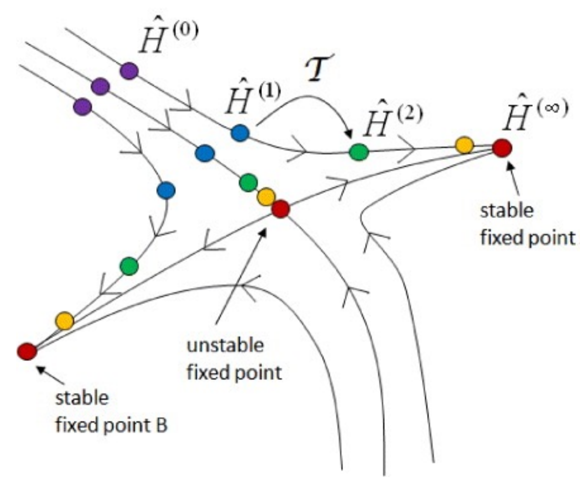


- Coarse-graining is important in theory/applications
- Deriving effective theories

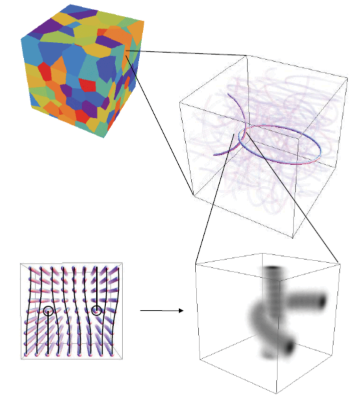


Schieber, Huetter, Physics Today (2020)

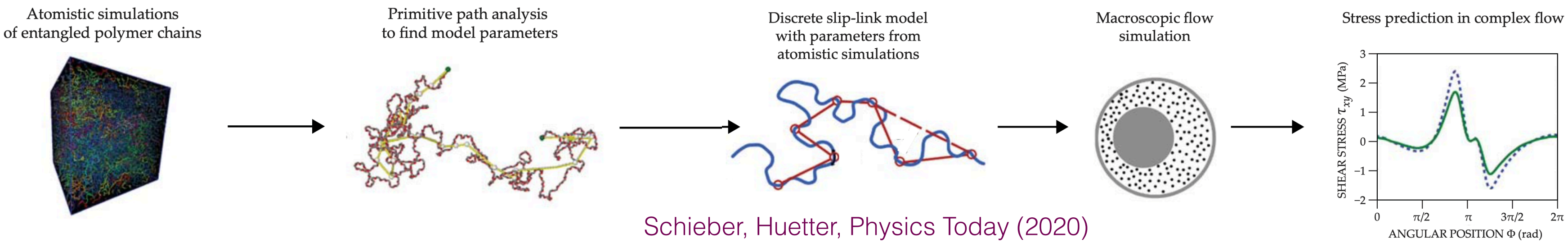
- Formalizes the notion of universality
- Many flavors: Wilsonian, DMRG, real-space,…



Renormalization Group

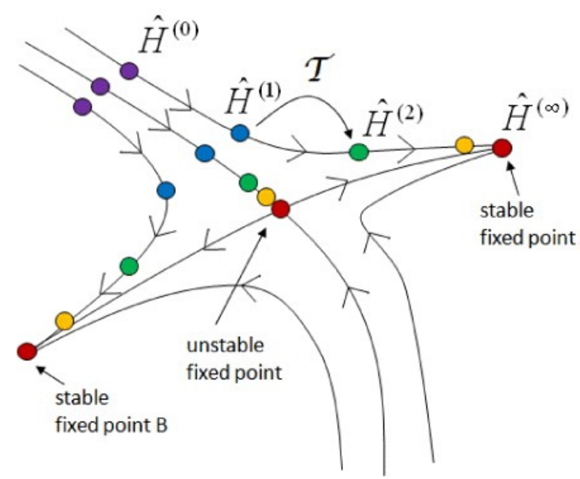


- Coarse-graining is important in theory/applications
- Deriving effective theories

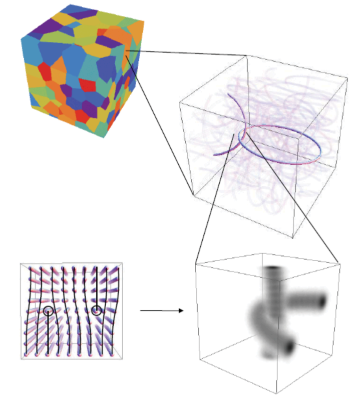


Schieber, Huetter, Physics Today (2020)

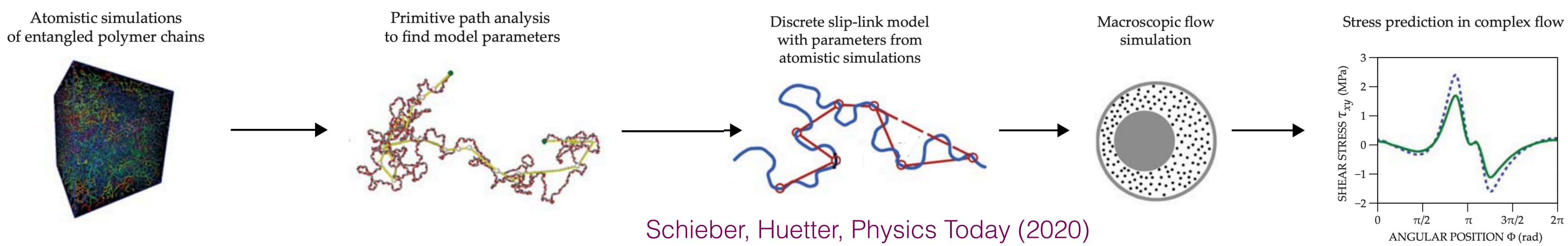
- Formalizes the notion of universality
- Many flavors: Wilsonian, DMRG, real-space,…
- Leitmotiv: integrate out DOFs to obtain effective theory of the remaining ones



Renormalization Group

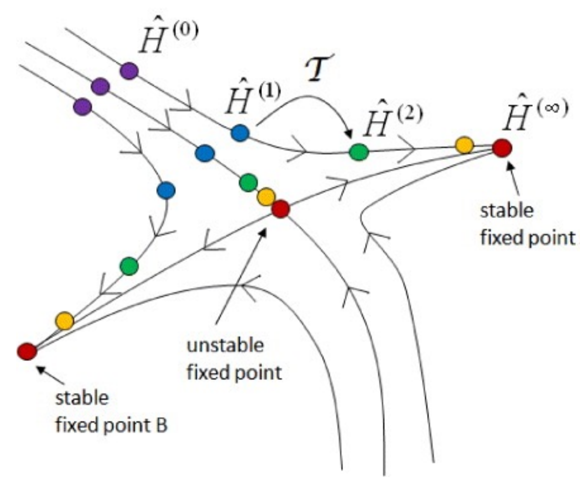


- Coarse-graining is important in theory/applications
- Deriving effective theories



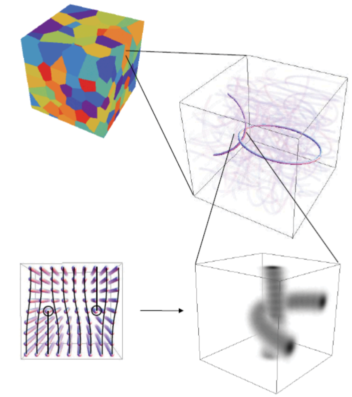
Schieber, Huetter, Physics Today (2020)

- Formalizes the notion of universality
- Many flavors: Wilsonian, DMRG, real-space,…
- Leitmotiv: integrate out DOFs to obtain effective theory of the remaining ones

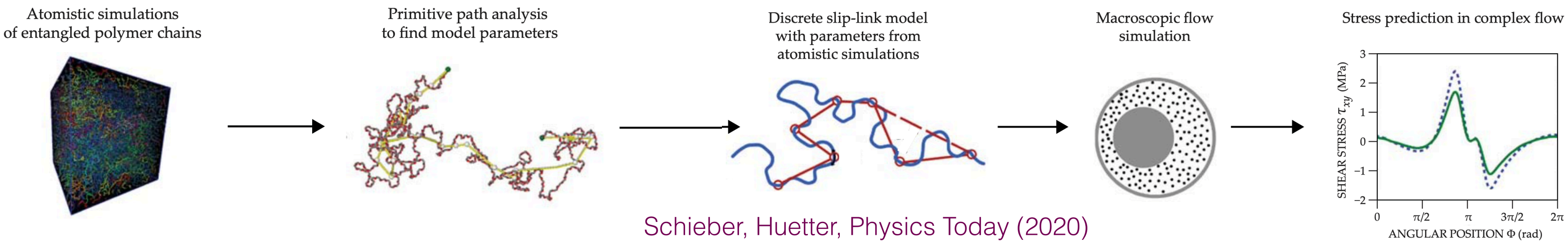


The relation of RG to information theory is obvious but:

Renormalization Group

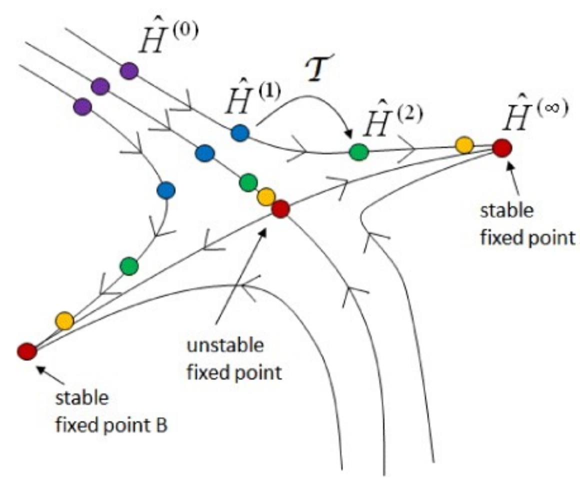


- Coarse-graining is important in theory/applications
- Deriving effective theories



Schieber, Huetter, Physics Today (2020)

- Formalizes the notion of universality
- Many flavors: Wilsonian, DMRG, real-space,…
- Leitmotiv: integrate out DOFs to obtain effective theory of the remaining ones



The relation of RG to information theory is obvious but:

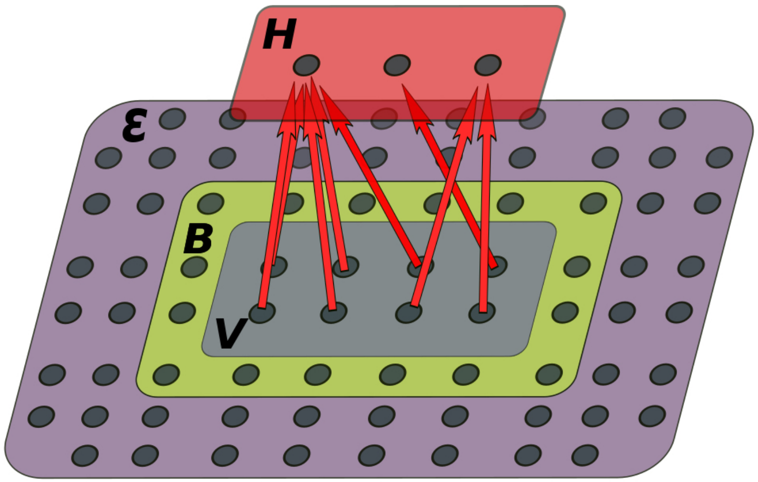
- Can this intuition be formalized?
- Is it useful in practice?

Outline

- The RSMI setup
- Efficient implementation with MINE-based methods
- Examples

- Relevance in RG and in information theory
- Optimality of RSMI
- Symmetries in IB and RG
- Future directions: Non-Eq?

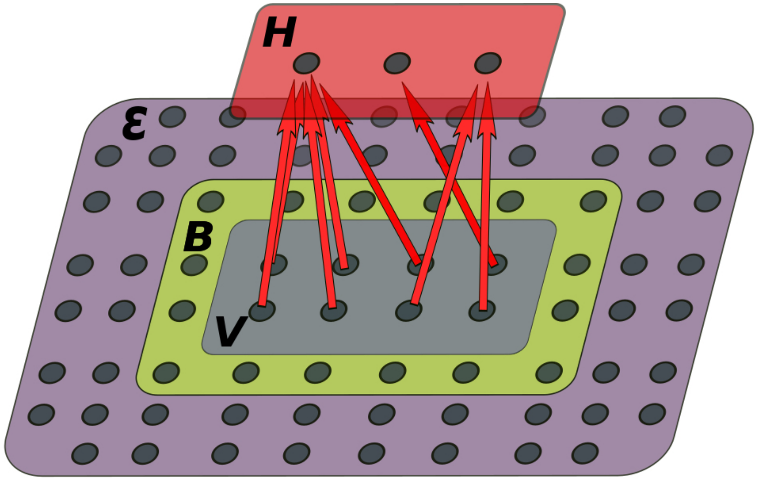
Real-space RG from Information Theory perspective



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

Real-space RG from Information Theory perspective

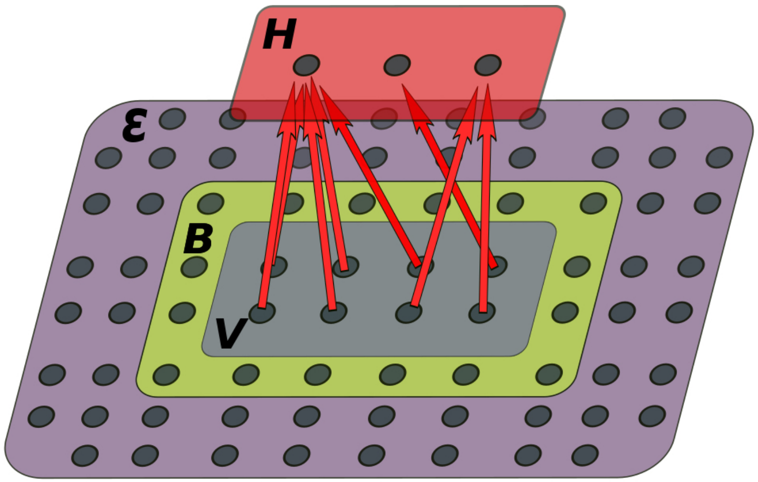
$$e^{\mathcal{K}'(\mathcal{X}')} = \sum_{\mathcal{X}} e^{\mathcal{K}(\mathcal{X})} P_\Lambda(\mathcal{X}'|\mathcal{X})$$



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

Real-space RG from Information Theory perspective

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

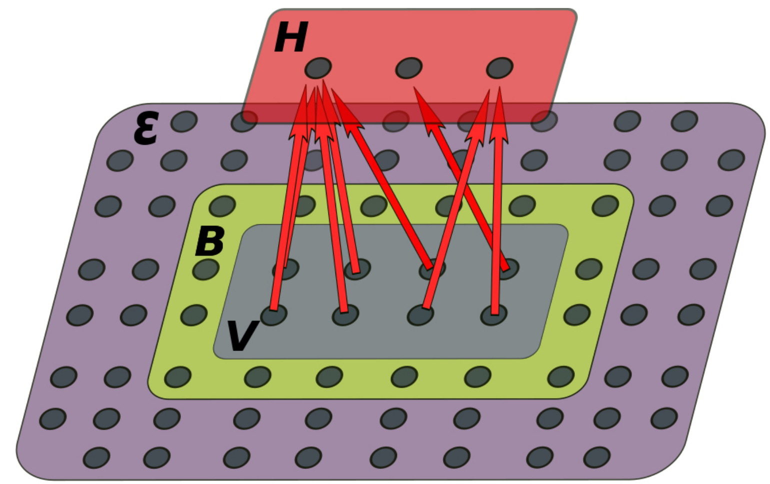


Task: Learn $P_\Lambda(\mathcal{H}|\mathcal{V})$ such that $\mathcal{H}$ tracks the most relevant degrees of freedom within region $\mathcal{V}$

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

Real-space RG from Information Theory perspective

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



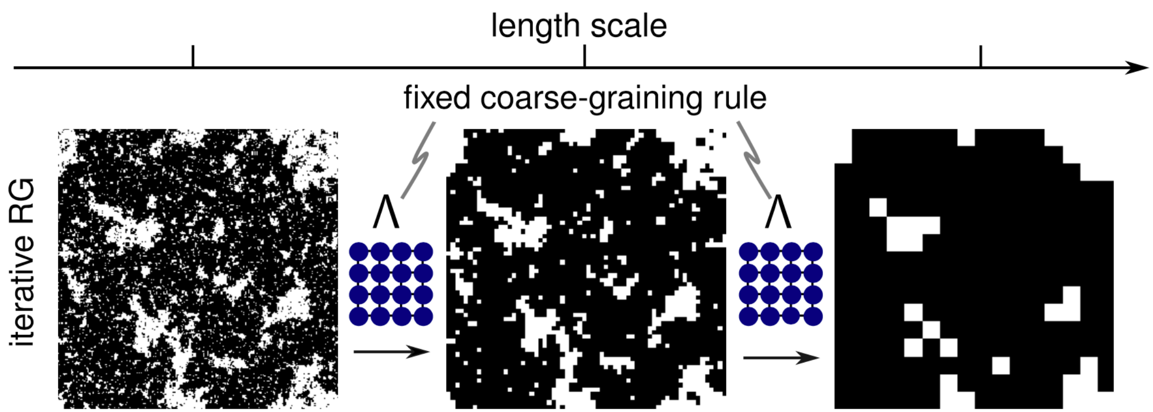
$$P(\mathcal{X}) = \frac{1}{Z} e^{\mathcal{K}(\mathcal{X})}$$

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

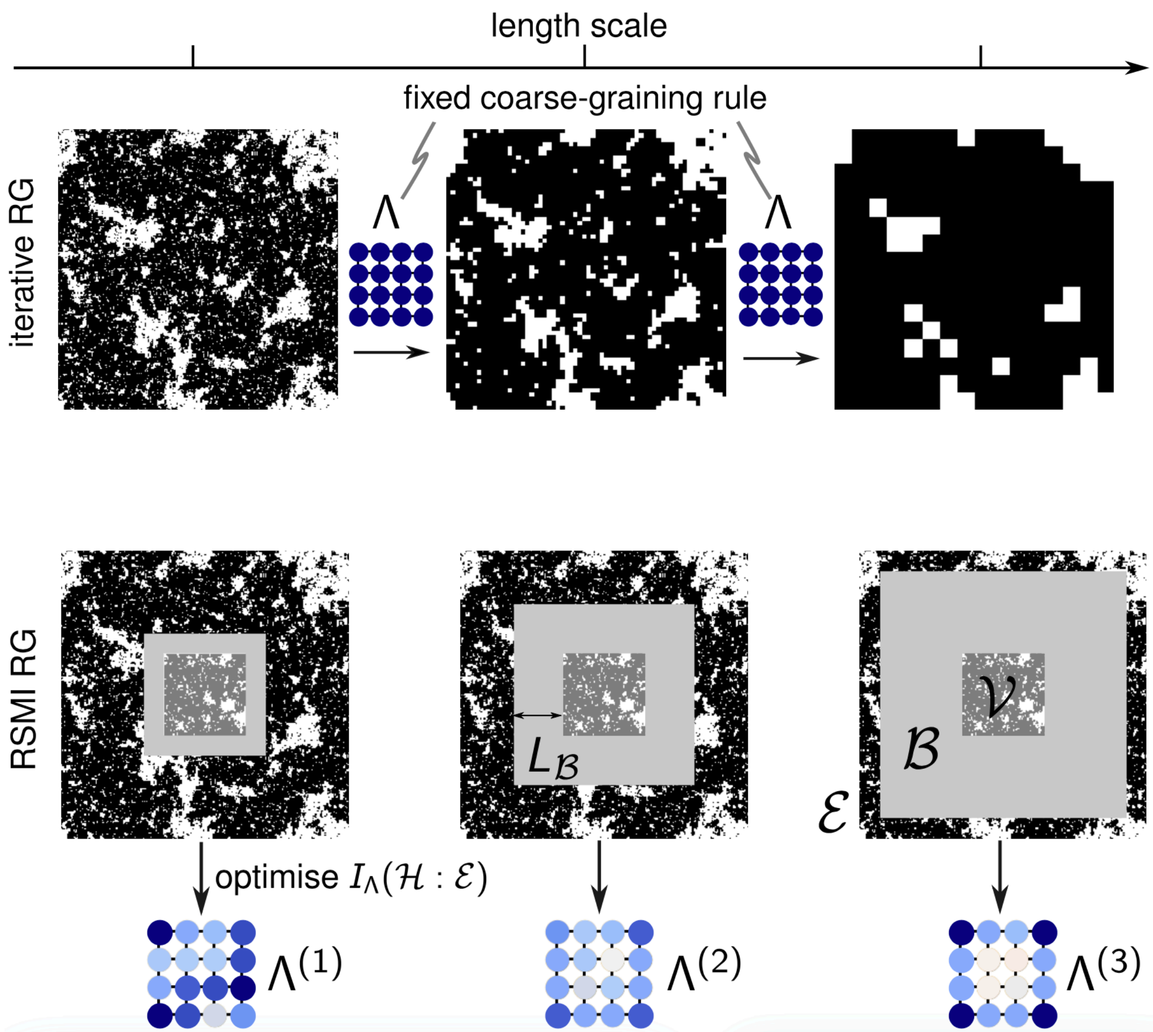
Method: find max[$I_\Lambda(\mathcal{H}:\mathcal{E})$] over parameters Λ

Iterative Fixed rs-RG vs RSMI

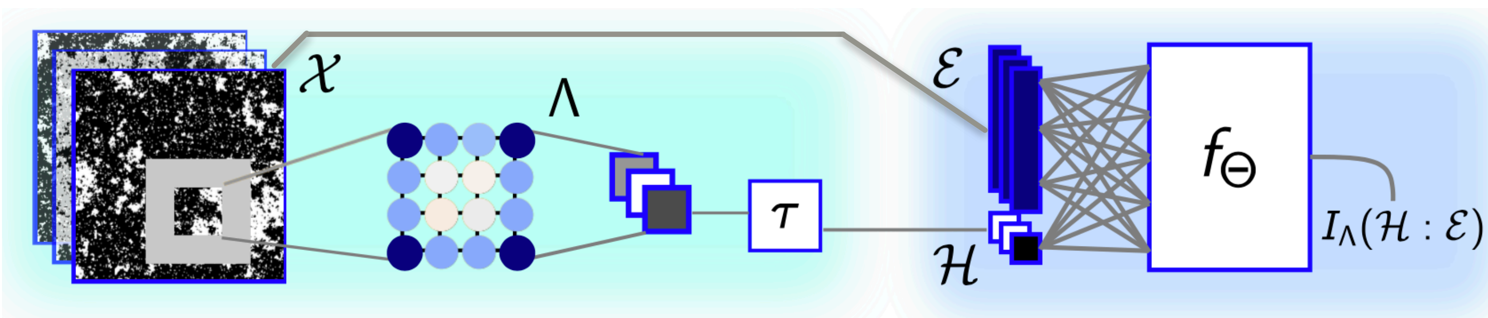
Iterative Fixed rs-RG vs RSMI



Iterative Fixed rs-RG vs RSMI

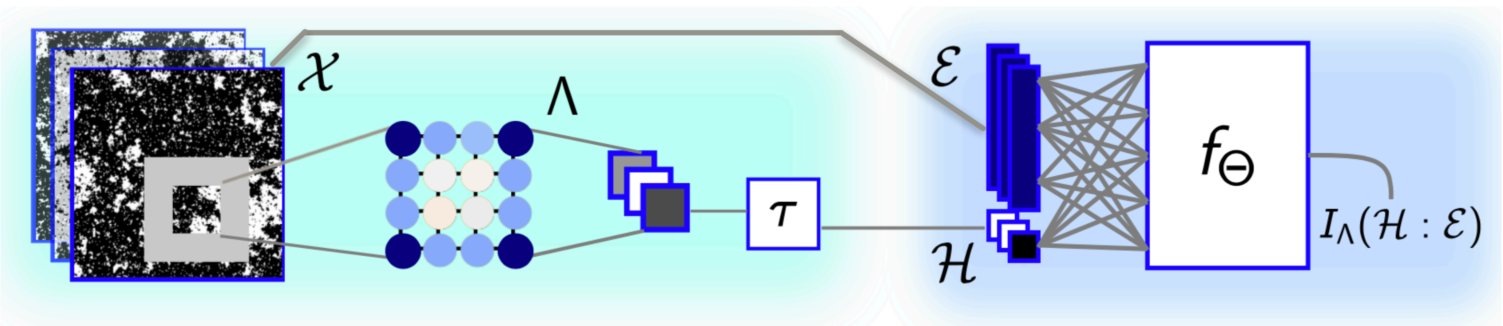


The essential ingredients



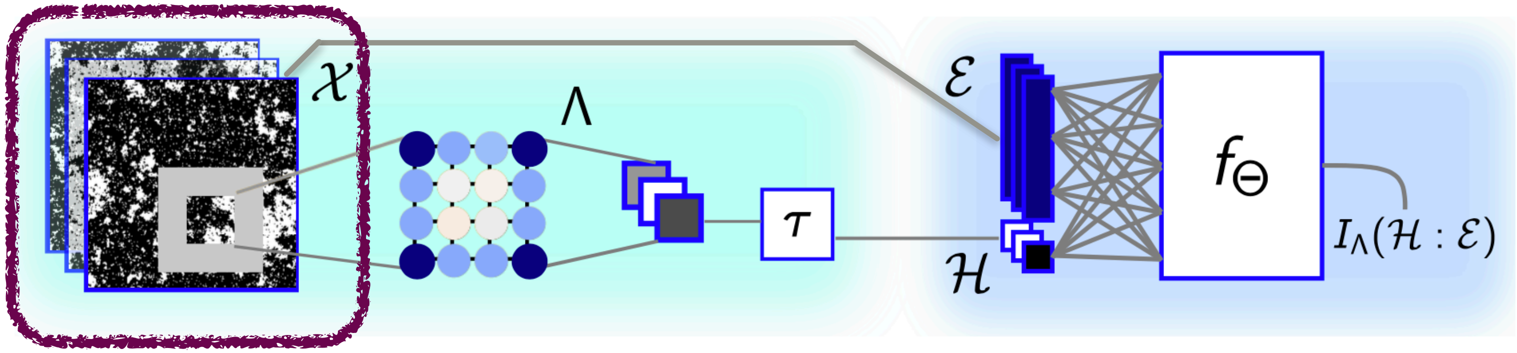
The essential ingredients

- The physical principle: “the cost function”



