Renormalisation Group and Machine Learning

The Wavelet Conditional Renormalisation Group

Giulio Biroli
Ecole Normale Supérieure
Paris

S. Mallat DI-ENS & CdF

M. Ozawa ENS

T. Marchand ENS

ENS
Département de Physique
École normale supérieure

PR[AI]RIE
PaRis Artificial Intelligence Research InstitutE

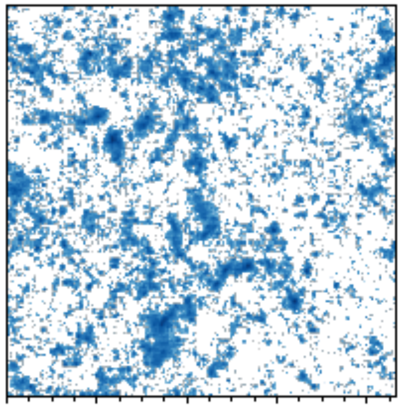
SIMONS FOUNDATION

Generative Models & Sampling

VAE, GAN, Normalizing flows,..

Generative Models

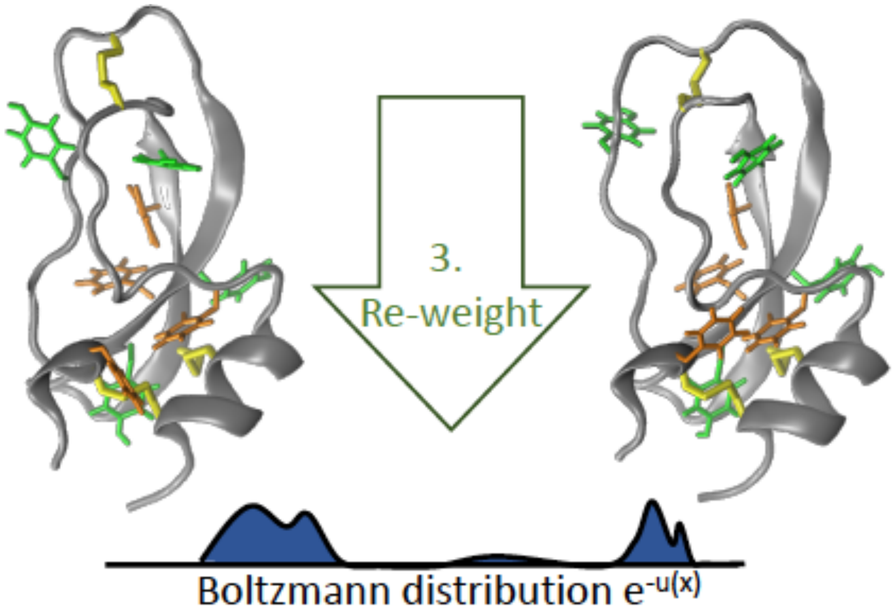
D'Angelo et al 2020



CosmoGAN 2017

New ways of Sampling

Sampling in "one-shot" vs MCMC



Noé et al, Science 2019

How faithful is the representation of the probability distribution?

Limitations (training efficiency) and scalability?

Theory: Why does it work?

Renormalisation Group & Machine Learning

Renormalisation Group: hierarchical coarse grain of the probability distribution from small to large scale. From small scale properties to large scale physics.

Deep neural networks also work in a hierarchical way and learn small scale structure at the first layers and hierarchically go to global features in the last layer.

Several works have proposed and investigated this analogy (eg Mehta Schwab 2014)

Outline

Focus on images from physical models (statphys & cosmology)

Aim: reconstruct an approximate representation of the probability distribution from data -> sampling and energy function

Combine idea from physics (renormalization group) and computer science (wavelet, sparse representations) to construct a method - ***the wavelet conditional RG*** - to approximate probability distributions

★Efficient training and sampling (“in one-shot”)

★Reconstruct the energy function

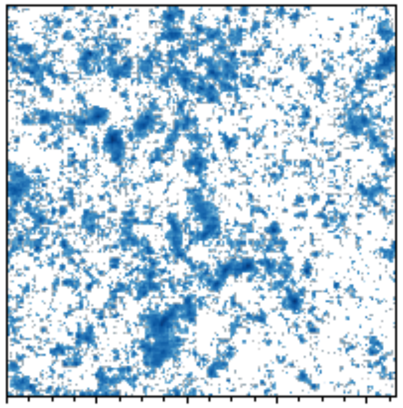
★Theoretical analysis: data structure vs high-dimensional distributions

Generative Models & Sampling

VAE, GAN, Normalizing flows,..

Generative Models

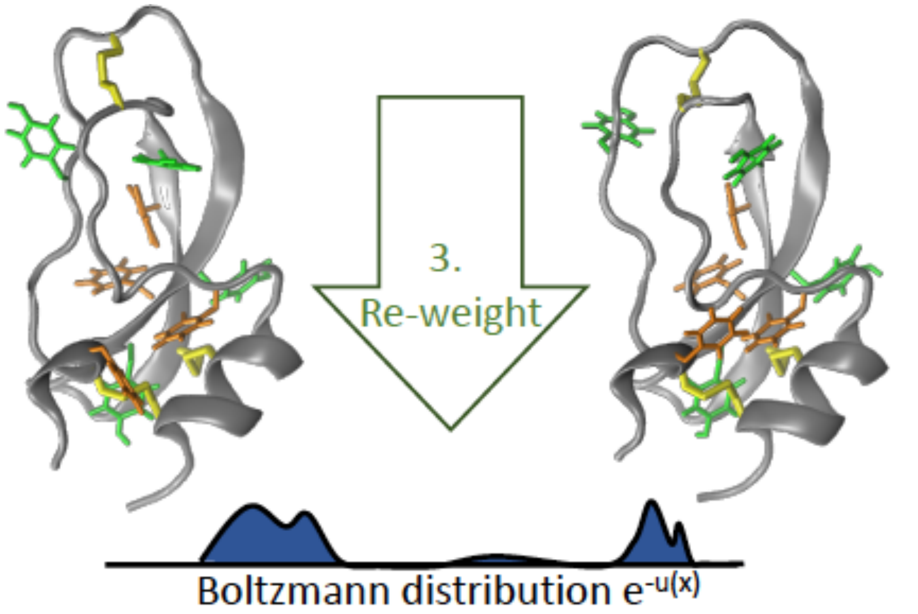
D'Angelo et al 2020



CosmoGAN 2017

New ways of Sampling

Sampling in "one-shot" vs MCMC



Noé et al, Science 2019

How faithful is the representation of the probability distribution?

