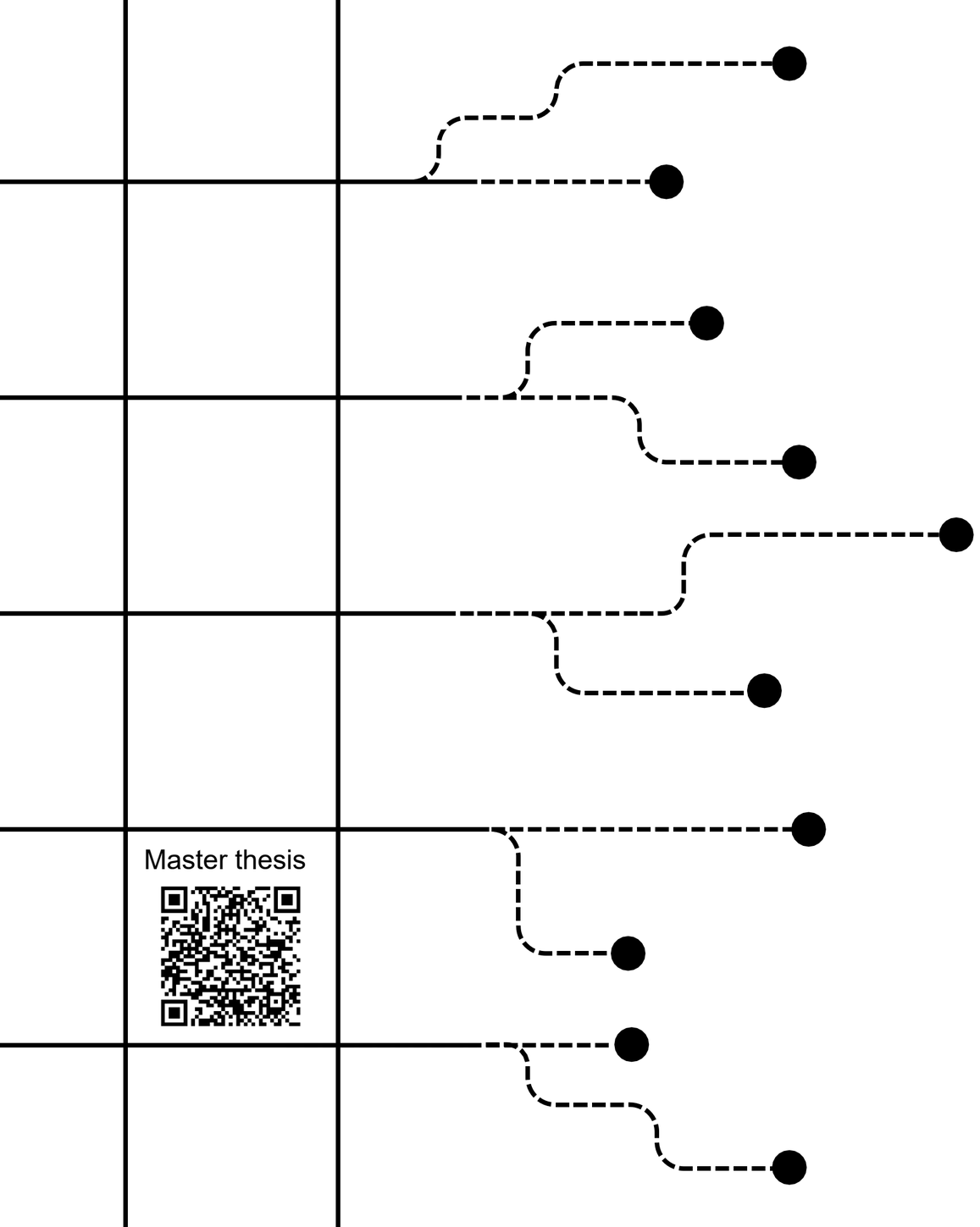


Master thesis

Swiss National Science Foundation

TU WIEN

b UNIVERSITÄT BERN

Machine learning a perfect lattice action for the O(3) non-linear sigma-model

Liane Backfried, Andreas Ipp, David Müller, Urs Wenger

Multi-canonical methods and lattice field theory
December 3 2025, ECT* Trento, Italy

Outline

1. Introduction
2. Non-linear sigma model
3. Fixed-point actions
4. Quantum perfect actions

Introduction

Consider an asymptotically free quantum field theory on the lattice,

e.g., the **O(3) non-linear σ-model**:

$$Z(\beta) = \int \mathcal{D}S \exp\{-\beta \mathcal{A}[S]\} \text{ with coupling } \beta = \frac{1}{g}$$

and expectation values for observables:

$$\langle \mathcal{O}(\beta) \rangle = \frac{1}{Z} \int \mathcal{D}S \exp\{-\beta \mathcal{A}[S]\} \mathcal{O}[S]$$

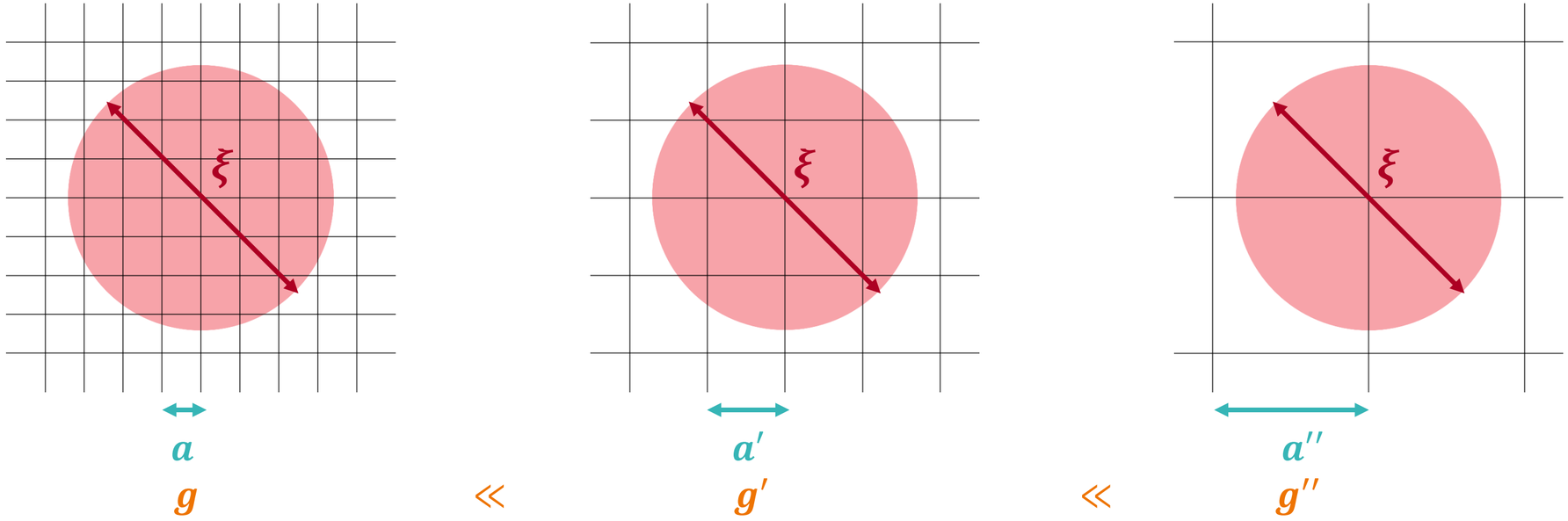
with a characteristic length scale ξ in units of the lattice spacing a:

$\xi/a \Rightarrow$ dimensionless

Introduction

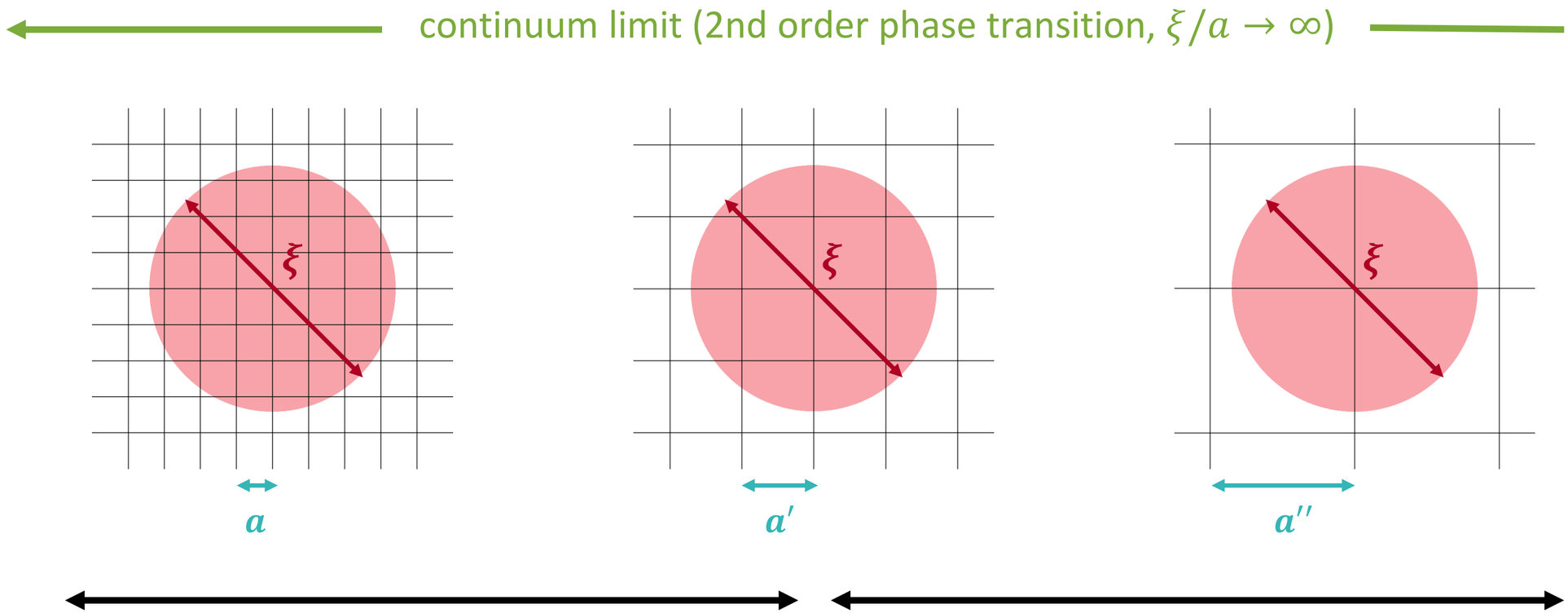
The lattice spacing a is determined by the coupling $\beta = 1/g$

continuum limit (2nd order phase transition, $\xi/a \to \infty$)



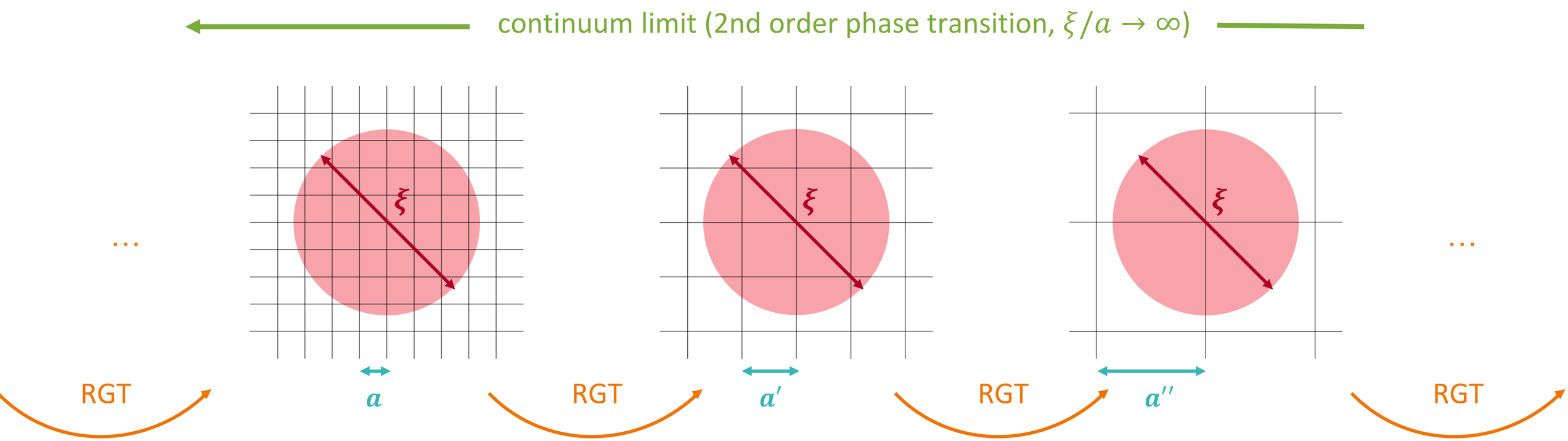
Introduction

The lattice spacing a is determined by the coupling $\beta = 1/g$

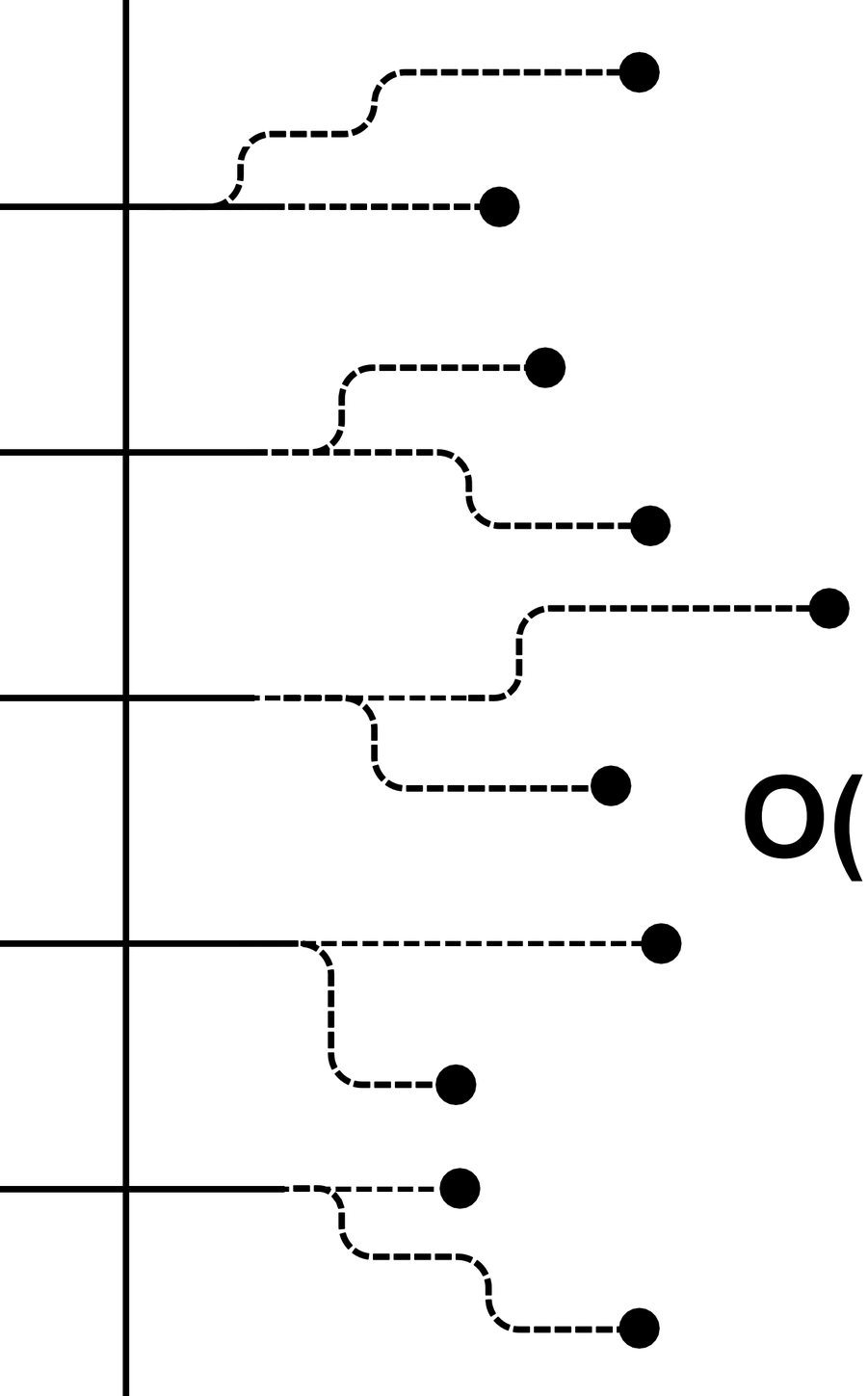


Renormalization group transformation

Introduce (real space) renormalization group transformation (RGT) :