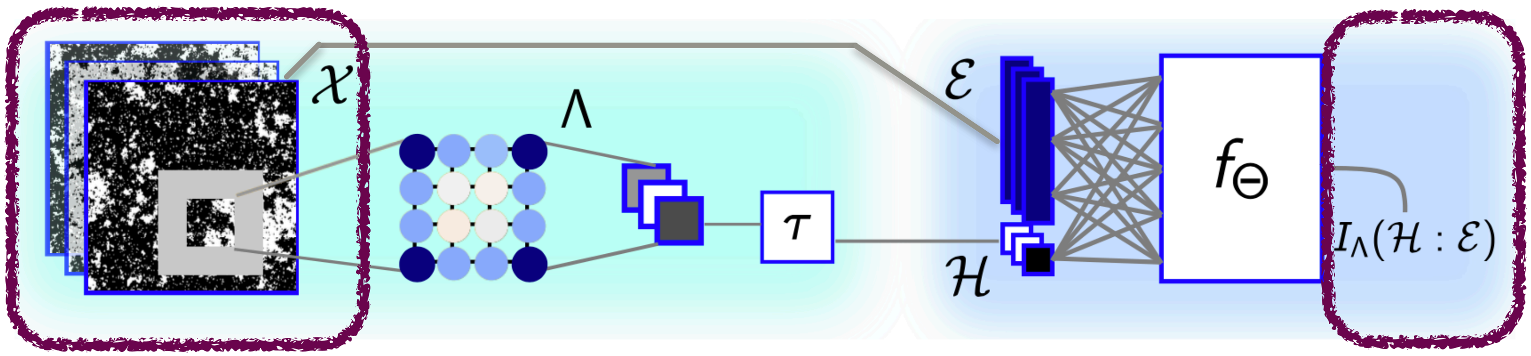
The essential ingredients

- The physical principle: “the cost function”



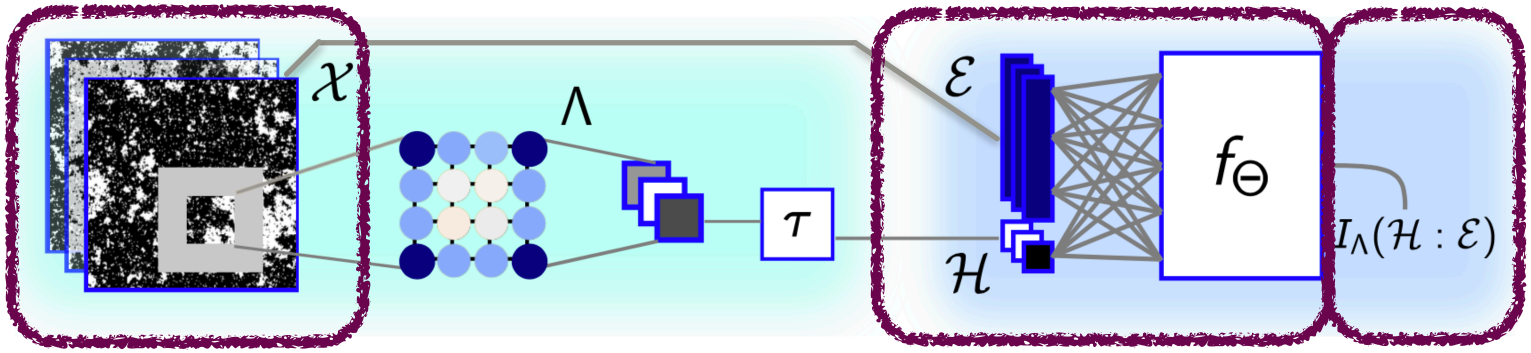
The essential ingredients

- The physical principle: “the cost function”



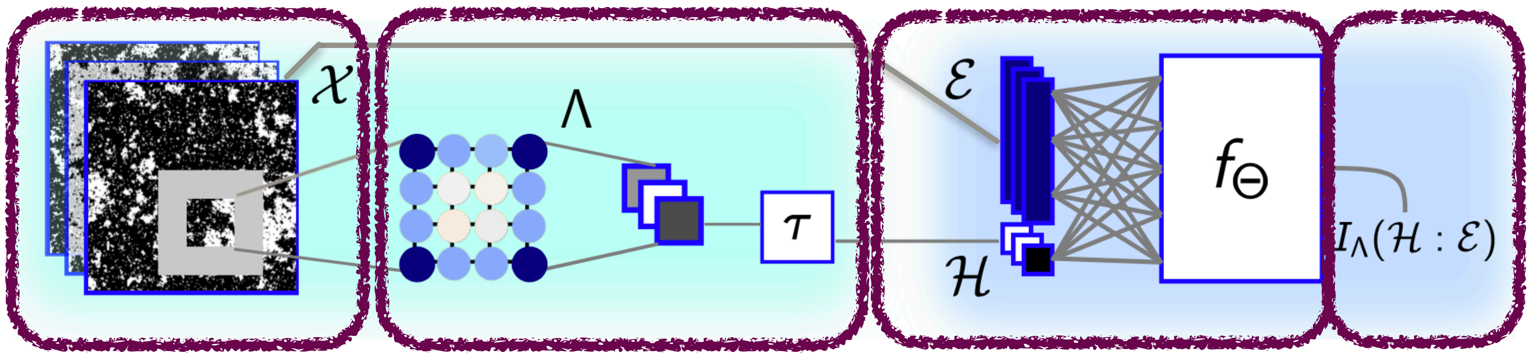
The essential ingredients

- The physical principle: “the cost function”
- The MI estimator



The essential ingredients

- The physical principle: “the cost function”
- The MI estimator
- The RG ansatz



$P_\Lambda(\mathcal{H} \mid \mathcal{V})$

Estimating MI with neural networks

Estimating MI with neural networks

$$I(X : Y) := H(X) - H(X|Y)$$

Estimating MI with neural networks

$$
\begin{aligned}
I(X:Y) &:= H(X) - H(X|Y) \\
&= \mathbb{E}_{p(x,y)}\left[\log \frac{p(x,y)}{p(x)p(y)}\right]
\end{aligned}
$$

Estimating MI with neural networks

$$
\begin{aligned}
I(X:Y) &:= H(X) - H(X|Y) \\
&= \mathbb{E}_{p(x,y)}\left[\log \frac{p(x,y)}{p(x)p(y)}\right] \\
&=: D_{\mathrm{KL}}\left[p(x,y)||p(x)p(y)\right]
\end{aligned}
$$

Estimating MI with neural networks

- Common methods are computationally demanding and/or not differentiable

$$
\begin{aligned}
I(X:Y) &:= H(X) - H(X|Y) \\
&= \mathbb{E}_{p(x,y)}\left[\log \frac{p(x,y)}{p(x)p(y)}\right] \\
&=: D_{\mathrm{KL}}\left[p(x,y) || p(x)p(y)\right]
\end{aligned}
$$

Estimating MI with neural networks

- Common methods are computationally demanding and/or not differentiable
- Estimation by optimising variational bounds

$$
\begin{aligned}
I(X:Y) &:= H(X) - H(X|Y) \\
&= \mathbb{E}_{p(x,y)}\left[\log \frac{p(x,y)}{p(x)p(y)}\right] \\
&=: D_{\mathrm{KL}}\left[p(x,y) || p(x)p(y)\right]
\end{aligned}
$$

Estimating MI with neural networks

- Common methods are computationally demanding and/or not differentiable
- Estimation by optimising variational bounds

$$D_{KL}(\mathbb{P} \,||\, \mathbb{Q}) = \sup_{T:\Omega\to\mathbb{R}} \mathbb{E}_{\mathbb{P}}[T] - \log(\mathbb{E}_{\mathbb{Q}}[e^T])$$

$$\begin{aligned} I(X:Y) &:= H(X) - H(X|Y) \\ &= \mathbb{E}_{p(x,y)}\left[\log \frac{p(x,y)}{p(x)p(y)}\right] \\ &=: D_{\mathrm{KL}}\left[p(x,y)||p(x)p(y)\right] \end{aligned}$$

Estimating MI with neural networks

- Common methods are computationally demanding and/or not differentiable

- Estimation by optimising variational bounds

$$D_{KL}(\mathbb{P} \,||\, \mathbb{Q}) = \sup_{T:\Omega\to\mathbb{R}} \mathbb{E}_{\mathbb{P}}[T] - \log(\mathbb{E}_{\mathbb{Q}}[e^T])$$

Make T a neural network! MINE, Belghazi et al. (2018)

$$\begin{aligned} I(X:Y) &:= H(X) - H(X|Y) \\ &= \mathbb{E}_{p(x,y)}\left[\log \frac{p(x,y)}{p(x)p(y)}\right] \\ &=: D_{\mathrm{KL}}\left[p(x,y)||p(x)p(y)\right] \end{aligned}$$

Estimating MI with neural networks

- Common methods are computationally demanding and/or not differentiable

$$I(X:Y) := H(X) - H(X|Y) = \mathbb{E}_{p(x,y)}\left[\log \frac{p(x,y)}{p(x)p(y)}\right] =: D_{\mathrm{KL}}\left[p(x,y)||p(x)p(y)\right]$$

- Estimation by optimising variational bounds

$$D_{KL}(\mathbb{P} \,||\, \mathbb{Q}) = \sup_{T:\Omega\to\mathbb{R}} \mathbb{E}_{\mathbb{P}}[T] - \log(\mathbb{E}_{\mathbb{Q}}[e^{T}])$$

Make T a neural network! MINE, Belghazi et al. (2018)

- MINE can be improved in many respects (e.g. variance)

Estimating MI with neural networks

- Common methods are computationally demanding and/or not differentiable

$$I(X:Y) := H(X) - H(X|Y)$$
$$= \mathbb{E}_{p(x,y)}\left[\log \frac{p(x,y)}{p(x)p(y)}\right]$$
$$=: D_{\mathrm{KL}}\left[p(x,y)||p(x)p(y)\right]$$

- Estimation by optimising variational bounds

$$D_{KL}(\mathbb{P} \,||\, \mathbb{Q}) = \sup_{T:\Omega\to\mathbb{R}} \mathbb{E}_{\mathbb{P}}[T] - \log(\mathbb{E}_{\mathbb{Q}}[e^T])$$

Make T a neural network! MINE, Belghazi et al. (2018)

- MINE can be improved in many respects (e.g. variance)

$$I_{\mathrm{NWJ}}(X:Y) := \mathbb{E}_{p(x,y)}[f(x,y)] - e^{-1}\mathbb{E}_{p(x)p(y)}[e^{f(x,y)}]$$

Nguyen, Wainwright, Jordan (2010)

Estimating MI with neural networks

- Common methods are computationally demanding and/or not differentiable

$$I(X:Y) := H(X) - H(X|Y)$$

$$= \mathbb{E}_{p(x,y)}\left[\log \frac{p(x,y)}{p(x)p(y)}\right]$$

$$=: D_{\mathrm{KL}}\left[p(x,y)||p(x)p(y)\right]$$

- Estimation by optimising variational bounds

$$D_{KL}(\mathbb{P}\,||\,\mathbb{Q}) = \sup_{T:\Omega\to\mathbb{R}} \mathbb{E}_{\mathbb{P}}[T] - \log(\mathbb{E}_{\mathbb{Q}}[e^T])$$

Make T a neural network! MINE, Belghazi et al. (2018)

- MINE can be improved in many respects (e.g. variance)

$$I_{\mathrm{NWJ}}(X:Y) := \mathbb{E}_{p(x,y)}[f(x,y)] - e^{-1}\mathbb{E}_{p(x)p(y)}[e^{f(x,y)}]$$

Nguyen, Wainwright, Jordan (2010)

- We end up with a parametric, differentiable, and tight lower bound

Estimating MI with neural networks

- Common methods are computationally demanding and/or not differentiable

$$I(X:Y) := H(X) - H(X|Y)$$
$$= \mathbb{E}_{p(x,y)}\left[\log \frac{p(x,y)}{p(x)p(y)}\right]$$
$$=: D_{\mathrm{KL}}\left[p(x,y)||p(x)p(y)\right]$$

- Estimation by optimising variational bounds

$$D_{KL}(\mathbb{P} \,||\, \mathbb{Q}) = \sup_{T:\Omega\to\mathbb{R}} \mathbb{E}_{\mathbb{P}}[T] - \log(\mathbb{E}_{\mathbb{Q}}[e^T])$$

Make T a neural network! MINE, Belghazi et al. (2018)

- MINE can be improved in many respects (e.g. variance)

$$I_{\mathrm{NWJ}}(X:Y) := \mathbb{E}_{p(x,y)}[f(x,y)] - e^{-1}\mathbb{E}_{p(x)p(y)}[e^{f(x,y)}]$$

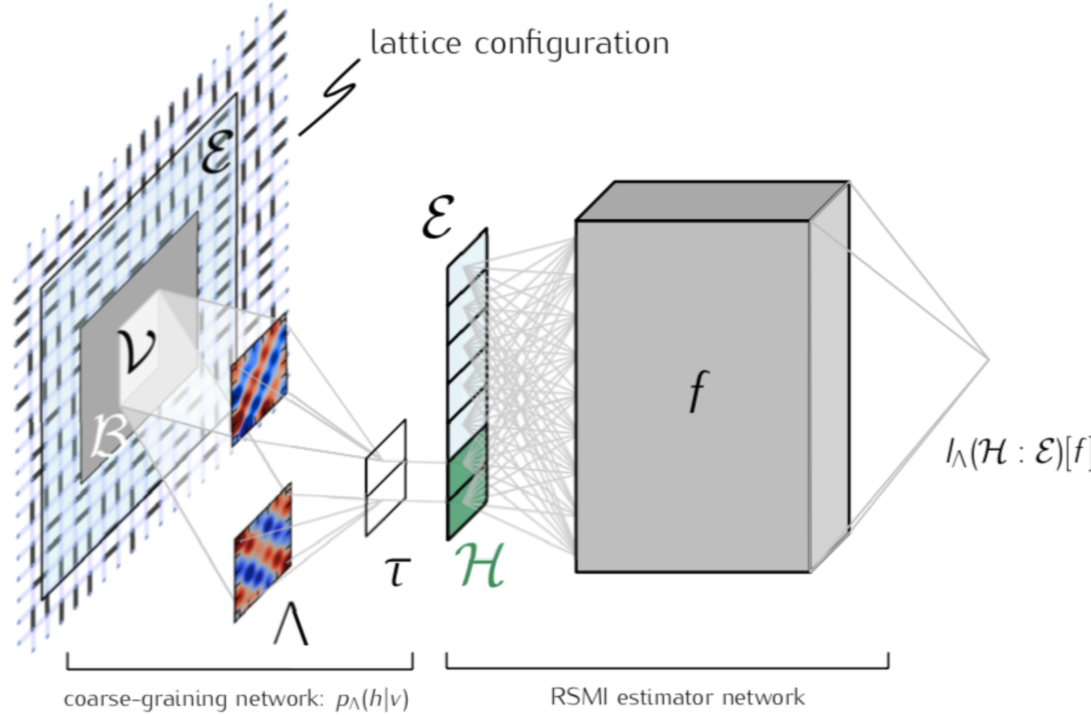
Nguyen, Wainwright, Jordan (2010)

- We end up with a parametric, differentiable, and tight lower bound

InfoNCE, van den Oord et al. (2018)

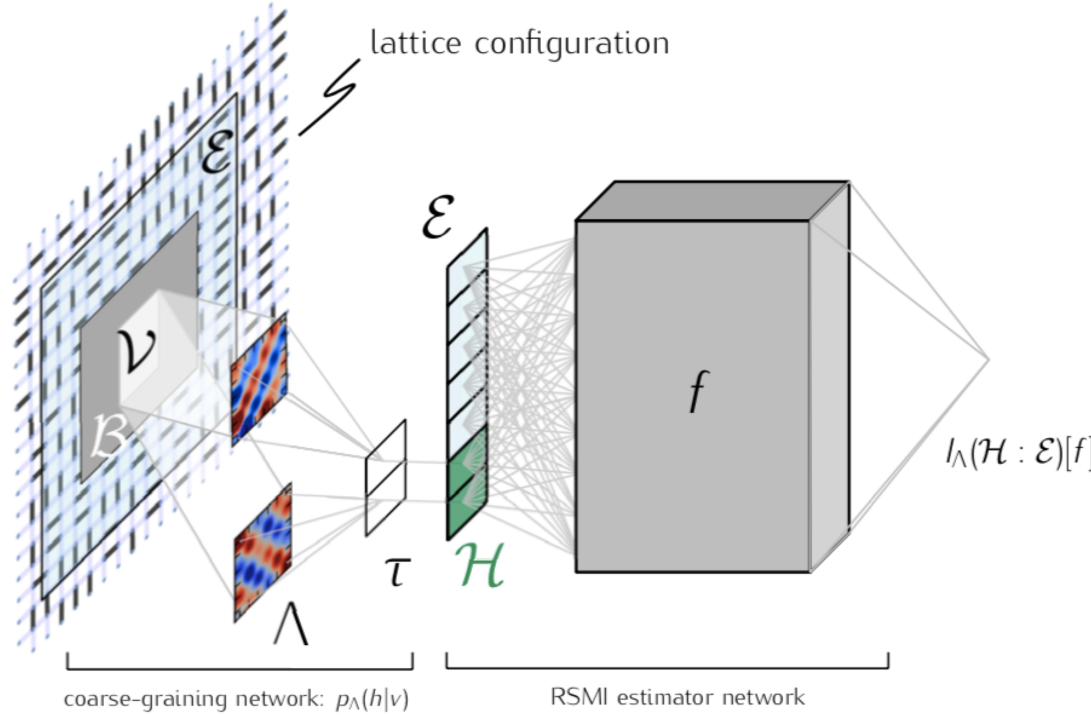
Poole et al. ICMLR (2019) "On variational bounds of mutual information"

The RSMI-NE network



The RSMI-NE network

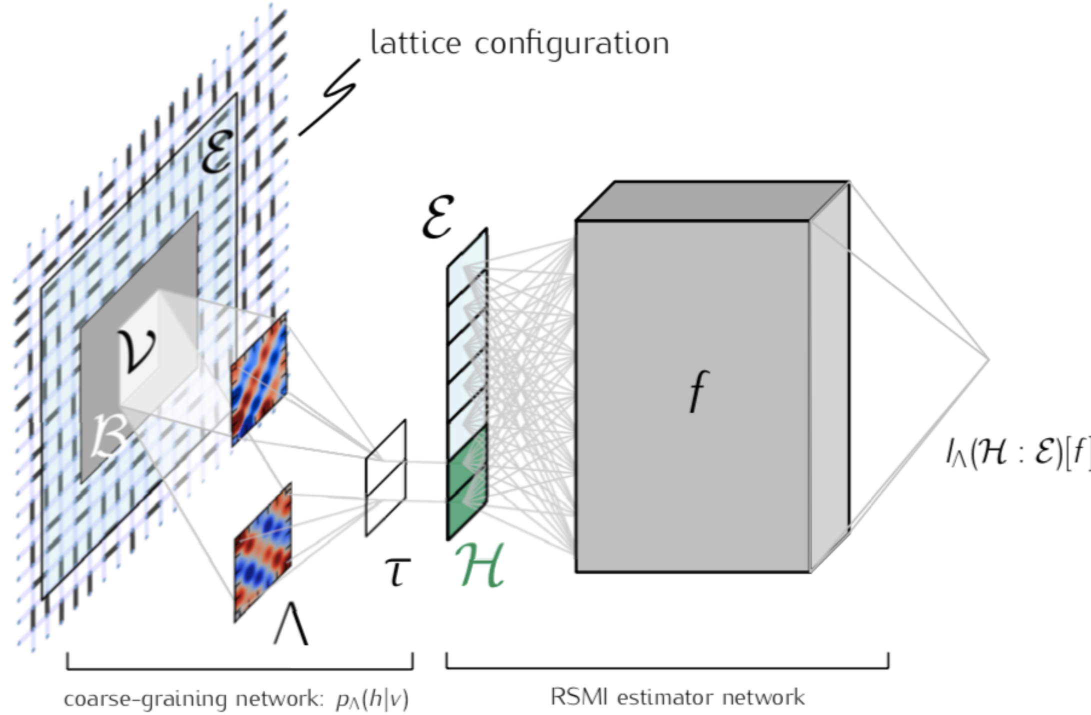
$$p_\Lambda(h|v) : v \mapsto h = \tau \circ (\Lambda \cdot v)$$



The RSMI-NE network

$$p_\Lambda(h|v) : v \mapsto h = \tau \circ (\Lambda \cdot v)$$

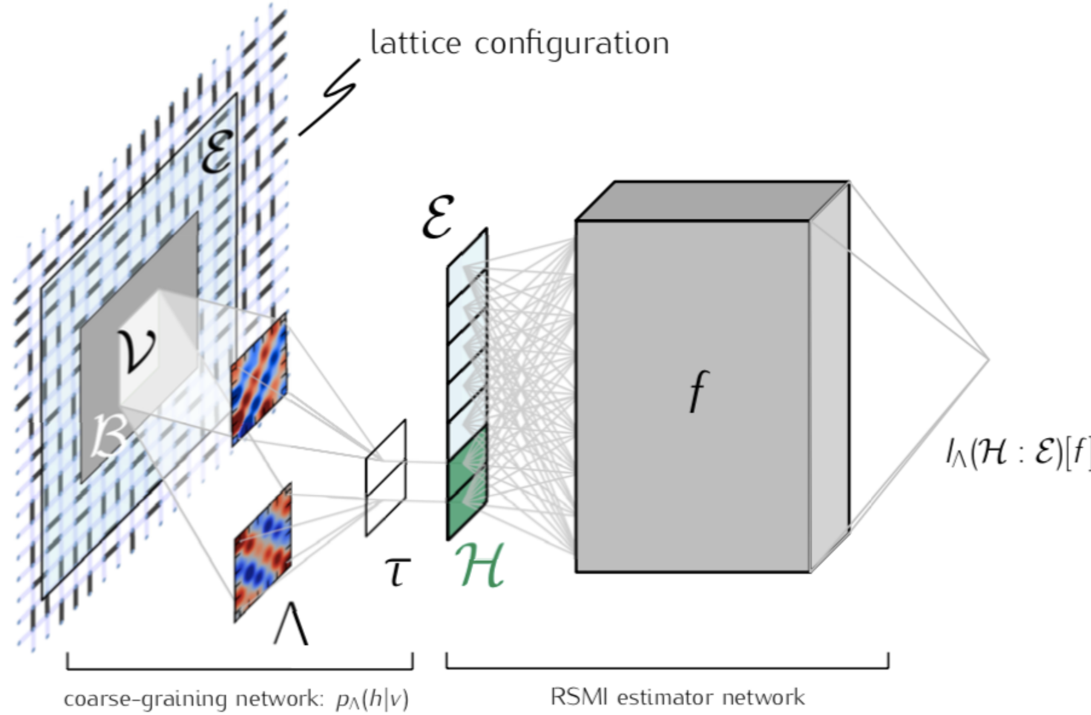
- The coarse-graining convolution(s)

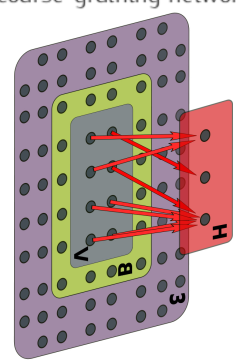


The RSMI-NE network

$$p_\Lambda(h|v) : v \mapsto h = \tau \circ (\Lambda \cdot v)$$

- The coarse-graining convolution(s)