Limitations (training efficiency) and scalability?

Theory: Why does it work?

Outline

Focus on images from physical models (statphys & cosmology)

Aim: reconstruct an approximate representation of the probability distribution from data -> sampling and energy function

Combine idea from physics (renormalization group) and computer science (wavelet, sparse representations) to construct a method - ***the wavelet conditional RG*** - to approximate probability distributions

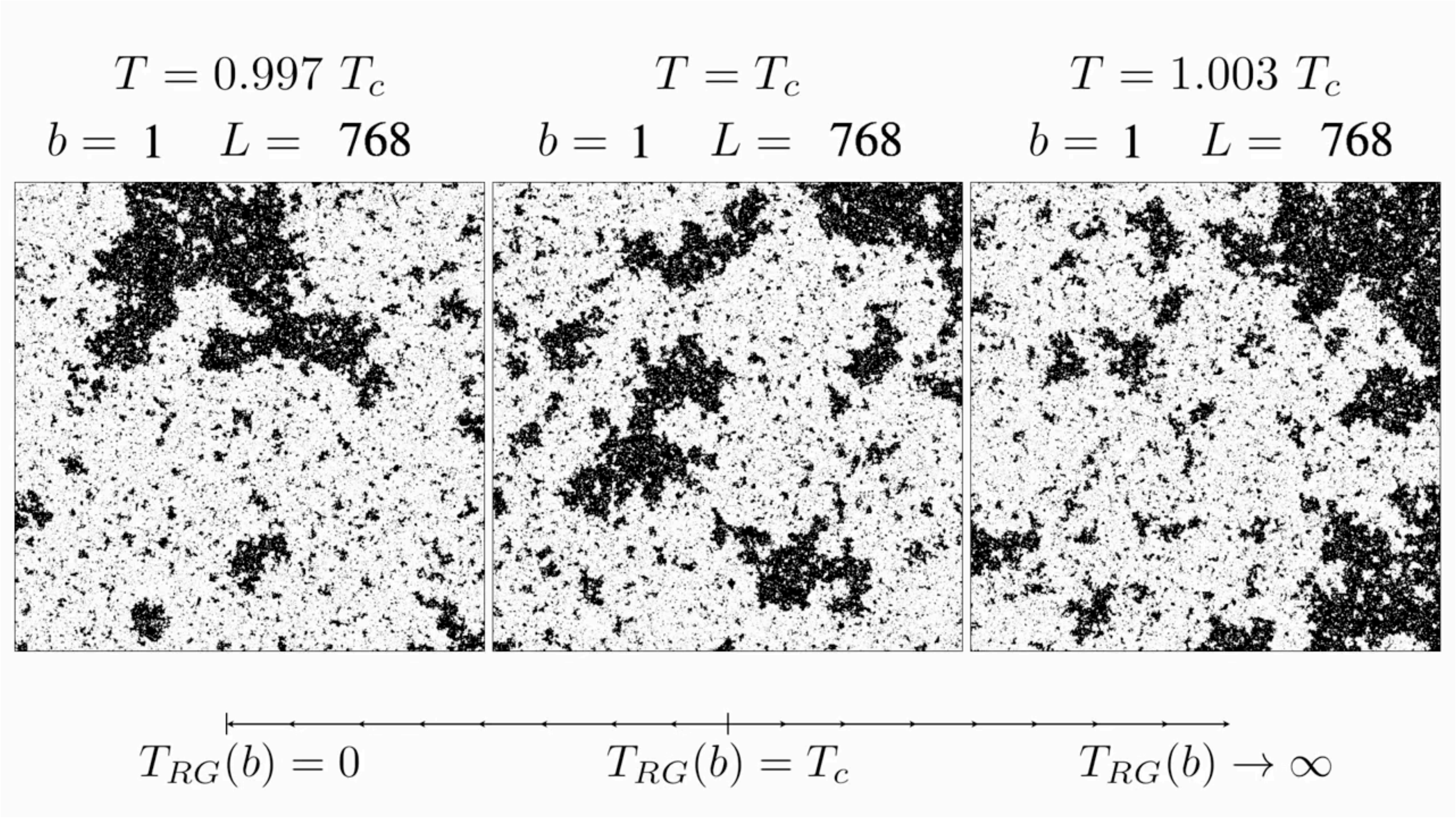
★Efficient training and sampling (“in one-shot”)

★Reconstruct the energy function

★Theoretical analysis: data structure vs high-dimensional distributions

Renormalisation group in a nutshell

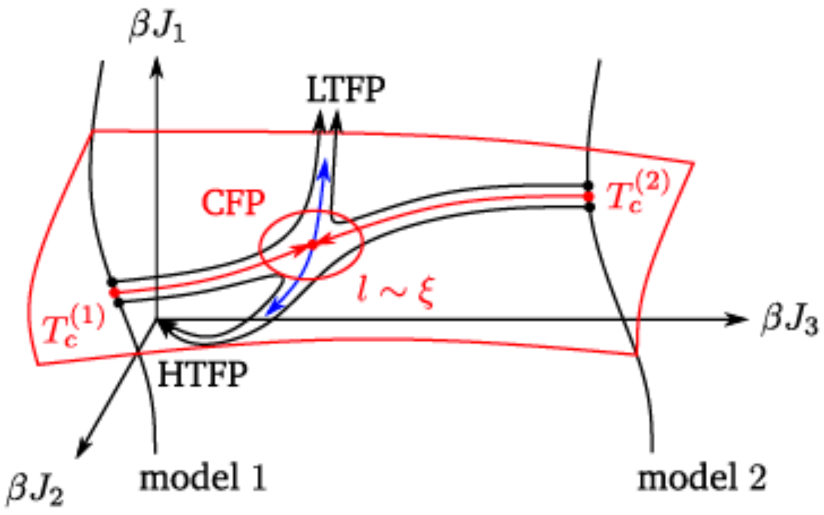
RG for
the Ising Model



- Integrate out the “fast” (or local) degrees of freedom and rescale

- RG leads to a flow in the space of models (or energy functions)

