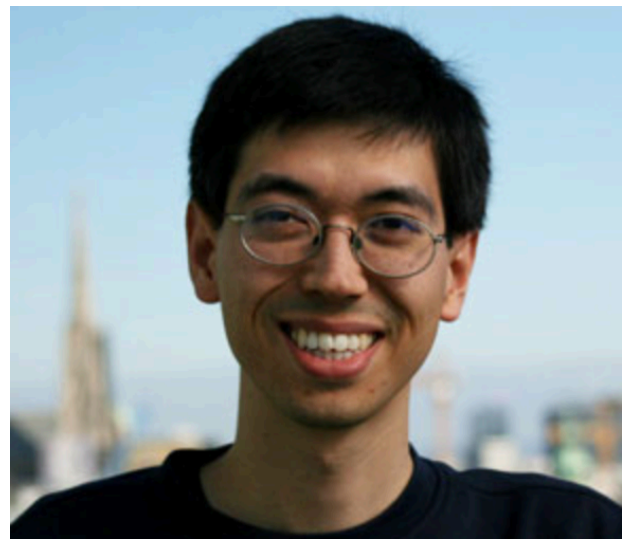
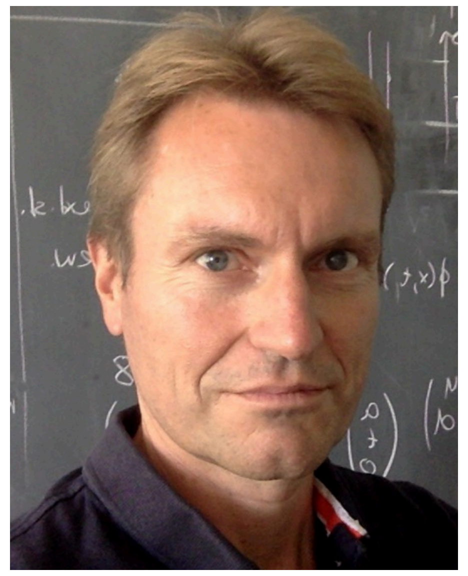
Machine learning RG actions

with a lattice gauge covariant convolutional neural network (L-CNN)





Kieran Holland (University of Pacific), Andreas Ipp and David Müller (TU Wien), Urs Wenger (University of Bern)

arXiv:2311.17816 [hep-lat], arXiv:2401.06481 [hep-lat]

Workshop on Machine Learning and the Renormalization Group, 29 May 2024 - ECT* Trento, Italy

Introduction

Consider an asymptotically free quantum field theory on the lattice, e.g., SU(N_c) lattice gauge theory:

$$Z(\beta) = \int \mathcal{D}U \exp\{-\beta A[U]\} \quad \text{with gauge coupling} \quad \beta = \frac{2N_c}{g^2}$$

and expectation values for observables:

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

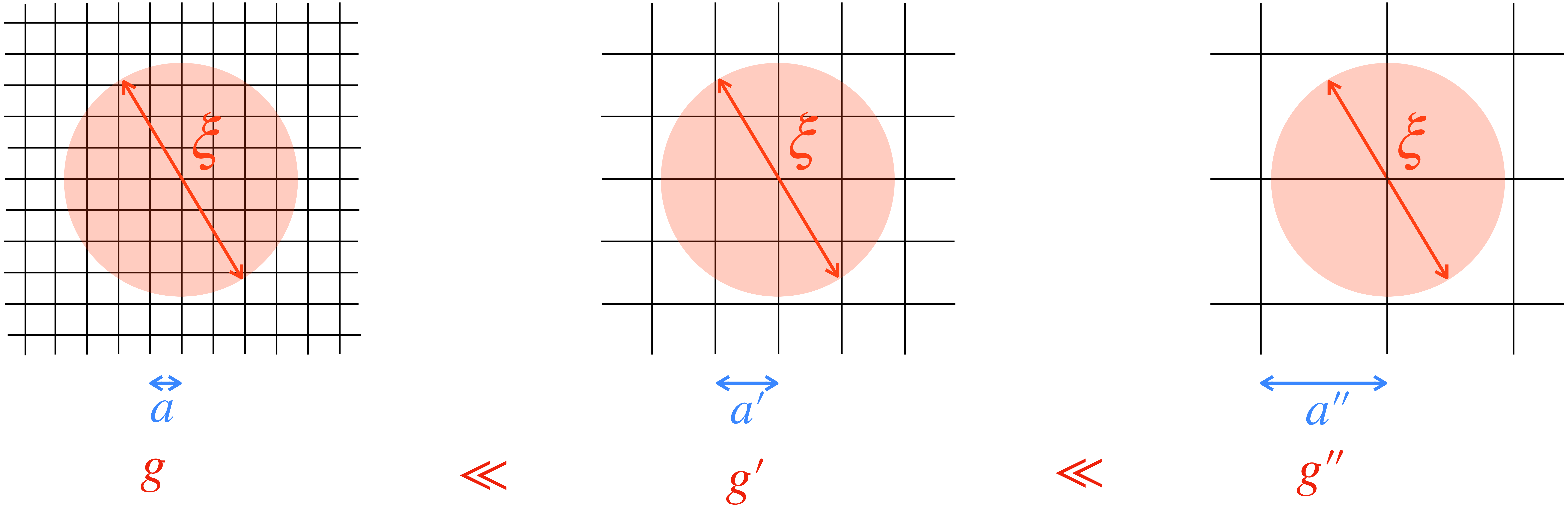
with a characteristic length scale ξ in units of the lattice spacing a:

$$\frac{\xi}{a} \quad \Rightarrow \quad \text{dimensionless}$$

Introduction

The lattice spacing a is determined by the gauge coupling: $\beta = \frac{2N_c}{g^2}$

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

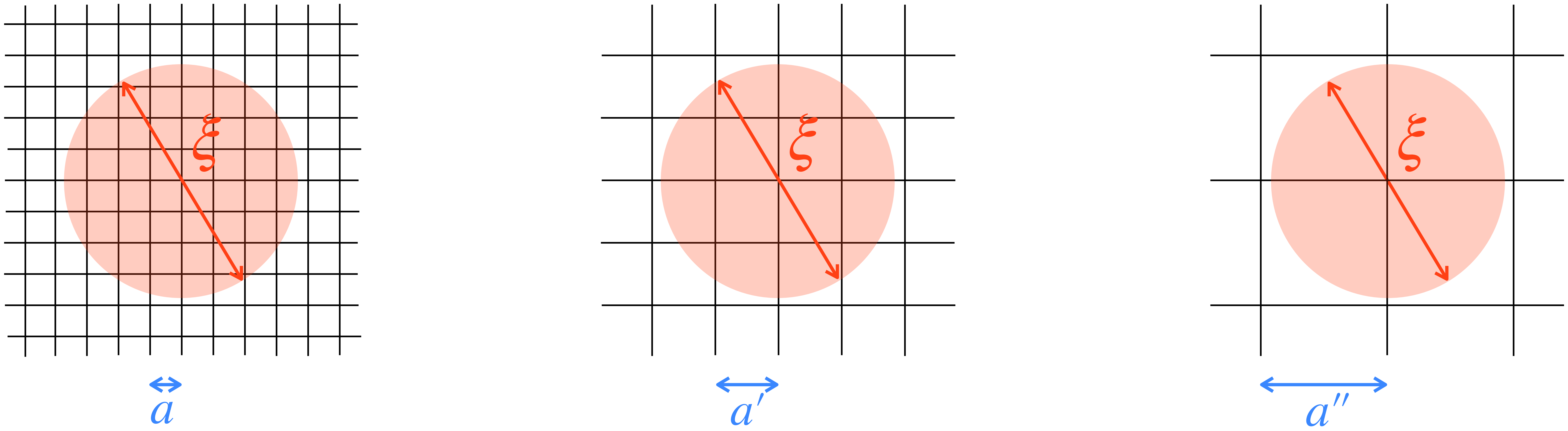


lattice spacing a from dimensionless g: $\Rightarrow$ dimensional transmutation

Introduction

The lattice spacing a is determined by the gauge coupling: $\beta = \dfrac{2N_c}{g^2}$

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



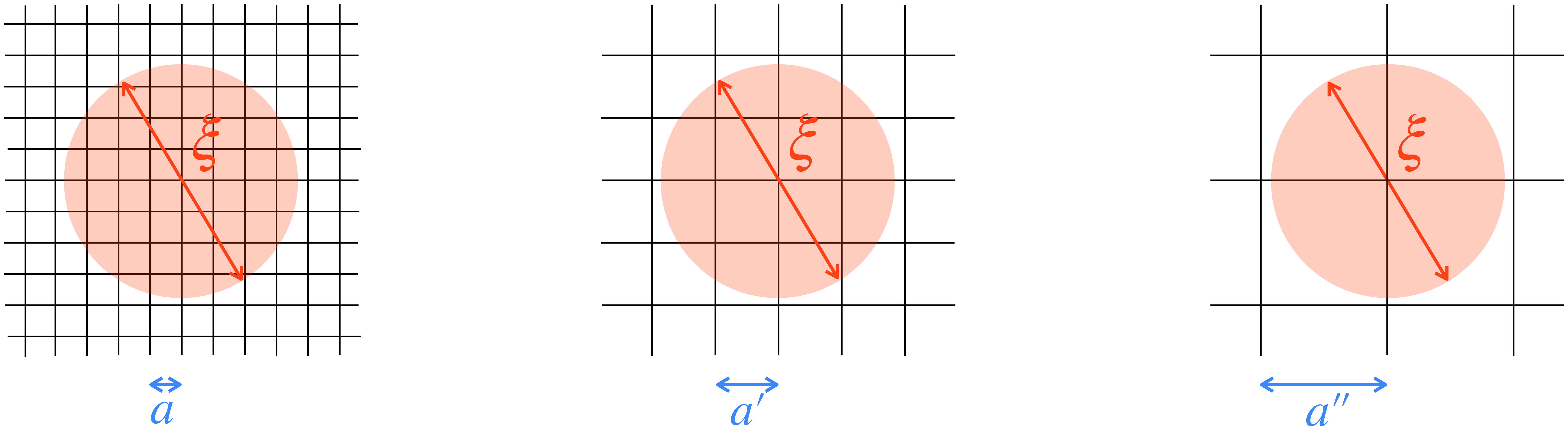
$$g \ll g' \ll g''$$

$$\beta \gg \beta' \gg \beta''$$

Introduction

The lattice spacing a is determined by the gauge coupling: $\beta = \dfrac{2N_c}{g^2}$

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



critical slowing down
(topological freezing) ?

large lattice artefacts?