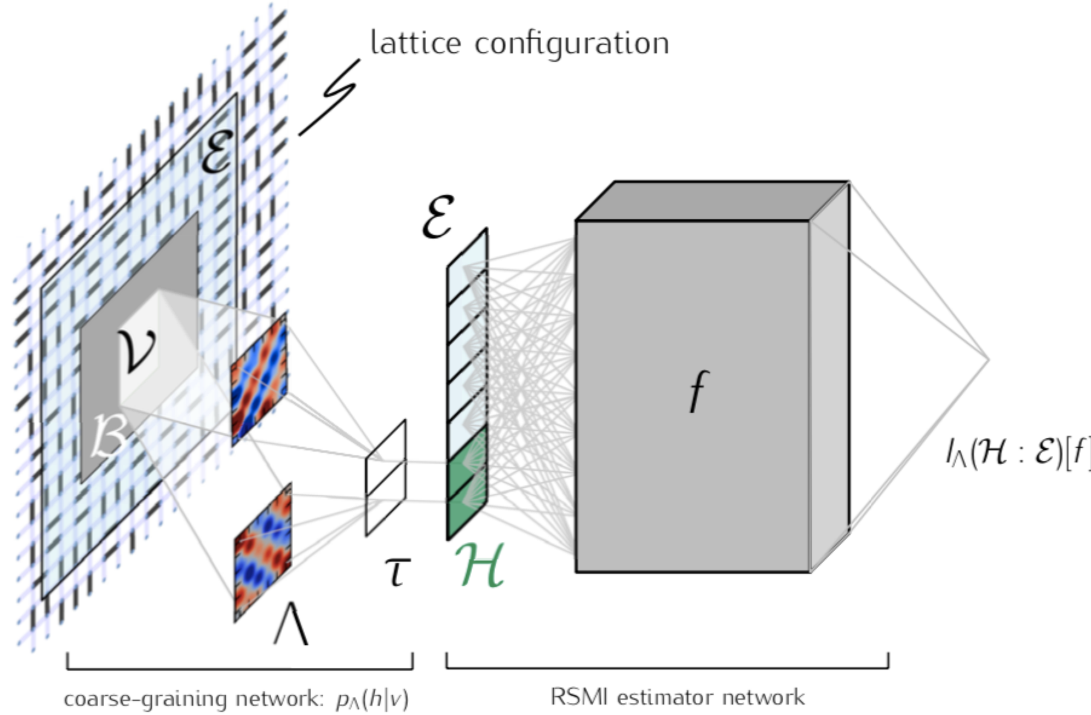


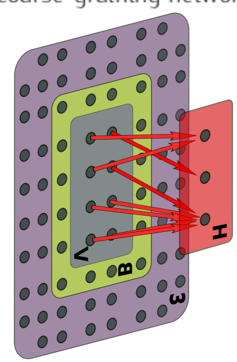
The RSMI-NE network

$$p_\Lambda(h|v) : v \mapsto h = \tau \circ (\Lambda \cdot v)$$

- The coarse-graining convolution(s)
- Differentiable ``discretizer``

Jang, Gu, Poole **ICLR (2017)**
Maddison, Mnih, Teh **ICLR (2017)**



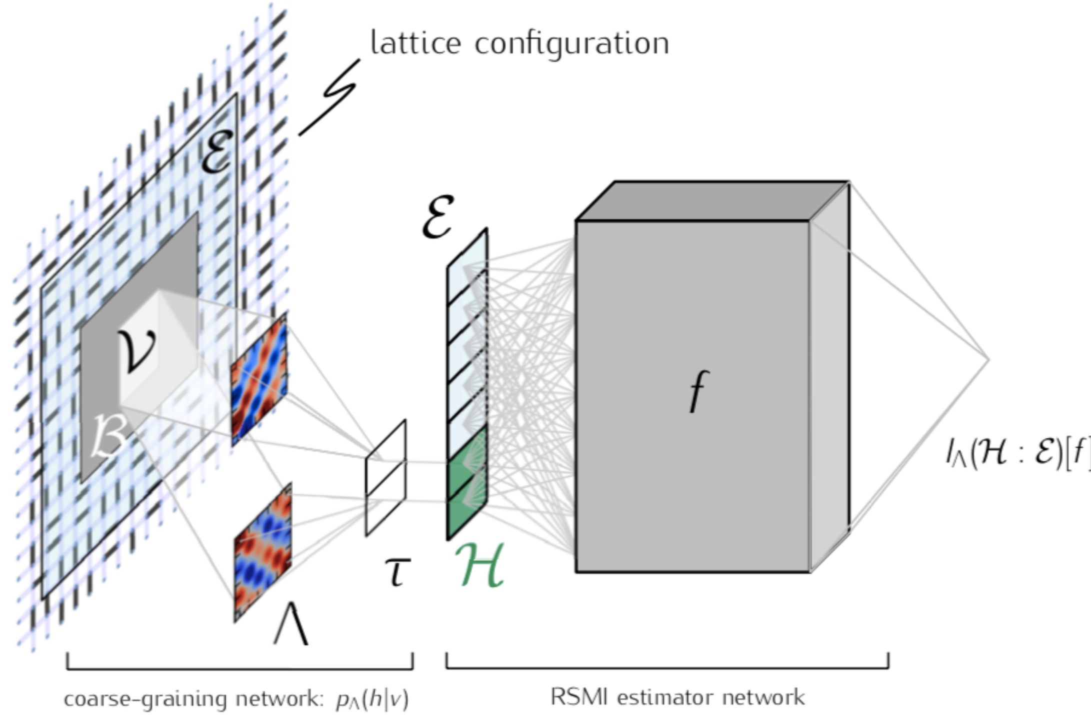


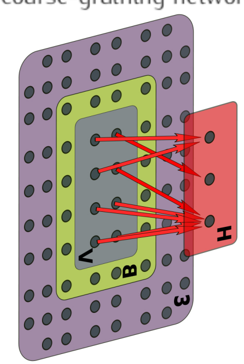
The RSMI-NE network

$$p_\Lambda(h|v) : v \mapsto h = \tau \circ (\Lambda \cdot v)$$

- The coarse-graining convolution(s)
- Differentiable ``discretizer``

  Jang, Gu, Poole **ICLR (2017)**
  Maddison, Mnih, Teh **ICLR (2017)**





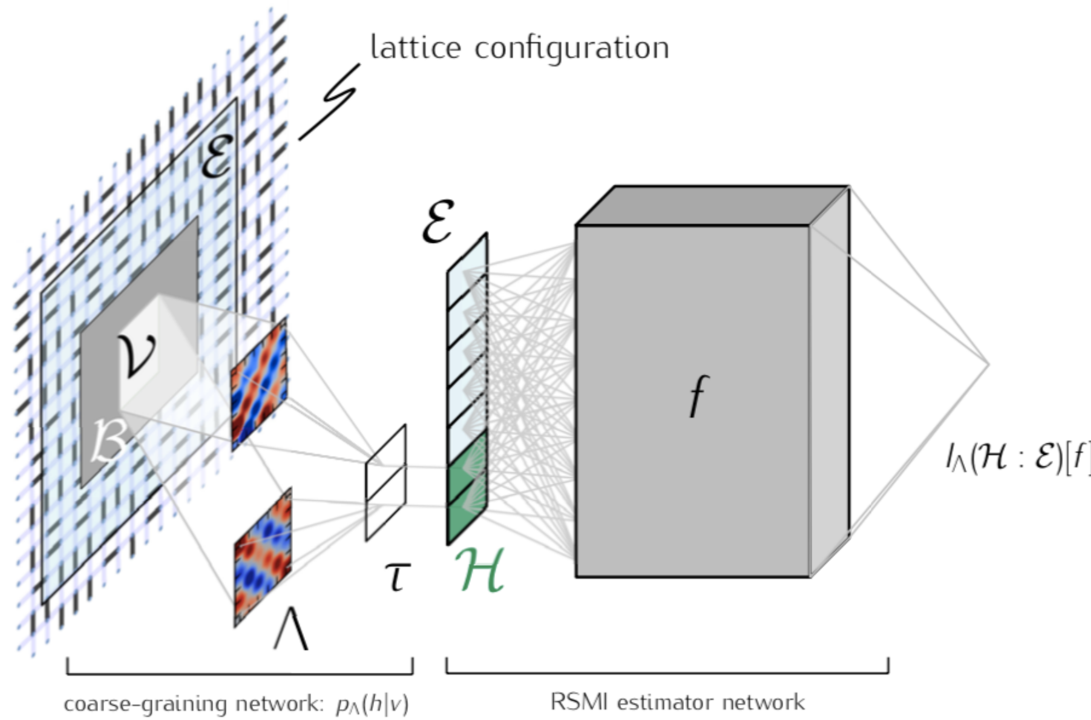
- The RSMI estimator and the coarse-graining Ansatz are chained together

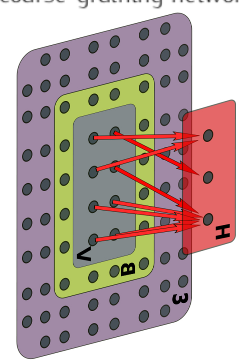
The RSMI-NE network

$$p_\Lambda(h|v) : v \mapsto h = \tau \circ (\Lambda \cdot v)$$

- The coarse-graining convolution(s)
- Differentiable ``discretizer``

  Jang, Gu, Poole **ICLR (2017)**
  Maddison, Mnih, Teh **ICLR (2017)**

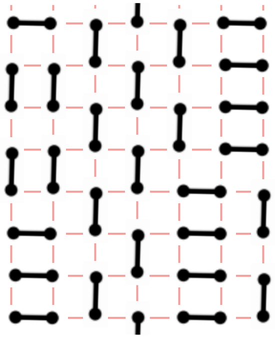




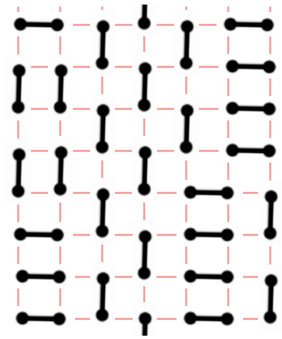
- The RSMI estimator and the coarse-graining Ansatz are chained together
- They are co-trained with SGD as a single network **(because: differentiable, upper bounded!)**

Example: interacting dimer model

Example: interacting dimer model



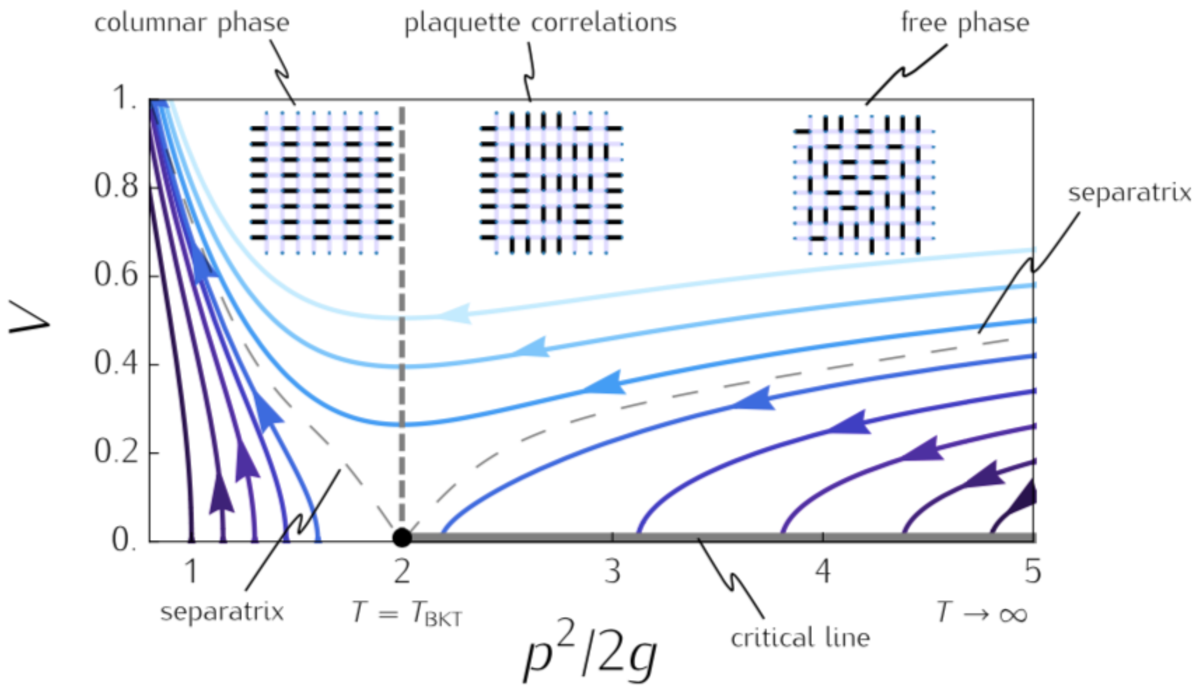
Example: interacting dimer model

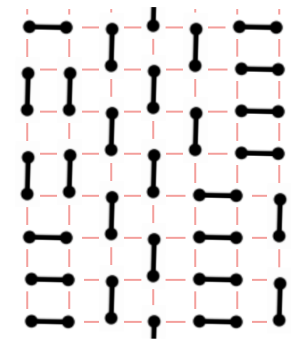


$$Z = \sum_c \exp(-E_c/T),$$

$$E_c = v[(N^c(=) + N^c(\|))].$$

Example: interacting dimer model





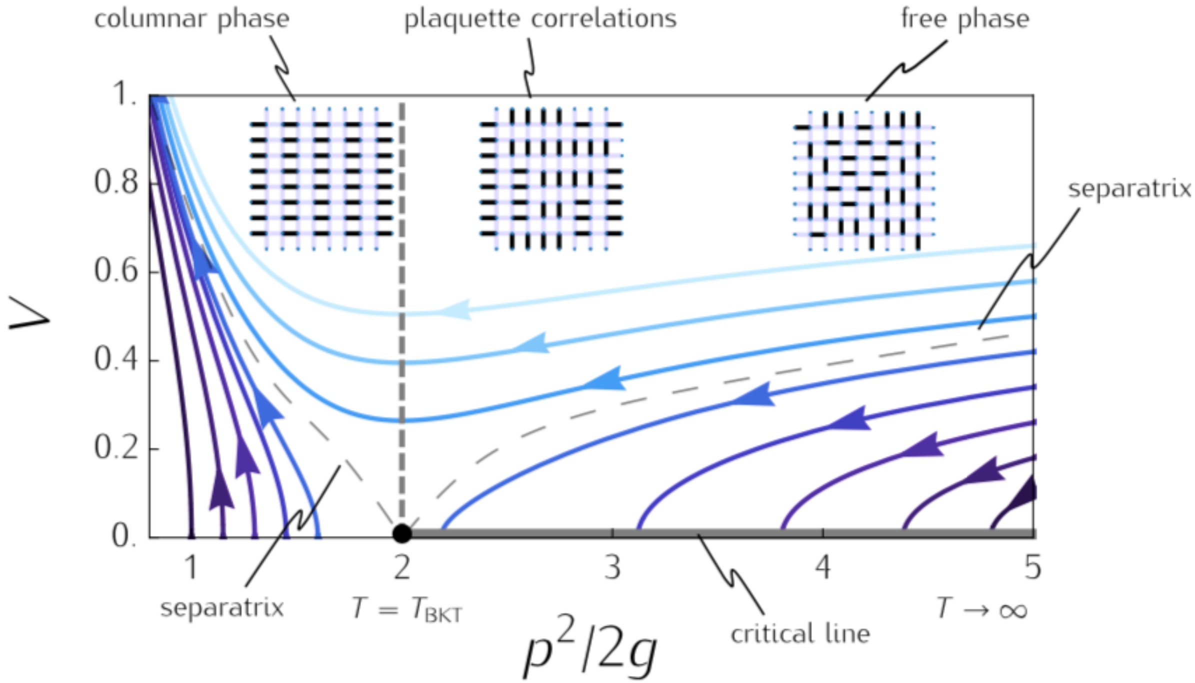
$$Z = \sum_c \exp(-E_c/T),$$

$$E_c = v[(N^c(\mathbf{=}) + N^c(\mathbf{||}))].$$

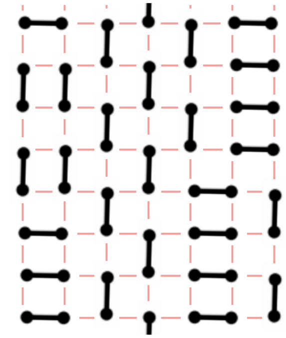
$$S[\varphi(\mathbf{r})] = \int \mathrm{d}^2\mathbf{r} \left[\frac{g(T)}{2} |\nabla\varphi(\mathbf{r})|^2 + V \cos\left(4\varphi(\mathbf{r})\right) \right]$$

Alet et al. **PRE 74**, 041124 (2006)

Example: interacting dimer model



$$S[\varphi(\mathbf{r})] = \int \mathrm{d}^2\mathbf{r} \left[\frac{g(T)}{2} |\nabla\varphi(\mathbf{r})|^2 + V \cos\left(4\varphi(\mathbf{r})\right) \right]$$

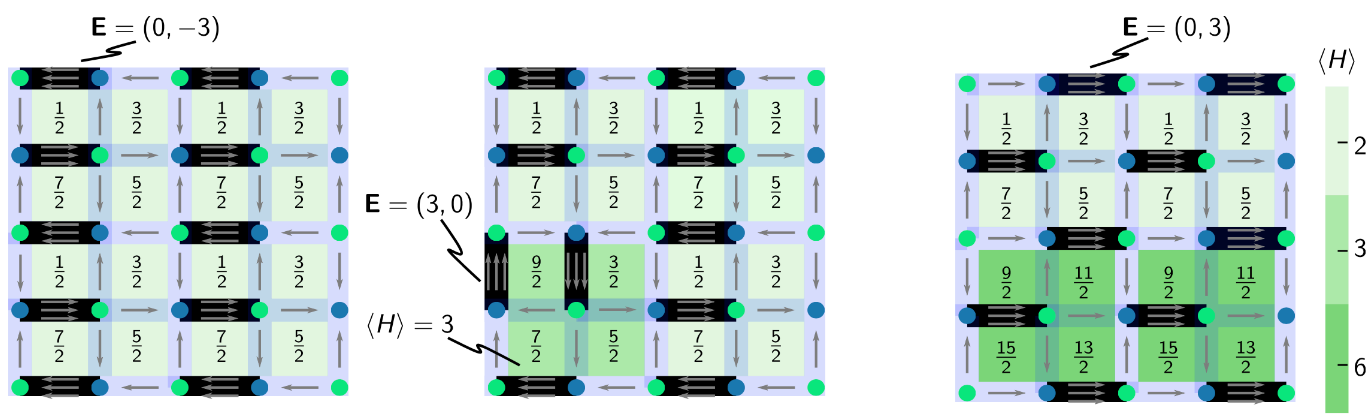


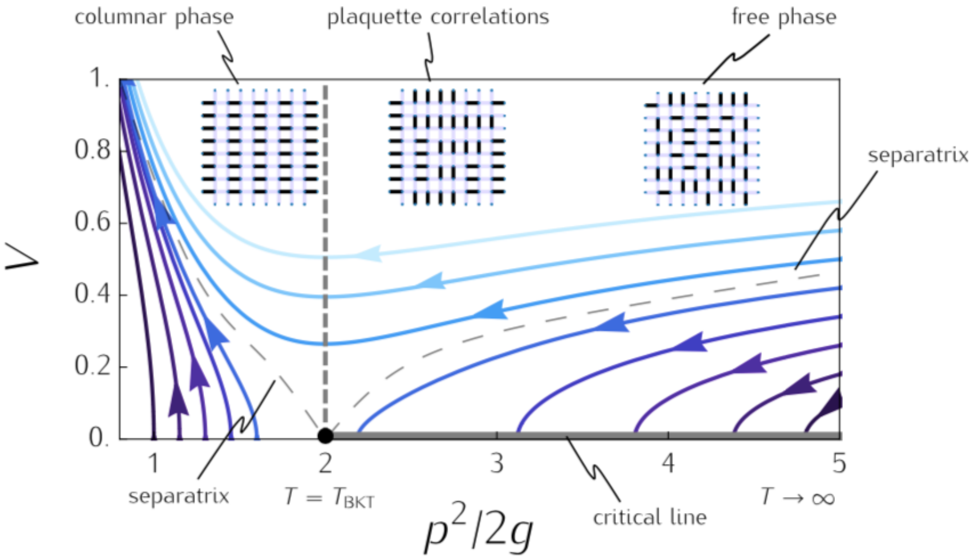
$$Z = \sum_c \exp(-E_c/T),$$

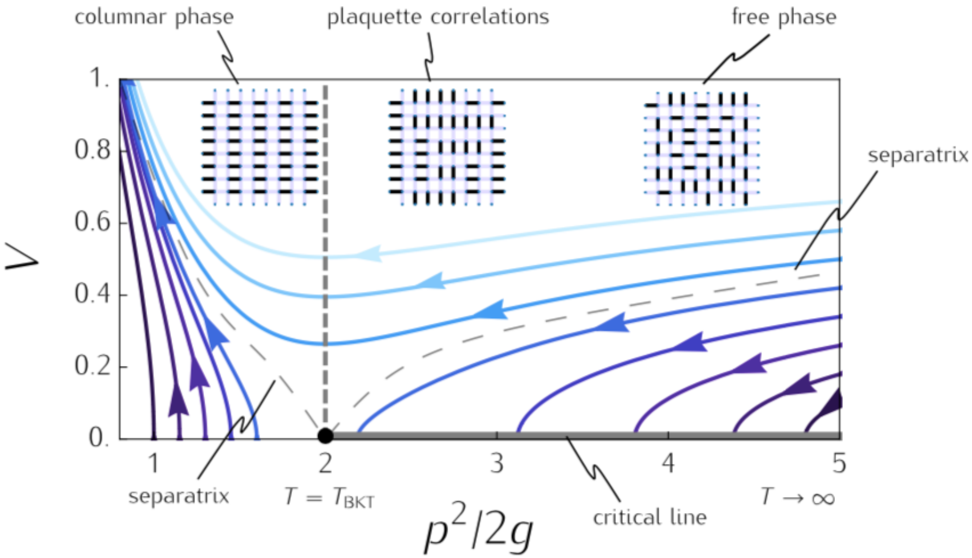
$$E_c = v[(N^c(=) + N^c(\|))].$$

Alet et al. **PRE 74**, 041124 (2006)

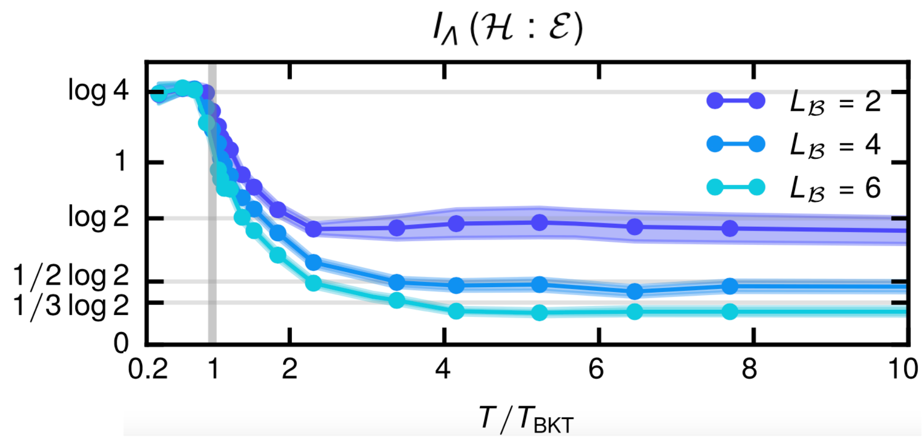
RG of dimer model: mapping to height field

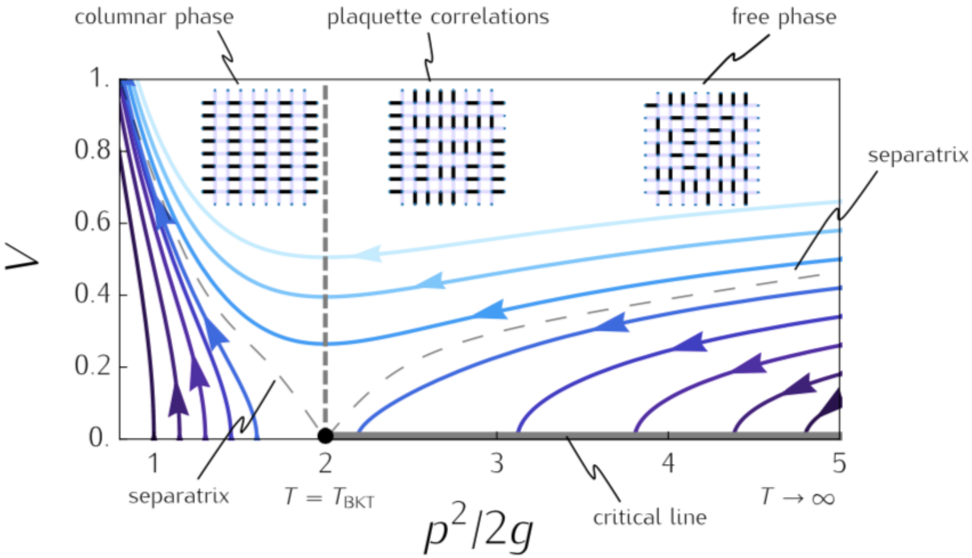



columnar phase
plaquette correlations
free phase
separatrix
separatrix
V
1.
0.8
0.6
0.4
0.2
0.
1
2
3
4
5
$T = T_{BKT}$
critical line
$T \to \infty$
$p^2/2g$