- Second order phase transition associated
  to non-trivial fixed points



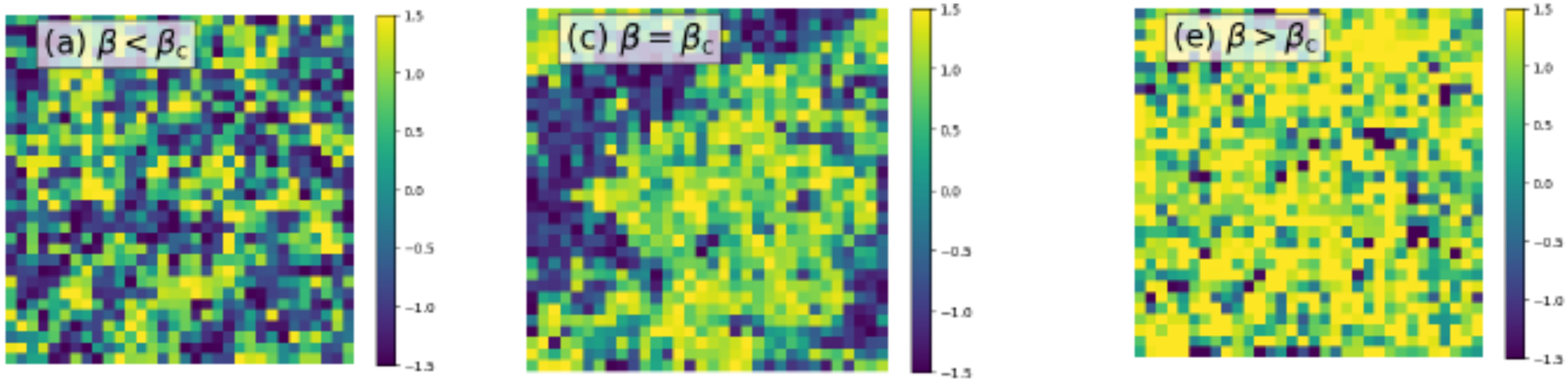
RG always works on “fast” (or local)
degrees of freedom scale by scale

-> no singular behaviour (divergencies), no instability,…

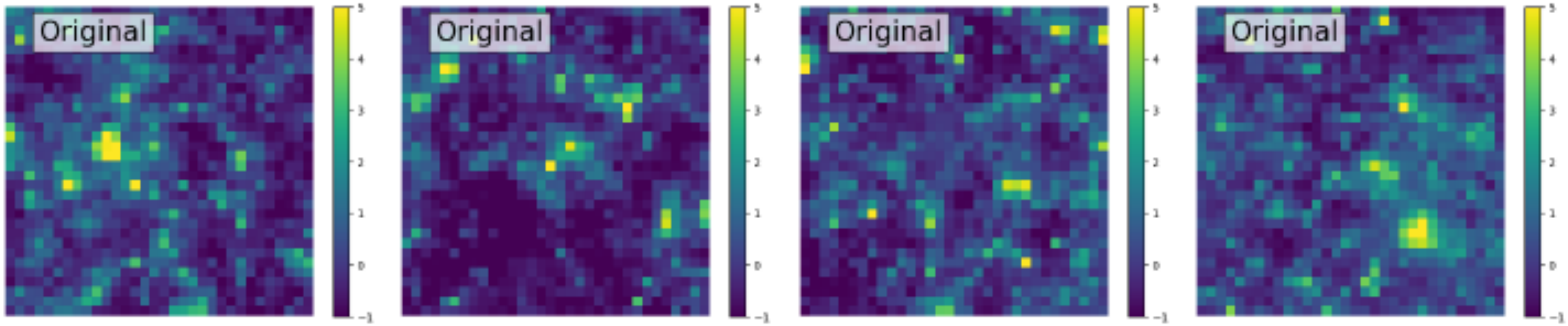
Data and Systems

- Discretized 2D field theory for the ferromagnetic phase transition

$$H(\varphi) = -\frac{1}{2}\sum \varphi(i)\Delta_{ij}\varphi(j) + \sum V(\varphi(i))$$



- Cosmological data (weak lensing maps from the Columbia group)



Data: 32 by 32 images, ~10000-70000 images (~250 parameters)