Part I:
The FP action

Renormalization group transformation

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

$\Rightarrow$ provides solution for avoiding critical slowing down and lattice artefacts

Renormalization group transformation

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

$$\exp\left\{-\beta' A'[V]\right\} = \int \mathcal{D}U \exp\left\{-\beta\left(A[U] + T[U,V]\right)\right\}$$

where $T[U,V]$ is a blocking kernel relating the fine gauge links U to the coarse gauge links $V \equiv U'$:

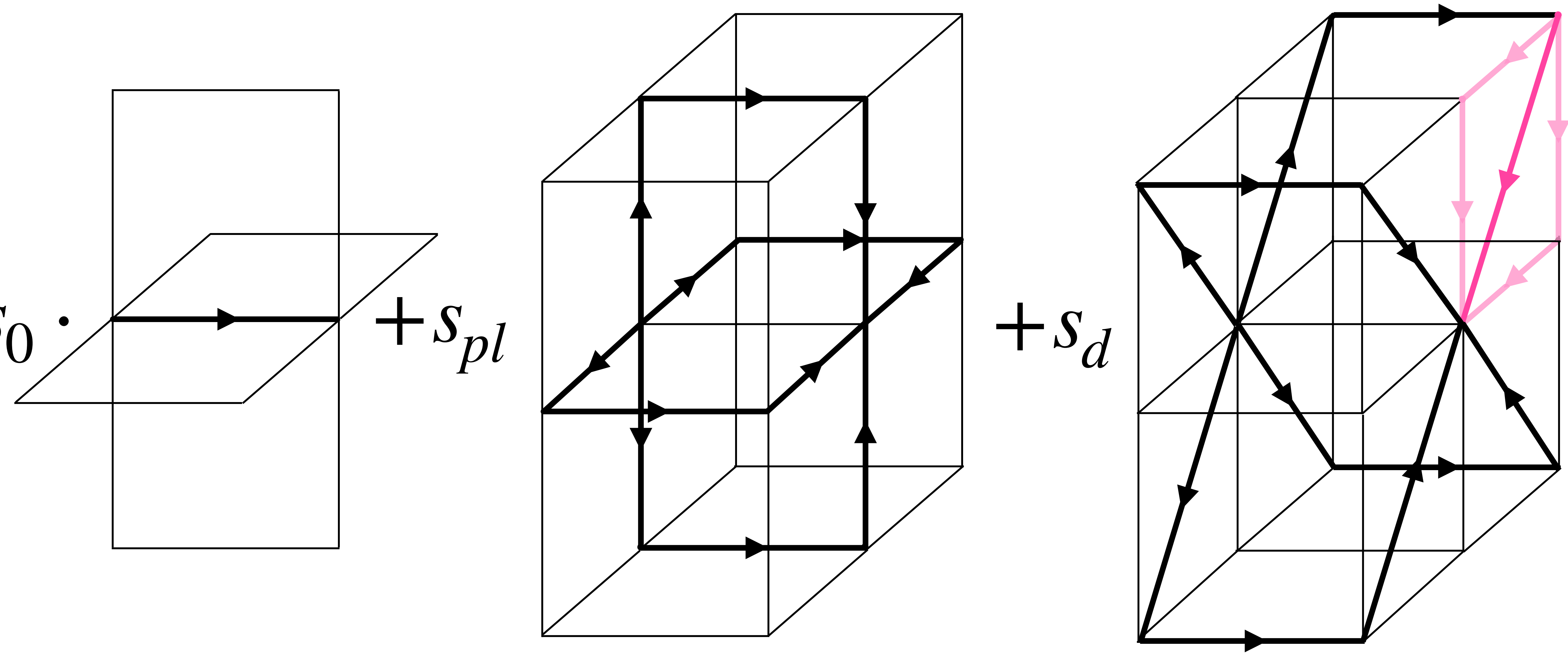
$$T[U,V] = -\frac{\kappa}{N_c}\sum_{x_B,\mu}\left\{\mathrm{ReTr}\left(V_\mu(x_B)\cdot Q_\mu^\dagger(x_B)\right) - \mathcal{N}_\mu^\beta\right\}$$

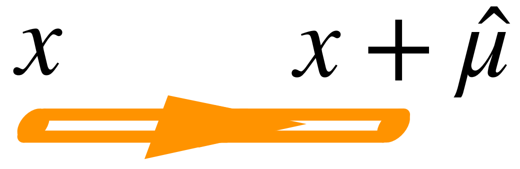
($\mathcal{N}_\mu^\beta$ is a normalization factor guaranteeing $Z(\beta') = Z(\beta)$, i.e., unchanged long-distance physics)

RGT blocking kernel

$$T[U,V] = -\frac{\kappa}{N_c}\sum_{x_B,\mu}\left\{ \mathrm{ReTr}\left(V_\mu(x_B)\cdot Q^\dagger_\mu(x_B)\right) - \mathcal{N}^\beta_\mu \right\}$$

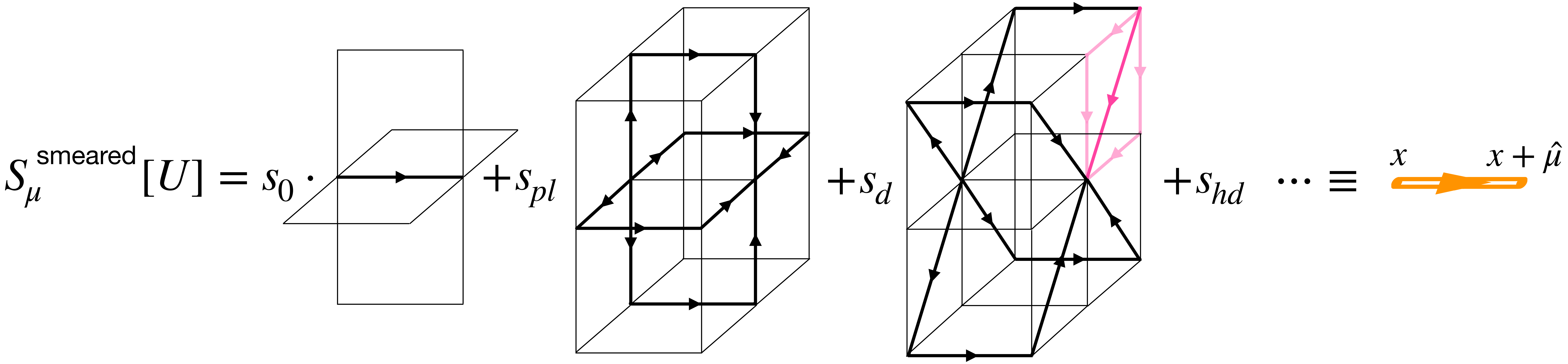
$S_\mu^{\text{smeared}}[U] = s_0 \cdot$ $+s_{pl}$ $+s_d$ $+s_{hd}$ $\cdots \equiv$

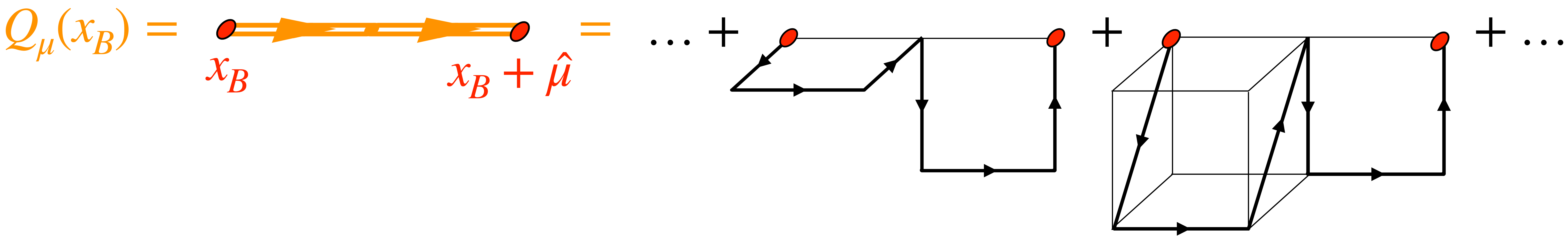




RGT blocking kernel

$$T[U,V] = -\frac{\kappa}{N_c}\sum_{x_B,\mu}\left\{\mathrm{ReTr}\left(V_\mu(x_B)\cdot Q_\mu^\dagger(x_B)\right) - \mathcal{N}_\mu^\beta\right\}$$