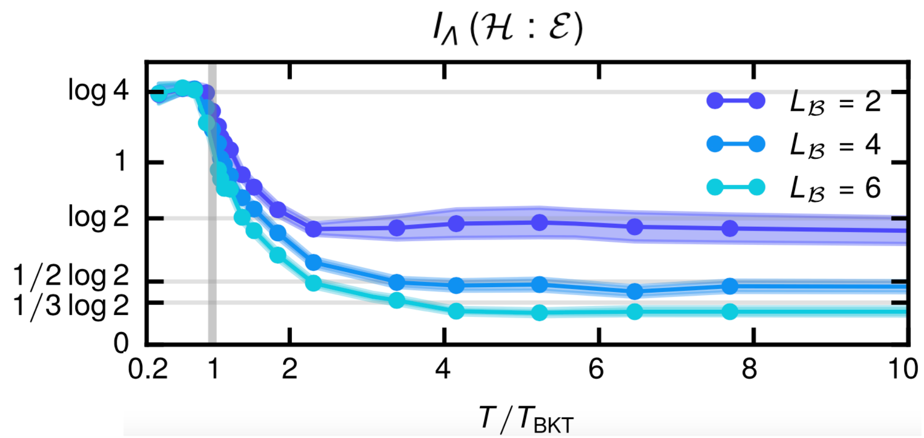


- Total RSMI with the *optimal* filter

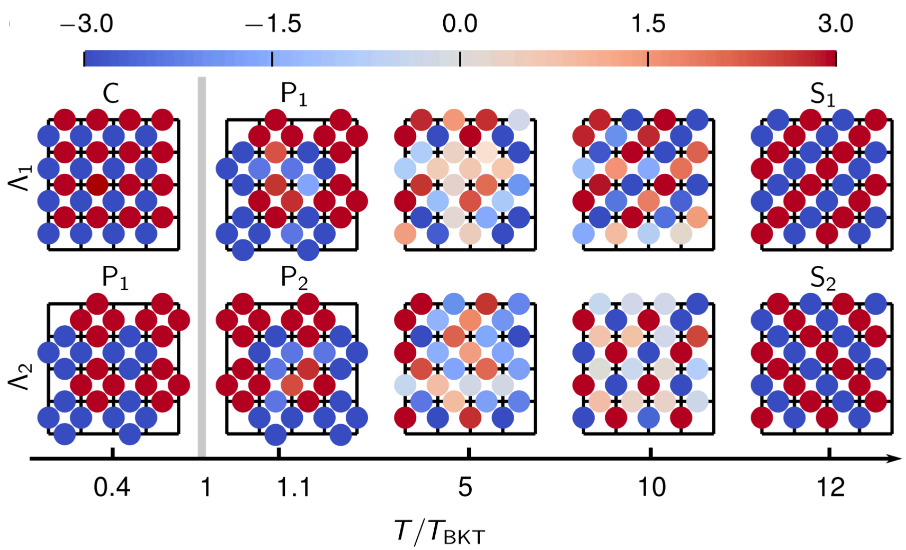


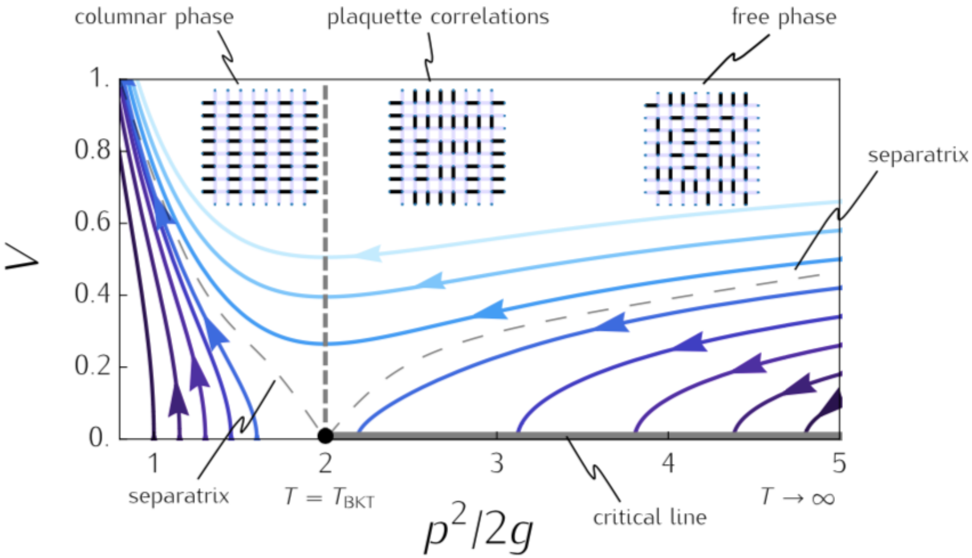


- Total RSMI with the *optimal* filter

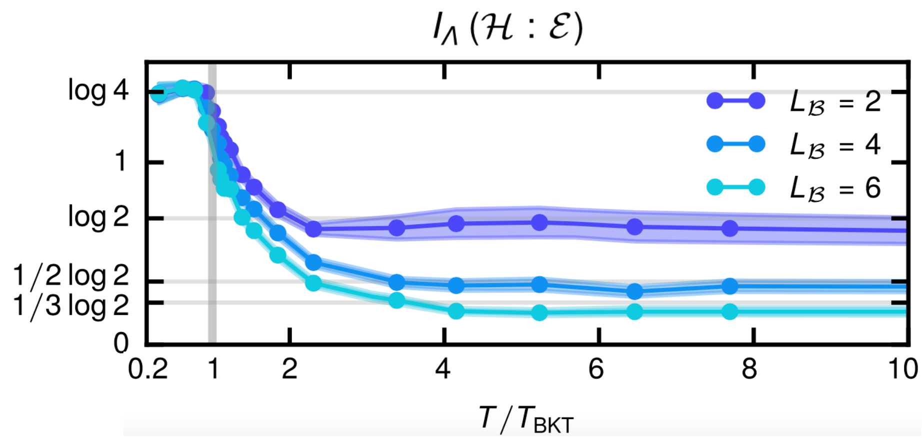


- The optimal filters ***depend*** on T

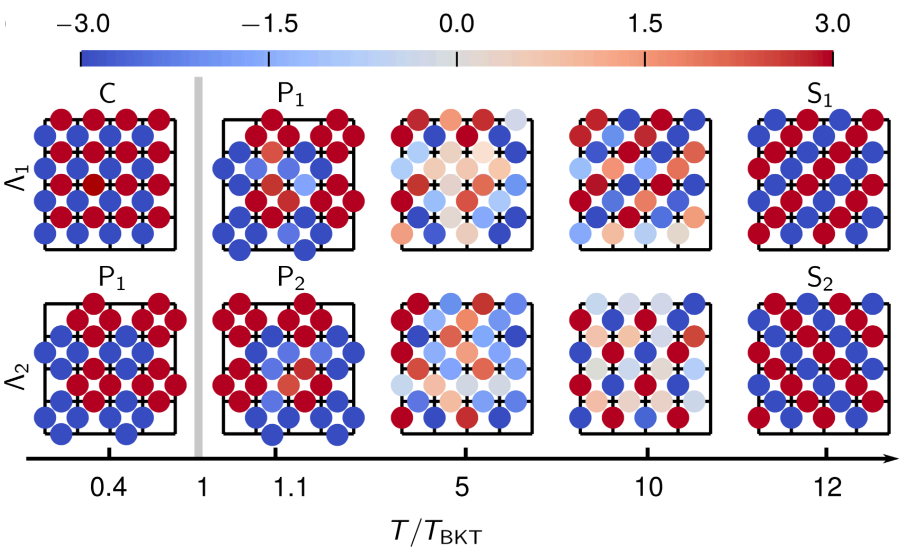




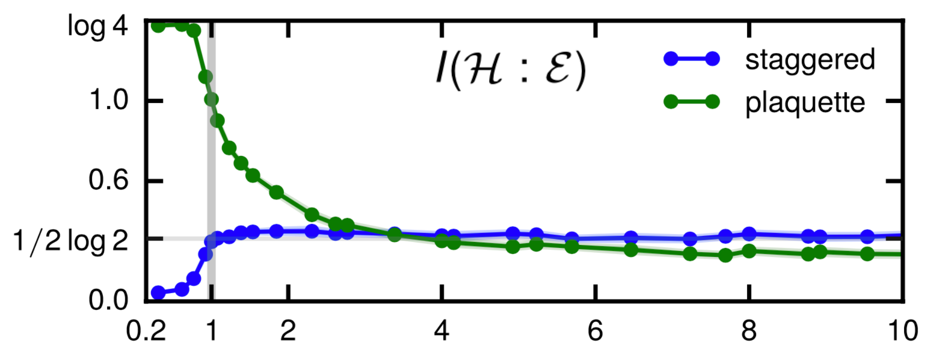
- Total RSMI with the *optimal* filter



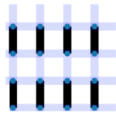
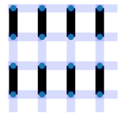
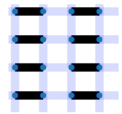
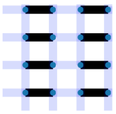
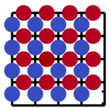
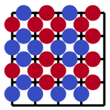
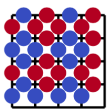
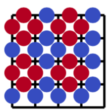
- The optimal filters ***depend*** on T



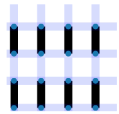
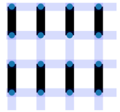
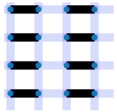
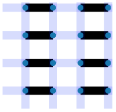
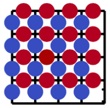
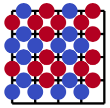
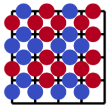
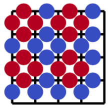
- RSMI retrieved by pristine filters reflect **competing correlations**



- Pairs of C/P filters **label broken symmetry** states

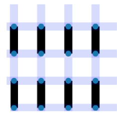
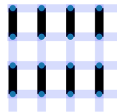
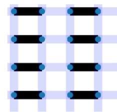
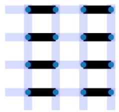
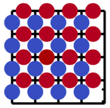
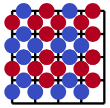
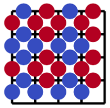
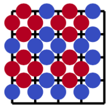
$C(\mathbf{r})$				
Λ_C  Λ_{P1} 	$(-1,-1)$	$(-1,+1)$	$(+1,-1)$	$(+1,+1)$
Λ_{P1}  Λ_{P2} 	$(-1,-1)$	$(+1,+1)$	$(-1,+1)$	$(+1,-1)$

- Pairs of C/P filters **label broken symmetry states**

$C(\mathbf{r})$				
Λ_C  Λ_{P1} 	$(-1,-1)$	$(-1,+1)$	$(+1,-1)$	$(+1,+1)$
Λ_{P1}  Λ_{P2} 	$(-1,-1)$	$(+1,+1)$	$(-1,+1)$	$(+1,-1)$

- Filters **define** order parameters:

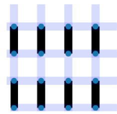
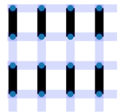
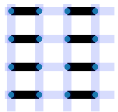
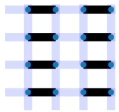
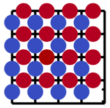
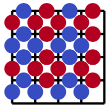
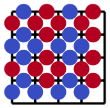
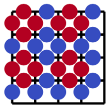
- Pairs of C/P filters **label broken symmetry states**

$C(\mathbf{r})$				
Λ_C  Λ_{P1} 	$(-1,-1)$	$(-1,+1)$	$(+1,-1)$	$(+1,+1)$
Λ_{P1}  Λ_{P2} 	$(-1,-1)$	$(+1,+1)$	$(-1,+1)$	$(+1,-1)$

- Filters **define** order parameters:

$$D_i := \mathbb{E}\left[\frac{1}{N_{\mathcal{V}}}\sum_k \tau \circ (\Lambda_i \cdot \mathcal{V}_k)\right]$$

- Pairs of C/P filters **label broken symmetry states**

$C(\mathbf{r})$				
Λ_{C}  Λ_{P1} 	$(-1,-1)$	$(-1,+1)$	$(+1,-1)$	$(+1,+1)$
Λ_{P1}  Λ_{P2} 	$(-1,-1)$	$(+1,+1)$	$(-1,+1)$	$(+1,-1)$

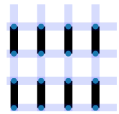
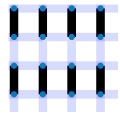
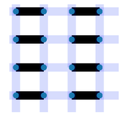
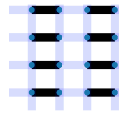
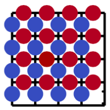
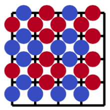
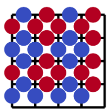
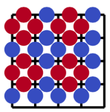
- Filters **define** order parameters:

$$D_i := \mathbb{E}\left[\frac{1}{N_{\mathcal{V}}}\sum_k \tau \circ (\Lambda_i \cdot \mathcal{V}_k)\right]$$

$$\mathrm{DSB} := \mathbb{E}\left[\sum_k \tau \circ (\Lambda_{\mathrm{C}} \cdot \mathcal{V}_k)\right]$$

Alet et al. **PRE 74**, 041124 (2006)

- Pairs of C/P filters **label broken symmetry states**

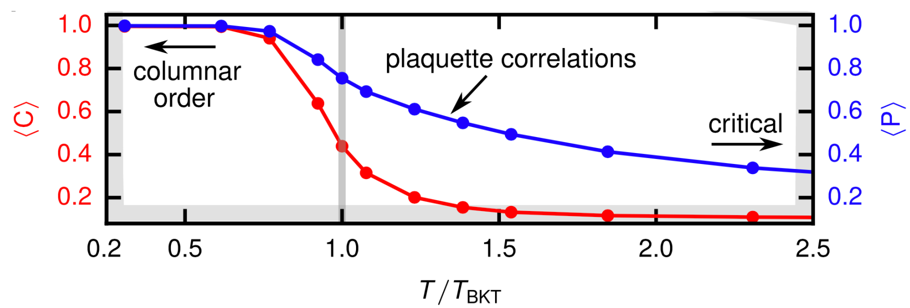
$C(\mathbf{r})$				
Λ_C  Λ_{P1} 	$(-1,-1)$	$(-1,+1)$	$(+1,-1)$	$(+1,+1)$
Λ_{P1}  Λ_{P2} 	$(-1,-1)$	$(+1,+1)$	$(-1,+1)$	$(+1,-1)$

- Filters **define** order parameters:

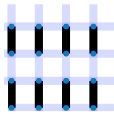
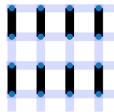
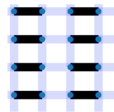
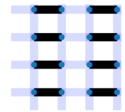
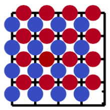
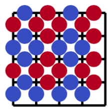
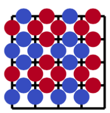
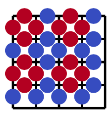
$$D_i := \mathbb{E}\left[\frac{1}{N_{\mathcal{V}}}\sum_k \tau \circ (\Lambda_i \cdot \mathcal{V}_k)\right]$$

$$\mathrm{DSB} := \mathbb{E}\left[\sum_k \tau \circ (\Lambda_C \cdot \mathcal{V}_k)\right]$$

Alet et al. **PRE 74**, 041124 (2006)



- Pairs of C/P filters **label broken symmetry states**

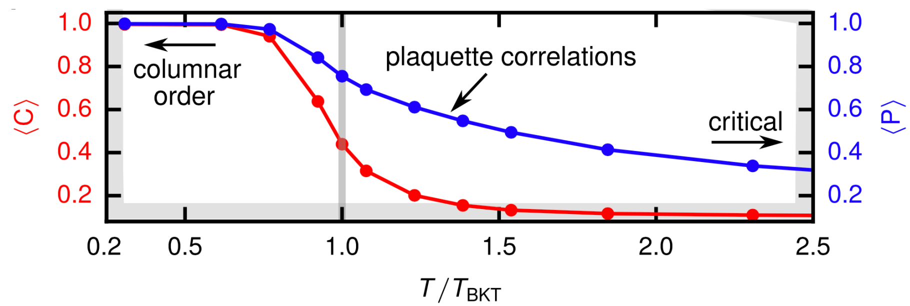
$C(\mathbf{r})$				
Λ_C  Λ_{P1} 	$(-1,-1)$	$(-1,+1)$	$(+1,-1)$	$(+1,+1)$
Λ_{P1}  Λ_{P2} 	$(-1,-1)$	$(+1,+1)$	$(-1,+1)$	$(+1,-1)$

- Filters **define** order parameters:

$$D_i := \mathbb{E}\left[\frac{1}{N_{\mathcal{V}}}\sum_k \tau \circ (\Lambda_i \cdot \mathcal{V}_k)\right]$$

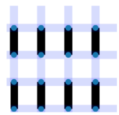
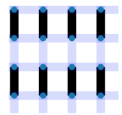
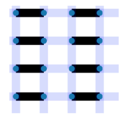
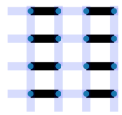
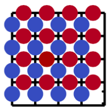
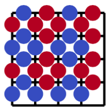
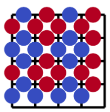
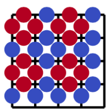
$$\mathrm{DSB} := \mathbb{E}\left[\sum_k \tau \circ (\Lambda_\mathrm{C} \cdot \mathcal{V}_k)\right]$$

Alet et al. **PRE 74**, 041124 (2006)



- Filters **are** relevant operators:

- Pairs of C/P filters **label broken symmetry states**

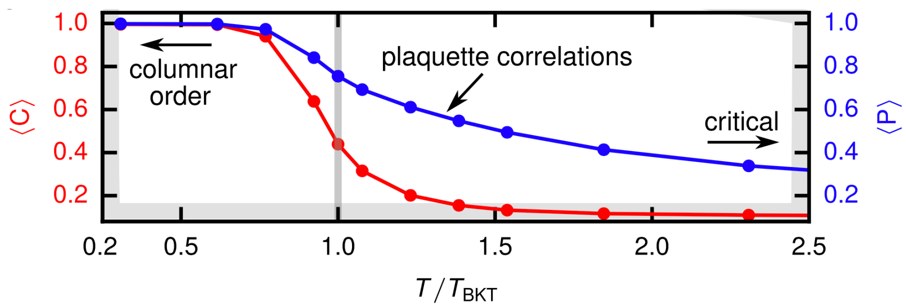
$C(\mathbf{r})$				
Λ_C  Λ_{P1} 	$(-1,-1)$	$(-1,+1)$	$(+1,-1)$	$(+1,+1)$
Λ_{P1}  Λ_{P2} 	$(-1,-1)$	$(+1,+1)$	$(-1,+1)$	$(+1,-1)$

- Filters **define** order parameters:

$$D_i := \mathbb{E}\left[\frac{1}{N_{\mathcal{V}}}\sum_k \tau \circ (\Lambda_i \cdot \mathcal{V}_k)\right]$$

$$\mathrm{DSB} := \mathbb{E}\left[\sum_k \tau \circ (\Lambda_C \cdot \mathcal{V}_k)\right]$$

Alet et al. **PRE 74**, 041124 (2006)

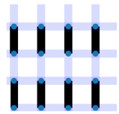
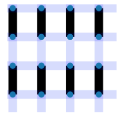
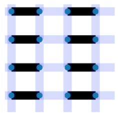
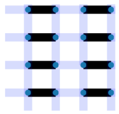
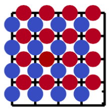
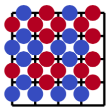
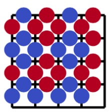
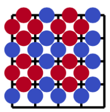


- Filters **are** relevant operators:

$$\mathcal{O}_n(\varphi) = (\cos(n\varphi), \sin(n\varphi))$$

Papanikolaou et al. **PRB 76**, 134514 (2007)

- Pairs of C/P filters **label broken symmetry states**

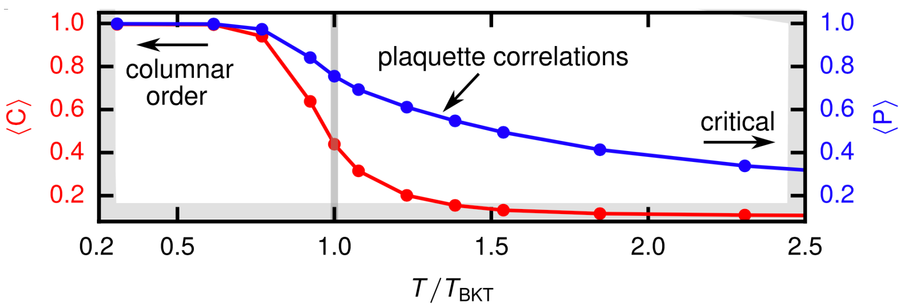
$C(\mathbf{r})$				
Λ_{C}  Λ_{P1} 	$(-1,-1)$	$(-1,+1)$	$(+1,-1)$	$(+1,+1)$
Λ_{P1}  Λ_{P2} 	$(-1,-1)$	$(+1,+1)$	$(-1,+1)$	$(+1,-1)$
$\varphi(\mathbf{r})$	$\frac{\pi}{2}$	$\frac{3\pi}{2}$	π	0
$\mathcal{O}_1(\varphi) = (\cos\varphi, \sin\varphi)$	$(0,1)$	$(0,-1)$	$(-1,0)$	$(+1,0)$
$\mathcal{O}_2(\varphi) = \cos(2\varphi)$	-1	-1	$+1$	$+1$

- Filters **define** order parameters:

$$D_i := \mathbb{E}\left[\frac{1}{N_{\mathcal{V}}}\sum_k \tau \circ (\Lambda_i \cdot \mathcal{V}_k)\right]$$

$$\mathrm{DSB} := \mathbb{E}\left[\sum_k \tau \circ (\Lambda_{\mathrm{C}} \cdot \mathcal{V}_k)\right]$$

Alet et al. **PRE 74**, 041124 (2006)

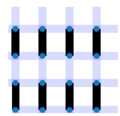
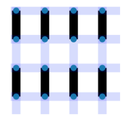
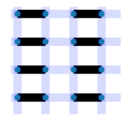
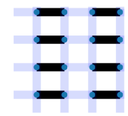
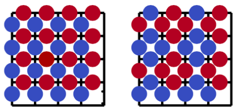
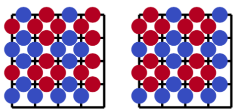


- Filters **are** relevant operators:

$$\mathcal{O}_n(\varphi) = (\cos(n\varphi), \sin(n\varphi))$$

Papanikolaou et al. **PRB 76**, 134514 (2007)

- Pairs of C/P filters **label broken symmetry states**

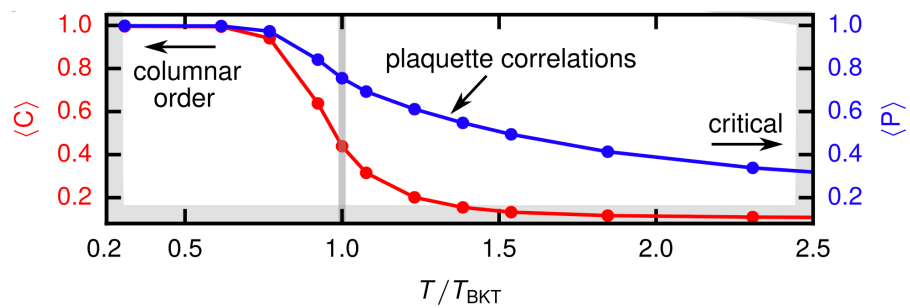
$C(\mathbf{r})$				
Λ_{C} Λ_{P1} 	$(-1,-1)$	$(-1,+1)$	$(+1,-1)$	$(+1,+1)$
Λ_{P1} Λ_{P2} 	$(-1,-1)$	$(+1,+1)$	$(-1,+1)$	$(+1,-1)$
$\varphi(\mathbf{r})$	$\frac{\pi}{2}$	$\frac{3\pi}{2}$	π	0
$\mathcal{O}_1(\varphi) = (\cos\varphi, \sin\varphi)$	$(0,1)$	$(0,-1)$	$(-1,0)$	$(+1,0)$
$\mathcal{O}_2(\varphi) = \cos(2\varphi)$	-1	-1	$+1$	$+1$

- Filters **define** order parameters:

$$D_i := \mathbb{E}\left[\frac{1}{N_{\mathcal{V}}}\sum_k \tau \circ (\Lambda_i \cdot \mathcal{V}_k)\right]$$

$$\mathrm{DSB} := \mathbb{E}\left[\sum_k \tau \circ (\Lambda_{\mathrm{C}} \cdot \mathcal{V}_k)\right]$$

Alet et al. **PRE 74**, 041124 (2006)



- Filters **are** relevant operators:

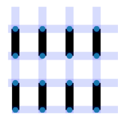
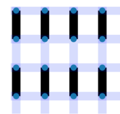
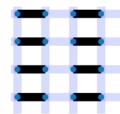
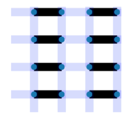
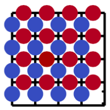
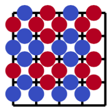
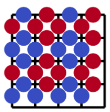
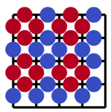
$$\mathcal{O}_n(\varphi) = (\cos(n\varphi), \sin(n\varphi))$$