⇒ provides solution for avoiding critical slowing down and lattice artefacts



Intermezzo:
O(3) non-linear sigma-model

O(3) non-linear σ-model (in d=2)

2d QFT of non-linearly interacting scalar fields = “spins”

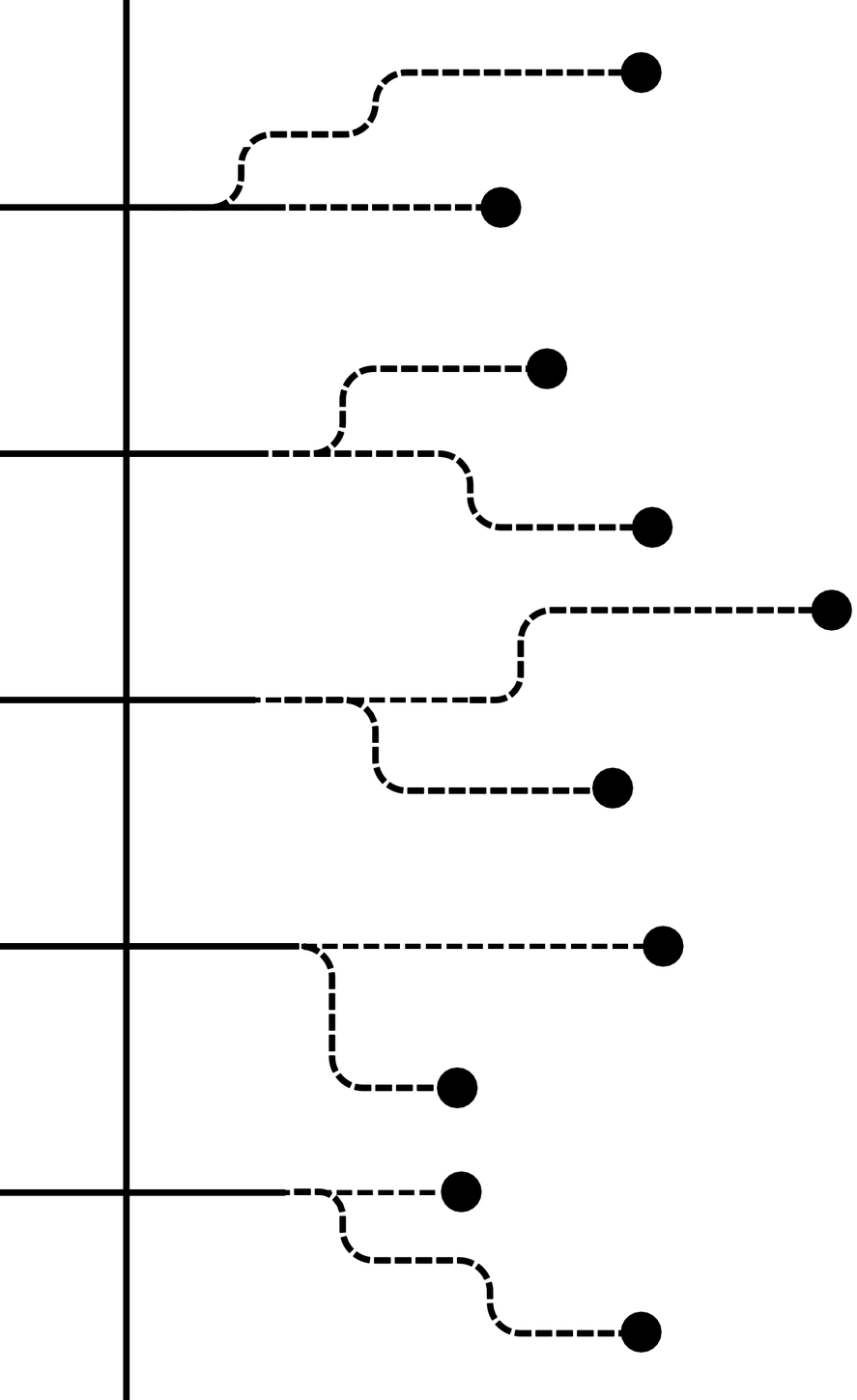
$\vec{S} = (S^1, S^2, S^3)^T$ with normalization $\vec{S} \cdot \vec{S} = 1$

with (Euclidean) continuum action:

$\beta \mathcal{A}[S] = {}^{\beta}/_{2} \int d^2x \, \partial_\mu S^i(x) \cdot \partial_\mu S^i(x)$ with coupling $\beta = \frac{1}{g}$

Use as toy model to test new machine learning approaches to RG-improved action studies

Hasenfratz, Niedermayer [hep-lat/9308004] – original FP study for non-linear σ-model
Holland, Ipp, Müller, Wenger [2401.06481] – recent machine learning study of FP actions for SU(3) gauge theory



Part I:
Fixed-point action

Renormalization group transformation

Introduce (real space) renormalization group transformation (RGT) :

Hasenfratz, Niedermayer [hep-lat/9308004]

$$\exp\{-\beta' \mathcal{A}'[R]\} = \int \mathcal{D}S \exp\{-\beta(\mathcal{A}[S] + T[R,S])\}$$

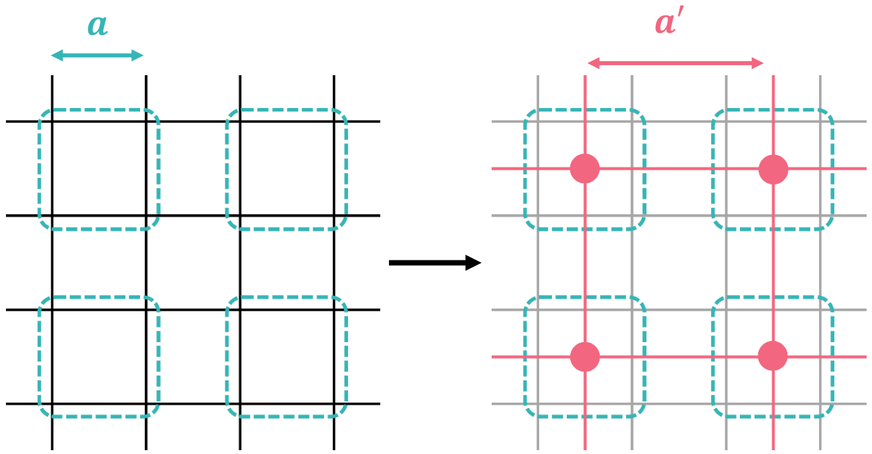
where $T[R,S]$ is a blocking kernel relating the fine spin configurations $\boldsymbol{S}$ to the coarse spins $\boldsymbol{R \equiv S'}$:

$$T[R,S] = \sum_{x_B}\left[-\kappa R_{x_B} \cdot \Sigma_{x_B} + \mathcal{N}\left(\left|\Sigma_{x_B}\right|\right)\right]$$

with normalization $\mathcal{N} = \frac{1}{\beta}\ln Y_3\left(\beta\kappa\left|\Sigma_{x_B}\right|\right)$,

blocked spins $\Sigma_{x_B} = \sum_{x \in x_B} S_x$,

and κ is a tuneable parameter of the RGT



blocking of fine $\boldsymbol{S}$ to coarse $\boldsymbol{\Sigma}$

Renormalization group transformation

Introduce (real space) renormalization group transformation (RGT) :

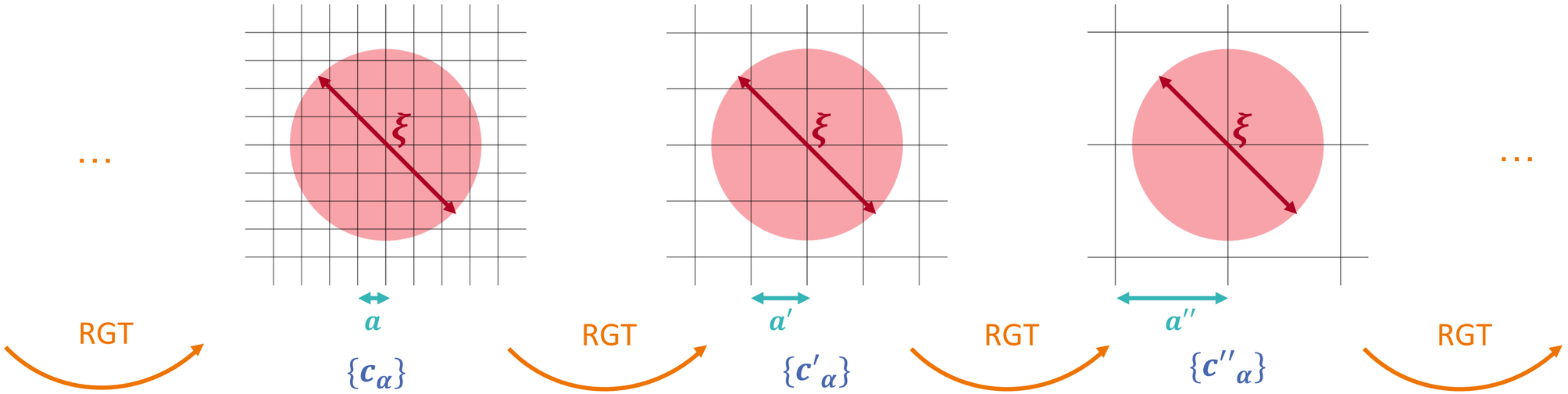
Hasenfratz, Niedermayer [hep-lat/9308004]

$$\exp\{-\beta'\mathcal{A}'[R]\} = \int \mathcal{D}S \exp\{-\beta(\mathcal{A}[S] + T[R,S])\}$$

e.g., $c_0 \sim \mathcal{O}(S^2)$, $c_1 \sim \mathcal{O}(S^4)$, …

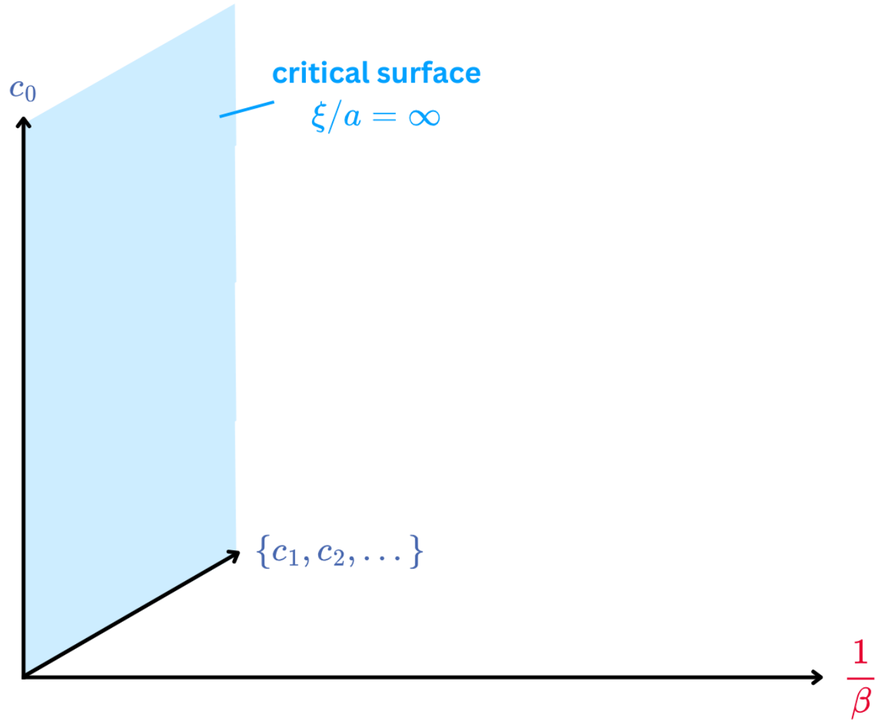
Let the original action be defined in terms of an (infinite) set of couplings $\mathcal{A}[S] = \mathcal{A}[S; c_0, c_1, \dots]$

⇒ The effective action $\beta'\mathcal{A}'[R]$ is described by infinitely many transformed couplings $\{c_\alpha{}'\}$:



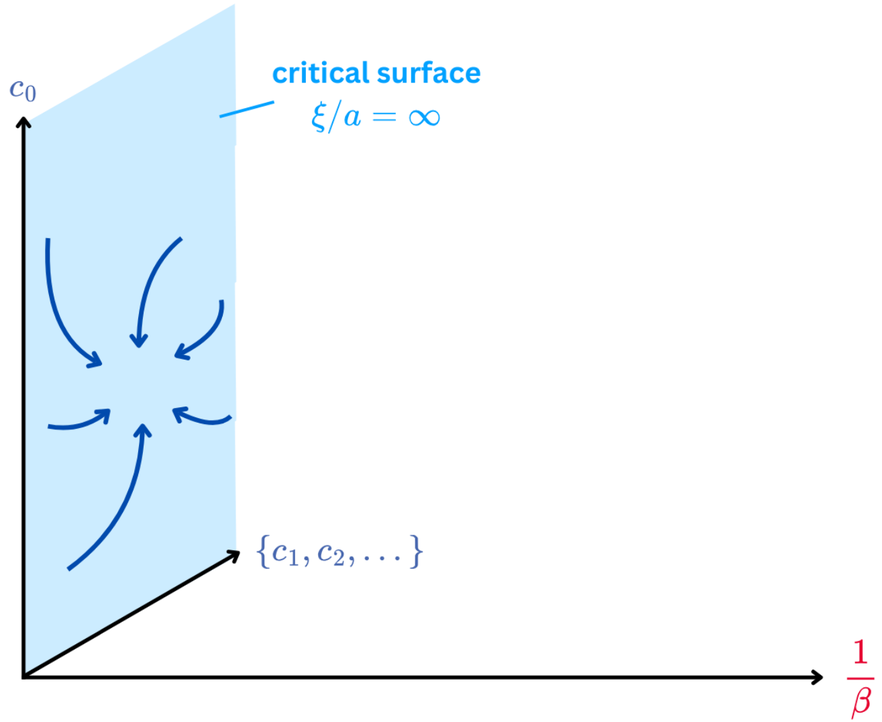
Renormalization group transformation

The effective action $\beta\mathcal{A}[R]$ is described by infinitely many transformed couplings $\{c_\alpha\}$:



Renormalization group transformation

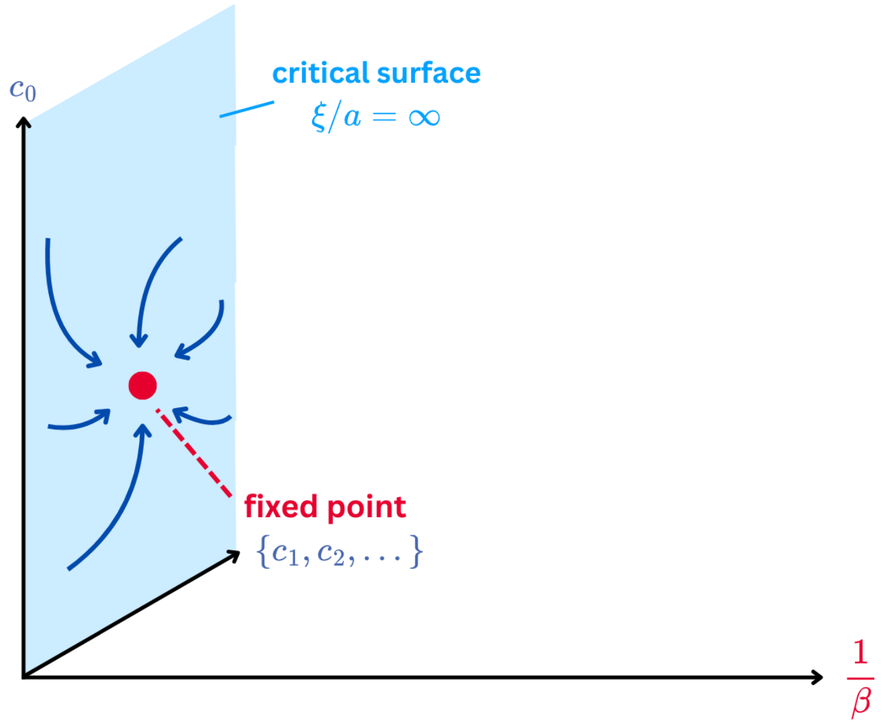
The effective action $\beta\mathcal{A}[R]$ is described by infinitely many transformed couplings $\{c_\alpha\}$:



$\Rightarrow$ at $\xi/a \to \infty$, fixed point of RGT: $\{c_\alpha^{\mathrm{FP}}\} \xrightarrow{\text{RGT}} \{c_\alpha^{\mathrm{FP}}\}$

Renormalization group transformation

The effective action $\beta\mathcal{A}[R]$ is described by infinitely many transformed couplings $\{c_\alpha\}$:

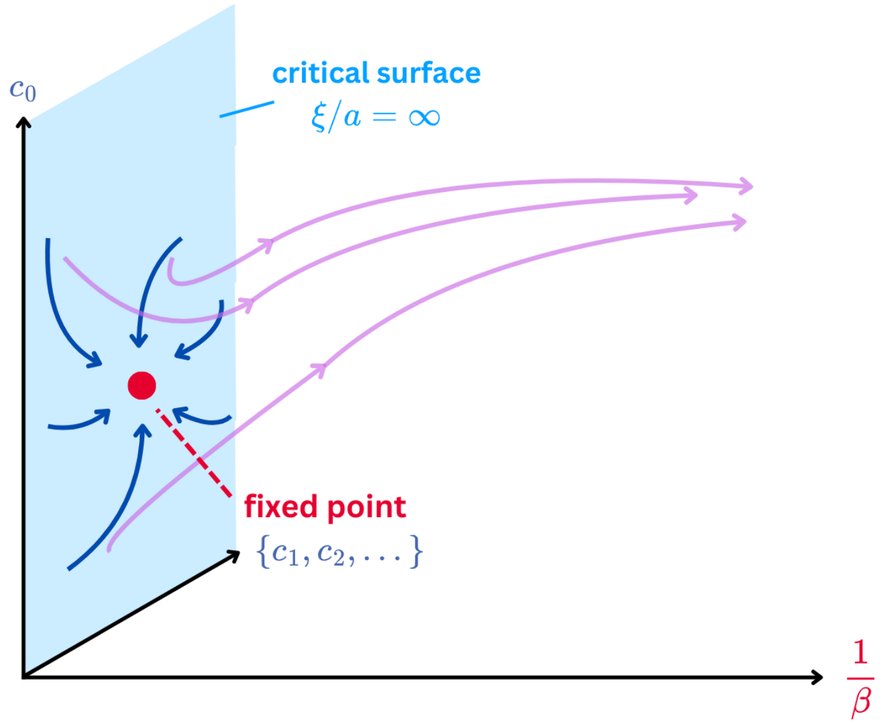


⇒ at $\xi/a \to \infty$, fixed point of RGT: $\{c_\alpha^{\mathrm{FP}}\} \xrightarrow{\text{RGT}} \{c_\alpha^{\mathrm{FP}}\}$

⇒ asymptotically free theories:
FP has one relevant coupling $g \sim 1/\beta$

Renormalization group transformation

The effective action $\beta\mathcal{A}[R]$ is described by infinitely many transformed couplings $\{c_\alpha\}$:

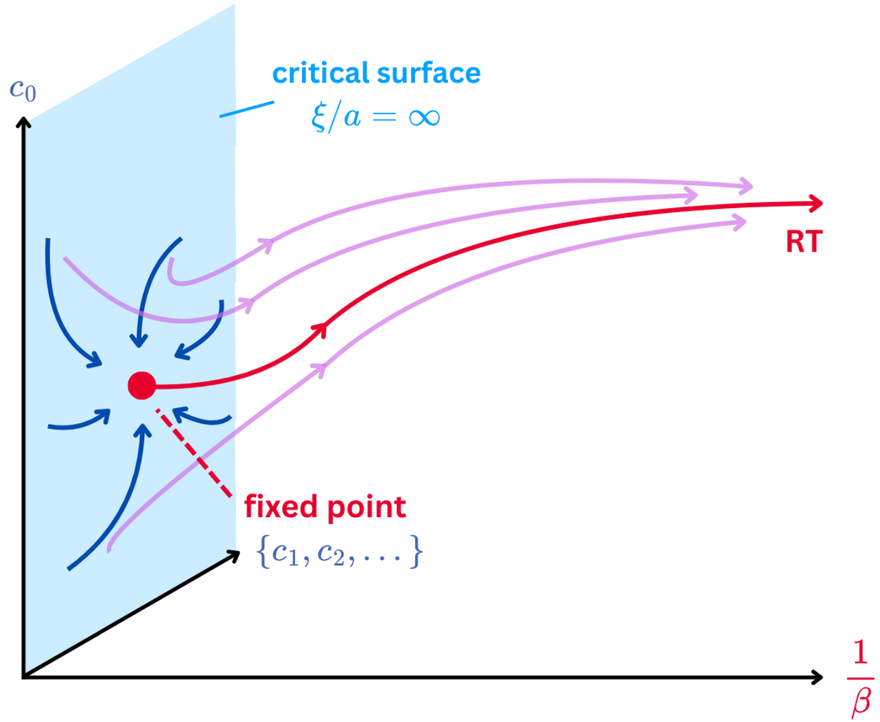


⇒ at $\xi/a \to \infty$, fixed point of RGT: $\{c_\alpha^{\mathrm{FP}}\} \xrightarrow{\text{RGT}} \{c_\alpha^{\mathrm{FP}}\}$

⇒ asymptotically free theories:
FP has one relevant coupling $g \sim 1/\beta$

Renormalization group transformation

The effective action $\beta\mathcal{A}[R]$ is described by infinitely many transformed couplings $\{c_\alpha\}$:



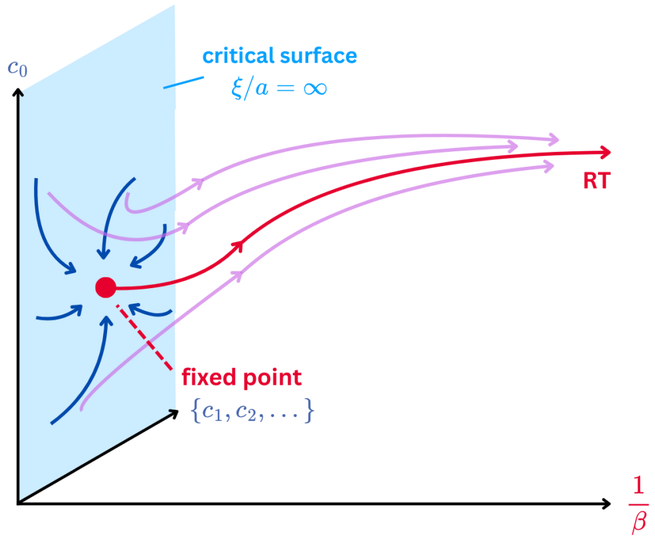
⇒ at $\xi/a \to \infty$, fixed point of RGT: $\{c_\alpha^{\mathrm{FP}}\} \xrightarrow{\text{RGT}} \{c_\alpha^{\mathrm{FP}}\}$

⇒ asymptotically free theories:
FP has one relevant coupling $g \sim 1/\beta$

⇒ renormalized trajectory (RT):
quantum perfect actions, no lattice artefacts

Renormalization group transformation

The effective action $\beta\mathcal{A}[R]$ is described by infinitely many transformed couplings $\{c_\alpha\}$:



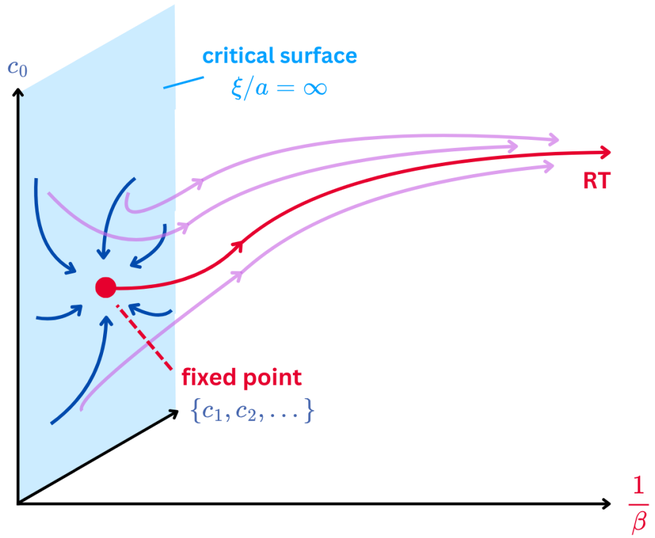
$$\exp\{-\beta'\mathcal{A}'[R]\} = \int \mathcal{D}S \exp\{-\beta(\mathcal{A}[S] + T[R,S])\}$$

Two practical problems:

- how to parametrize RT $\leftrightarrow$ which set $\{c_\alpha\}$?
- how to determine $\{c_\alpha^{\mathrm{RT}}\}$ or $\{c_\alpha^{\mathrm{FP}}\}$?

Renormalization group transformation

The effective action $\beta\mathcal{A}[R]$ is described by infinitely many transformed couplings $\{c_\alpha\}$:



$$\exp\{-\beta'\mathcal{A}'[R]\} = \int \mathcal{D}S \exp\{-\beta(\mathcal{A}[S] + T[R,S])\}$$

Two practical problems:

- how to parametrize RT $\leftrightarrow$ which set $\{c_\alpha\}$?
- how to determine $\{c_\alpha^{\mathrm{RT}}\}$ or $\{c_\alpha^{\mathrm{FP}}\}$?

P.Hasenfratz, F.Niedermayer [hep-lat/9308004]

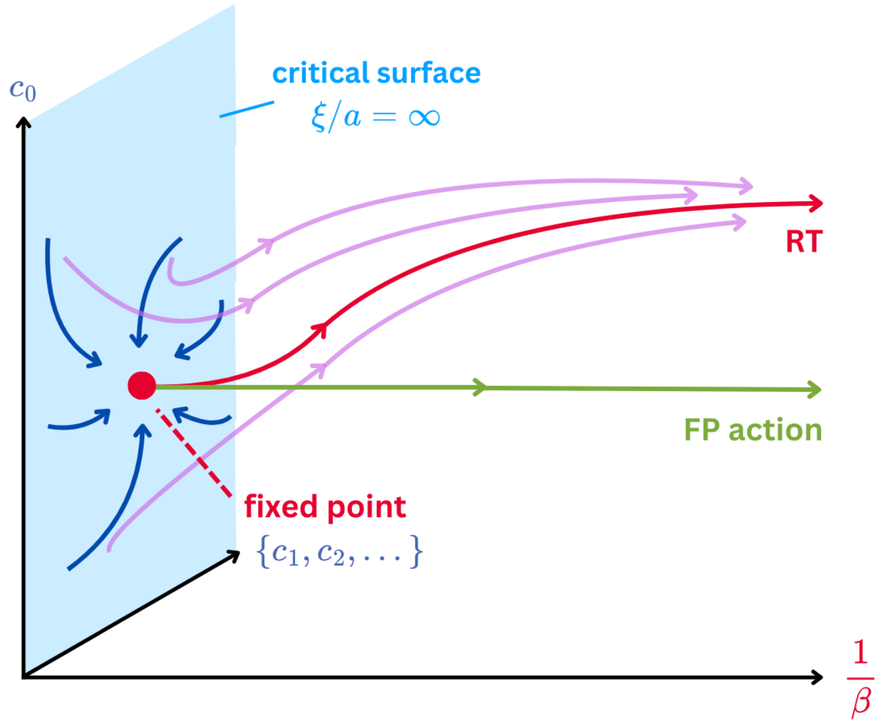
for $\beta \to \infty$ (on critical surface) the RGT becomes a classical saddle point problem:

$$\mathcal{A}^{\mathrm{FP}}[R] = \min_{\{S\}}\{\mathcal{A}^{\mathrm{FP}}[S] + T[R,S]\}$$