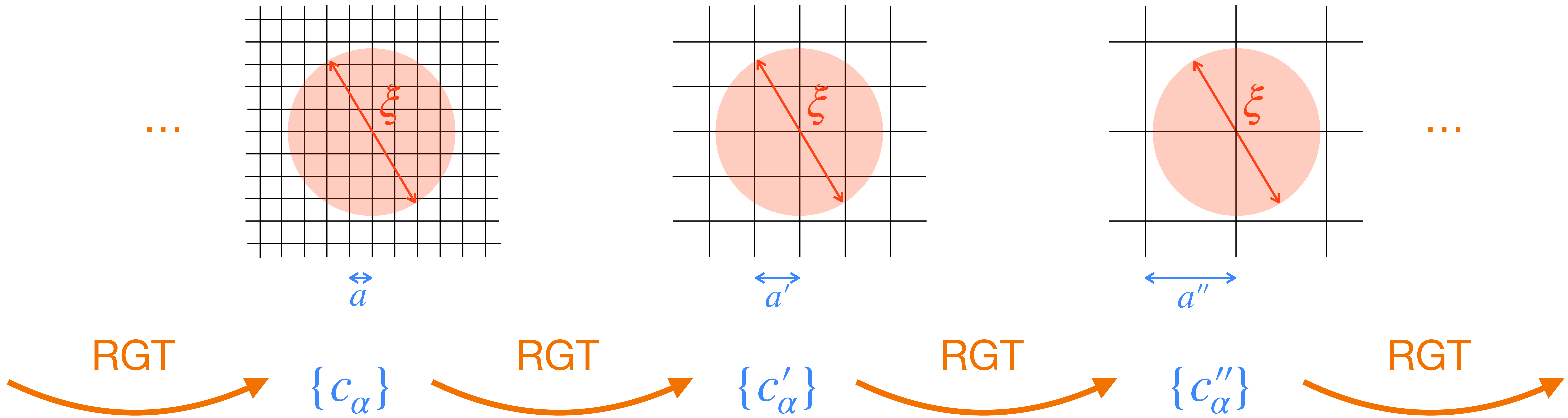


Renormalization group transformation

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

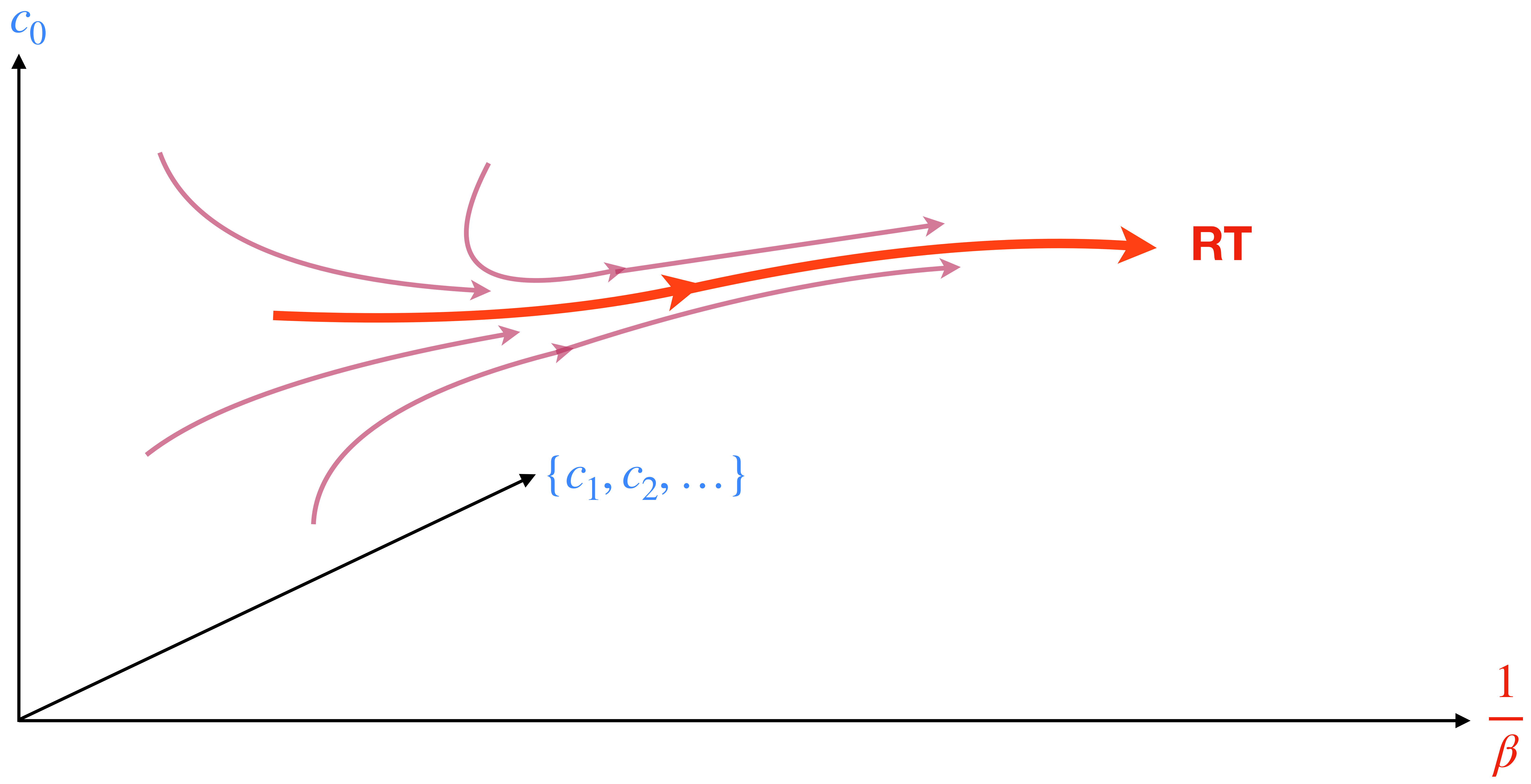
$$\exp\left\{-\beta' A'[V]\right\} = \int \mathscr{D}U \exp\left\{-\beta\left(A[U] + T[U,V]\right)\right\}$$

The effective action $\beta' A'[V]$ is described by infinitely many couplings $\{c'_\alpha\}$:



Renormalization group transformation

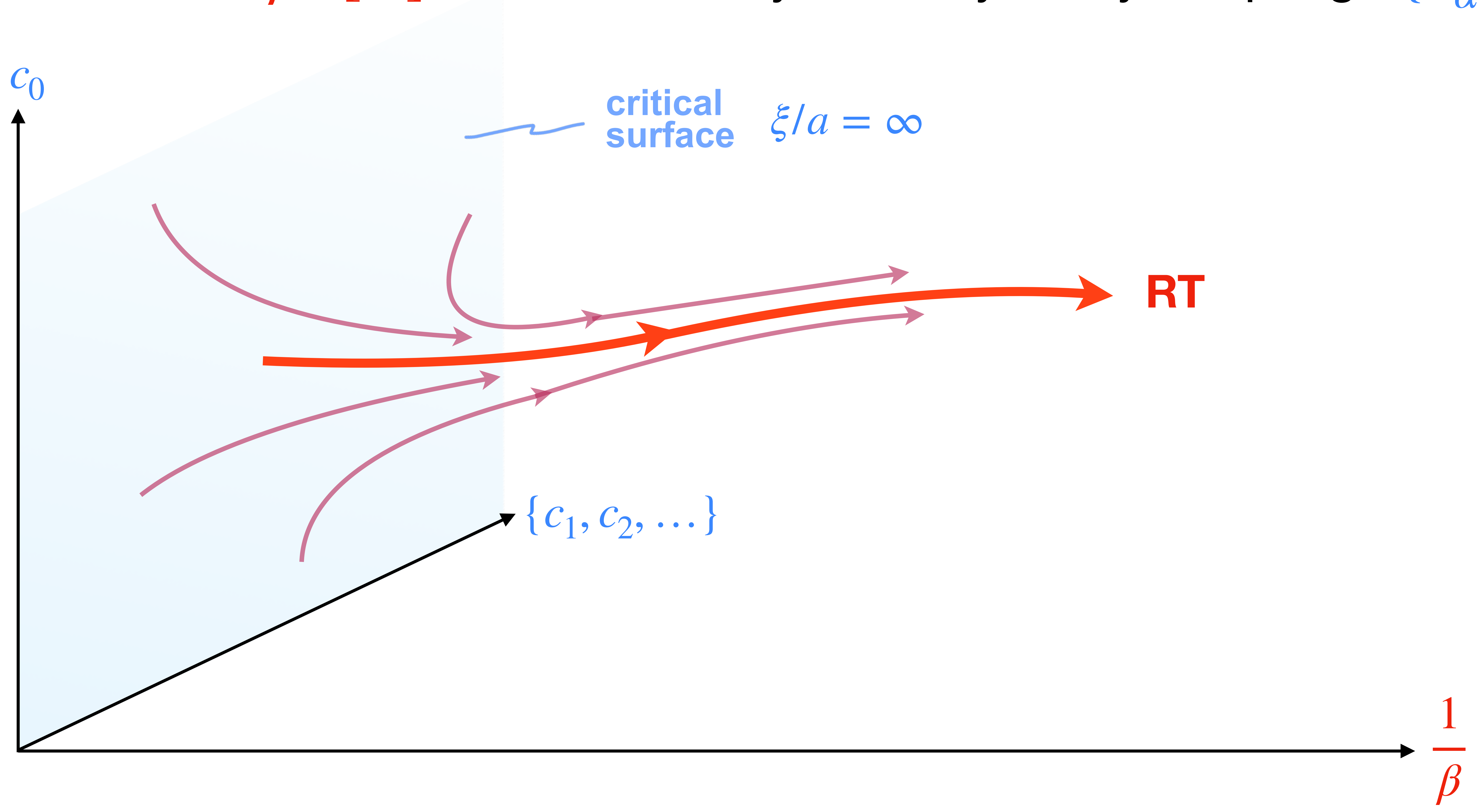
The effective action $\beta A[V]$ is described by infinitely many couplings $\{c_\alpha\}$:



$\Rightarrow$ for asymptotically free gauge theories, there is one relevant direction $1/\beta$