approximate faithfully the probability distribution
obtain the energy function

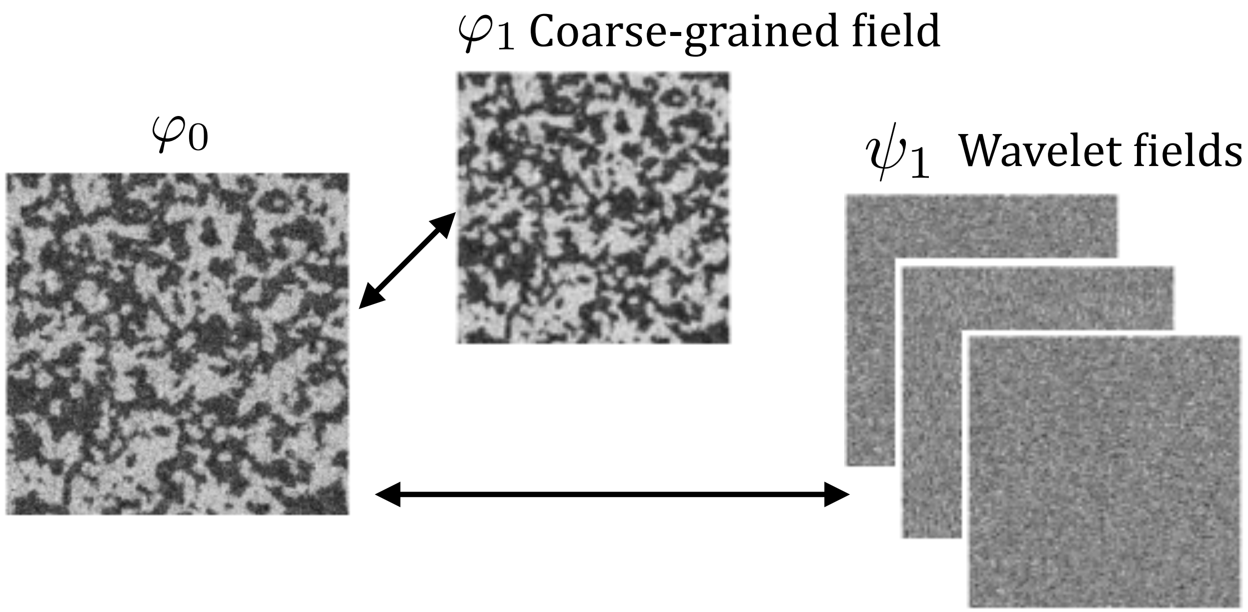
Wavelet Conditional RG

- Reconstruct and sample the probability distribution scale by scale from coarse to fine scales

$$\ell_j = 2^j \rightarrow \ell_{j-1} = 2^{j-1}$$

$$\varphi_{j-1}(i) = L^T \varphi_j(i) + G^T \psi_j(i)$$

$$p_{j-1}(\varphi_{j-1}) = p_j(\psi_j|\varphi_j)p_j(\varphi_j)$$



$$p(\varphi_0) = p_1(\psi_1|\varphi_1)p_2(\psi_2|\varphi_2)...p_{J-1}(\psi_{J-1}|\varphi_{J-1})p_J(\varphi_J)$$

Reconstruction scale by scale

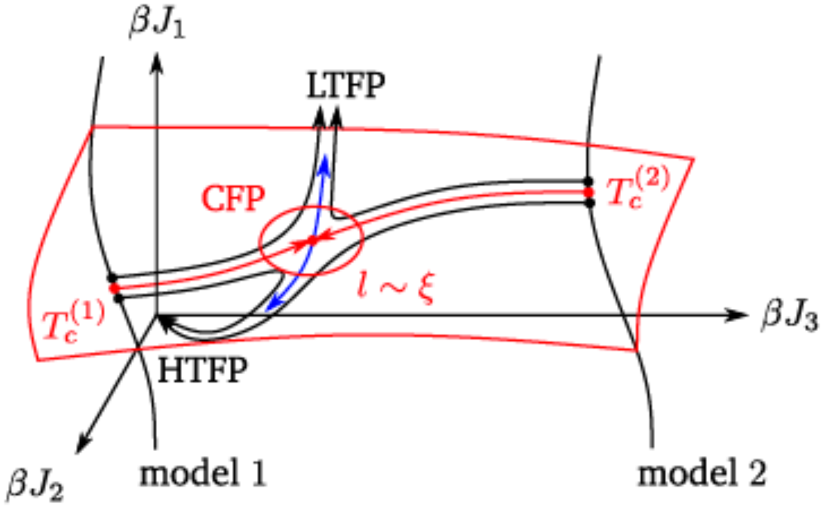
$$p^A_{j-1}(\varphi_{j-1}) = p^A_j(\underset{\uparrow}{\psi_j|\varphi_j})p^A_j(\varphi_j) \leftarrow$$

known

estimate from data

Sampling scale by scale $\varphi_J \rightarrow \psi_{J-1} ... \rightarrow \psi_2 \rightarrow \psi_1$

Inverse RG



Stable representation of the probability distribution

It always works on "fast" (or local)
degrees of freedom scale by scale

Model for the conditional probability

Usual RG flow: model (or Ansatz) for the energy function over which "project" the flow

$$\mathcal{H}_{j-1}(\varphi_{j-1}) = -\frac{1}{2}\sum \varphi_{j-1}(i)K^A_{j-1}(i-i')\varphi_{j-1}(i') + \sum V^A_{j-1}(\varphi_{j-1}(i)) \quad \mathcal{H}_{j-1} \to \mathcal{H}_j$$

$$p_j(\psi_j|\varphi_j) = p_{j-1}(\varphi_{j-1})/p_j(\varphi_j)$$

Model $p^A_j(\psi_j|\varphi_j) = e^{-\overline{\mathcal{H}}_{j-1}(\varphi_{j-1})+F_j(\varphi_j)}$

$$\overline{\mathcal{H}}_{j-1}(\varphi_{j-1}) = \mathcal{H}_{j-1}(\varphi_{j-1}) - \mathcal{H}_{j-1}(\varphi_{j-1})|_{\psi_j=0} \qquad F_j(\varphi_j) = -\ln\int d\psi_j e^{-\overline{\mathcal{H}}_{j-1}(\varphi_{j-1})}$$

To reconstruct the microscopic energy $\mathcal{H} = \sum\left(\overline{\mathcal{H}}_{j-1} - F_j\right) + \mathcal{H}_J$

Model $F_j(\varphi_j) = \underline{-\mathcal{H}_{j-1}(\varphi_{j-1})|_{\psi_j=0} + \mathcal{H}_j(\varphi_j)} + \sum \underline{\tilde{V}^A_j(\varphi_j(i))}$

$$\downarrow$$

New term needed
for cosmological data

$$\mathcal{H}(\varphi) = \underline{-\frac{1}{2}\sum \varphi(i)K^A_0(i-i')\varphi(i') + \sum V^A_0(\varphi(i))} + \sum_j\sum_i \underline{\tilde{V_j}^A(\varphi_j(i))}$$