Classically perfect FP actions

The classical fixed-point action $\mathcal{A}^{\mathrm{FP}}$ defines an action for all β:

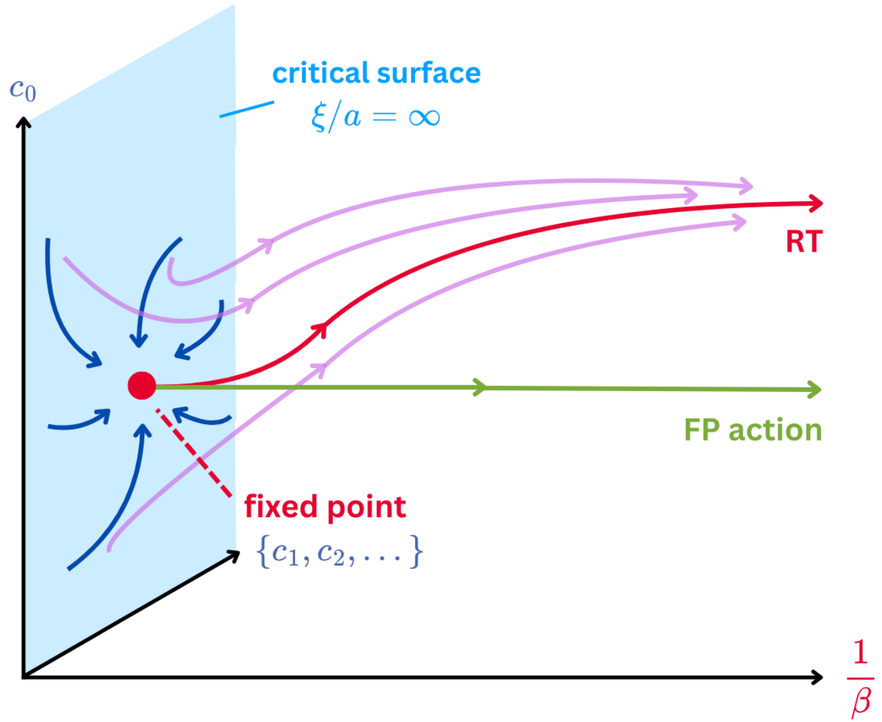
P.Hasenfratz, F.Niedermayer [hep-lat/9308004]



Classically perfect FP actions

The classical fixed-point action $\mathcal{A}^{\mathrm{FP}}$ defines an action for all β:

P.Hasenfratz, F.Niedermayer [hep-lat/9308004]



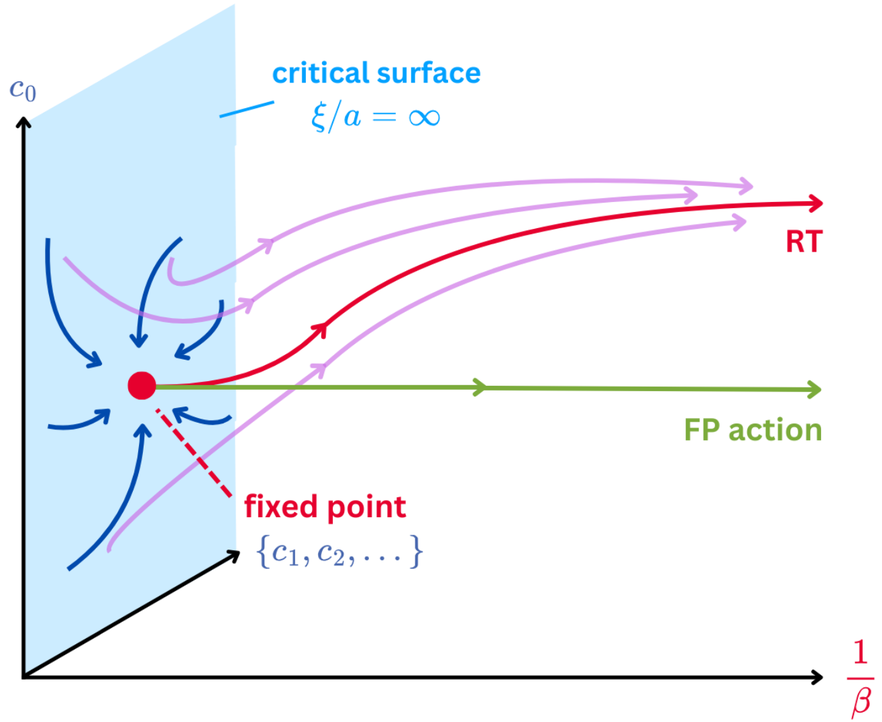
⇒ FP action values for rough configurations through iterations of the FP equation:

$$\mathcal{A}^{\mathrm{FP}}[R] = \min_{\{S\}}\{\mathcal{A}^{\mathrm{FP}}[S] + T[R,S]\}$$
$$= \min_{\{S,S'\}}\{\mathcal{A}^{\mathrm{FP}}[S'] + T[R,S'] + T[R,S]\} = \cdots$$

Classically perfect FP actions

The classical fixed-point action $\mathcal{A}^{\mathrm{FP}}$ defines an action for all β:

P.Hasenfratz, F.Niedermayer [hep-lat/9308004]



⇒ FP action values for rough configurations through iterations of the FP equation:

$$\mathcal{A}^{\mathrm{FP}}[R] = \min_{\{S\}}\{\mathcal{A}^{\mathrm{FP}}[S] + T[R,S]\}$$
$$= \min_{\{S,S'\}}\{\mathcal{A}^{\mathrm{FP}}[S'] + T[R,S'] + T[R,S]\} = \cdots$$

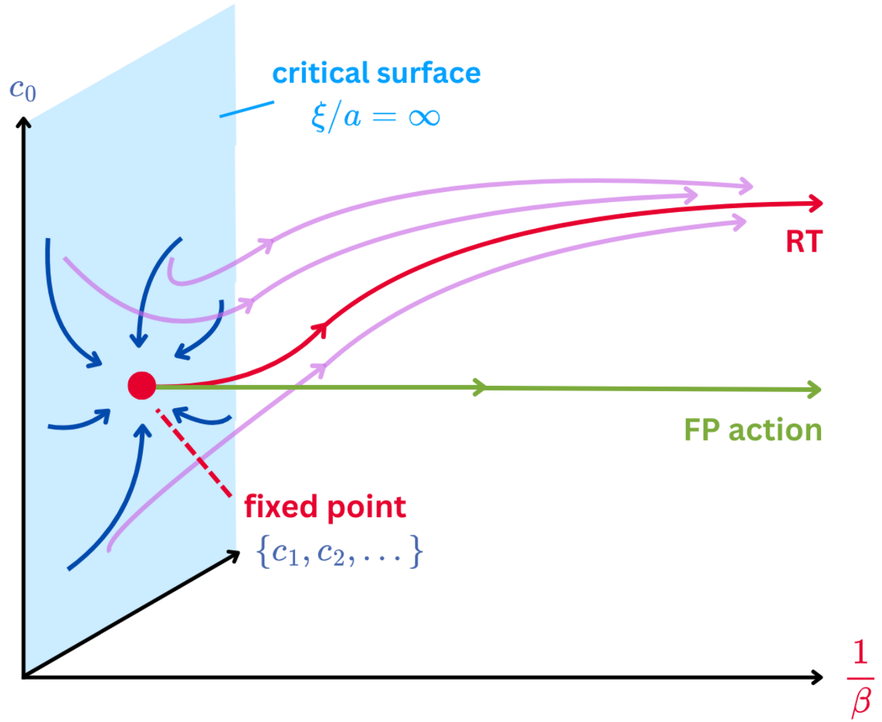
⇒ classically perfect action:
no lattice artefacts on classical configurations

$$\frac{\delta \mathcal{A}^{\mathrm{FP}}[R]}{\delta R} = 0 \;\Rightarrow\; \frac{\delta \mathcal{A}^{\mathrm{FP}}[S]}{\delta S} = 0$$

Classically perfect FP actions

The classical fixed-point action $\mathcal{A}^{\mathrm{FP}}$ defines an action for all β:

P.Hasenfratz, F.Niedermayer [hep-lat/9308004]



⇒ FP action values for rough configurations through iterations of the FP equation:

$$\mathcal{A}^{\mathrm{FP}}[R] = \min_{\{S\}}\{\mathcal{A}^{\mathrm{FP}}[S] + T[R,S]\}$$
$$= \min_{\{S,S'\}}\{\mathcal{A}^{\mathrm{FP}}[S'] + T[R,S'] + T[R,S]\} = \cdots$$

⇒ classically perfect action:
no lattice artefacts on classical configurations

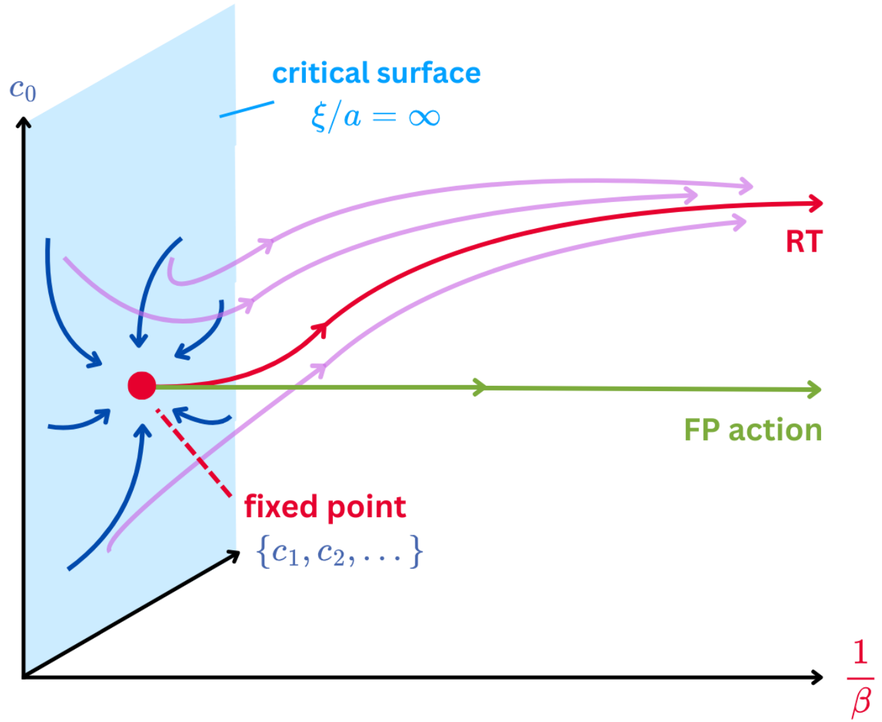
$$\frac{\delta \mathcal{A}^{\mathrm{FP}}[R]}{\delta R} = 0 \;\Rightarrow\; \frac{\delta \mathcal{A}^{\mathrm{FP}}[S]}{\delta S} = 0$$

$$\frac{\delta \mathcal{A}^{\mathrm{FP}}}{\delta R} = \left[\frac{\delta}{\delta S}\{\mathcal{A}^{\mathrm{FP}}[S] + T[R,S]\}\frac{\delta S}{\delta R} + \frac{\delta T[R,S]}{\delta R}\right]_{S_{\min}} = 0 \Rightarrow \left.\frac{\delta T[R,S]}{\delta R}\right|_{S_{\min}} = 0 \Rightarrow T[R,S_{\min}] = 0$$

Classically perfect FP actions

The classical fixed-point action $\mathcal{A}^{\mathrm{FP}}$ defines an action for all β:

P.Hasenfratz, F.Niedermayer [hep-lat/9308004]



⇒ FP action values for rough configurations through iterations of the FP equation:

$$\mathcal{A}^{\mathrm{FP}}[R] = \min_{\{S\}}\{\mathcal{A}^{\mathrm{FP}}[S] + T[R,S]\}$$
$$= \min_{\{S,S'\}}\{\mathcal{A}^{\mathrm{FP}}[S'] + T[R,S'] + T[R,S]\} = \cdots$$

⇒ classically perfect action:
no lattice artefacts on classical configurations

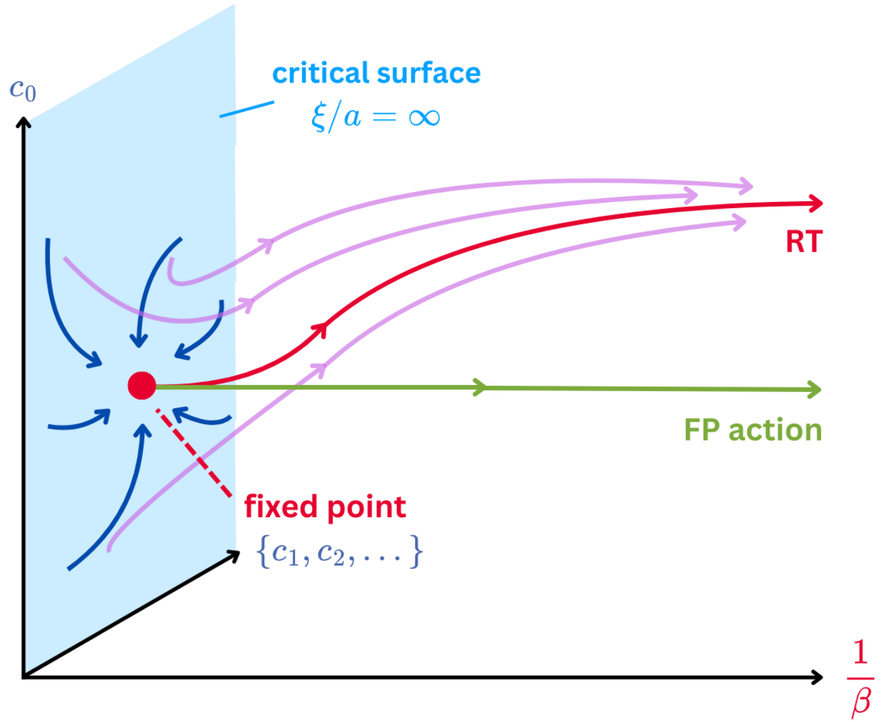
$$\frac{\delta\mathcal{A}^{\mathrm{FP}}[R]}{\delta R} = 0 \;\Rightarrow\; \frac{\delta\mathcal{A}^{\mathrm{FP}}[S]}{\delta S} = 0$$

⇒ $\mathcal{A}^{\mathrm{FP}}$ has scale invariant instanton solutions
$(\mathcal{A}^{\mathrm{FP}}[R] = \mathcal{A}^{\mathrm{FP}}[S_{\mathrm{min}}]$ and $\delta\mathcal{A}^{\mathrm{FP}}[S]/\delta S|_{S_{\mathrm{min}}}=0)$

Classically perfect FP actions

The classical fixed-point action $\mathcal{A}^{\mathrm{FP}}$ defines an action for all β:

P.Hasenfratz, F.Niedermayer [hep-lat/9308004]