Renormalization group transformation

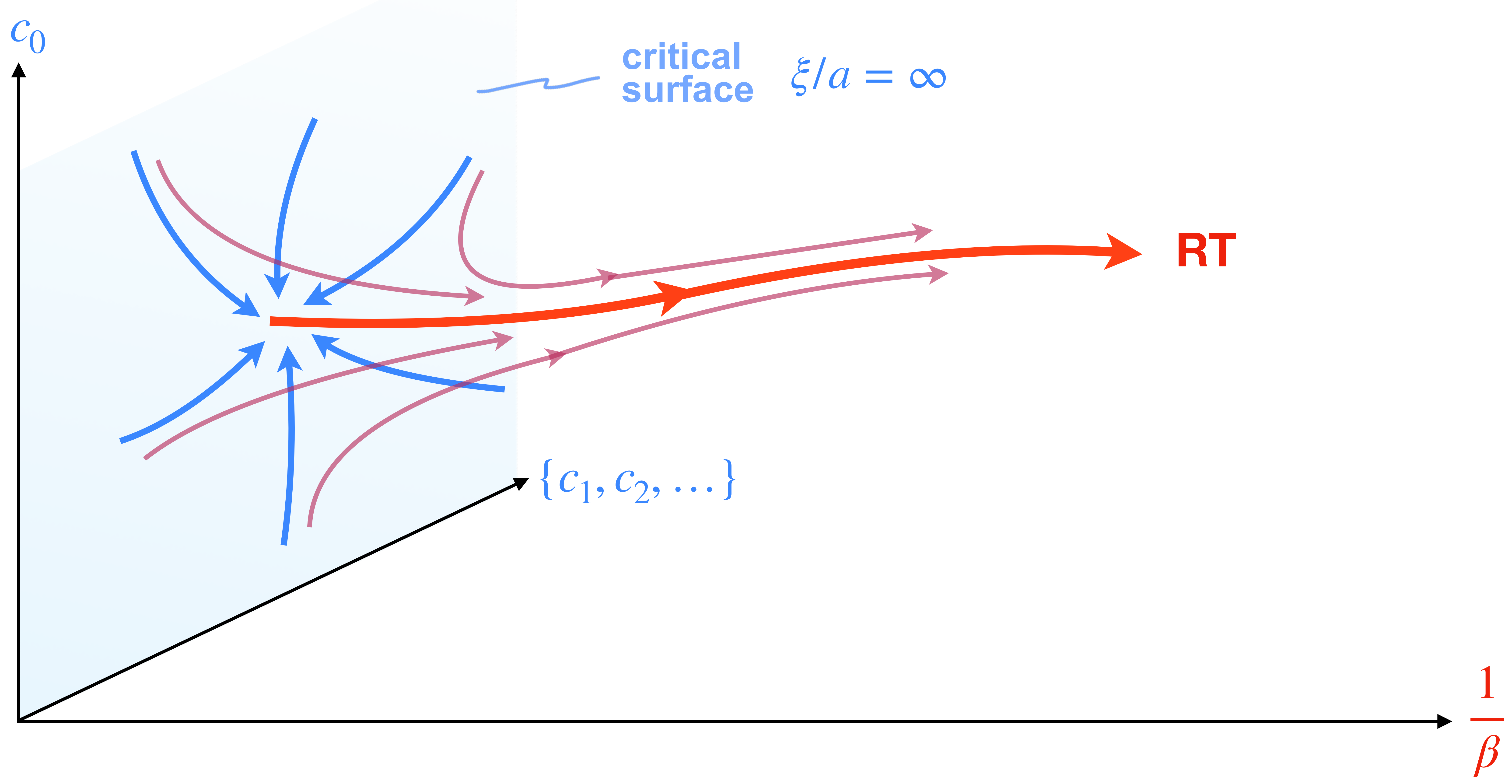
The effective action $\beta A[V]$ is described by infinitely many couplings $\{c_\alpha\}$:



$\Rightarrow$ for asymptotically free gauge theories, there is one relevant direction $1/\beta$

Renormalization group transformation

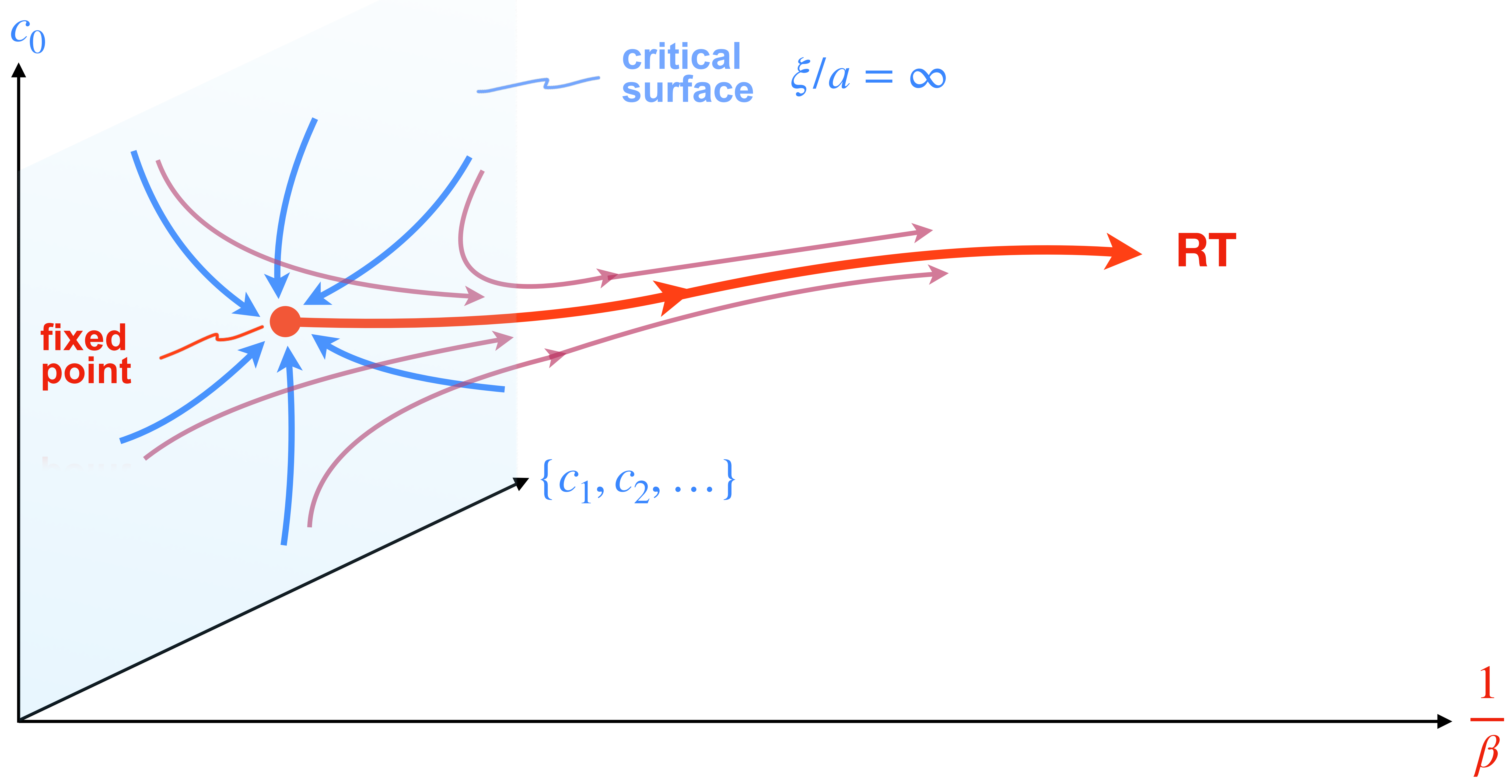
The effective action $\beta A[V]$ is described by infinitely many couplings $\{c_\alpha\}$:



$\Rightarrow$ for asymptotically free gauge theories, there is one relevant direction $1/\beta$