Estimation from data

$$p_j^A(\psi_j|\varphi_j) = e^{-\overline{\mathcal{H}}_{j-1}(\varphi_{j-1})+F_j(\varphi_j)} \longrightarrow p_j^A(\psi_j|\varphi_j) = e^{-\sum_l g_l^{(j)} U_l(\varphi_{j-1})+F_j(\varphi_j)}$$

To obtain the parameters: minimise the Kullback-Leibler divergence $D_{KL}(p_j^{true}||p_j^A)$

Gradient descent: $$g_l^{(j)}(t+1) - g_l^{(j)}(t) = \eta\left(\langle U_l(\phi_{j-1})\rangle_{p_j^A} - \langle U_l(\phi_{j-1})\rangle_{p_j^{true}}\right)$$

From Montecarlo simulation

From data

To obtain the approximated free-energy: regress the true free-energy to the model one

$$F_j(\varphi_j) = -\ln\int d\psi_j e^{-\overline{\mathcal{H}}_{j-1}(\varphi_{j-1})} \xrightarrow[\text{regression}]{} F_j^A(\varphi_j)$$

From Montecarlo simulation

Gradient descent and MCMC well-conditioned & fast because they always work with the wavelet field (“fast degrees of freedom)

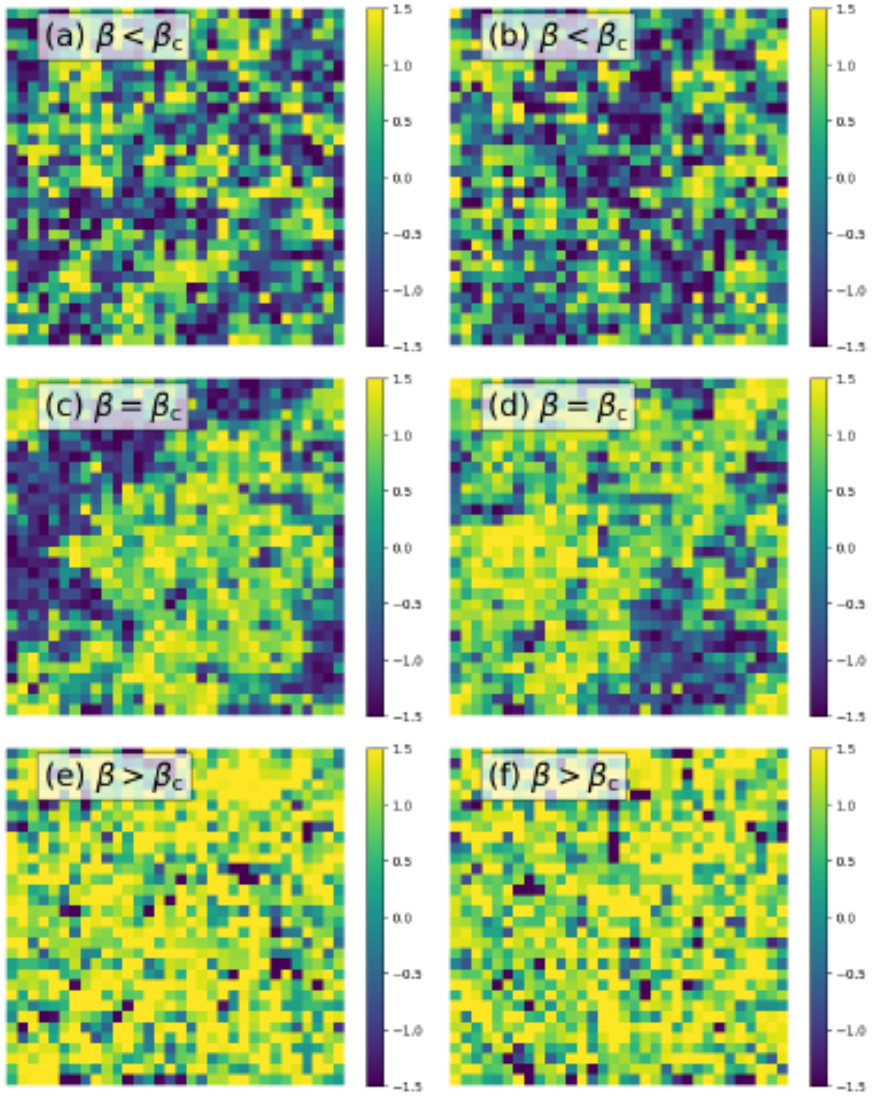
Convergence in a time O(N) for N degrees of freedom (optimal)

Rigorous analysis in the Gaussian case,
RG/perturbation theory for the more general cases