Papanikolaou et al. **PRB 76**, 134514 (2007)

$$(\Lambda_{\mathrm{P1}}, \Lambda_{\mathrm{P1}}) \circ \varphi = (\cos(\varphi + \pi/4), \sin(\varphi + \pi/4))$$

$$\Lambda_{\mathrm{C}} \circ \varphi = \cos(2\varphi)$$

- Pairs of C/P filters **label broken symmetry states**

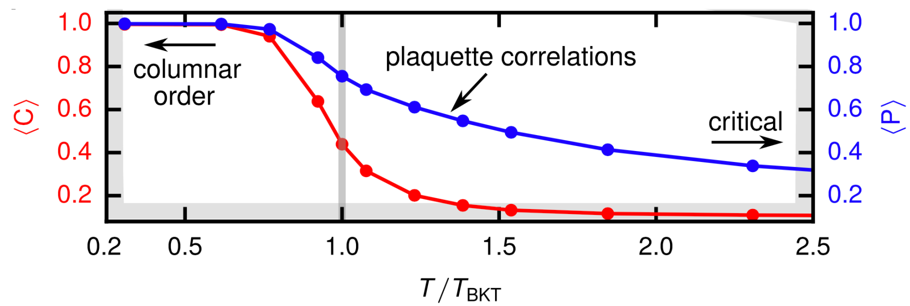
$C(\mathbf{r})$				
Λ_C  Λ_{P1} 	$(-1,-1)$	$(-1,+1)$	$(+1,-1)$	$(+1,+1)$
Λ_{P1}  Λ_{P2} 	$(-1,-1)$	$(+1,+1)$	$(-1,+1)$	$(+1,-1)$
$\varphi(\mathbf{r})$	$\frac{\pi}{2}$	$\frac{3\pi}{2}$	π	0
$\mathcal{O}_1(\varphi) = (\cos\varphi, \sin\varphi)$	$(0,1)$	$(0,-1)$	$(-1,0)$	$(+1,0)$
$\mathcal{O}_2(\varphi) = \cos(2\varphi)$	-1	-1	$+1$	$+1$

- Filters **define** order parameters:

$$D_i := \mathbb{E}\left[\frac{1}{N_{\mathcal{V}}}\sum_k \tau \circ (\Lambda_i \cdot \mathcal{V}_k)\right]$$

$$\mathrm{DSB} := \mathbb{E}\left[\sum_k \tau \circ (\Lambda_C \cdot \mathcal{V}_k)\right]$$

Alet et al. **PRE 74**, 041124 (2006)



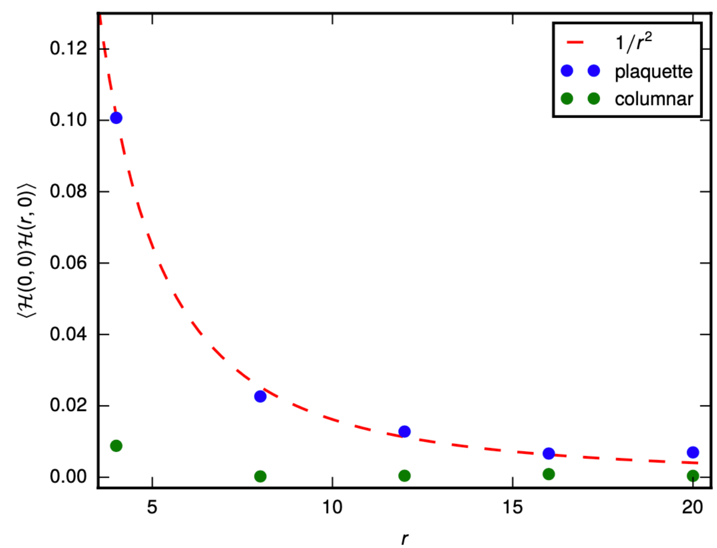
- Filters **are** relevant operators:

$$\mathcal{O}_n(\varphi) = (\cos(n\varphi), \sin(n\varphi))$$

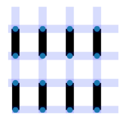
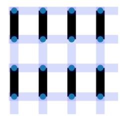
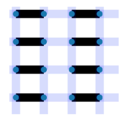
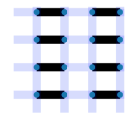
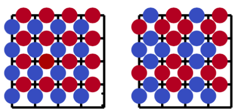
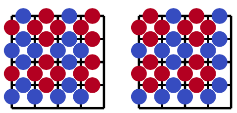
Papanikolaou et al. **PRB 76**, 134514 (2007)

$$(\Lambda_{P1}, \Lambda_{P1}) \circ \varphi = (\cos(\varphi + \pi/4), \sin(\varphi + \pi/4))$$

$$\Lambda_C \circ \varphi = \cos(2\varphi)$$



- Pairs of C/P filters **label broken symmetry states**

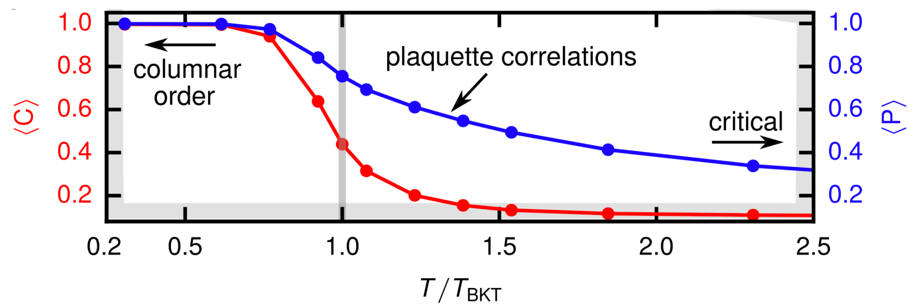
$C(\mathbf{r})$				
Λ_C Λ_{P1} 	$(-1,-1)$	$(-1,+1)$	$(+1,-1)$	$(+1,+1)$
Λ_{P1} Λ_{P2} 	$(-1,-1)$	$(+1,+1)$	$(-1,+1)$	$(+1,-1)$
$\varphi(\mathbf{r})$	$\frac{\pi}{2}$	$\frac{3\pi}{2}$	π	0
$\mathcal{O}_1(\varphi) = (\cos\varphi, \sin\varphi)$	$(0,1)$	$(0,-1)$	$(-1,0)$	$(+1,0)$
$\mathcal{O}_2(\varphi) = \cos(2\varphi)$	-1	-1	$+1$	$+1$

- Filters **define** order parameters:

$$D_i := \mathbb{E}\left[\frac{1}{N_{\mathcal{V}}}\sum_k \tau \circ (\Lambda_i \cdot \mathcal{V}_k)\right]$$

$$\mathrm{DSB} := \mathbb{E}\left[\sum_k \tau \circ (\Lambda_{\mathrm{C}} \cdot \mathcal{V}_k)\right]$$

Alet et al. **PRE 74**, 041124 (2006)



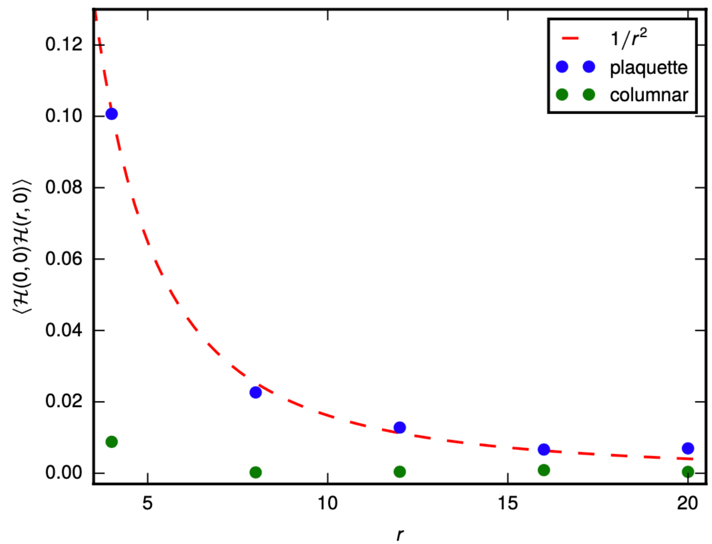
- Filters **are** relevant operators:

$$\mathcal{O}_n(\varphi) = (\cos(n\varphi), \sin(n\varphi))$$

Papanikolaou et al. **PRB 76**, 134514 (2007)

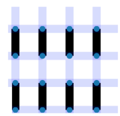
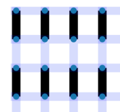
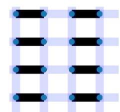
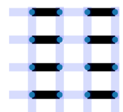
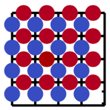
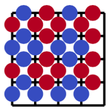
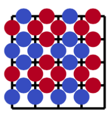
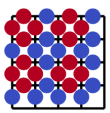
$$(\Lambda_{\mathrm{P1}}, \Lambda_{\mathrm{P1}}) \circ \varphi = (\cos(\varphi + \pi/4), \sin(\varphi + \pi/4))$$

$$\Lambda_{\mathrm{C}} \circ \varphi = \cos(2\varphi)$$



- They can be assigned scaling dimensions

- Pairs of C/P filters **label broken symmetry states**

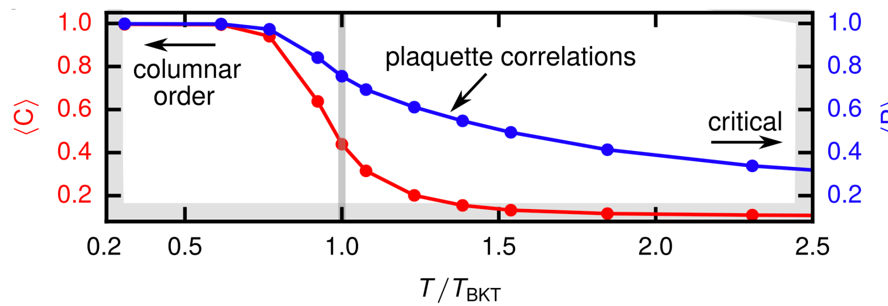
$C(\mathbf{r})$				
Λ_C  Λ_{P1} 	$(-1,-1)$	$(-1,+1)$	$(+1,-1)$	$(+1,+1)$
Λ_{P1}  Λ_{P2} 	$(-1,-1)$	$(+1,+1)$	$(-1,+1)$	$(+1,-1)$
$\varphi(\mathbf{r})$	$\frac{\pi}{2}$	$\frac{3\pi}{2}$	π	0
$\mathcal{O}_1(\varphi) = (\cos\varphi, \sin\varphi)$	$(0,1)$	$(0,-1)$	$(-1,0)$	$(+1,0)$
$\mathcal{O}_2(\varphi) = \cos(2\varphi)$	-1	-1	$+1$	$+1$

- Filters **define** order parameters:

$$D_i := \mathbb{E}\left[\frac{1}{N_{\mathcal{V}}}\sum_k \tau \circ (\Lambda_i \cdot \mathcal{V}_k)\right]$$

$$\mathrm{DSB} := \mathbb{E}\left[\sum_k \tau \circ (\Lambda_{\mathrm{C}} \cdot \mathcal{V}_k)\right]$$

Alet et al. **PRE 74**, 041124 (2006)



- Filters **are** relevant operators:

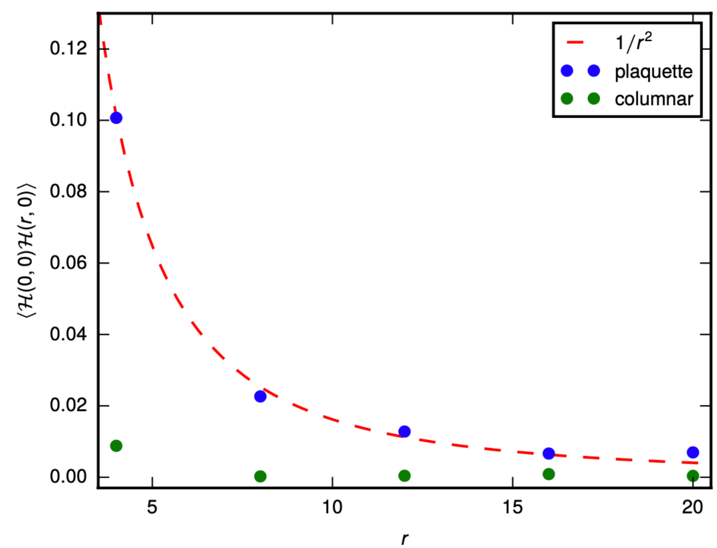
$$\mathcal{O}_n(\varphi) = (\cos(n\varphi), \sin(n\varphi))$$

Papanikolaou et al. **PRB 76**, 134514 (2007)

$$(\Lambda_{\mathrm{P1}}, \Lambda_{\mathrm{P1}}) \circ \varphi = (\cos(\varphi + \pi/4), \sin(\varphi + \pi/4))$$

$$\Lambda_{\mathrm{C}} \circ \varphi = \cos(2\varphi)$$

(Also: staggered filters are gradients of the height field)



- They can be assigned scaling dimensions

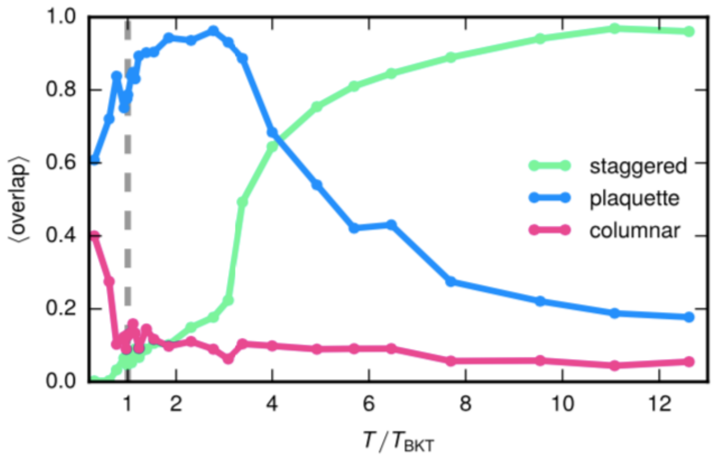
Data analysis of RMSI filter ensemble

Data analysis of RMSI filter ensemble

- In intermediate regimes and finite systems competing correlations yield mixtures of filters

Data analysis of RMSI filter ensemble

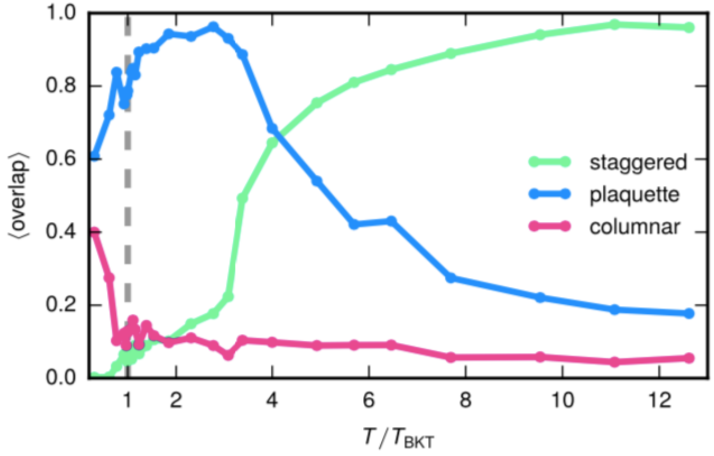
- In intermediate regimes and finite systems competing correlations yield mixtures of filters



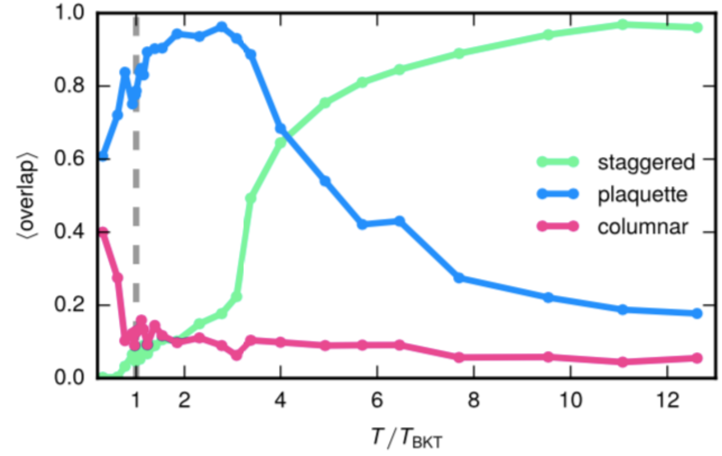
Data analysis of RMSI filter ensemble

- In intermediate regimes and finite systems competing correlations yield mixtures of filters

- They can be decoupled and recovered

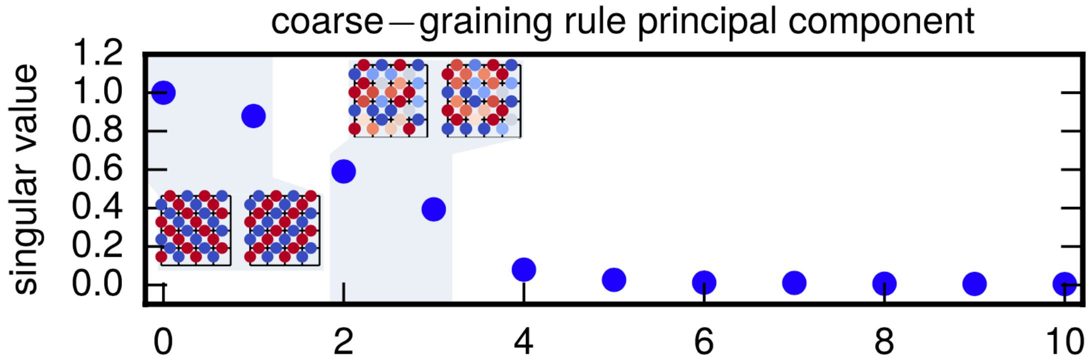


Data analysis of RMSI filter ensemble



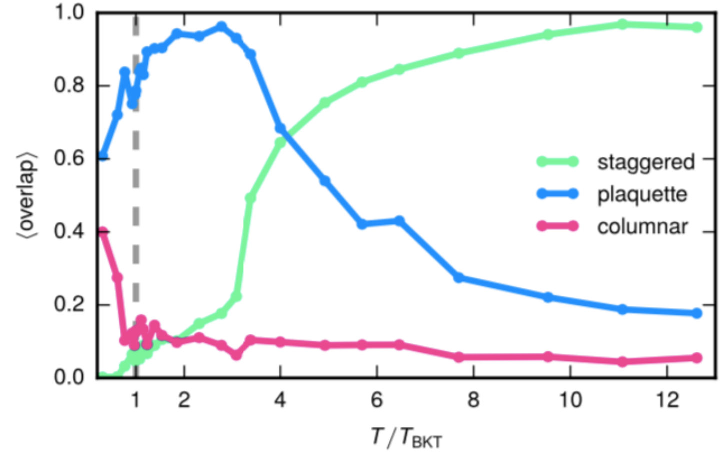
- In intermediate regimes and finite systems competing correlations yield mixtures of filters

- They can be decoupled and recovered

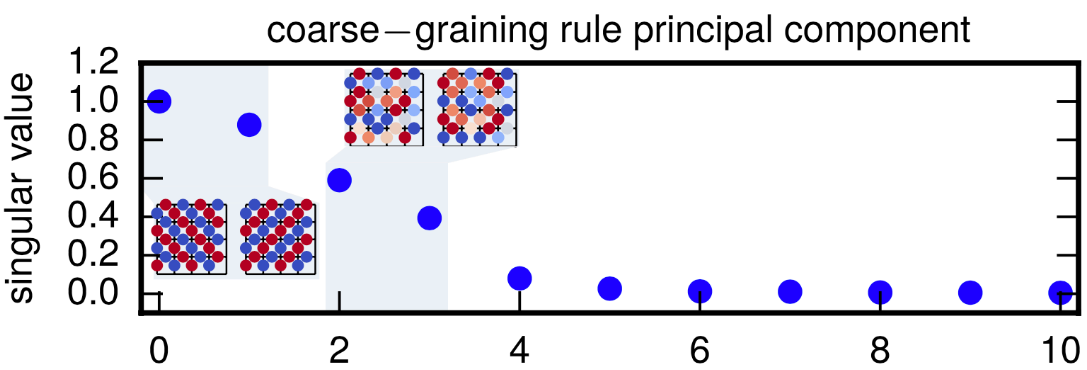


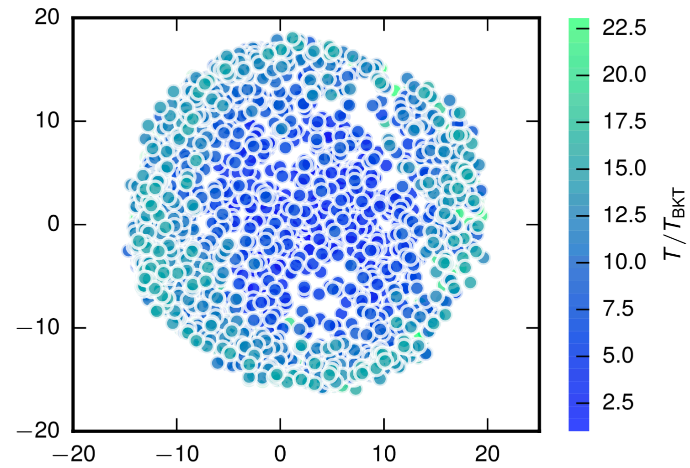
Data analysis of RMSI filter ensemble

- In intermediate regimes and finite systems competing correlations yield mixtures of filters



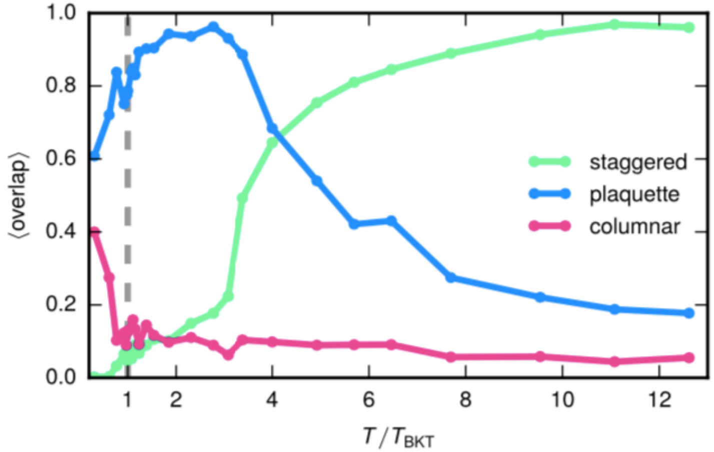
- They can be decoupled and recovered



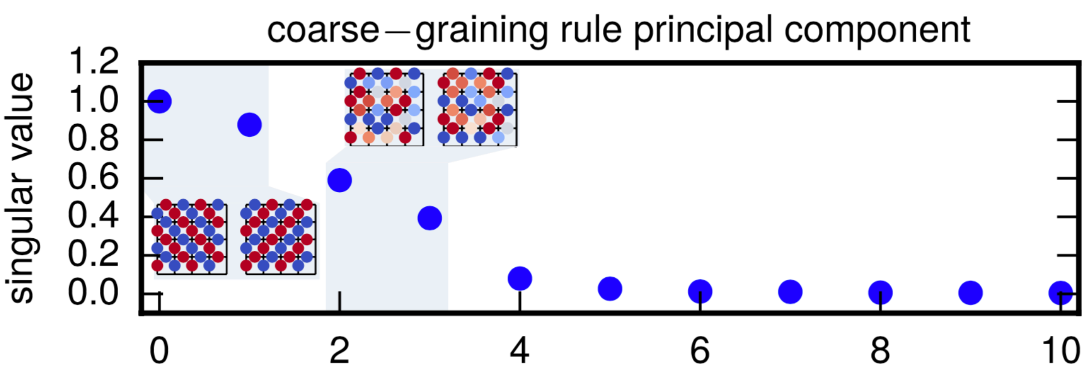


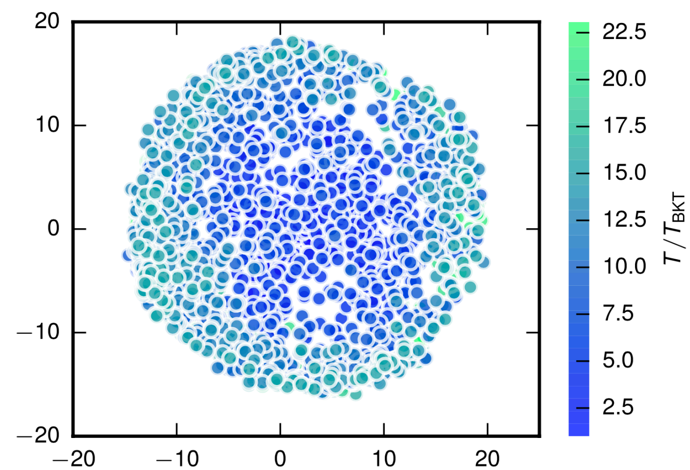
Data analysis of RMSI filter ensemble

- In intermediate regimes and finite systems competing correlations yield mixtures of filters