Renormalization group transformation

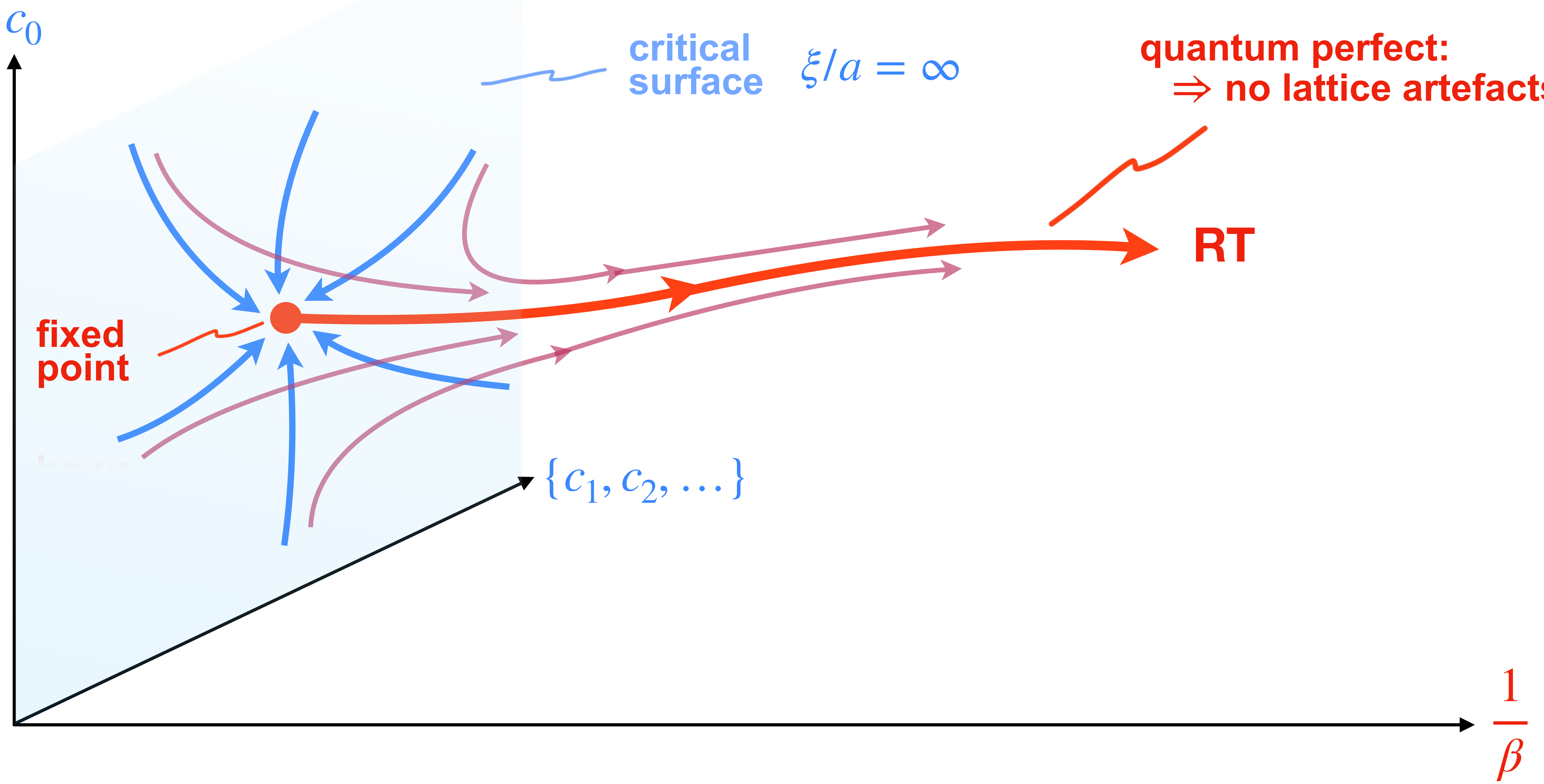
The effective action $\beta A[V]$ is described by infinitely many couplings $\{c_\alpha\}$:



$\Rightarrow$ fixed point of RGT iterations (when $\xi/a \to \infty$): $\{c^*_\alpha\}$ $\xrightarrow{\text{RGT}}$ $\{c^*_\alpha\}$

Renormalization group transformation

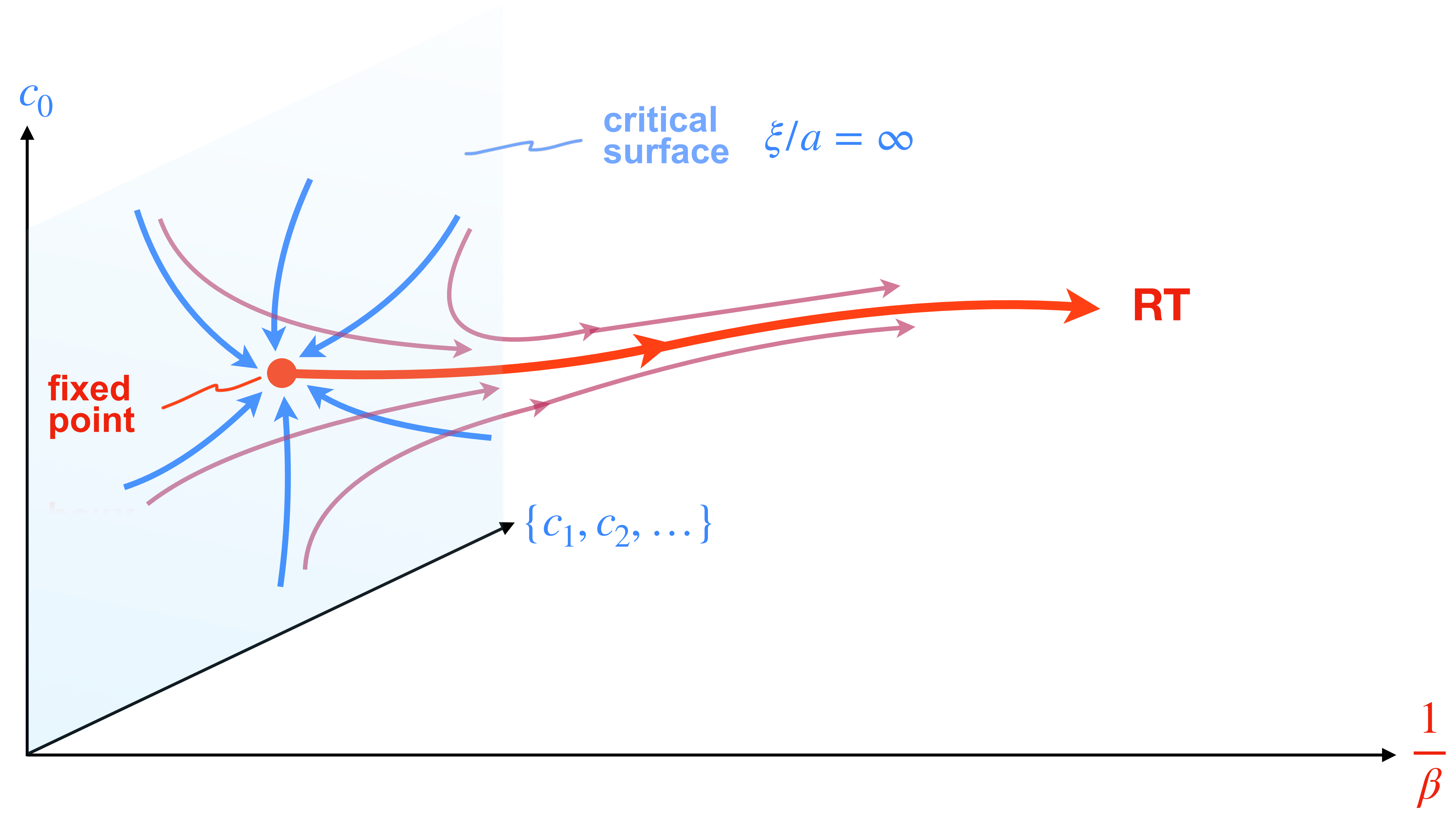
The effective action $\beta A[V]$ is described by infinitely many couplings $\{c_\alpha\}$:



⇒ fixed point of RGT iterations (when $\xi/a \to \infty$): $\{c_\alpha^*\}$ $\xrightarrow{\text{RGT}}$ $\{c_\alpha^*\}$

Renormalization group transformation

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



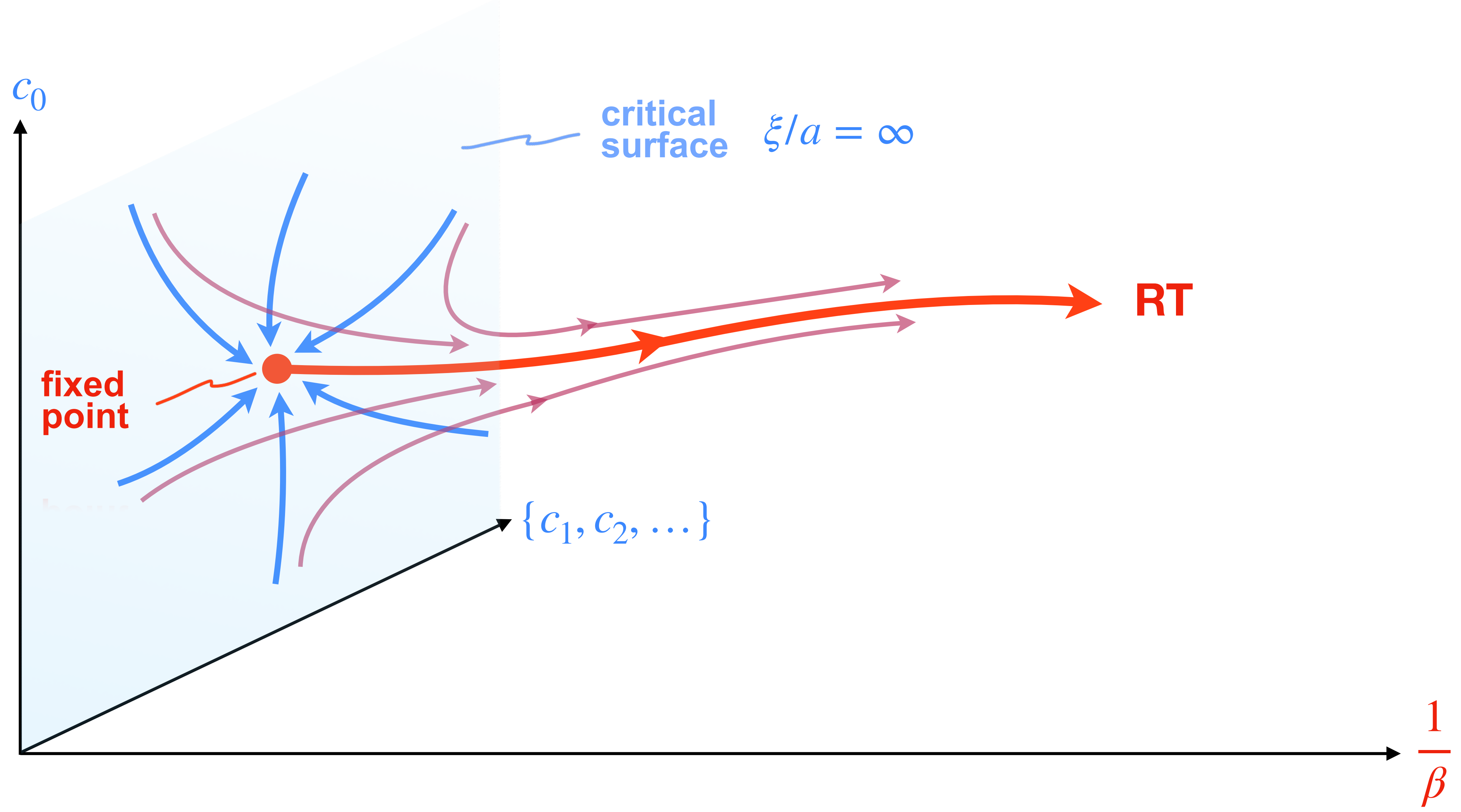
$$\exp\left\{-\beta' A'[V]\right\} = \int \mathscr{D}U \exp\left\{-\beta\left(A[U] + T[U,V]\right)\right\}$$

Two practical problems:

- how to parametrize RT, i.e., which set $\{c_\alpha\}$?
- how to determine $\{c_\alpha^{\mathrm{RT}}\}$ or $\{c_\alpha^{\mathrm{FP}}\}$?