⇒ FP action values for rough configurations through iterations of the FP equation:

$$\mathcal{A}^{\mathrm{FP}}[R] = \min_{\{S\}}\{\mathcal{A}^{\mathrm{FP}}[S] + T[R,S]\}$$
$$= \min_{\{S,S'\}}\{\mathcal{A}^{\mathrm{FP}}[S'] + T[R,S'] + T[R,S]\} = \cdots$$

⇒ classically perfect action:
no lattice artefacts on classical configurations

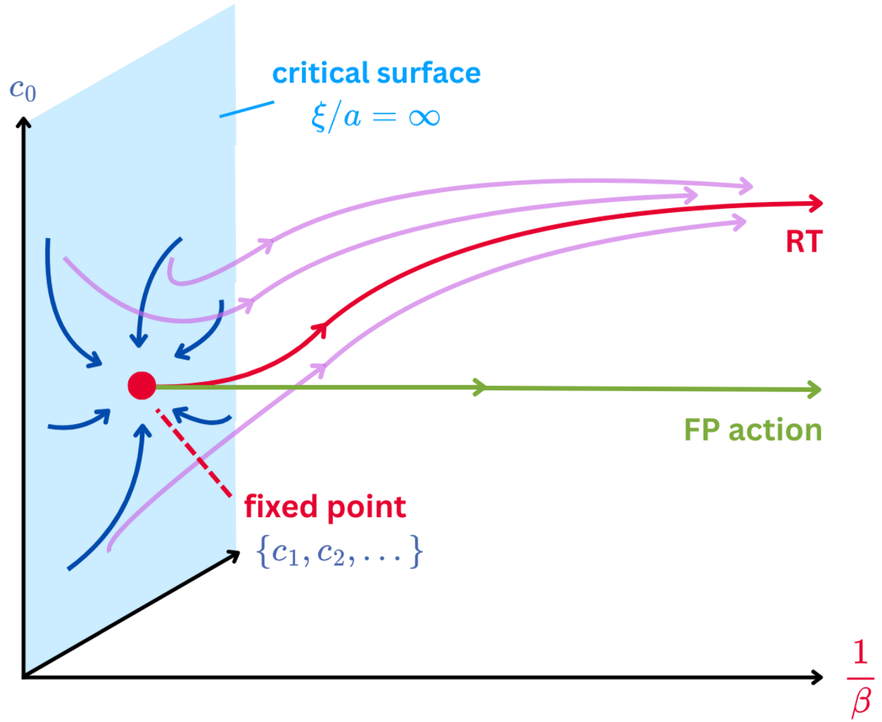
⇒ $\mathcal{A}^{\mathrm{FP}}$ has scale invariant instanton solutions

⇒ for large β, $\mathcal{A}^{\mathrm{FP}}[R]$ is very close to $\mathcal{A}^{\mathrm{RT}}[R]$

Classically perfect FP actions

The classical fixed-point action $\mathcal{A}^{\mathrm{FP}}$ defines an action for all β:

P.Hasenfratz, F.Niedermayer [hep-lat/9308004]



⇒ FP action values for rough configurations through iterations of the FP equation:

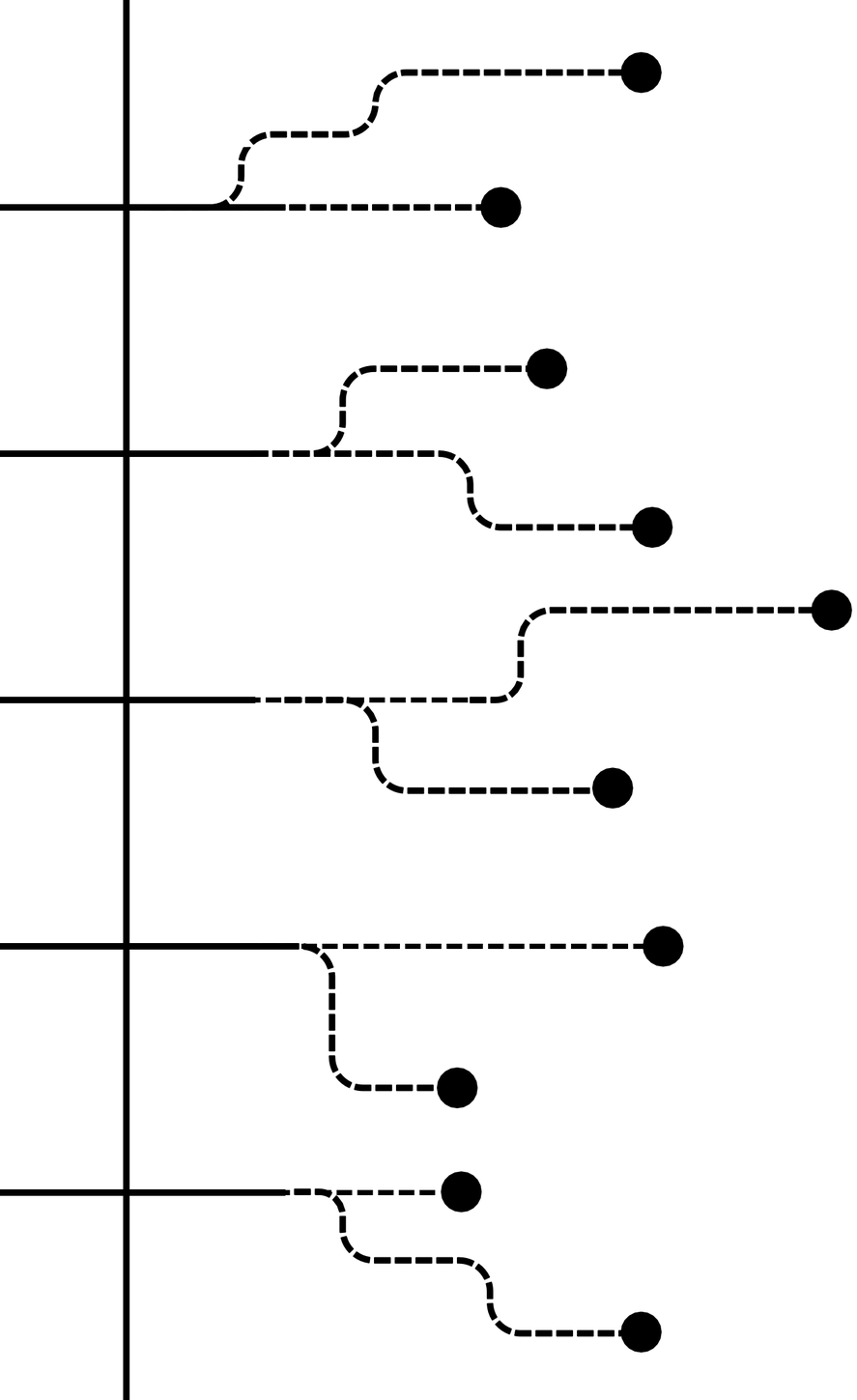
$$\mathcal{A}^{\mathrm{FP}}[R] = \min_{\{S\}}\{\mathcal{A}^{\mathrm{FP}}[S] + T[R,S]\}$$
$$= \min_{\{S,S'\}}\{\mathcal{A}^{\mathrm{FP}}[S'] + T[R,S'] + T[R,S]\} = \cdots$$

⇒ classically perfect action:
no lattice artefacts on classical configurations

⇒ $\mathcal{A}^{\mathrm{FP}}$ has scale invariant instanton solutions

⇒ for large β, $\mathcal{A}^{\mathrm{FP}}[R]$ is very close to $\mathcal{A}^{\mathrm{RT}}[R]$

⇒ $\mathcal{A}^{\mathrm{FP}}$ can be made local through parameters of RGT



Part II: Machine learning the FP action

FP data

For given coarse ensembles $\{R\}$, FP action values through iterations of the FP equation:

$$\mathcal{A}^{\mathrm{FP}}[R] = \min_{\{S\}}\{\mathcal{A}^{\mathrm{FP}}[S] + T[R,S]\} = \min_{\{S,S'\}}\{\mathcal{A}^{\mathrm{FP}}[S'] + T[R,S'] + T[R,S]\} = \cdots$$

FP data

For given coarse ensembles $\{R\}$, FP action values through iterations of the FP equation:

$$\mathcal{A}^{\mathrm{FP}}[R] = \min_{\{S\}}\{\mathcal{A}^{\mathrm{FP}}[S] + T[R,S]\} = \min_{\{S,S'\}}\{\mathcal{A}^{\mathrm{FP}}[S'] + T[R,S'] + T[R,S]\} = \cdots$$

Alternatively:
use approximate FP action $\mathcal{A}^{\mathrm{FP}}[S]$ in first iteration

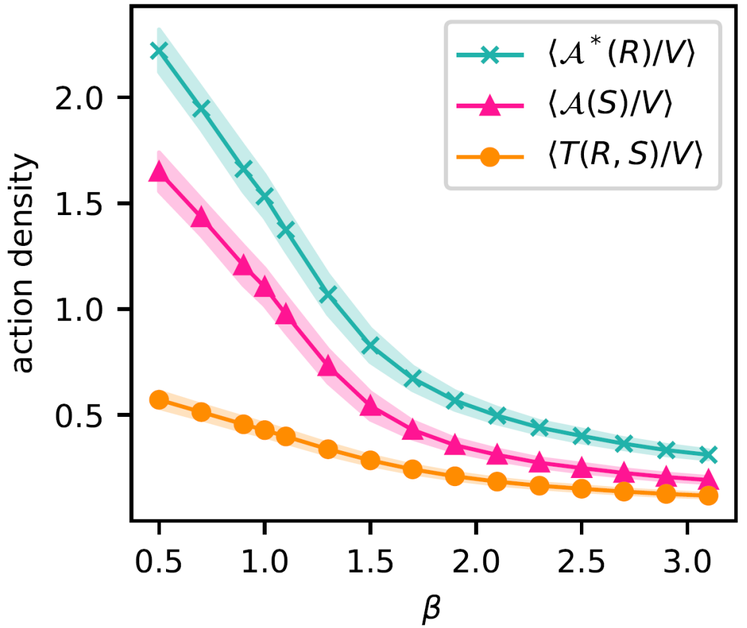
Minimizing RHS crucial for generating the training data:

⇒ generate ensembles of coarse spin configurations for range of field fluctuations

⇒ find minimizing fine configurations

⇒ use $S_{\min}$ in FP equation to obtain:
FP action values $\mathcal{A}^{\mathrm{FP}}[R]$ and
FP action derivatives $\frac{\delta \mathcal{A}^{\mathrm{FP}}[R]}{\delta R_x} = \frac{\delta T\,[R,S]}{\delta R_x}\Big|_{S_{\min}}$

} data set for supervised ML



Network architecture

Similar to M.Favoni, A.Ipp, D.Müller, D.Schuh [2012.12901]:

Convolutional neural networks as universal function approximators

Encode in architecture symmetries of the theory:

a) translation invariance ⇒ convolution-based layers

b) global O(3) invariance ⇒ consider scalar products $\langle R_x, R_y \rangle$

Network architecture

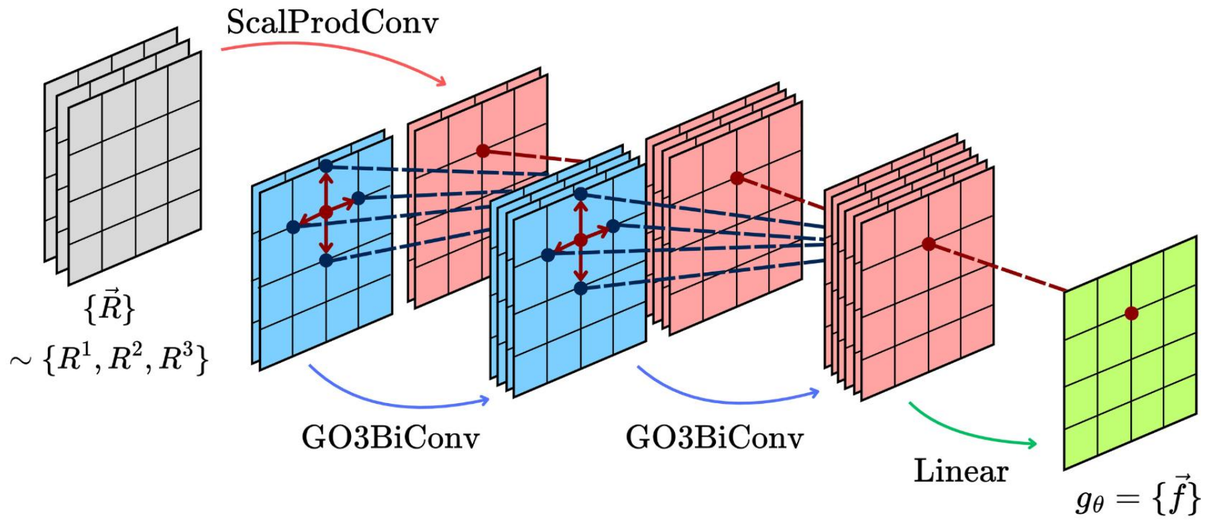
Encode symmetries in architecture:

a) translation invariance

⇒ convolution-based layers

b) global O(3) invariance

⇒ propagate scalar products $\langle R_x, R_y \rangle$



Network couples local spin products to build up higher order interactions $\sim \left(1 - R_x \cdot R_y\right)^n$

Network architecture

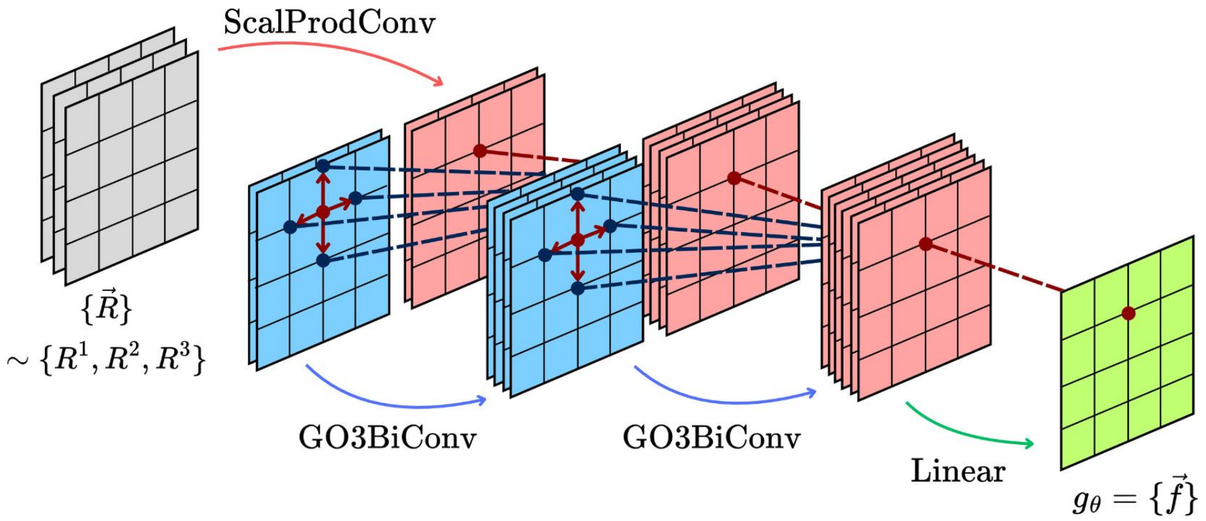
Encode symmetries in architecture:

a) translation invariance

⇒ convolution-based layers

b) global O(3) invariance

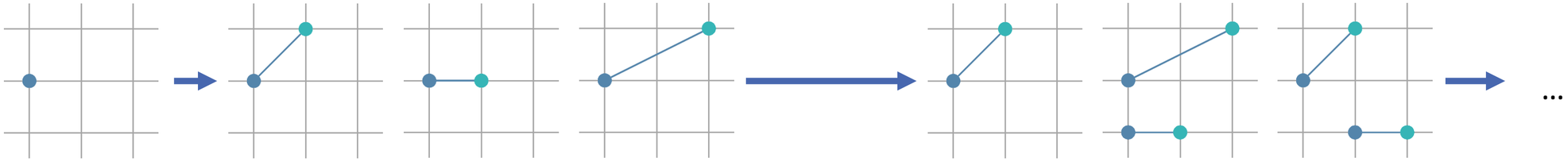
⇒ propagate scalar products $\langle R_x, R_y \rangle$