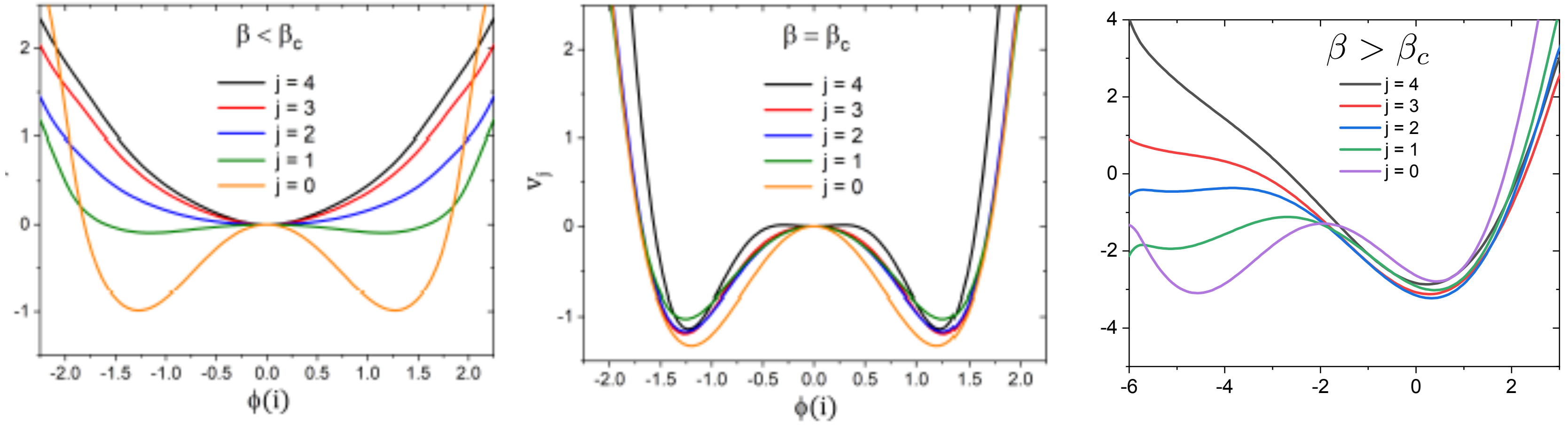
Numerical Applications I

Discretized 2D field theory for the ferromagnetic phase transition

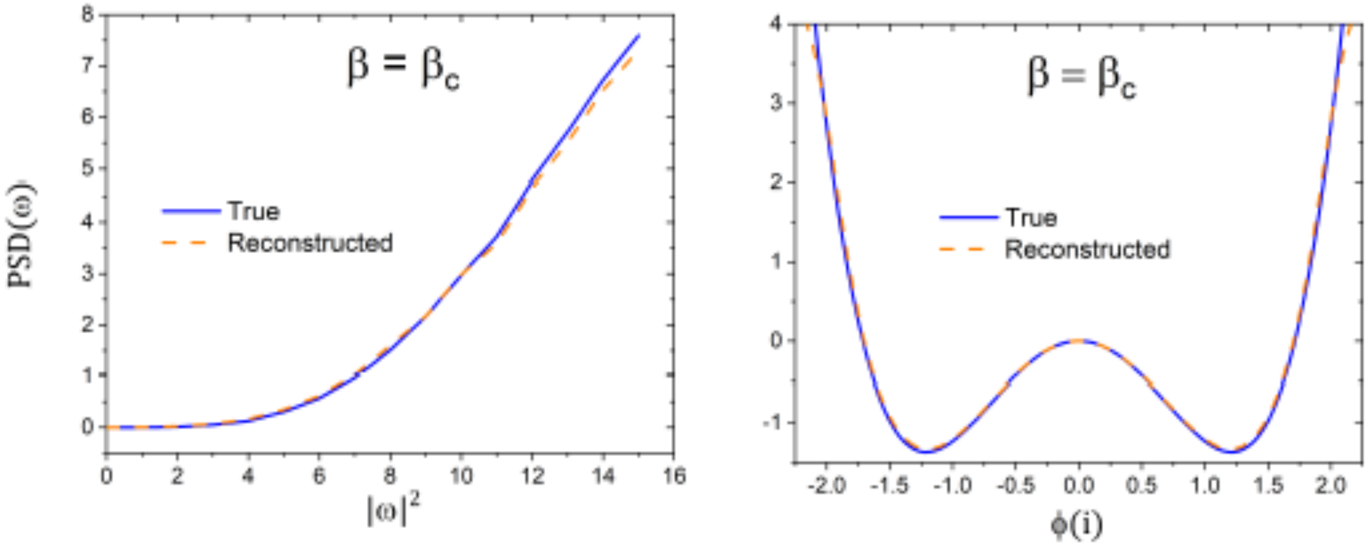
True Generated



Inverse RG Flow



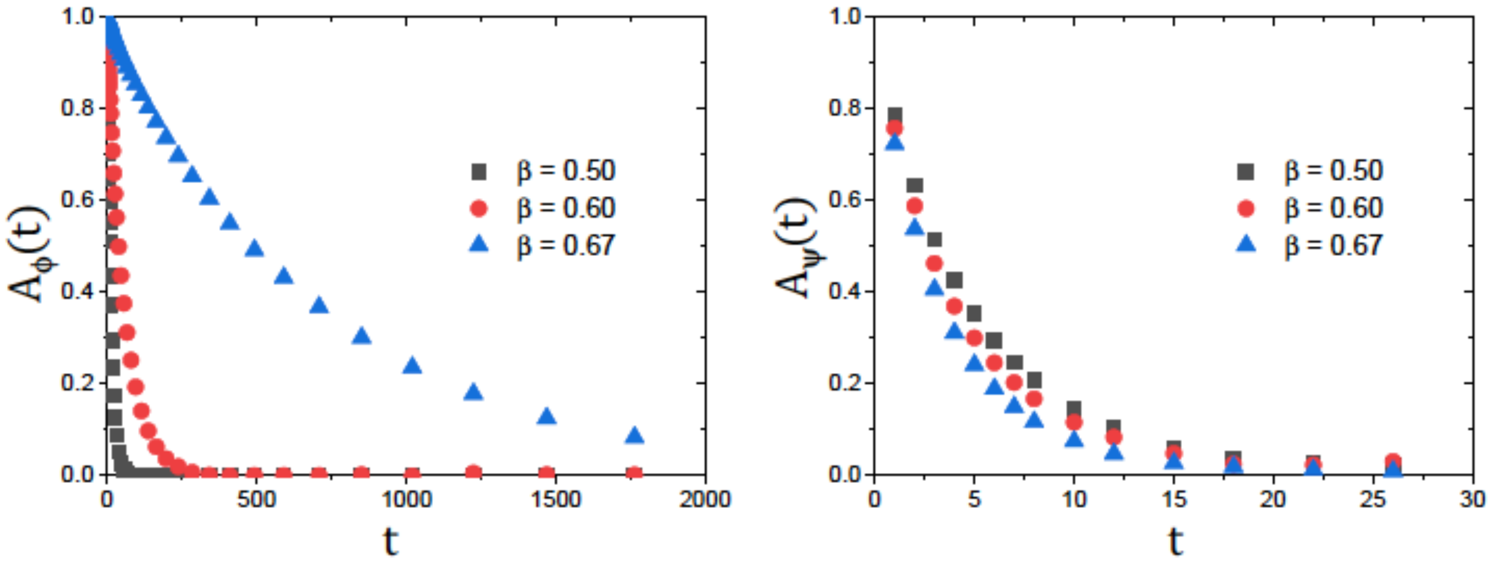
Comparison true & reconstructed



Numerical Applications I

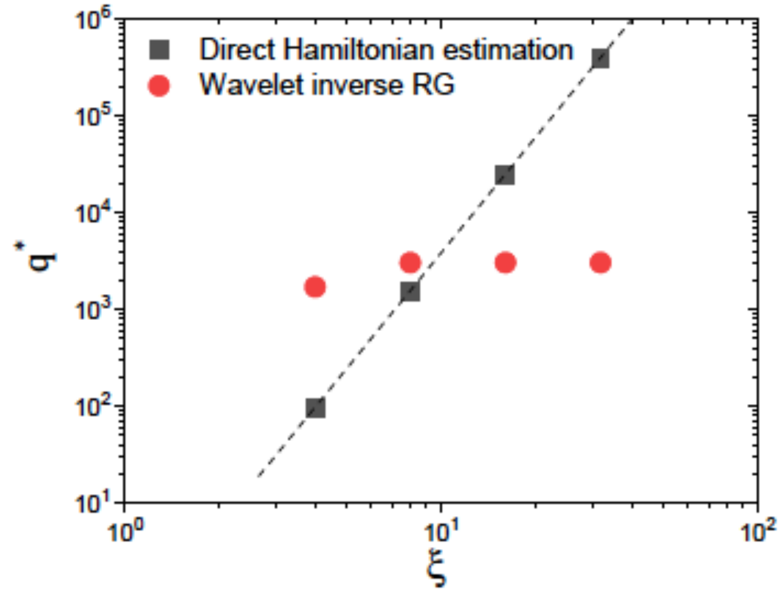