Renormalization group transformation

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



$$\exp\left\{-\beta' A'[V]\right\} = \int \mathscr{D}U \exp\left\{-\beta\left(A[U] + T[U,V]\right)\right\}$$

Two practical problems:

- how to parametrize RT, i.e., which set $\{c_\alpha\}$?
- how to determine $\{c_\alpha^{\mathrm{RT}}\}$ or $\{c_\alpha^{\mathrm{FP}}\}$?

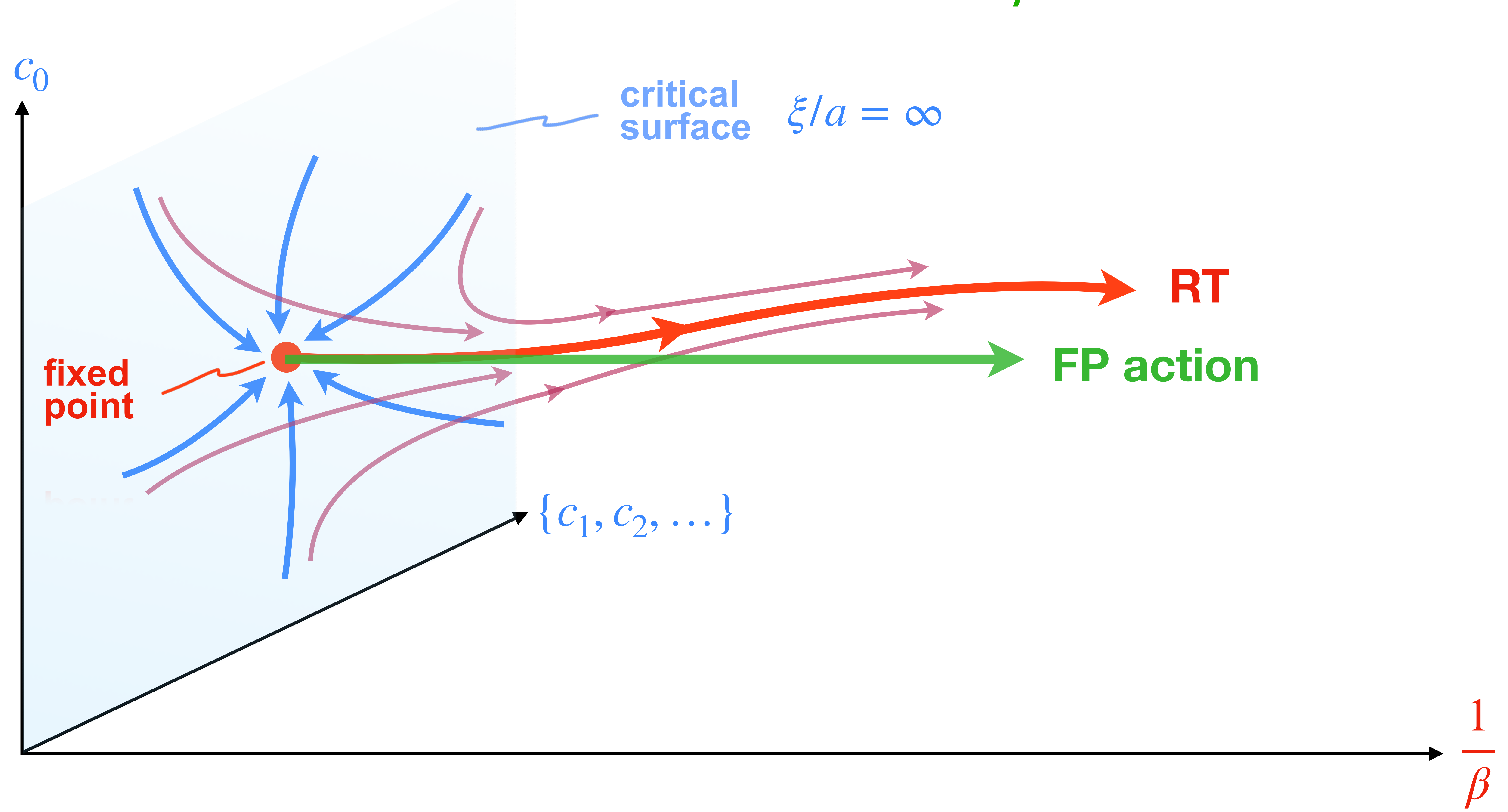
P. Hasenfratz, F. Niedermayer [Nucl. Phys. B414 (1994) 785, hep-lat/9308004]

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

$$A^{\mathrm{FP}}[V] = \min_{\{U\}} \left\{A^{\mathrm{FP}}[U] + T[U,V]\right\}$$

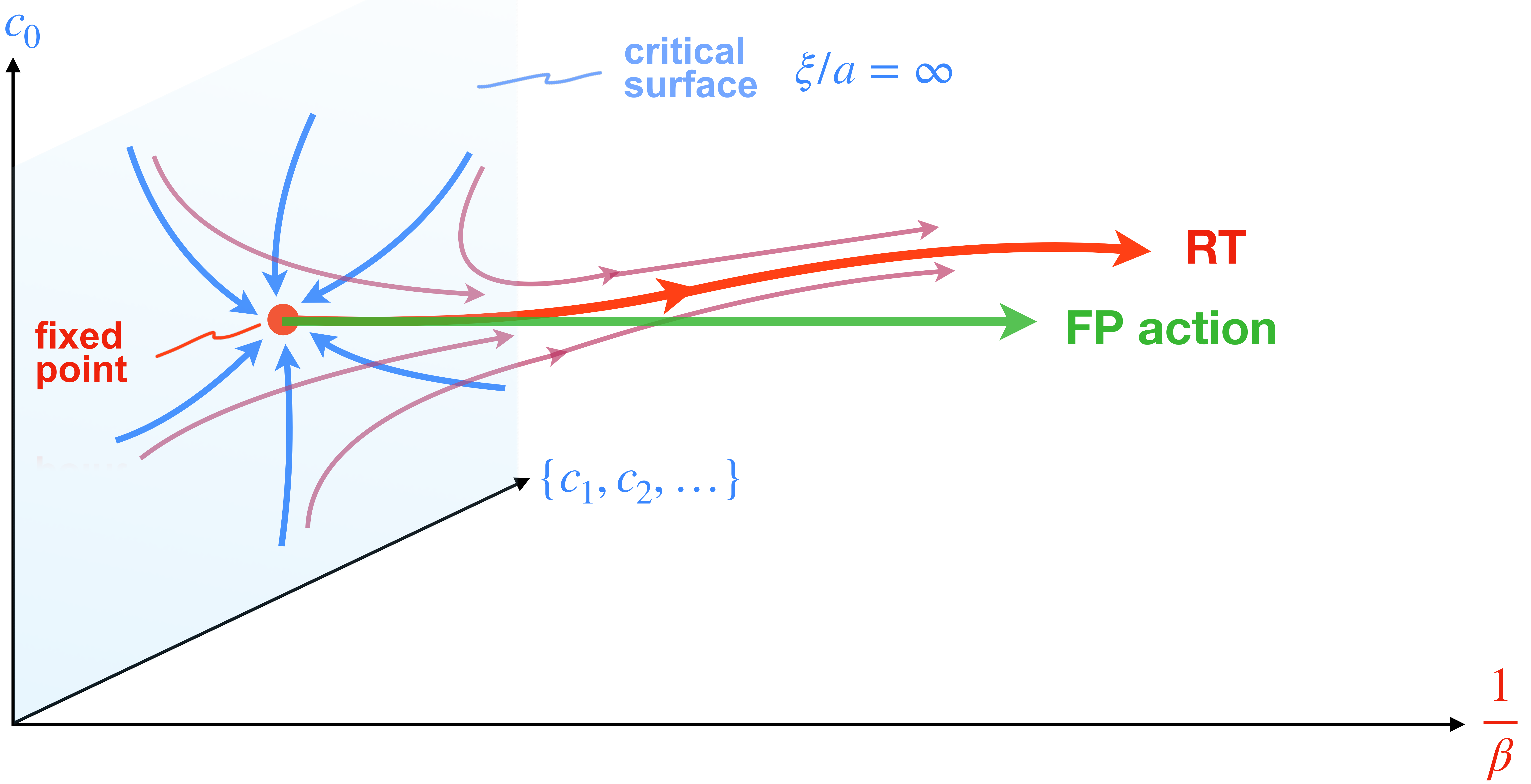
Classically perfect FP actions

The classical FP action A^{FP} defines an action for all β:



Classically perfect FP actions

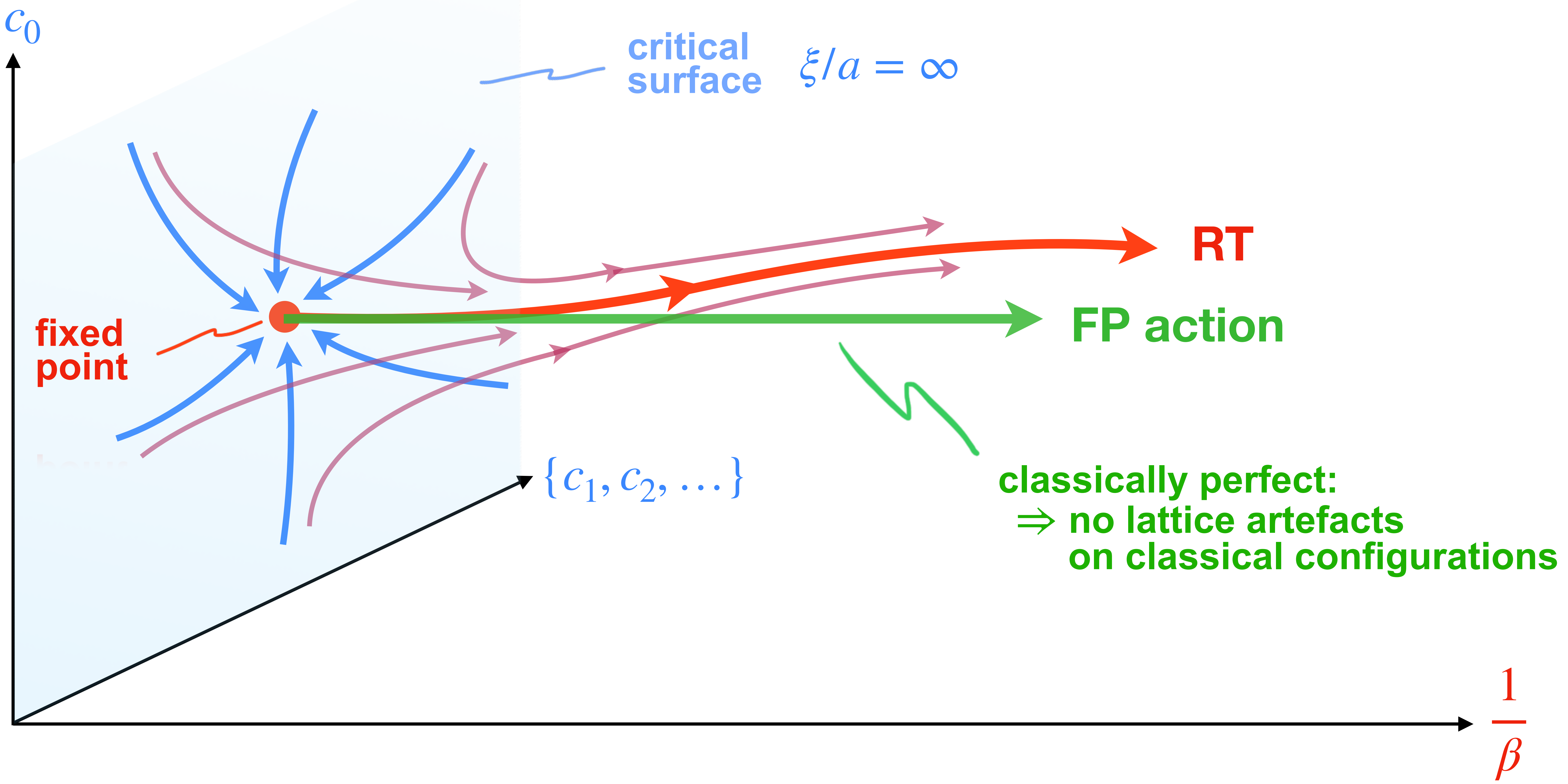
The FP action values for rough configurations defined through an inception procedure:



$$A^{\mathrm{FP}}[V] = \min_{\{U\}} \{A^{\mathrm{FP}}[U] + T[U,V]\} = \min_{\{U',U\}} \{A^{\mathrm{FP}}[U'] + T[U',U] + T[U,V]\}$$

Classically perfect FP actions

The FP action values for rough configurations defined through an inception procedure:



$$A^{\mathrm{FP}}[V] = \min_{\{U\}} \{A^{\mathrm{FP}}[U] + T[U,V]\} = \min_{\{U',U\}} \{A^{\mathrm{FP}}[U'] + T[U',U] + T[U,V]\}$$

Classically perfect FP actions

The classical FP equation can be iterated:

$$A^{\mathrm{FP}}[V] = \min_{\{U\}} \{A^{\mathrm{FP}}[U] + T[U,V]\} = \min_{\{U',U\}} \{A^{\mathrm{FP}}[U'] + T[U',U] + T[U,V]\}$$

- There are no lattice artefacts on classical configurations:

$$\frac{\delta A^{\mathrm{FP}}[V]}{\delta V} = 0 \quad \Rightarrow \quad \frac{\delta A^{\mathrm{FP}}[U]}{\delta U} = 0$$

Classically perfect FP actions

The classical FP equation can be iterated:

$$A^{\mathrm{FP}}[V] = \min_{\{U\}} \{A^{\mathrm{FP}}[U] + T[U,V]\} = \min_{\{U',U\}} \{A^{\mathrm{FP}}[U'] + T[U',U] + T[U,V]\}$$

- There are no lattice artefacts on classical configurations:

$$\frac{\delta A^{\mathrm{FP}}[V]}{\delta V} = 0 \quad \Rightarrow \quad \frac{\delta A^{\mathrm{FP}}[U]}{\delta U} = 0$$

- Proof using the chain rule:

$$\frac{\delta A^{\mathrm{FP}}[V]}{\delta V} = \left[\frac{\delta}{\delta U}(A^{\mathrm{FP}}[U] + T[U,V])\frac{\delta U}{\delta V} + \frac{\delta T[U,V]}{\delta V}\right]_{U_{\mathrm{min}}}$$

$$\Rightarrow \left.\frac{\delta T[U,V]}{\delta V}\right|_{U_{\mathrm{min}}} = 0 \quad \text{hence} \quad T[U_{\mathrm{min}}, V] = 0$$

Classically perfect FP actions

The classical FP equation can be iterated:

$$A^{\mathrm{FP}}[V] = \min_{\{U\}} \{A^{\mathrm{FP}}[U] + T[U,V]\} = \min_{\{U',U\}} \{A^{\mathrm{FP}}[U'] + T[U',U] + T[U,V]\}$$

- There are no lattice artefacts on classical configurations:

$$\frac{\delta A^{\mathrm{FP}}[V]}{\delta V} = 0 \quad \Rightarrow \quad \frac{\delta A^{\mathrm{FP}}[U]}{\delta U} = 0$$

- Proof using the chain rule:

$$\frac{\delta A^{\mathrm{FP}}[V]}{\delta V} = \left[\frac{\delta}{\delta U}(A^{\mathrm{FP}}[U] + T[U,V])\frac{\delta U}{\delta V} + \frac{\delta T[U,V]}{\delta V}\right]_{U_{\mathrm{min}}}$$

$$A^{\mathrm{FP}}[V] = A^{\mathrm{FP}}[U_{\mathrm{min}}] \quad \text{and} \quad \left.\frac{\delta A^{\mathrm{FP}}[U]}{\delta U}\right|_{U_{\mathrm{min}}} = 0$$

Classically perfect FP actions

The classical FP equation can be iterated:

$$A^{\mathrm{FP}}[V] = \min_{\{U\}} \{A^{\mathrm{FP}}[U] + T[U,V]\} = \min_{\{U',U\}} \{A^{\mathrm{FP}}[U'] + T[U',U] + T[U,V]\}$$

- There are no lattice artefacts on classical configurations:

$$\frac{\delta A^{\mathrm{FP}}[V]}{\delta V} = 0 \quad \Rightarrow \quad \frac{\delta A^{\mathrm{FP}}[U]}{\delta U} = 0$$

$\Rightarrow A^{\mathrm{FP}}[V]$ has scale invariant instanton solutions

- Proof using the chain rule:

$$\frac{\delta A^{\mathrm{FP}}[V]}{\delta V} = \left[\frac{\delta}{\delta U}(A^{\mathrm{FP}}[U] + T[U,V])\frac{\delta U}{\delta V} + \frac{\delta T[U,V]}{\delta V}\right]_{U_{\mathrm{min}}}$$

$$A^{\mathrm{FP}}[V] = A^{\mathrm{FP}}[U_{\mathrm{min}}] \quad \text{and} \quad \left.\frac{\delta A^{\mathrm{FP}}[U]}{\delta U}\right|_{U_{\mathrm{min}}} = 0$$