- They can be decoupled and recovered

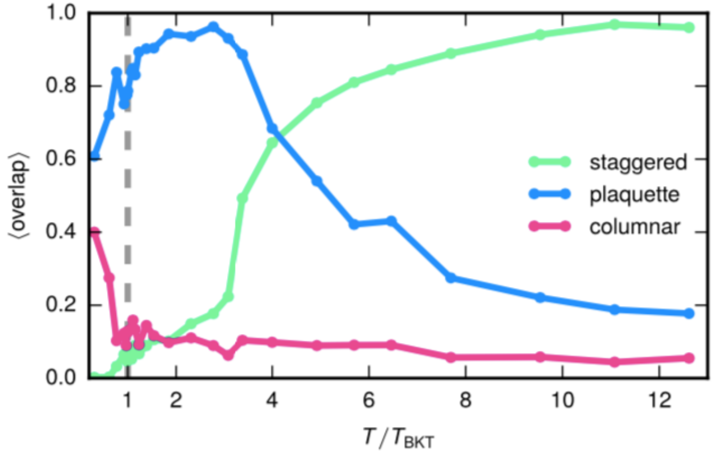




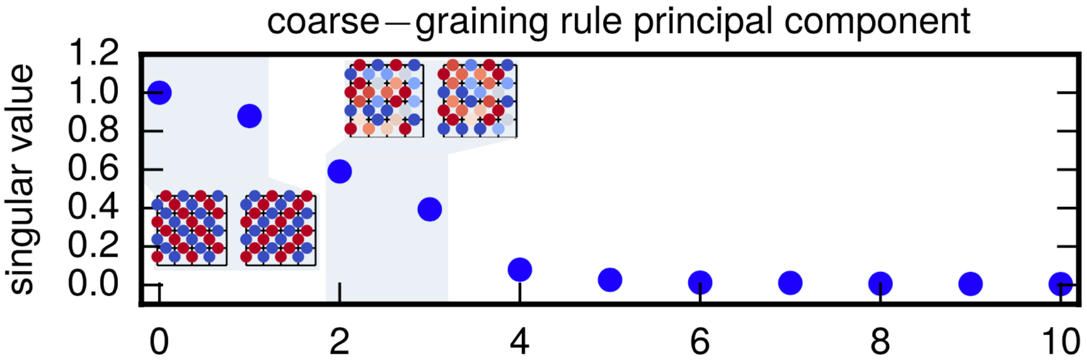
- Works also from partial data

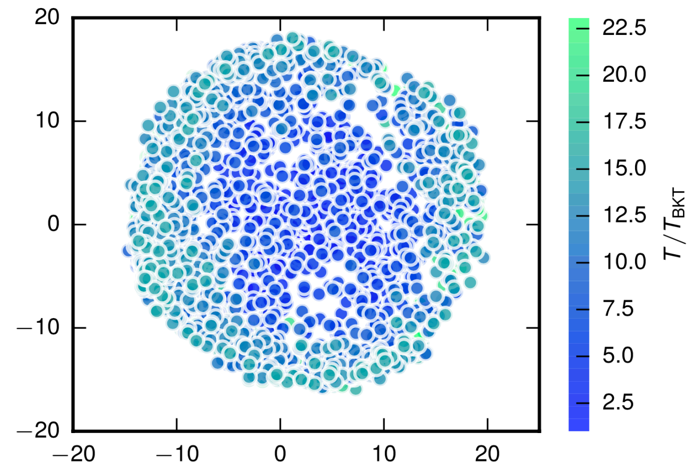
Data analysis of RMSI filter ensemble

- In intermediate regimes and finite systems competing correlations yield mixtures of filters



- They can be decoupled and recovered





- Works also from partial data

- As a function of disorder distribution?

The information bottleneck (IB) compression

The information bottleneck (IB) compression

- “Relevance” defined implicitly, by correlations with a signal variable

The information bottleneck (IB) compression

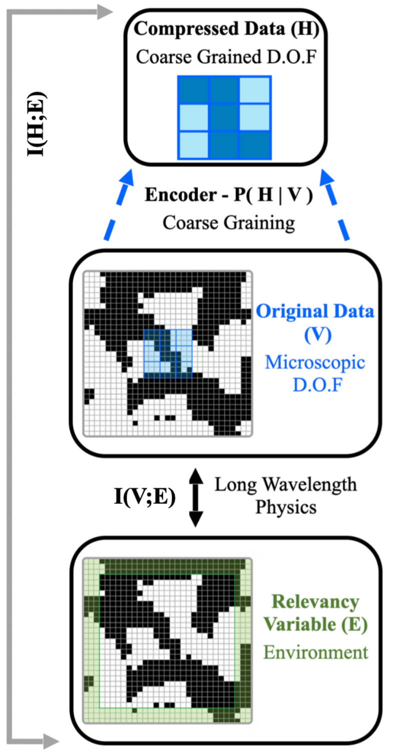
- “Relevance” defined implicitly, by correlations with a signal variable

Tishby, Pereira, Bialek (1999)
Tishby Slonim **NeurIPS** (2000)

The information bottleneck (IB) compression

- “Relevance” defined implicitly, by correlations with a signal variable

Tishby, Pereira, Bialek (1999)
Tishby Slonim **NeurIPS** (2000)

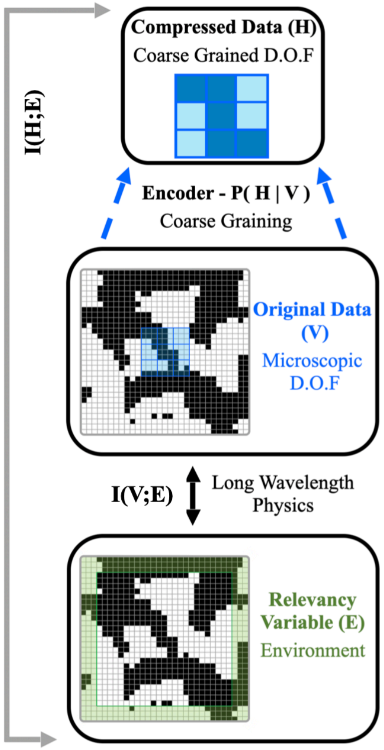


The information bottleneck (IB) compression

- “Relevance” defined implicitly, by correlations with a signal variable
- Optimal compression of relevant information is a variational problem

Tishby, Pereira, Bialek (1999)
Tishby Slonim **NeurIPS** (2000)

$$\min_{P(H|V)} \mathcal{L}_{IB}[P(H|V)] \equiv \min_{P(H|V)} I(V;H) - \beta I(H;E)$$



The information bottleneck (IB) compression

- "Relevance" defined implicitly, by correlations with a signal variable
- Optimal compression of relevant information is a variational problem

Tishby, Pereira, Bialek (1999)
Tishby Slonim **NeurIPS** (2000)

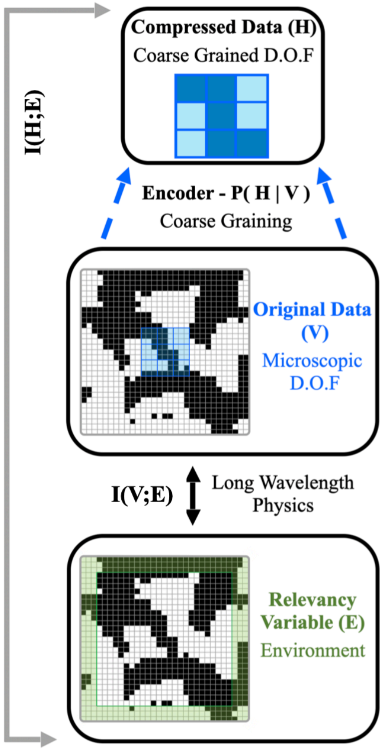
$$\min_{P(H|V)} \mathcal{L}_{IB}[P(H|V)] \equiv \min_{P(H|V)} I(V;H) - \beta I(H;E)$$

- IB equations:

$$P(h|v) = \frac{P(h)}{Z}\exp\left(-\beta\sum_e P(e|v)\log\left[\frac{P(e|v)}{P(e|h)}\right]\right)$$

$$P(h) = \sum_v P(h|v)P(v)$$

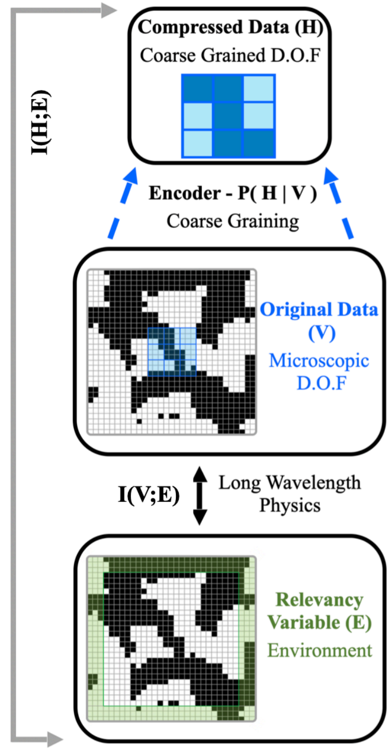
$$P(e|h) = \sum_v P(e|v)P(v|h)$$



The information bottleneck (IB) compression

- "Relevance" defined implicitly, by correlations with a signal variable
- Optimal compression of relevant information is a variational problem

Tishby, Pereira, Bialek (1999)
Tishby Slonim **NeurIPS** (2000)



$$\min_{P(H|V)} \mathcal{L}_{IB}[P(H|V)] \equiv \min_{P(H|V)} I(V;H) - \beta I(H;E)$$

- IB equations:

$$P(h|v) = \frac{P(h)}{Z} \exp\left(-\beta \sum_e P(e|v) \log\left[\frac{P(e|v)}{P(e|h)}\right]\right)$$

$$P(h) = \sum_v P(h|v) P(v)$$

$$P(e|h) = \sum_v P(e|v) P(v|h)$$

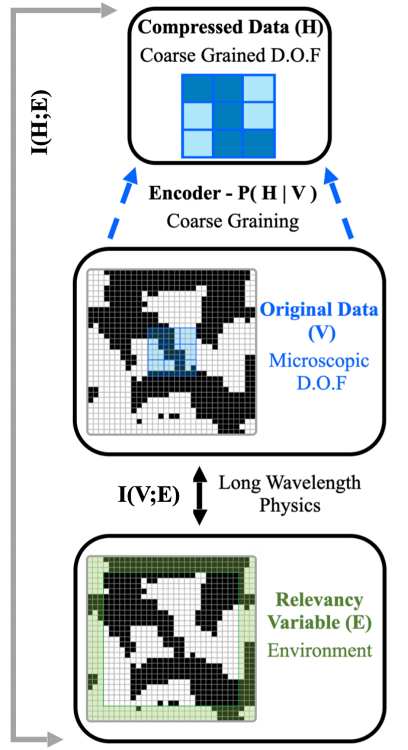
- Optimal IB encoder goes through a sequence of permutation **symmetry breaking** transitions

Gedeon et al. **Entropy (2012)**, 14(3) 456-479

The information bottleneck (IB) compression

- "Relevance" defined implicitly, by correlations with a signal variable
- Optimal compression of relevant information is a variational problem

Tishby, Pereira, Bialek (1999)
Tishby Slonim **NeurIPS** (2000)



$$\min_{P(H|V)} \mathcal{L}_{IB}[P(H|V)] \equiv \min_{P(H|V)} I(V;H) - \beta I(H;E)$$

- IB equations:

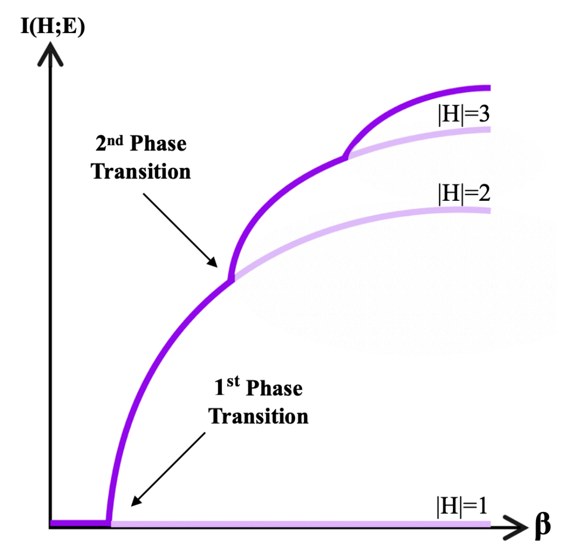
$$P(h|v) = \frac{P(h)}{Z} \exp\left(-\beta \sum_e P(e|v) \log\left[\frac{P(e|v)}{P(e|h)}\right]\right)$$

$$P(h) = \sum_v P(h|v)P(v)$$

$$P(e|h) = \sum_v P(e|v)P(v|h)$$

- Optimal IB encoder goes through a sequence of permutation **symmetry breaking** transitions

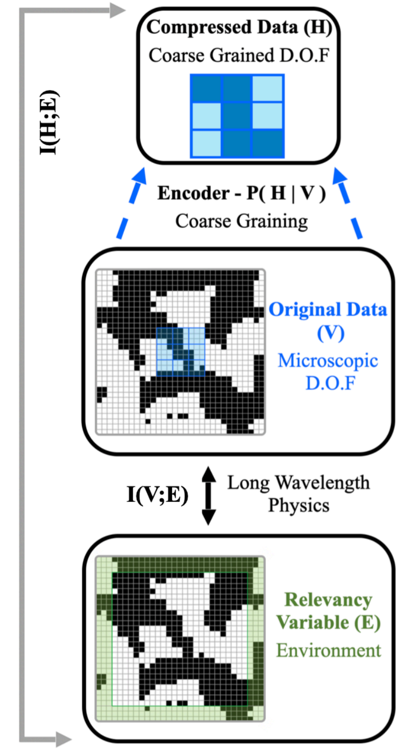
Gedeon et al. **Entropy (2012)**, 14(3) 456-479



The information bottleneck (IB) compression

- "Relevance" defined implicitly, by correlations with a signal variable
- Optimal compression of relevant information is a variational problem

Tishby, Pereira, Bialek (1999)
Tishby Slonim **NeurIPS** (2000)



$$\min_{P(H|V)} \mathcal{L}_{IB}[P(H|V)] \equiv \min_{P(H|V)} I(V;H) - \beta I(H;E)$$

- IB equations:

$$P(h|v) = \frac{P(h)}{Z} \exp\left(-\beta \sum_e P(e|v) \log\left[\frac{P(e|v)}{P(e|h)}\right]\right)$$

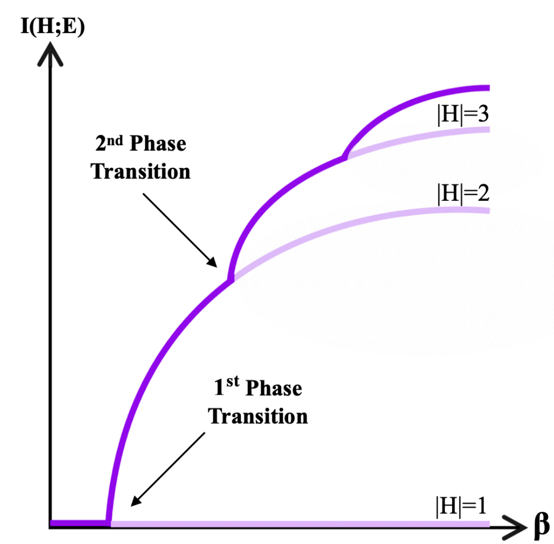
$$P(h) = \sum_v P(h|v) P(v)$$

$$P(e|h) = \sum_v P(e|v) P(v|h)$$

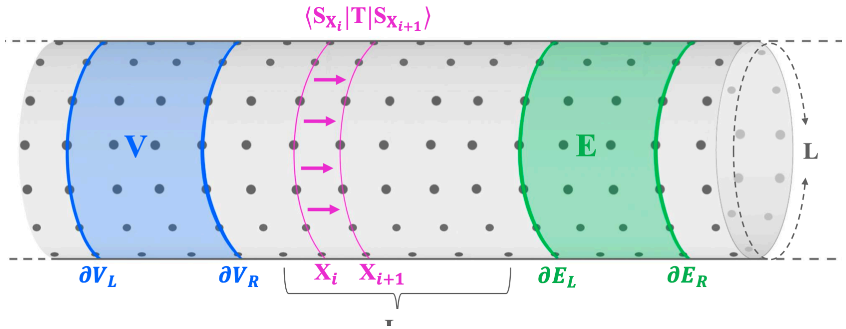
- Optimal IB encoder goes through a sequence of permutation **symmetry breaking** transitions

Gedeon et al. **Entropy (2012)**, 14(3) 456-479

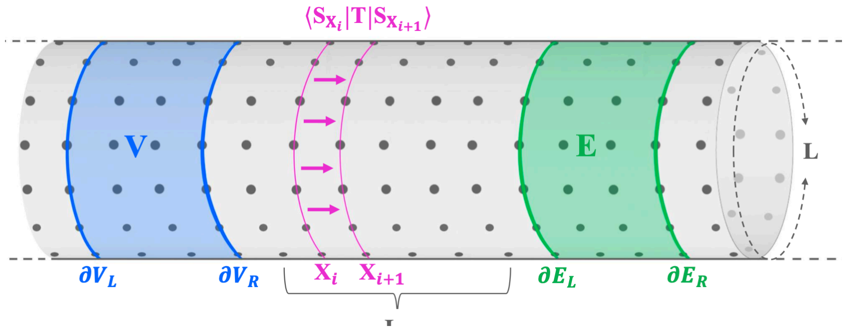
- RSMI arises in the infinite β limit, and finite alphabet



IB and the transfer matrix



IB and the transfer matrix

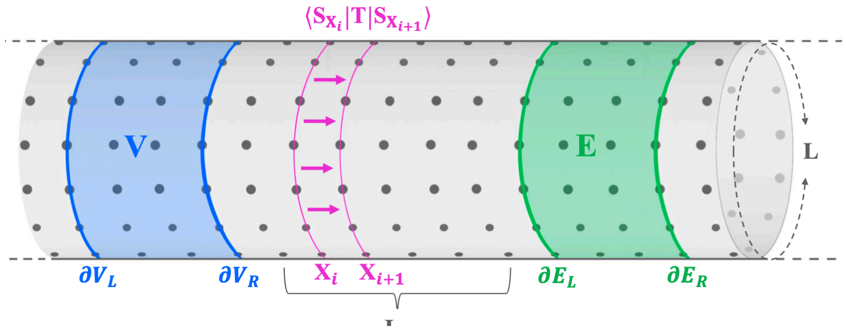


- We want to solve the IB eqs. for encoder

$$P(h|x) \propto P(h)e^{\beta \sum_e P(e|x) \log(P(e|h))}$$

$$P(e|h) = \sum_x p(e|x)p(x|h)$$

IB and the transfer matrix



- We want to solve the IB eqs. for encoder

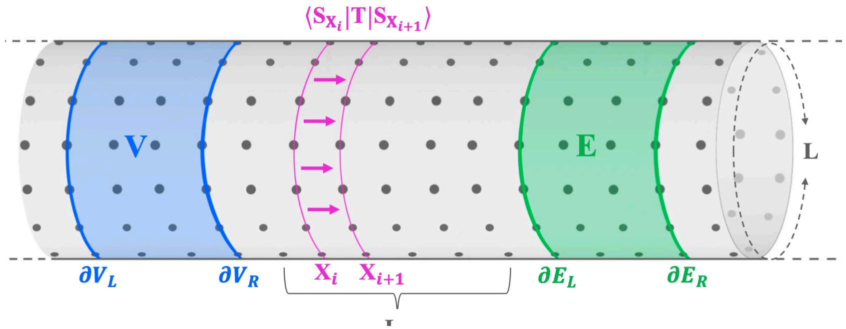
$$P(h|x) \propto P(h)e^{\beta \sum_e P(e|x)\log(P(e|h))}$$

$$P(e|h) = \sum_x p(e|x)p(x|h)$$

- Distributions entering IB equations can be expressed using transfer matrices
- Eigenvectors/eigenvalues of transfer matrices have direct relation to CFT operator content

Cardy **J. Phys. A: Math. Gen. 17,** L385 (1984)
Bloete et al. **Phys. Rev.Lett.** 56, 742 (1986)

IB and the transfer matrix



- We want to solve the IB eqs. for encoder

$$P(h|x) \propto P(h) e^{\beta \sum_e P(e|x) \log(P(e|h))}$$

$$P(e|h) = \sum_x p(e|x) p(x|h)$$

- Distributions entering IB equations can be expressed using transfer matrices
- Eigenvectors/eigenvalues of transfer matrices have direct relation to CFT operator content

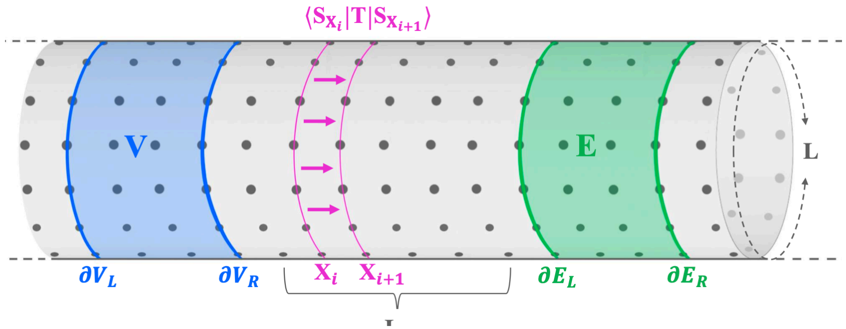
Cardy **J. Phys. A: Math. Gen. 17,** L385 (1984)
Bloete et al. **Phys. Rev.Lett.** 56, 742 (1986)

- All quantities of interest are functions of matrix elements of relevant operators:

$$P(v|e) = N^{-1} P(v) \left[1 + \frac{\langle \partial V_R | 1 \rangle \langle 1 | \partial E_L \rangle}{\langle \partial V_R | 0 \rangle \langle 0 | \partial E_L \rangle} \left(\frac{\lambda_1}{\lambda_0} \right)^{L_B} \right]$$

$$= N^{-1} P(v) \left[1 + \epsilon r_e r_v \right]$$

IB and the transfer matrix



- Distributions entering IB equations can be expressed using transfer matrices
- Eigenvectors/eigenvalues of transfer matrices have direct relation to CFT operator content

Cardy **J. Phys. A: Math. Gen. 17,** L385 (1984)
Bloete et al. **Phys. Rev.Lett.** 56, 742 (1986)

- We want to solve the IB eqs. for encoder

$$P(h|x) \propto P(h) e^{\beta \sum_e P(e|x) \log(P(e|h))}$$

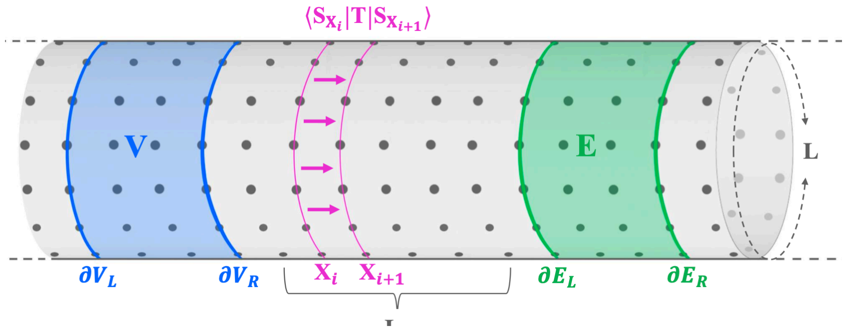
$$P(e|h) = \sum_x p(e|x) p(x|h)$$

- All quantities of interest are functions of matrix elements of relevant operators:

$$P(v|e) = N^{-1} P(v) \left[1 + \frac{\langle \partial V_R | 1 \rangle \langle 1 | \partial E_L \rangle}{\langle \partial V_R | 0 \rangle \langle 0 | \partial E_L \rangle} \left(\frac{\lambda_1}{\lambda_0} \right)^{L_B} \right]$$
$$= N^{-1} P(v) \left[1 + \epsilon r_e r_v \right]$$

$$r_v = \frac{\langle 1 | \partial V_R \rangle}{\langle 0 | \partial V_R \rangle} = \frac{\langle 0 | \phi_{\Delta_1} | \partial V_R \rangle}{\langle 0 | \partial V_R \rangle}$$
$$r_e = \frac{\langle \partial E_L | 1 \rangle}{\langle \partial E_L | 0 \rangle} = \frac{\langle \partial E_L | \phi_{\Delta_1} | 0 \rangle}{\langle \partial E_L | 0 \rangle}$$

IB and the transfer matrix



- Distributions entering IB equations can be expressed using transfer matrices
- Eigenvectors/eigenvalues of transfer matrices have direct relation to CFT operator content

Cardy **J. Phys. A: Math. Gen. 17,** L385 (1984)
Bloete et al. **Phys. Rev.Lett.** 56, 742 (1986)