Discretized 2D field theory for the ferromagnetic phase transition

Critical slowing down avoided
Montecarlo on wavelet field is fast at each scale



Normalized time-dependent correlation functions

Numerically stable estimation
by gradient descent

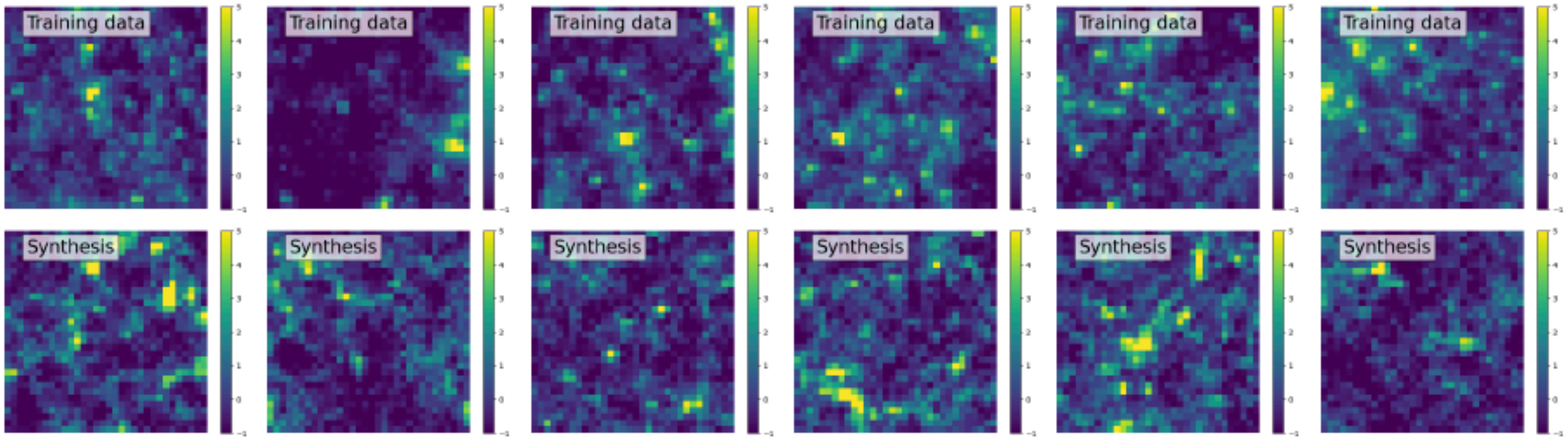


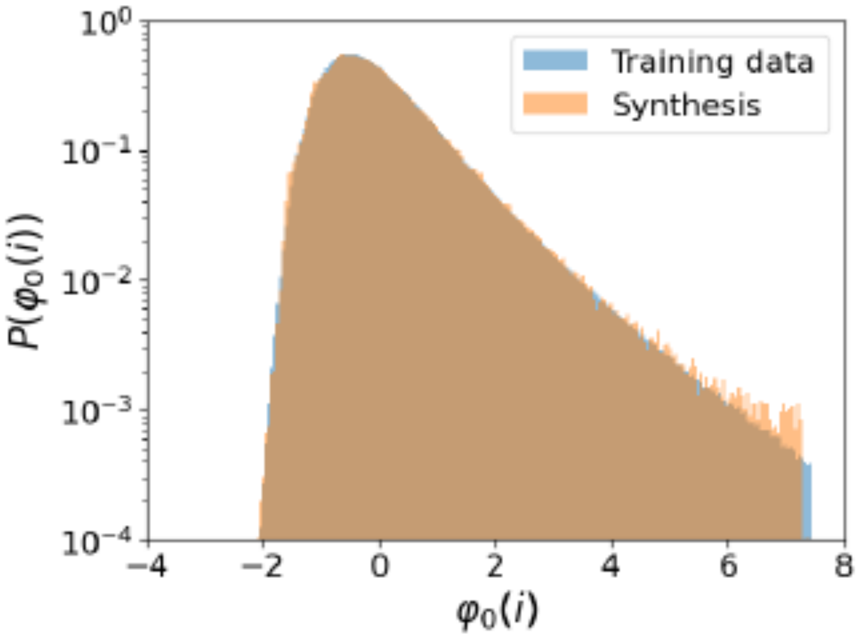
Iterations to reach a prescribed accuracy