Network couples local spin products to build up higher order interactions $\sim \left(1 - R_x \cdot R_y\right)^n$

ScalProdConv: $f_x^a = \sum_r \rho_r^a (1 - R_x \cdot R_{x+r}) + \delta^a$

GO3BiConv: $\tilde{f}_x^a = \delta^a + \sum_{b,r} \gamma_r^{ab} f_{n+r}^b + \sum_{b,c,r} \omega_r^{abc} f_n^b f_{n+r}^c$



Network architecture

Physical ansatz: neural network as correction to some lattice action to ensure correct continuum limit

$$\mathcal{A}^{\mathrm{NN}}[R] = \sum_x \mathcal{A}_x^{\mathrm{pre}}[R] \sum_{n=0}^{\infty} b^{(n)} (g_x[R] - g_x[1])^n$$

Network architecture

Physical ansatz: neural network as correction to some lattice action to ensure correct continuum limit

$$\mathcal{A}^{\mathrm{NN}}[R] = \sum_x \mathcal{A}_x^{\mathrm{pre}}[R] \sum_{n=0}^{\infty} b^{(n)} (g_x[R] - g_x[1])^n$$

For local prefactor action use, e.g., existing FP parametrization

g_x is local output of a neural network, difference ensures normalization towards the continuum

Choose coefficients, e.g., as $b^{(n)} = \frac{1}{n!}$ $\Rightarrow$ final ansatz:

$$\mathcal{A}^{\mathrm{NN}}[R] = \sum_x \mathcal{A}_x^{\mathrm{pre}}[R] \exp(g_x[R] - g_x[1])$$

Loss function

ML loss function has two weighted contributions:

$$\mathcal{L}_1 = \frac{1}{L^2 N_{\text{cfg}}} \sum_{i=1}^{N_{\text{cfg}}} \left| \mathcal{A}^{\text{FP}}[R_i] - \mathcal{A}^{NN}[R_i] \right|$$

$$\mathcal{L}_2 = \frac{1}{L^2 N_{\text{cfg}}} \sum_{i=1}^{N_{\text{cfg}}} \sum_x \left\| D_x \mathcal{A}^{\text{FP}}[R_i] - D_x \mathcal{A}^{NN}[R_i] \right\|^2$$

$$D_x \mathcal{A}[R] = R_{x,i} \times (\nabla_x \mathcal{A}[R])$$

$$\mathcal{L} = w_1 \mathcal{L}_1 + w_2 \mathcal{L}_2$$

Derivatives of the neural network wrt. input are computed exact through backpropagation

Architecture search

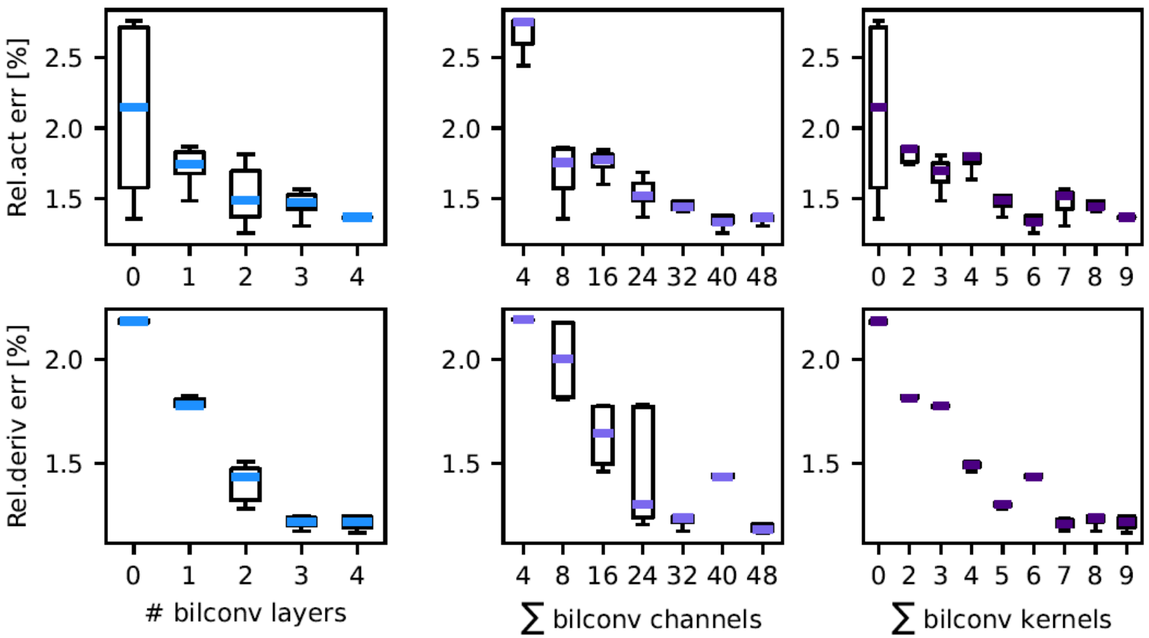
Compare performance for different hyperparameters

of the model (width, depth, „interaction range“)

⇒ final architecture:

1 ScalProdConv layer with
max. interaction range 3
3 GO3BiConv layer
with out-channels {8,12,12},
kernel sizes {3,2,2}
1 linear layer

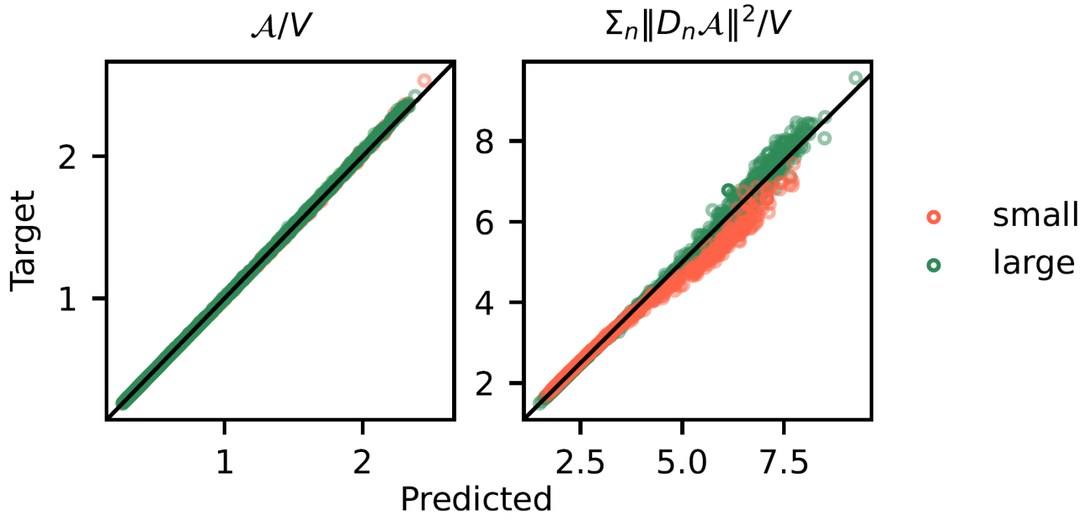
with 52k parameters



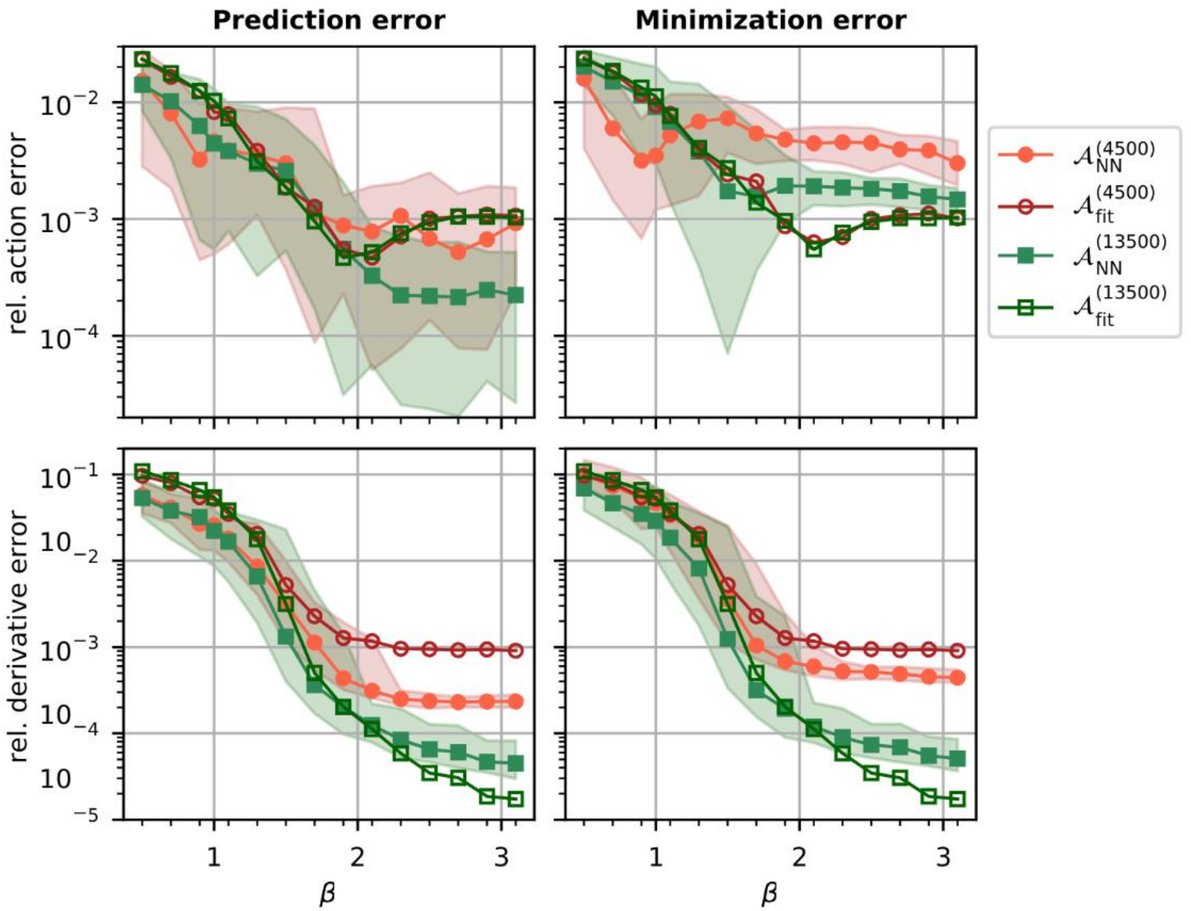
Training results

Two models, from smaller and larger data sets

Compare to existing parametrization $\mathcal{A}_{\mathrm{fit}}^{\mathrm{FP}}$



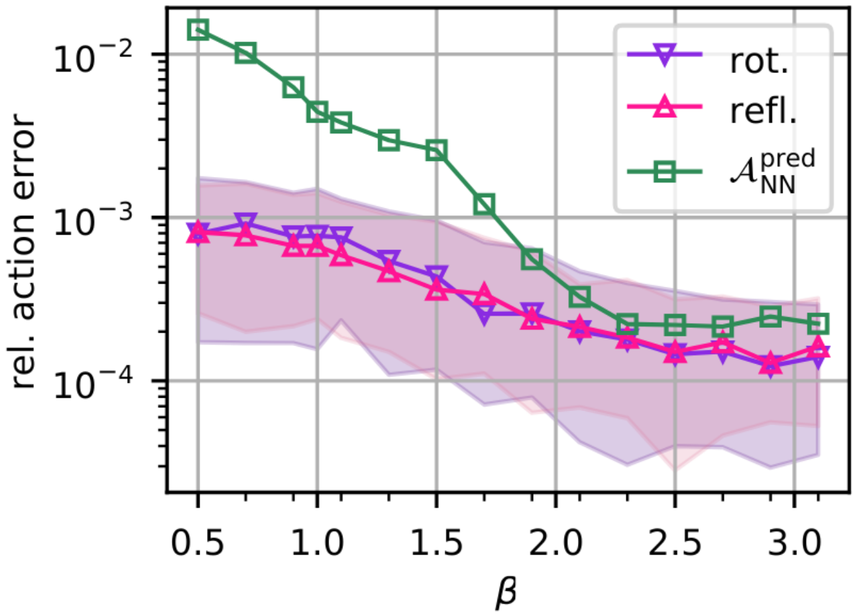
How well is FP equation fulfilled?