- We want to solve the IB eqs. for encoder

$$P(h|x) \propto P(h)e^{\beta \sum_e P(e|x)\log(P(e|h))}$$
$$P(e|h) = \sum_x p(e|x)p(x|h)$$

- All quantities of interest are functions of matrix elements of relevant operators:

$$P(v|e) = N^{-1}P(v)\left[1 + \frac{\langle\partial V_R|1\rangle\langle 1|\partial E_L\rangle}{\langle\partial V_R|0\rangle\langle 0|\partial E_L\rangle}\left(\frac{\lambda_1}{\lambda_0}\right)^{L_B}\right]$$
$$= N^{-1}P(v)\left[1 + \epsilon r_e r_v\right]$$

$$r_v = \frac{\langle 1|\partial V_R\rangle}{\langle 0|\partial V_R\rangle} = \frac{\langle 0|\phi_{\Delta_1}|\partial V_R\rangle}{\langle 0|\partial V_R\rangle}$$
$$r_e = \frac{\langle\partial E_L|1\rangle}{\langle\partial E_L|0\rangle} = \frac{\langle\partial E_L|\phi_{\Delta_1}|0\rangle}{\langle\partial E_L|0\rangle}$$

- The encoder (needs to be solved self-consistently):

$$P(h|v) = P(h|r_v) \propto P(h)e^{\beta\epsilon^2 r_v \langle r_v\rangle_h}$$

- Expand around first transition $\beta = \beta_{c,1} + t$:

- Expand around first transition $\beta = \beta_{c,1} + t$:

$$P(h|r_v) = \frac{1}{|H|} + t b_{r_v}(h)$$

$$\sum_h b_{r_v}(h) = 0,$$

- Expand around first transition $\beta = \beta_{c,1} + t$:

$$P(h|r_v) = \frac{1}{|H|} + t b_{r_v}(h)$$

Differentiation from trivial encoder

$$\sum_h b_{r_v}(h) = 0,$$

- Expand around first transition $\beta = \beta_{c,1} + t$:

$$P(h|r_v) = \frac{1}{|H|} + t b_{r_v}(h)$$

Differentiation from trivial encoder

$$\sum_h b_{r_v}(h) = 0,$$

- Trivial solution exists always, but a nontrivial one appears when: $\beta_{c,1}^{-1} = \epsilon^2 + o(\epsilon^2)$

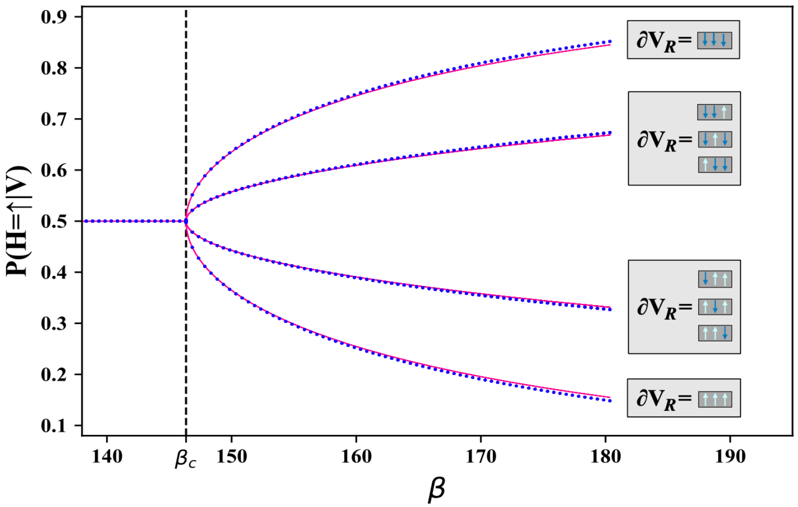
- Expand around first transition $\beta = \beta_{c,1} + t$:

$$P(h|r_v) = \frac{1}{|H|} + t b_{r_v}(h)$$

$$\sum_h b_{r_v}(h) = 0,$$

Differentiation from trivial encoder

- Trivial solution exists always, but a nontrivial one appears when: $\beta_{c,1}^{-1} = \epsilon^2 + o(\epsilon^2)$



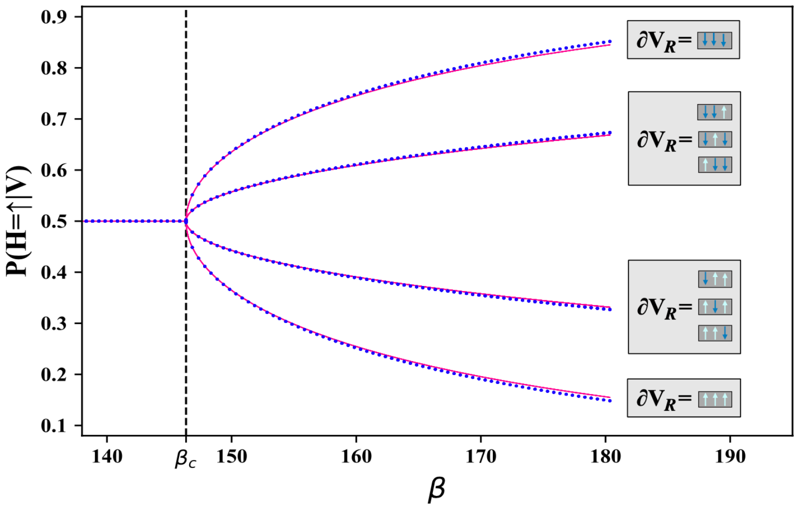
- Expand around first transition $\beta = \beta_{c,1} + t$:

Differentiation from trivial encoder

$$P(h|r_v) = \frac{1}{|H|} + t b_{r_v}(h)$$

$$\sum_h b_{r_v}(h) = 0,$$

- Trivial solution exists always, but a nontrivial one appears when: $\beta_{c,1}^{-1} = \epsilon^2 + o(\epsilon^2)$



Symmetries in IB and RG

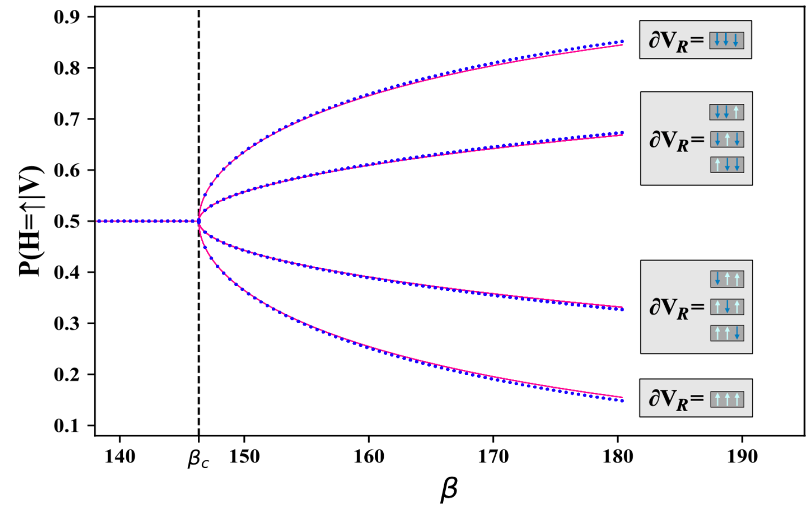
- Expand around first transition $\beta = \beta_{c,1} + t$:

$$P(h|r_v) = \frac{1}{|H|} + t b_{r_v}(h)$$

Differentiation from trivial encoder

$$\sum_h b_{r_v}(h) = 0,$$

- Trivial solution exists always, but a nontrivial one appears when: $\beta_{c,1}^{-1} = \epsilon^2 + o(\epsilon^2)$



Symmetries in IB and RG

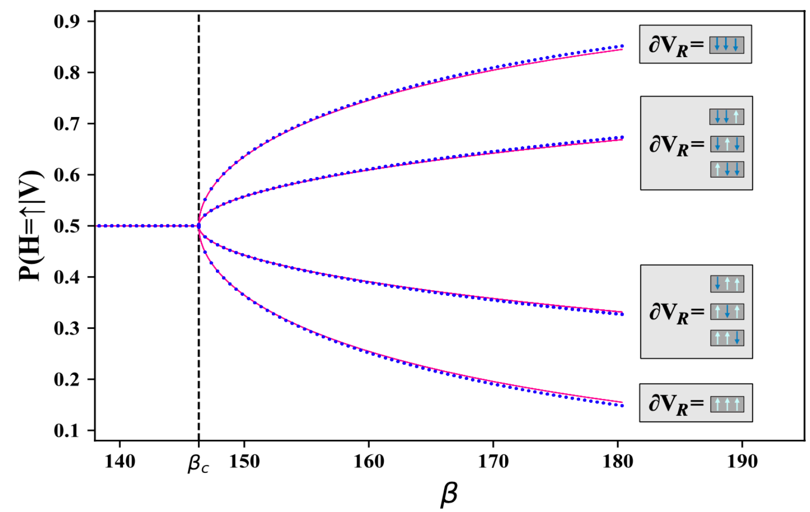
- Physical symmetries of the data: $P(e, v) = P(se, sv)$

- Expand around first transition $\beta = \beta_{c,1} + t$: $$P(h|r_v) = \frac{1}{|H|} + t b_{r_v}(h)$$

Differentiation from trivial encoder

$$\sum_h b_{r_v}(h) = 0,$$

- Trivial solution exists always, but a nontrivial one appears when: $\beta_{c,1}^{-1} = \epsilon^2 + o(\epsilon^2)$



Symmetries in IB and RG

- Physical symmetries of the data: $P(e, v) = P(se, sv)$
- We expect the optimal encoder to maintain it: $P_\beta(h|v) = P_\beta(\phi_s h|sv)$

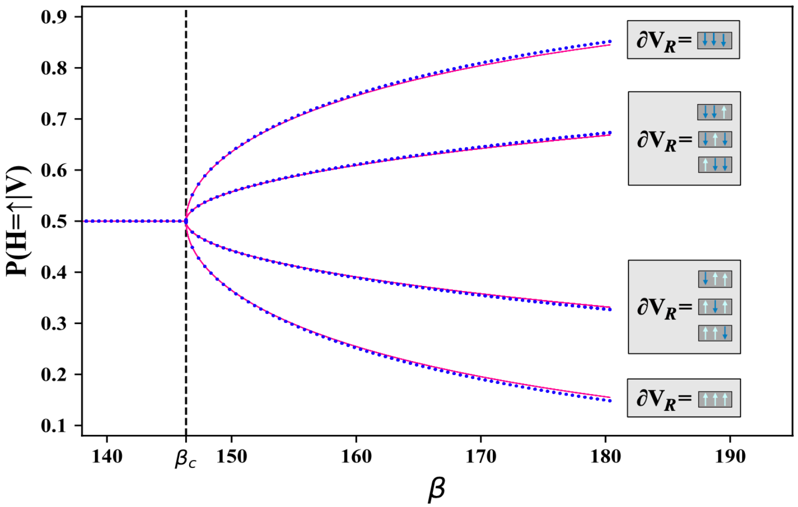
- Expand around first transition $\beta = \beta_{c,1} + t$:

$$P(h|r_v) = \frac{1}{|H|} + t b_{r_v}(h)$$

Differentiation from trivial encoder

$$\sum_h b_{r_v}(h) = 0,$$

- Trivial solution exists always, but a nontrivial one appears when: $\beta_{c,1}^{-1} = \epsilon^2 + o(\epsilon^2)$



Symmetries in IB and RG

- Physical symmetries of the data: $P(e,v) = P(se,sv)$
- We expect the optimal encoder to maintain it: $P_\beta(h|v) = P_\beta(\phi_s h|sv)$
  - Below first transition the encoder is trivially symmetric
  - We can explicitly construct an optimal symmetric encoder after first transition

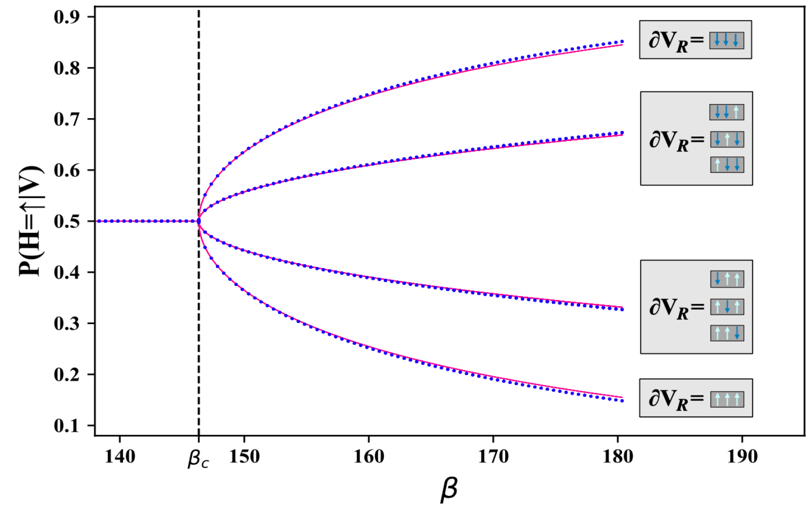
- Expand around first transition $\beta = \beta_{c,1} + t$:

Differentiation from trivial encoder

$$P(h|r_v) = \frac{1}{|H|} + t b_{r_v}(h)$$

$$\sum_h b_{r_v}(h) = 0,$$

- Trivial solution exists always, but a nontrivial one appears when: $\beta_{c,1}^{-1} = \epsilon^2 + o(\epsilon^2)$



Symmetries in IB and RG

- Physical symmetries of the data: $P(e, v) = P(se, sv)$
- We expect the optimal encoder to maintain it: $P_\beta(h|v) = P_\beta(\phi_s h|sv)$
  - Below first transition the encoder is trivially symmetric
  - We can explicitly construct an optimal symmetric encoder after first transition
- Can potentially be used to identify symmetries.

Optimality of the RSMI approach

Optimality of the RSMI approach

- The upper bound on compressed information

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

Optimality of the RSMI approach

- The upper bound on compressed information

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

- Assume first perfect RSMI capture, then:

Optimality of the RSMI approach

- The upper bound on compressed information

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

- Assume first perfect RSMI capture, then:

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

Optimality of the RSMI approach

- The upper bound on compressed information

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

- Assume first perfect RSMI capture, then:

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

→ The range of the Hamiltonian does not increase!

Optimality of the RSMI approach

- The upper bound on compressed information

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

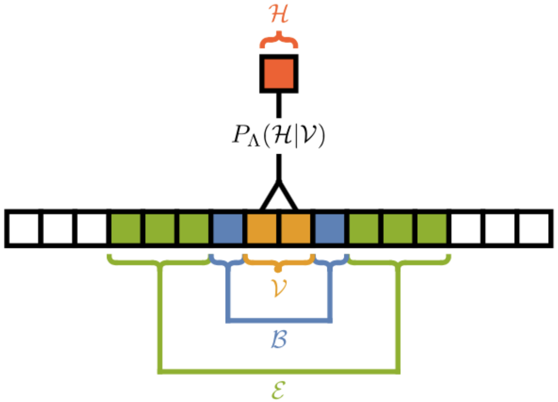
- Assume first perfect RSMI capture, then:

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

→ The range of the Hamiltonian does not increase!

- This assumption is rarely true. But the factoring property holds also in practice:

Optimality of the RSMI approach



- The upper bound on compressed information

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

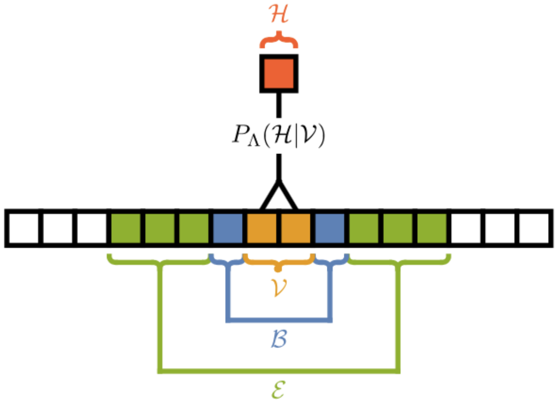
- Assume first perfect RSMI capture, then:

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

The range of the Hamiltonian does not increase!

- This assumption is rarely true. But the factoring property holds also in practice:

Optimality of the RSMI approach



- The upper bound on compressed information

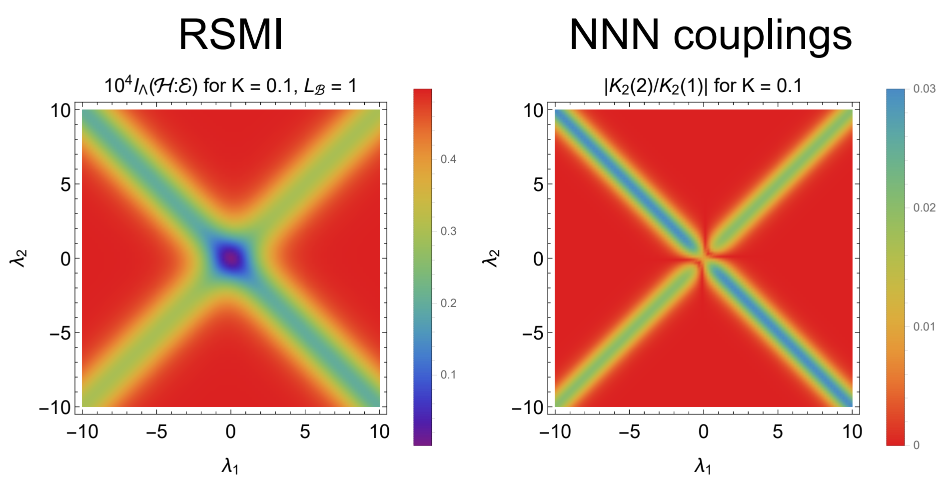
$$I_\Lambda(\mathcal{H}:\mathcal{E}) \leq I(\mathcal{V}:\mathcal{E})$$

- Assume first perfect RSMI capture, then:

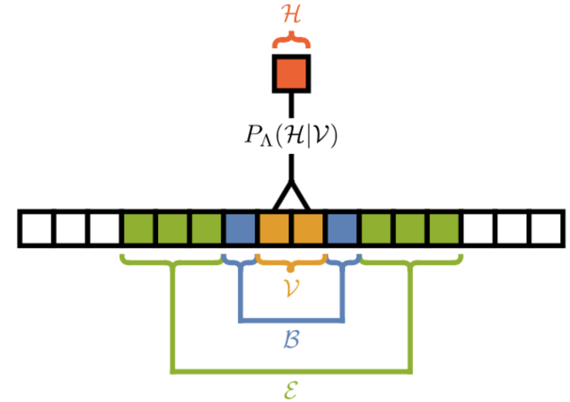
$$P(\mathcal{H}_{j\leq 2}, \mathcal{H}_{j\geq 2}|\mathcal{H}_0) = P(\mathcal{H}_{j\leq -2}|\mathcal{H}_0) \cdot P(\mathcal{H}_{j\geq 2}|\mathcal{H}_0)$$

→ The range of the Hamiltonian does not increase!

- This assumption is rarely true. But the factoring property holds also in practice:



Optimality of the RSMI approach



- The upper bound on compressed information

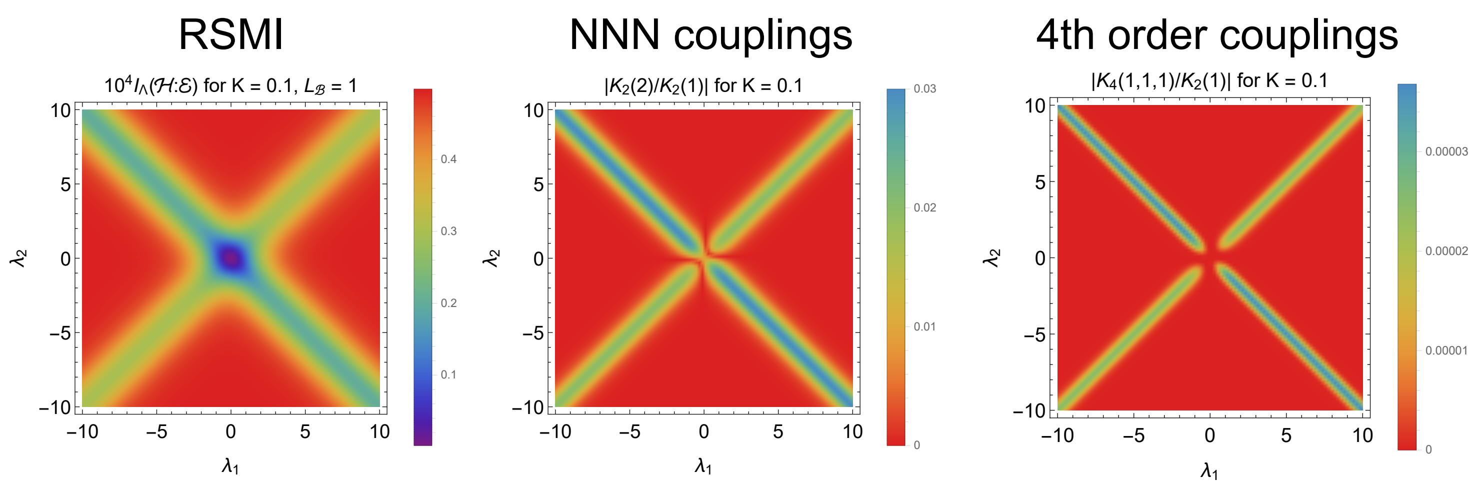
$$I_\Lambda(\mathcal{H} : \mathcal{E}) \leq I(\mathcal{V} : \mathcal{E})$$

- Assume first perfect RSMI capture, then:

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

→ The range of the Hamiltonian does not increase!