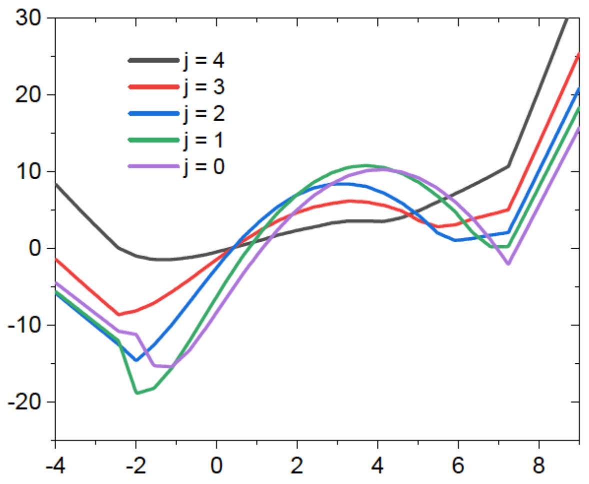
- The Wavelet Conditional RG is a well-conditioned method
- It can generates typical samples in times O(N) for N degrees of freedom even at criticality

Numerical Applications II

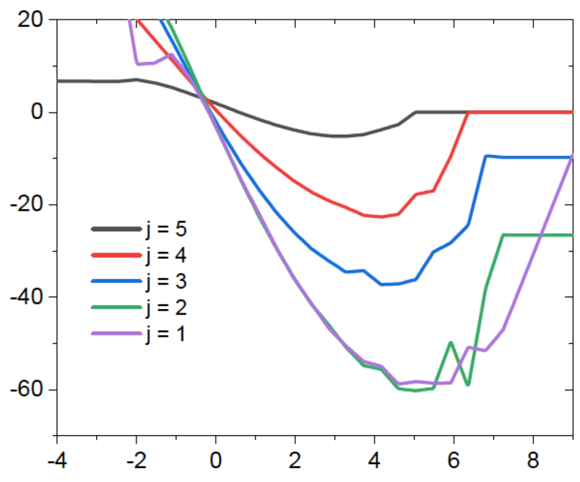
- Cosmological data (weak lensing maps from the Columbia group)







$V_j^A(\varphi_j)$

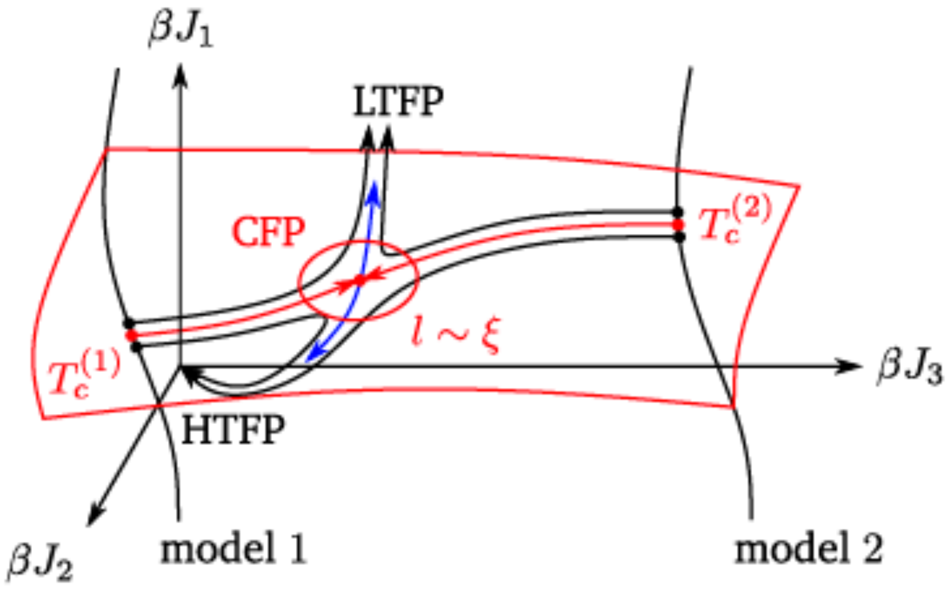


$\tilde{V}_j^A(\varphi_j)$

Wavelet Conditional RG

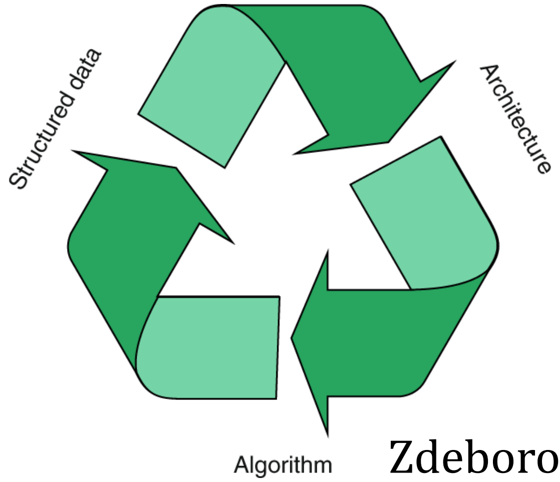
Reconstructing the microscopic probability distribution scale by scale as an inverse RG flow

It always works on “fast” degrees of freedom at each scale like RG



- <u>Sampling and training in “one-shot”</u>:not affected by long-range correlations and critical slowing down
- <u>Reconstruct the Hamiltonian</u>

A first step to understand theoretically how the interplay of data structure and architecture allows to avoid the curse of dimensionality and enable efficient training and sampling



Zdeborová,
Nat Phys 2020