Training results

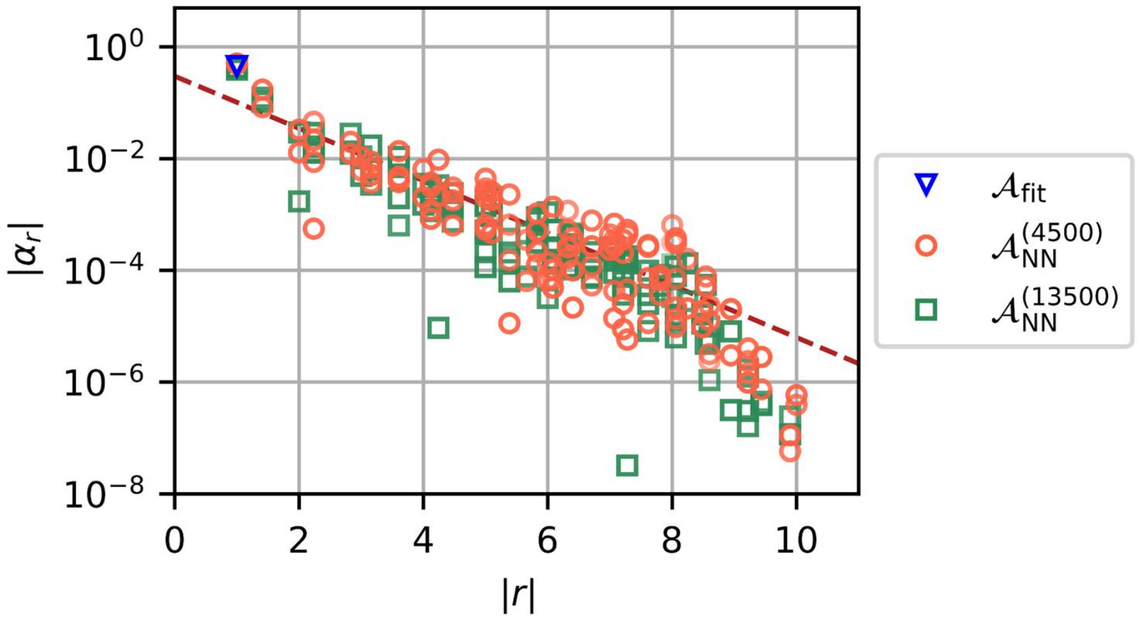
What about…

… rotational symmetry?



two-spin interaction coupling

… locality? ($\alpha_r(r) \propto \rho(r)$)



FP actions from ML

Checkpoint:

- ✓ Can the FP action be parametrized sufficiently well?
- ✓ Is the FP action sufficiently local?

Can model accuracy be further improved by changing approach to learning?

After supervised training ⇒ unsupervised training

Unsupervised training

Consider the FP equation:

$$\mathcal{A}^{\mathrm{FP}}[R] = \min_{\{S\}}\{\mathcal{A}^{\mathrm{FP}}[S] + T[R,S]\}$$

Solve iteratively, starting from final supervised $\mathcal{A}^{\mathrm{NN}}[R_i]$ and pre-minimized $S_i \approx S_{i,\mathrm{min}}$:

1) assume that $\mathcal{A}^{\mathrm{FP}}[R_i] \approx \mathcal{A}^{\mathrm{NN}}[S_i] + T[R_i, S_i]$
2) assume that $\mathcal{D}_{x,R}\mathcal{A}^{\mathrm{FP}}[R_i] \approx \mathcal{D}_{x,R}T[R_i, S_i]$ ⟵ $D_{x,R}\mathcal{A}[R] = R_{x,i} \times (\nabla_x \mathcal{A}[R])$
3) assume that $\mathcal{D}_{x,S}\left(\mathcal{A}^{\mathrm{NN}}[S_i] + T[R_i, S_i]\right) = 0$ (holds if fine configuration minimizes expression)

⇒ update S and NN simultaneously

Unsupervised training

1) assume that $\mathcal{A}^{\mathrm{FP}}[R_i] \approx \mathcal{A}^{\mathrm{NN}}[S_i] + T[R_i, S_i]$
2) assume that $\mathcal{D}_{x,R}\mathcal{A}^{\mathrm{FP}}[R_i] \approx \mathcal{D}_{x,R}T[R_i, S_i]$
3) assume that $\mathcal{D}_{x,S}\left(\mathcal{A}^{\mathrm{NN}}[S_i] + T[R_i, S_i]\right) = 0$

$$\mathcal{A}^{\mathrm{FP}}[R] = \min_{\{S\}}\{\mathcal{A}^{\mathrm{FP}}[S] + T[R, S]\}$$

Update loss function: (θ-index represents model parameters)

$$\mathcal{L}_1(S;\theta) = \frac{1}{L^2 N_{\mathrm{cfg}}} \sum_{i=1}^{N_{\mathrm{cfg}}} \left|\mathcal{A}_\theta^{\mathrm{NN}}[R_i] - \left(\mathcal{A}_\theta^{\mathrm{NN}}[S_i] + T[R_i, S_i]\right)\right|$$

$$\mathcal{L}_2(S;\theta) = \frac{1}{L^2 N_{\mathrm{cfg}}} \sum_{i=1}^{N_{\mathrm{cfg}}} \sum_x \left\|\mathcal{D}_{x,S}\left(\mathcal{A}_\theta^{\mathrm{NN}}[S_i] + T[R_i, S_i]\right)\right\|^2$$

$$\mathcal{L}_3(S;\theta) = \frac{1}{L^2 N_{\mathrm{cfg}}} \sum_{i=1}^{N_{\mathrm{cfg}}} \sum_x \left\|\mathcal{D}_{x,R}\mathcal{A}_\theta^{\mathrm{NN}}[R_i] - \mathcal{D}_{x,R}T[R_i, S_i]\right\|^2$$

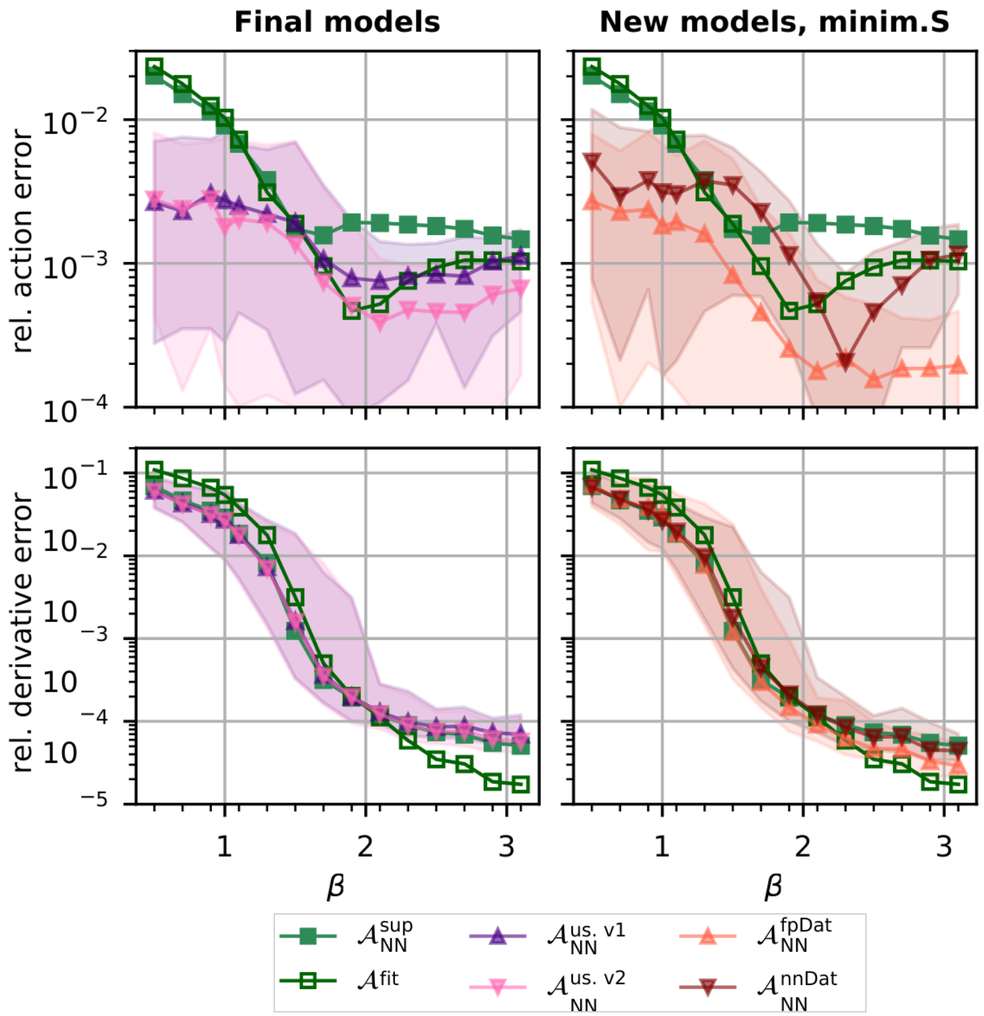
$$\mathcal{L}(S;\theta) = w_1\mathcal{L}_1 + w_2\mathcal{L}_2 + w_3\mathcal{L}_3$$

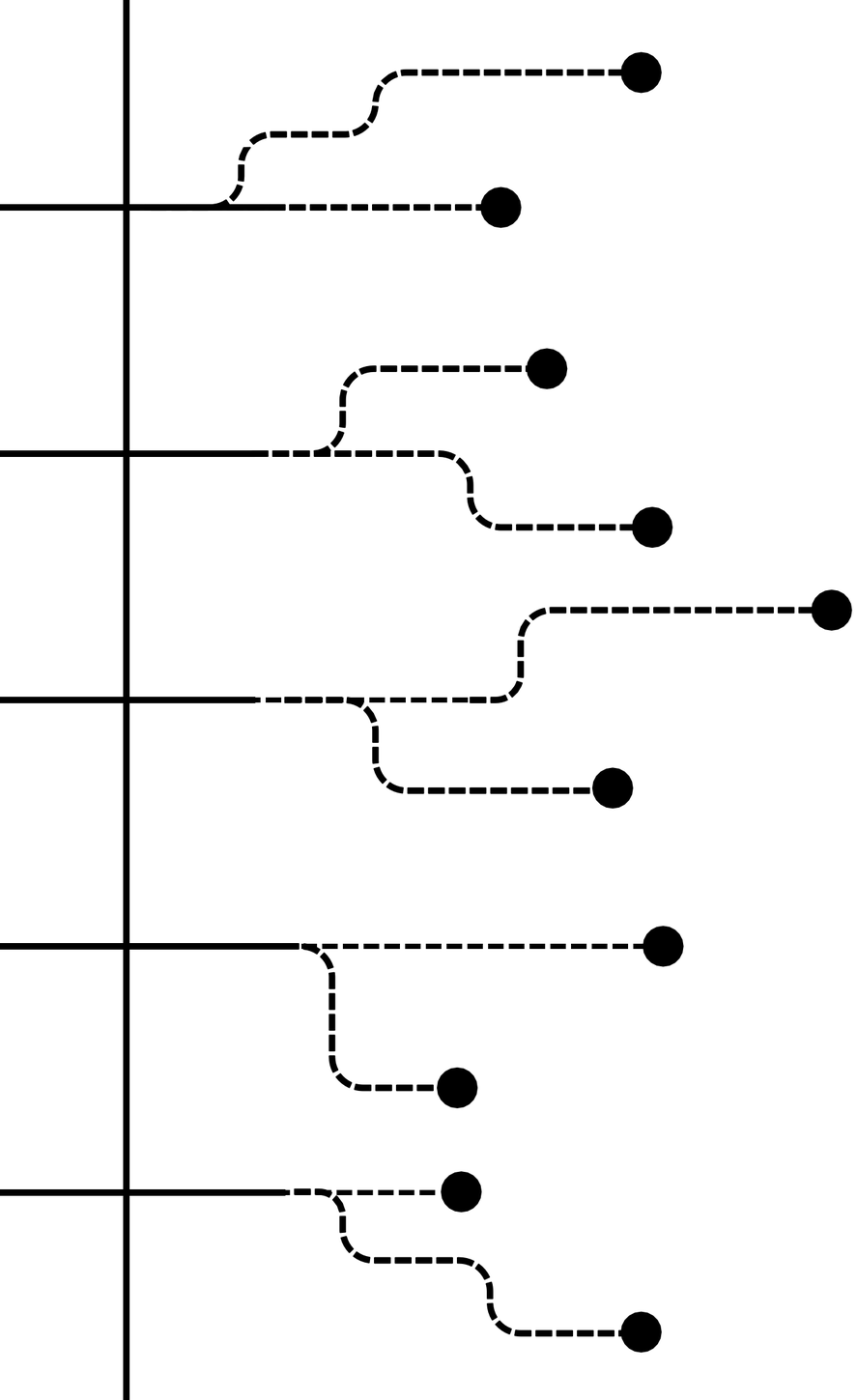
Unsupervised training results

Different unsupervised NNs:

- continue training from supervised NNs
- initialize NN randomly, start with pre-minimized $S_{i,\min}$

⇒ improvement for action value

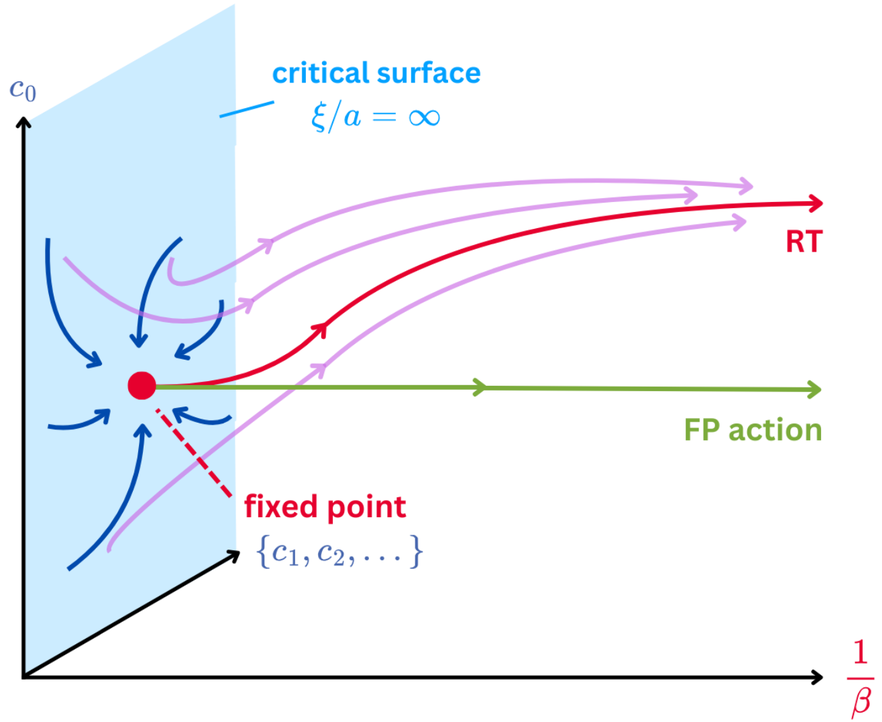




Part III:
Quantum perfect action

Quantum perfect actions

Apply RGT to FP action to reconstruct renormalized trajectory:

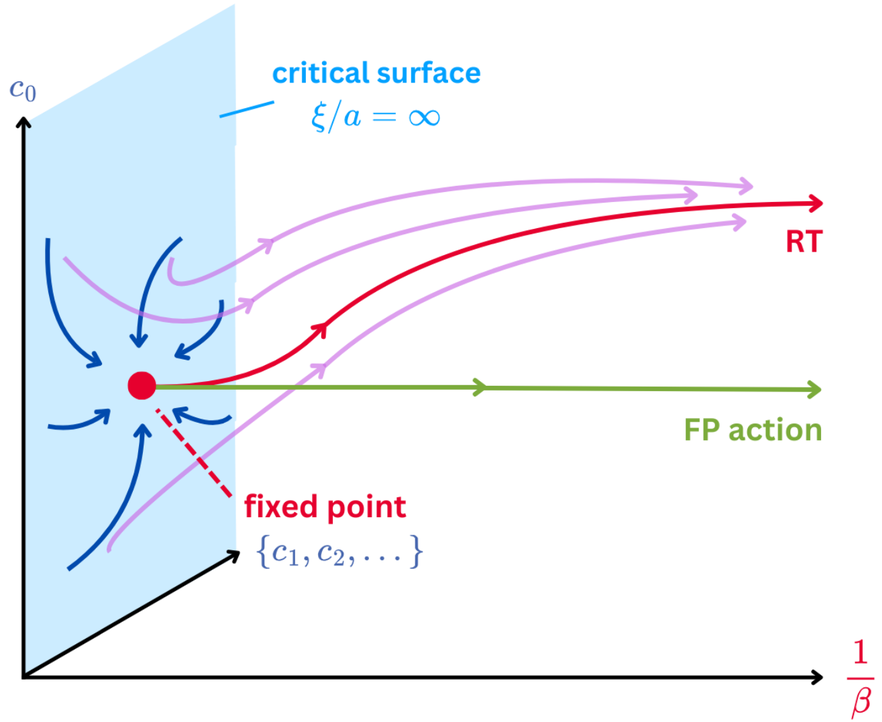


$$\exp\{-\beta' \mathcal{A}^{\mathrm{RT}}[R]\} = \int \mathcal{D}S \exp\left\{-\beta^{\mathrm{FP}}\left(\mathcal{A}^{\mathrm{FP}}[S] + T[R,S]\right)\right\}$$

$\Rightarrow$ for sufficiently large β: $\{c_\alpha^{\mathrm{FP}}\} \xrightarrow{\mathrm{RGT}} \{c_\alpha^{\mathrm{RT}}\}$

Quantum perfect actions

From FP action, reconstruct renormalized trajectory:



$$\exp\{-\beta' \mathcal{A}^{\mathrm{RT}}[R]\} = \int \mathcal{D}S \exp\left\{-\beta^{\mathrm{FP}}\left(\mathcal{A}^{\mathrm{FP}}[S] + T[R,S]\right)\right\}$$

$\Rightarrow$ for sufficiently large β: $\{c_\alpha^{\mathrm{FP}}\} \xrightarrow{\mathrm{RGT}} \{c_\alpha^{\mathrm{RT}}\}$

$\Rightarrow$ iterate: 1) apply RGT to (ML) action

2) parametrize RT action through ML

Quantum perfect actions

Data from RTG: ⇒ no explicit information on action values $\mathcal{A}^{\mathrm{RT}}[R]$

⇒ BUT: information on $\delta\mathcal{A}^{\mathrm{RT}}[R]/\delta R$

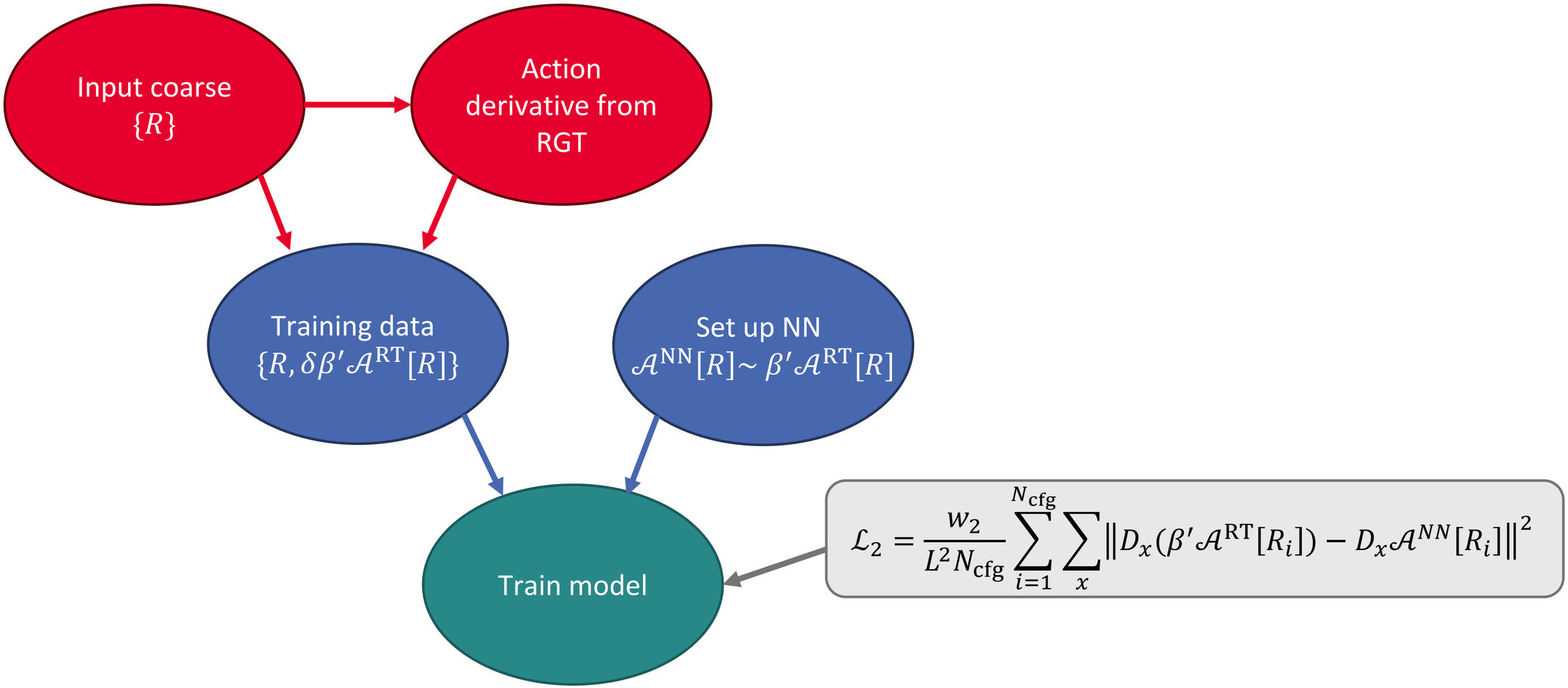
Step I:
$$\exp\{-\beta'(\mathcal{A}^{\mathrm{RT}}[R^{\delta}]-\mathcal{A}^{\mathrm{RT}}[R])\}=\frac{\int \mathcal{D}S\,\exp\{-\beta^{\mathrm{FP}}(\mathcal{A}^{\mathrm{FP}}[S]+T[R^{\delta},S]+T[R,S]-T[R,S])\}}{\int \mathcal{D}S\,\exp\{-\beta^{\mathrm{FP}}(\mathcal{A}^{\mathrm{FP}}[S]+T[R,S])\}}$$

$$=\langle\exp\{-\beta^{\mathrm{FP}}(T[R^{\delta},S]-T[R,S])\}\rangle_{RGT}$$

Step II: let $R^{\delta}=R+\delta R$ for infinitesimal δR and expand both sides of the RGT to obtain

$$\frac{\delta(\beta'\mathcal{A}^{\mathrm{RT}}[R])}{\delta R}=\left\langle\frac{\delta(\beta^{\mathrm{FP}}T[R,S])}{\delta R}\right\rangle_{RGT}$$

Machine learning an RT action



Outlook and Summary

- Discussed approaches to the machine learning of RG-improved actions for toy model
- Fixed-point action:
  - Supervised and unsupervised training
  - Sufficiently good parametrization
- Quantum perfect action
  - proof-of-concept study, ongoing
- Next steps
  - towards a machine learned quantum perfect action for SU(3) gauge theory

Thank you!

FP data - minimization

Minimization evolution uses global updates

For every fixed R, across a range of coarser and smoother ensembles, obtain corresponding $S_{\min}$

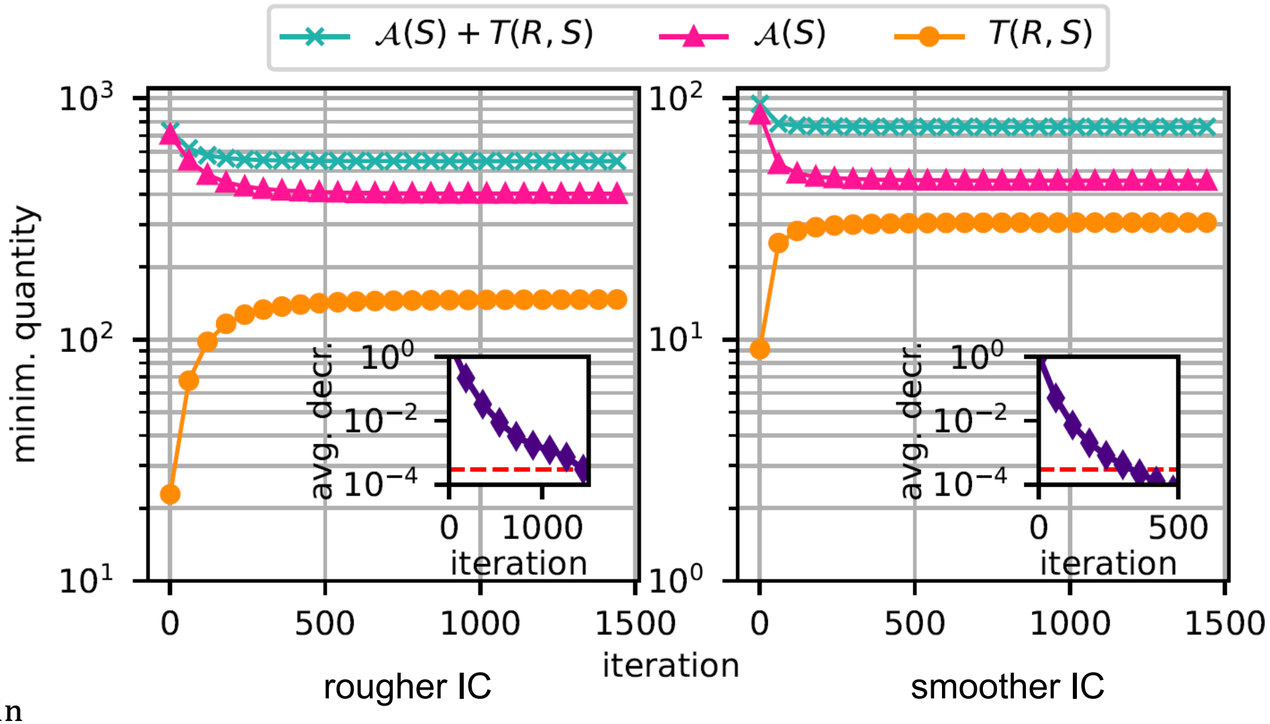
From minimizing configurations:

- FP action value:

  $$\mathcal{A}^{\mathrm{FP}}[R] = \{\mathcal{A}^{\mathrm{FP}}[S] + T[R,S]\}_{S_{\min}}$$

- FP action derivatives from FP equation:

  $$\frac{\delta \mathcal{A}^{\mathrm{FP}}[R]}{\delta R_x} = \frac{\delta T\,[R,S]}{\delta R_x}\Big|_{S_{\min}} = [-\kappa \textstyle\sum_{n\in x} S_n]_{S_{\min}}$$

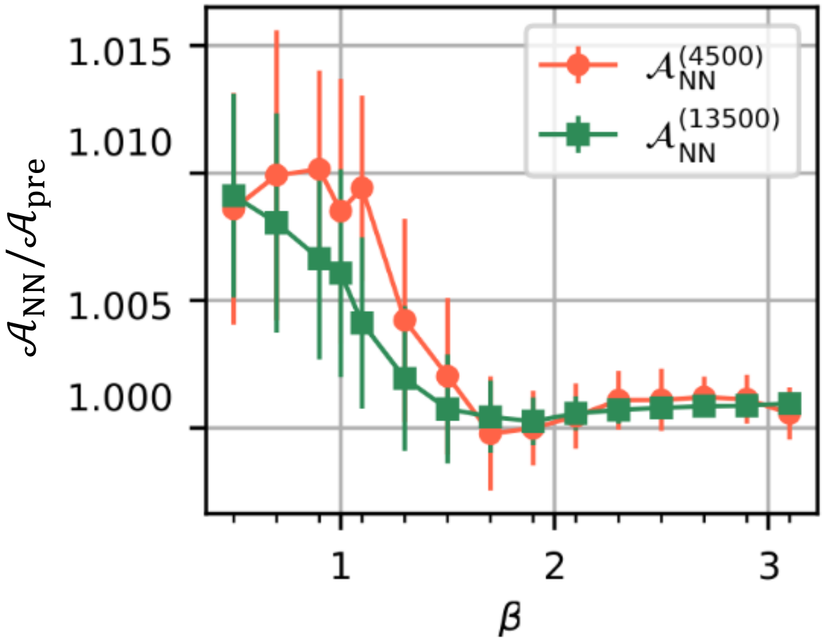


⇒ {FP action values, FP action derivatives} form data set for supervised ML

Training results – NN as correction

Our ansatz: NN as correction to some prefactor lattice action

Question: How strong is correction?



RT action derivatives

From RTG, derive training data for ML ⇒ no explicit information on action values $\mathcal{A}^{\mathrm{RT}}[R]$

⇒ BUT: information on $\delta\mathcal{A}^{\mathrm{RT}}[R]/\delta R$

Step I: RGT $\exp\{-\beta'(\mathcal{A}^{\mathrm{RT}}[R^{\delta}] - \mathcal{A}^{\mathrm{RT}}[R])\} = \frac{\int \mathcal{D}S \exp\{-\beta^{\mathrm{FP}}(\mathcal{A}^{\mathrm{FP}}[S]+T[R^{\delta},S])\}}{\int \mathcal{D}S \exp\{-\beta^{\mathrm{FP}}(\mathcal{A}^{\mathrm{FP}}[S]+T[R,S])\}}$

$$= \frac{\int \mathcal{D}S \exp\{-\beta^{\mathrm{FP}}(\mathcal{A}^{\mathrm{FP}}[S]+T[R^{\delta},S]+T[R,S]-T[R,S])\}}{\int \mathcal{D}S \exp\{-\beta^{\mathrm{FP}}(\mathcal{A}^{\mathrm{FP}}[S]+T[R,S])\}}$$

$$:= \frac{1}{Z_R}\int \mathcal{D}S \exp\{-\beta^{\mathrm{FP}}(\mathcal{A}^{\mathrm{FP}}[S] + T[R,S])\}\exp\{-\beta^{\mathrm{FP}}(T[R^{\delta},S] - T[R,S])\}$$

$$= \langle \exp\{-\beta^{\mathrm{FP}}(T[R^{\delta},S] - T[R,S])\}\rangle_R$$

Step II: let $R^{\delta} = R + \delta R$ for infinitesimal δR and expand both sides of the RGT to obtain

$$\frac{\delta\beta'\mathcal{A}^{\mathrm{RT}}[R]}{\delta R} = \left\langle \frac{\delta\beta^{\mathrm{FP}}T[R,S]}{\delta R} \right\rangle_R$$