- This assumption is rarely true. But the factoring property holds also in practice:



Disorder distributions

Disorder distributions

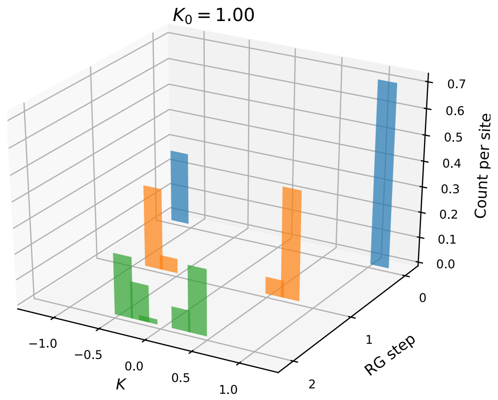
- RG acts on distributions of couplings:

Disorder distributions

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

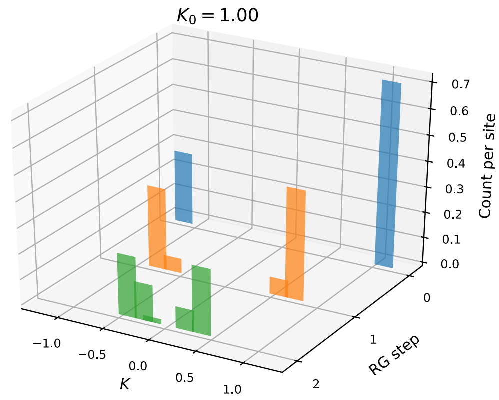
Disorder distributions

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

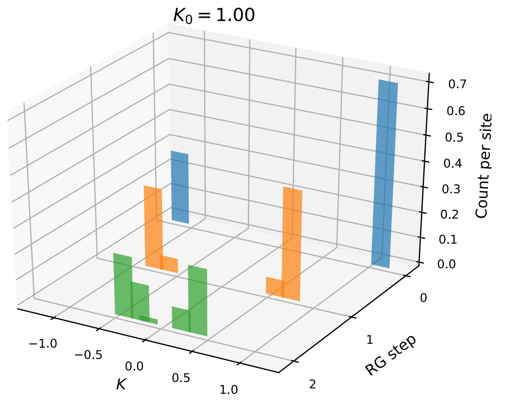


Disorder distributions

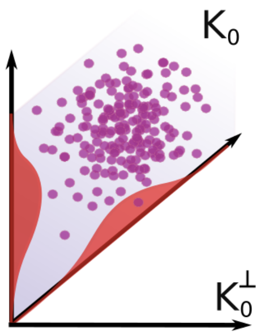
- RG acts on distributions of couplings: $\mathcal{P}(\mathbf{K}) \rightarrow \mathcal{P}'(\mathbf{K})$
- Under perfect RSMI capture factorability of disorder distribution is maintained



Disorder distributions

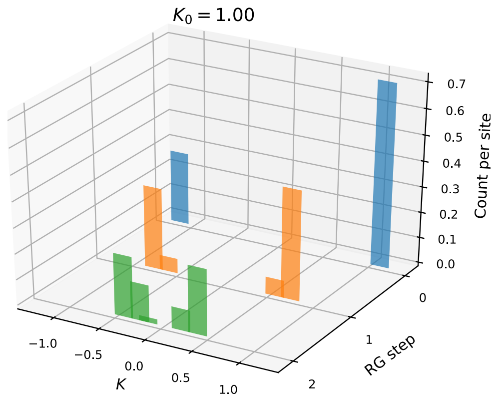


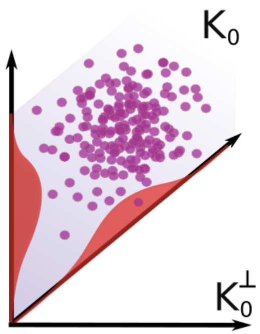
- RG acts on distributions of couplings: $\mathcal{P}(\mathbf{K}) \to \mathcal{P}'(\mathbf{K})$
- Under perfect RSMI capture factorability of disorder distribution is maintained



Disorder distributions

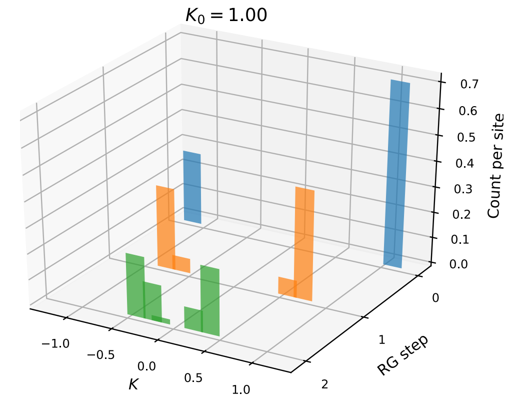
- RG acts on distributions of couplings: $\mathcal{P}(\mathbf{K}) \to \mathcal{P}'(\mathbf{K})$
- Under perfect RSMI capture factorability of disorder distribution is maintained



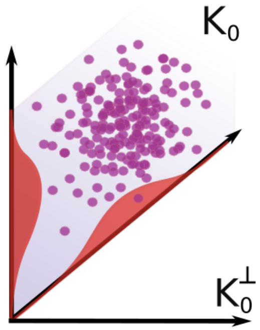


$$\mathcal{P}(\mathbf{K}) \to \mathcal{P}'(\mathbf{K})$$

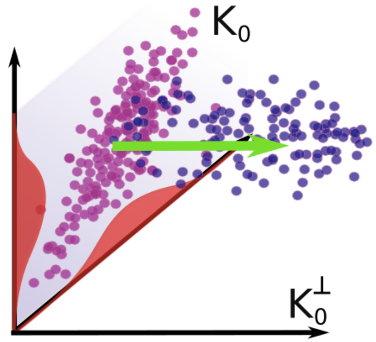
Disorder distributions



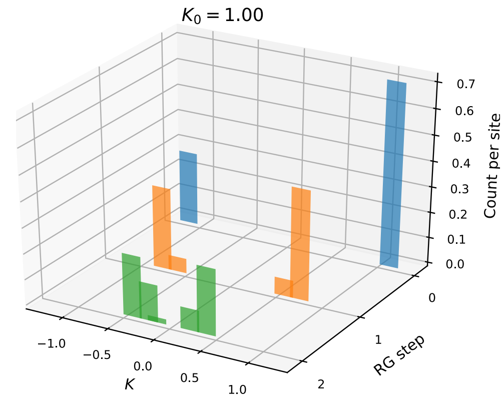
- RG acts on distributions of couplings: $\mathcal{P}(\mathbf{K}) \to \mathcal{P}'(\mathbf{K})$
- Under perfect RSMI capture factorability of disorder distribution is maintained



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

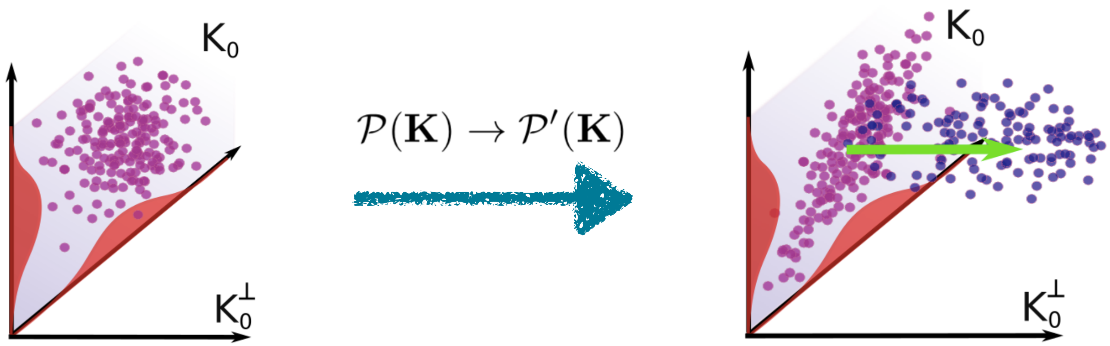


Disorder distributions



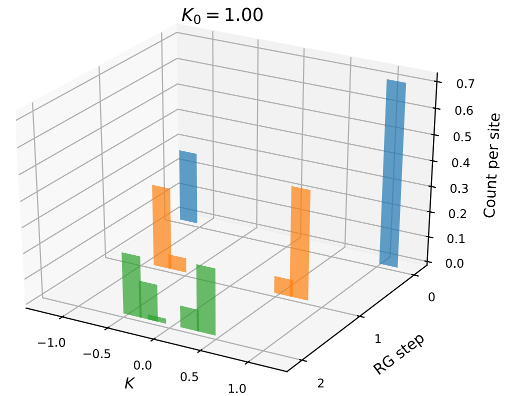
- RG acts on distributions of couplings: $\mathcal{P}(\mathbf{K}) \to \mathcal{P}'(\mathbf{K})$

- Under perfect RSMI capture factorability of disorder distribution is maintained

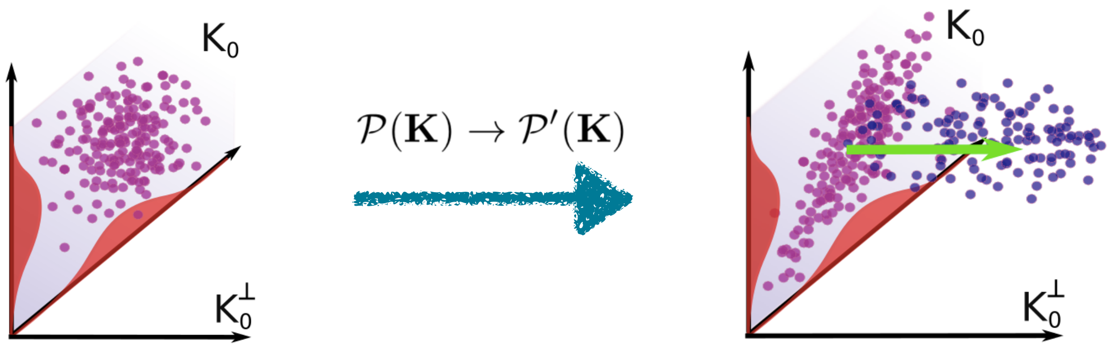


- Example: disordered 1D Ising model

Disorder distributions



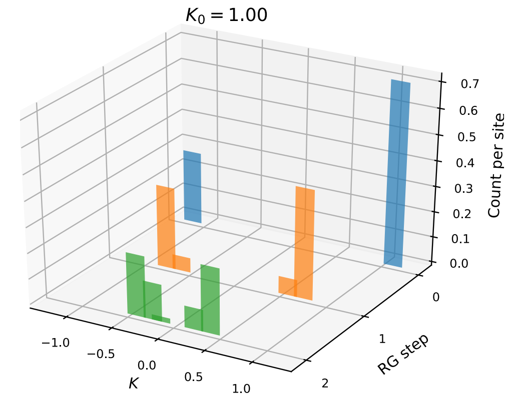
- RG acts on distributions of couplings: $\mathcal{P}(\mathbf{K}) \to \mathcal{P}'(\mathbf{K})$
- Under perfect RSMI capture factorability of disorder distribution is maintained



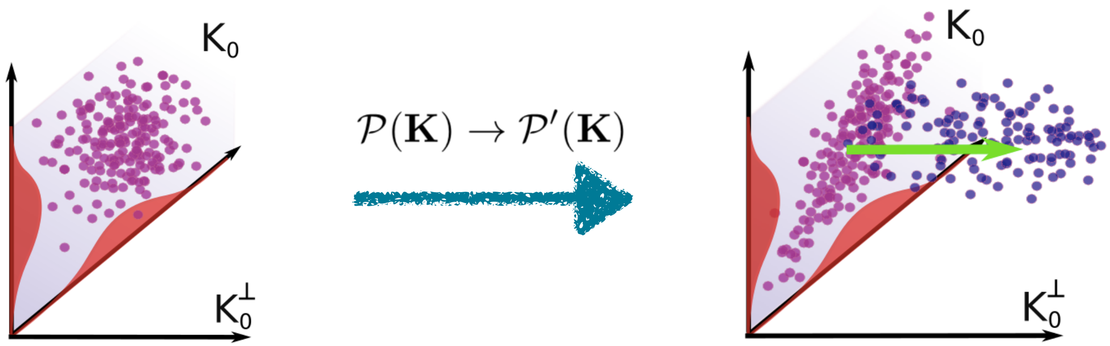
- Example: disordered 1D Ising model

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

Disorder distributions



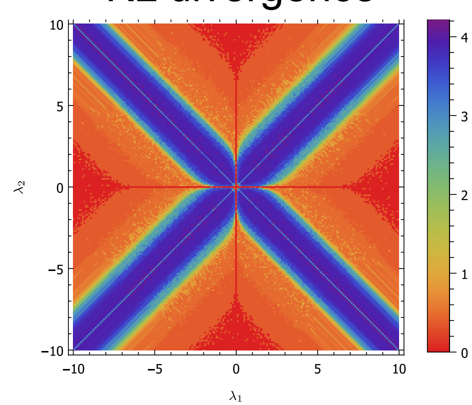
- RG acts on distributions of couplings: $\mathcal{P}(\mathbf{K}) \to \mathcal{P}'(\mathbf{K})$
- Under perfect RSMI capture factorability of disorder distribution is maintained



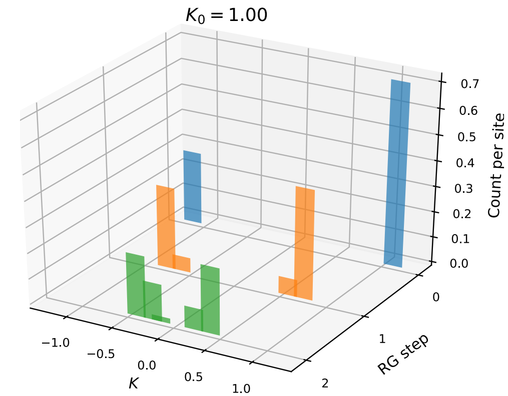
- Example: disordered 1D Ising model

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

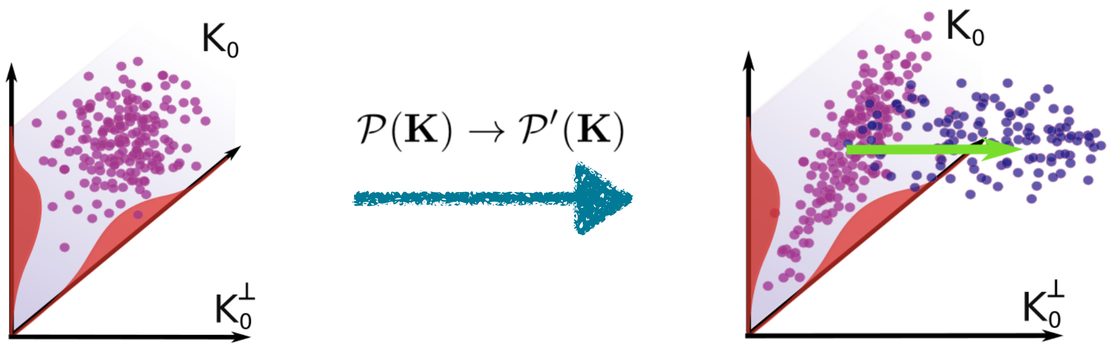
KL divergence



Disorder distributions



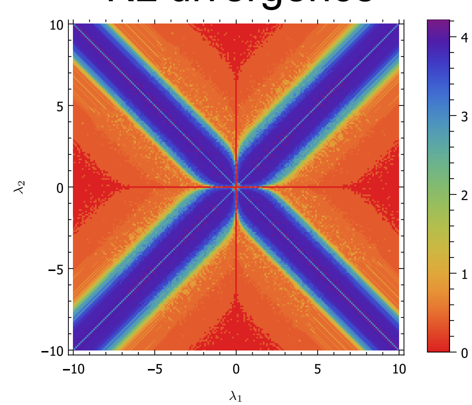
- RG acts on distributions of couplings: $\mathcal{P}(\mathbf{K}) \to \mathcal{P}'(\mathbf{K})$
- Under perfect RSMI capture factorability of disorder distribution is maintained



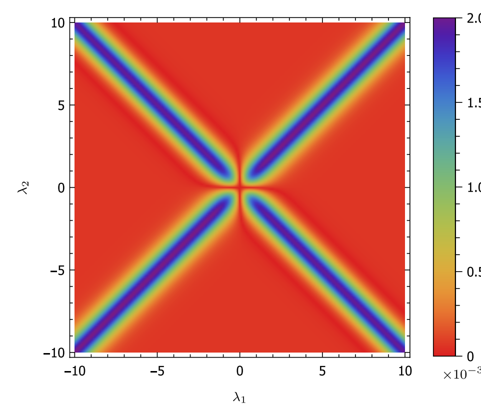
- Example: disordered 1D Ising model

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

KL divergence

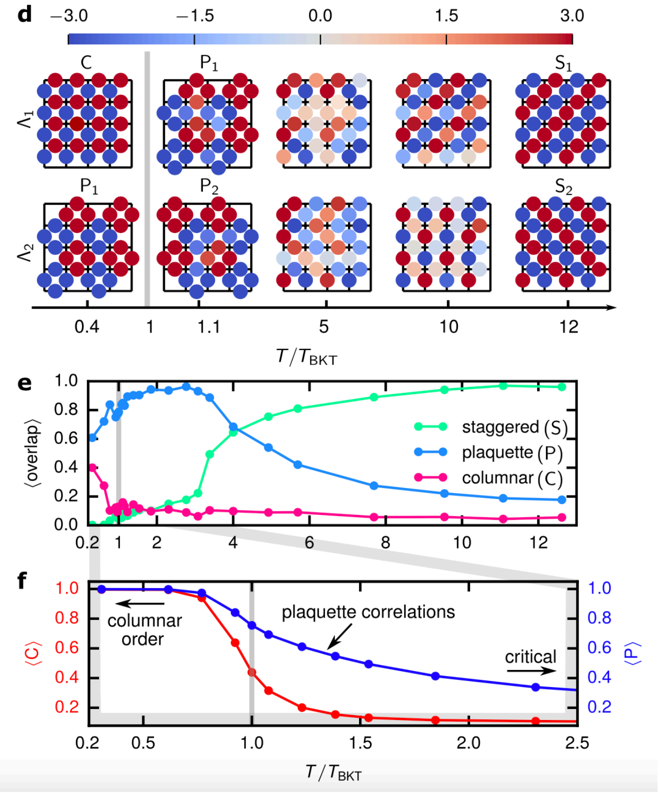


Centre of mass shift



Conclusions

- Constructing an optimal RG transformation as a variation problem in information theory
- https://github.com/RSMI-NE/RSMI-NE
- Comprehensive view of long-distance physics
- Constructing the relevant operators
- Formal connection between compression theory and field theory formalism



Outlook

- Extension to non-equilibrium
- Statistical models in 3D
- Correlations in experimental data: surface measurements, meteo, …

ETH zürich

Thank you!

$$\Lambda_{\mathrm{S1}} \cdot \mathcal{V}(\mathbf{r}) = (-1)^{x+y} N_{|}(\mathbf{r}),$$
$$\Lambda_{\mathrm{S2}} \cdot \mathcal{V}(\mathbf{r}) = (-1)^{x+y+1} N_{-}(\mathbf{r})$$

$$N_{|}(\mathbf{r}) = \frac{1}{4} + \frac{(-1)^{x+y+1}}{2\pi}\partial_x \varphi(\mathbf{r}) + (-1)^y \sin\varphi(\mathbf{r})$$
$$N_{-}(\mathbf{r}) = \frac{1}{4} + \frac{(-1)^{x+y}}{2\pi}\partial_y \varphi(\mathbf{r}) + (-1)^x \cos\varphi(\mathbf{r})$$

Papanikolaou et al. **PRB 76**, 134514 (2007)

$$\mathcal{H}_1 \sim \sum_{\mathbf{r}\in\mathcal{V}} \Lambda_{\mathrm{S1}} \cdot \mathcal{V}(\mathbf{r}) = \sum_{\mathbf{r}\in\mathcal{V}} \left[\frac{(-1)^{x+y}}{4} + \frac{\partial_x\varphi(\mathbf{r})}{2\pi} + (-1)^x \sin\varphi(\mathbf{r})\right] = \sum_{\mathbf{r}\in\mathcal{V}} \left[\frac{\partial_x\varphi(\mathbf{r})}{2\pi} + (-1)^x \sin\varphi(\mathbf{r})\right]$$
$$\mathcal{H}_2 \sim \sum_{\mathbf{r}\in\mathcal{V}} \Lambda_{\mathrm{S2}} \cdot \mathcal{V}(\mathbf{r}) = \sum_{\mathbf{r}\in\mathcal{V}} \left[\frac{(-1)^{x+y}}{4} + \frac{\partial_y\varphi(\mathbf{r})}{2\pi} + (-1)^y \cos\varphi(\mathbf{r})\right] = \sum_{\mathbf{r}\in\mathcal{V}} \left[\frac{\partial_y\varphi(\mathbf{r})}{2\pi} + (-1)^y \cos\varphi(\mathbf{r})\right]$$

$$\mathcal{H} \propto \tau \circ \nabla \left\langle \varphi(\mathbf{r}) \right\rangle_{\mathbf{r}\in\mathcal{V}}$$

- The staggered filters:

$$\Lambda_{\mathrm{S1}} \cdot \mathcal{V}(\mathbf{r}) = (-1)^{x+y} N_{|}(\mathbf{r}),$$
$$\Lambda_{\mathrm{S2}} \cdot \mathcal{V}(\mathbf{r}) = (-1)^{x+y+1} N_{-}(\mathbf{r})$$

$$N_{|}(\mathbf{r}) = \frac{1}{4} + \frac{(-1)^{x+y+1}}{2\pi}\partial_x \varphi(\mathbf{r}) + (-1)^y \sin\varphi(\mathbf{r})$$
$$N_{-}(\mathbf{r}) = \frac{1}{4} + \frac{(-1)^{x+y}}{2\pi}\partial_y \varphi(\mathbf{r}) + (-1)^x \cos\varphi(\mathbf{r})$$

Papanikolaou et al. PRB 76, 134514 (2007)

$$\mathcal{H}_1 \sim \sum_{\mathbf{r}\in\mathcal{V}} \Lambda_{\mathrm{S1}} \cdot \mathcal{V}(\mathbf{r}) = \sum_{\mathbf{r}\in\mathcal{V}} \left[\frac{(-1)^{x+y}}{4} + \frac{\partial_x \varphi(\mathbf{r})}{2\pi} + (-1)^x \sin\varphi(\mathbf{r})\right] = \sum_{\mathbf{r}\in\mathcal{V}} \left[\frac{\partial_x \varphi(\mathbf{r})}{2\pi} + (-1)^x \sin\varphi(\mathbf{r})\right]$$
$$\mathcal{H}_2 \sim \sum_{\mathbf{r}\in\mathcal{V}} \Lambda_{\mathrm{S2}} \cdot \mathcal{V}(\mathbf{r}) = \sum_{\mathbf{r}\in\mathcal{V}} \left[\frac{(-1)^{x+y}}{4} + \frac{\partial_y \varphi(\mathbf{r})}{2\pi} + (-1)^y \cos\varphi(\mathbf{r})\right] = \sum_{\mathbf{r}\in\mathcal{V}} \left[\frac{\partial_y \varphi(\mathbf{r})}{2\pi} + (-1)^y \cos\varphi(\mathbf{r})\right]$$

$$\mathcal{H} \propto \tau \circ \nabla \left\langle \varphi(\mathbf{r}) \right\rangle_{\mathbf{r}\in\mathcal{V}}$$

- The staggered filters:

$$\Lambda_{S1} \cdot \mathcal{V}(\mathbf{r}) = (-1)^{x+y} N_{|}(\mathbf{r}),$$
$$\Lambda_{S2} \cdot \mathcal{V}(\mathbf{r}) = (-1)^{x+y+1} N_{-}(\mathbf{r})$$

$$N_{|}(\mathbf{r}) = \frac{1}{4} + \frac{(-1)^{x+y+1}}{2\pi}\partial_x \varphi(\mathbf{r}) + (-1)^y \sin\varphi(\mathbf{r})$$
$$N_{-}(\mathbf{r}) = \frac{1}{4} + \frac{(-1)^{x+y}}{2\pi}\partial_y \varphi(\mathbf{r}) + (-1)^x \cos\varphi(\mathbf{r})$$

Papanikolaou et al. **PRB 76**, 134514 (2007)

$$\mathcal{H}_1 \sim \sum_{\mathbf{r}\in\mathcal{V}} \Lambda_{S1} \cdot \mathcal{V}(\mathbf{r}) = \sum_{\mathbf{r}\in\mathcal{V}} \left[\frac{(-1)^{x+y}}{4} + \frac{\partial_x \varphi(\mathbf{r})}{2\pi} + (-1)^x \sin\varphi(\mathbf{r})\right] = \sum_{\mathbf{r}\in\mathcal{V}} \left[\frac{\partial_x \varphi(\mathbf{r})}{2\pi} + (-1)^x \sin\varphi(\mathbf{r})\right]$$
$$\mathcal{H}_2 \sim \sum_{\mathbf{r}\in\mathcal{V}} \Lambda_{S2} \cdot \mathcal{V}(\mathbf{r}) = \sum_{\mathbf{r}\in\mathcal{V}} \left[\frac{(-1)^{x+y}}{4} + \frac{\partial_y \varphi(\mathbf{r})}{2\pi} + (-1)^y \cos\varphi(\mathbf{r})\right] = \sum_{\mathbf{r}\in\mathcal{V}} \left[\frac{\partial_y \varphi(\mathbf{r})}{2\pi} + (-1)^y \cos\varphi(\mathbf{r})\right]$$

- Expanding sin/cos and averaging we obtain:

$$\mathcal{H} \propto \tau \circ \nabla \left\langle \varphi(\mathbf{r}) \right\rangle_{\mathbf{r}\in\mathcal{V}}$$

Algorithm 3.1 One epoch for the unsupervised learning procedure for the RSMI-net using InfoNCE lower-bound

η = learning rate
ϵ = relaxation parameter for Gumbel-softmax distribution
$\mathbf{w}^0 \leftarrow$ random hyperparameter tensor ▷ initialise InfoNCE *ansatz* $f(h, e)$
$\Lambda^0 \leftarrow$ random hyperparameter tensor ▷ initialise coarse-graining filter
for s in $1:n$ **do** ▷ loop over all n K-replica samples for $(\mathcal{V}, \mathcal{E})$
  $\epsilon^s \leftarrow$ reduce Gumbel-softmax relaxation parameter
  $\tau^s \leftarrow \tau(\epsilon^s)$ ▷ Anneal Gumbel-softmax layer
  for i in $1:K$ **do**
    for j in $1:K$ **do**
      $h_i^s[\Lambda^s] \leftarrow \tau^s(\Lambda^s \cdot v_i^s)$ ▷ Coarse-grain visible degrees of freedom
      $F_{ij}(\mathbf{w}^s, \Lambda^s) \leftarrow f(h_i^s[\Lambda^s], e_j^s; \mathbf{w}^s)$ ▷ ij'th element of scores matrix
    end for
  end for
  $Q(x_{1:K}, y_{1:K}; \mathbf{w}^s, \Lambda^s) \leftarrow \sum_{j=1}^{K} \frac{F_{jj}(\mathbf{w}^s,\Lambda^s)}{\sum_{ij=1}^{K} \exp F_{ij}(\mathbf{w}^s,\Lambda^s)}$ ▷ InfoNCE "prediction"
  ▷ Update parameters of the RSMI estimator network:
  $\Delta \mathbf{w}^s \leftarrow \eta \nabla_{\mathbf{w}} [\log Q(\mathbf{w}, \Lambda^s)] \Big|_{\mathbf{w}=\mathbf{w}^s}$ ▷ automatic differentiation
  $\mathbf{w}^s \leftarrow \mathbf{w}^s + \Delta \mathbf{w}^s$ ▷ stochastic gradient-ascent
  ▷ Update parameters of the coarse-grainer network:
  $\Delta \Lambda^s \leftarrow \eta \nabla_{\Lambda} [\log Q(\mathbf{w}^s, \Lambda)] \Big|_{\Lambda=\Lambda^s}$
  $\Lambda^s \leftarrow \Lambda^s + \Delta \Lambda^s$
end for
$\tilde{I}_{\Lambda}(\mathcal{H} : \mathcal{E}) = \frac{1}{n} \sum_{t=1}^{n} \log Q(x_{1:K}, y_{1:K}; \mathbf{w}^t, \Lambda^t) + \log K$ ▷ average over n samples
return $\tilde{I}_{\Lambda}(\mathcal{H} : \mathcal{E}), \Lambda^n$

RSMI training