Classically perfect FP actions

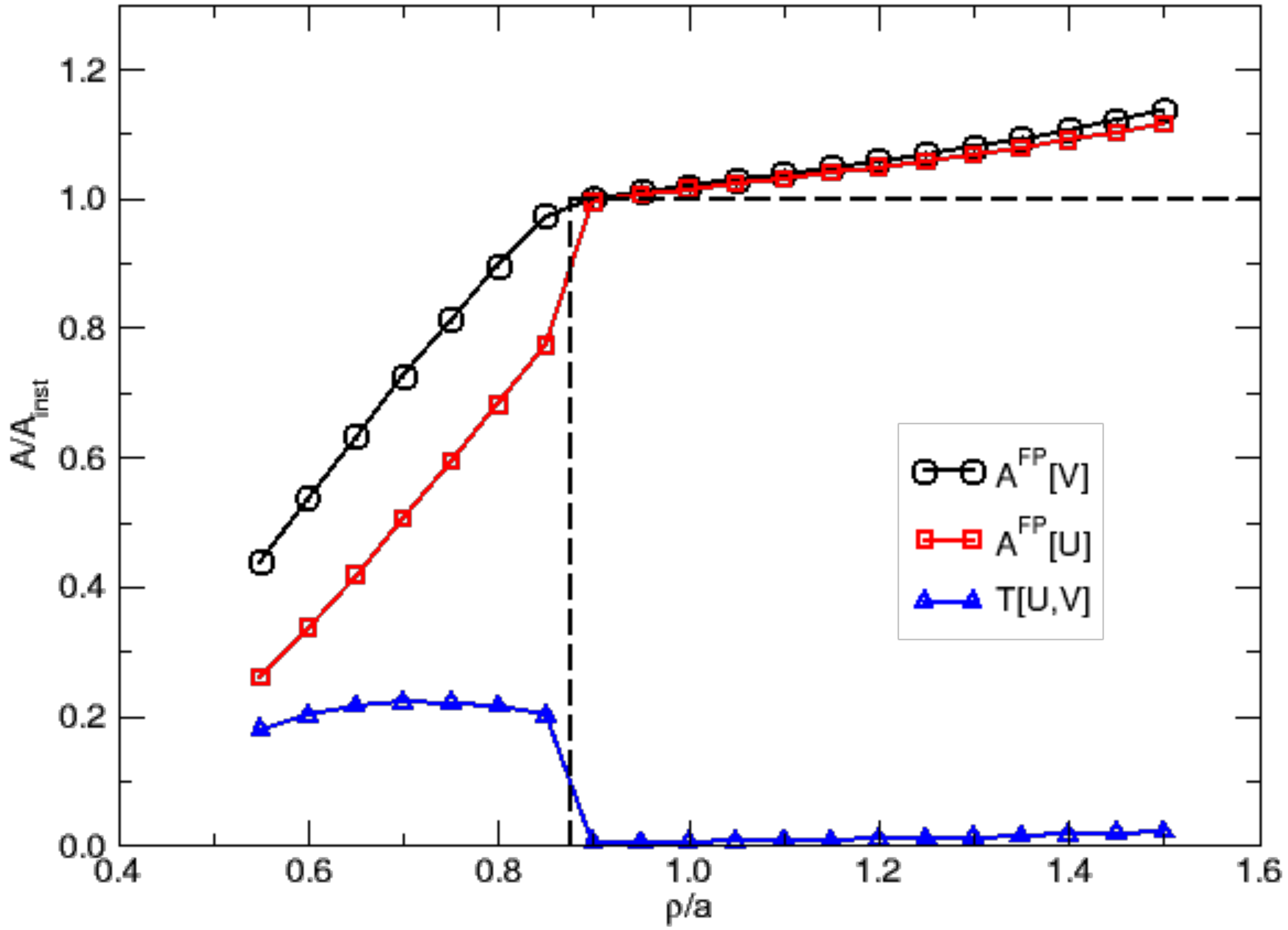
The classical FP equation can be iterated:

$$A^{\mathrm{FP}}[V] = \min_{\{U\}} \{A^{\mathrm{FP}}[U] + T[U,V]\} = \min_{\{U',U\}} \{A^{\mathrm{FP}}[U'] + T[U',U] + T[U,V]\}$$

- There are no lattice artefacts on classical configurations:

$$\frac{\delta A^{\mathrm{FP}}[V]}{\delta V} = 0 \quad \Rightarrow \quad \frac{\delta A^{\mathrm{FP}}[U]}{\delta U} = 0$$

$\Rightarrow A^{\mathrm{FP}}[V]$ has scale invariant instanton solutions



Classically perfect FP actions

The classical FP equation can be iterated:

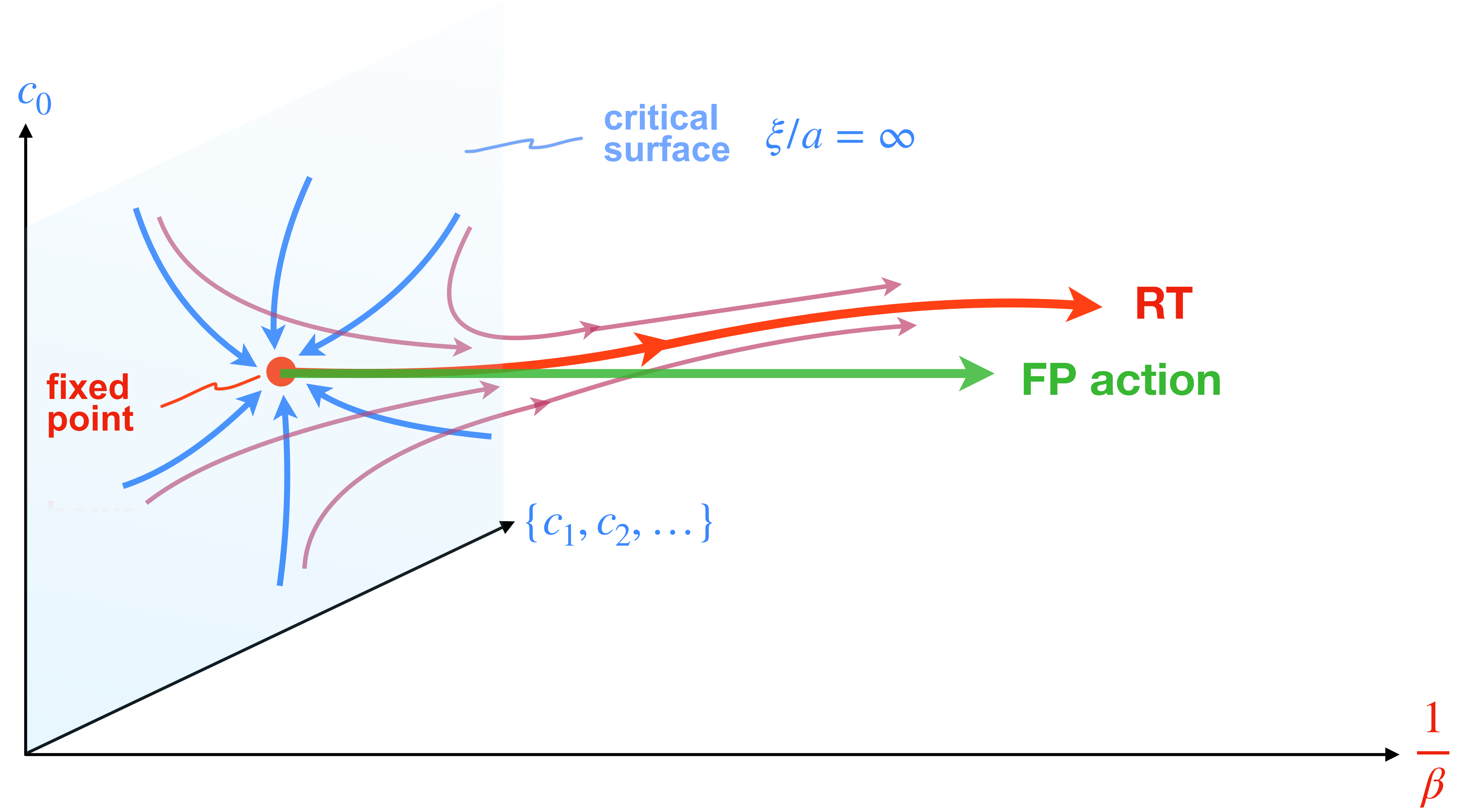
$$A^{\mathrm{FP}}[V] = \min_{\{U\}} \{A^{\mathrm{FP}}[U] + T[U,V]\} = \min_{\{U',U\}} \{A^{\mathrm{FP}}[U'] + T[U',U] + T[U,V]\}$$

- There are no lattice artefacts on classical configurations:

$$\frac{\delta A^{\mathrm{FP}}[V]}{\delta V} = 0 \quad \Rightarrow \quad \frac{\delta A^{\mathrm{FP}}[U]}{\delta U} = 0$$

$\Rightarrow A^{\mathrm{FP}}[V]$ has scale invariant instanton solutions

- For large $\beta, A^{\mathrm{FP}}[V]$ is very close to $A^{\mathrm{RT}}[V]$:



Classically perfect FP actions

The classical FP equation can be iterated:

$$A^{\mathrm{FP}}[V] = \min_{\{U\}} \{A^{\mathrm{FP}}[U] + T[U, V]\} = \min_{\{U',U\}} \{A^{\mathrm{FP}}[U'] + T[U', U] + T[U, V]\}$$

- There are no lattice artefacts on classical configurations:

$$\frac{\delta A^{\mathrm{FP}}[V]}{\delta V} = 0 \quad \Rightarrow \quad \frac{\delta A^{\mathrm{FP}}[U]}{\delta U} = 0$$

$\Rightarrow$ $A^{\mathrm{FP}}[V]$ has scale invariant instanton solutions

- For large $\beta, A^{\mathrm{FP}}[V]$ is very close to $A^{\mathrm{RT}}[V]$:

$\Rightarrow$ lattice artefacts expected to be substantially reduced:

$$\cancel{\mathcal{O}(a^{2n})}, \mathcal{O}(g^2 a^{2n}) \quad n = 1,2,\ldots$$

Classically perfect FP actions

The classical FP equation can be iterated:

$$A^{\mathrm{FP}}[V] = \min_{\{U\}} \{A^{\mathrm{FP}}[U] + T[U,V]\} = \min_{\{U',U\}} \{A^{\mathrm{FP}}[U'] + T[U',U] + T[U,V]\}$$

- There are no lattice artefacts on classical configurations:

$$\frac{\delta A^{\mathrm{FP}}[V]}{\delta V} = 0 \quad \Rightarrow \quad \frac{\delta A^{\mathrm{FP}}[U]}{\delta U} = 0$$

$\Rightarrow$ $A^{\mathrm{FP}}[V]$ has scale invariant instanton solutions

- For large $\beta, A^{\mathrm{FP}}[V]$ is very close to $A^{\mathrm{RT}}[V]$:

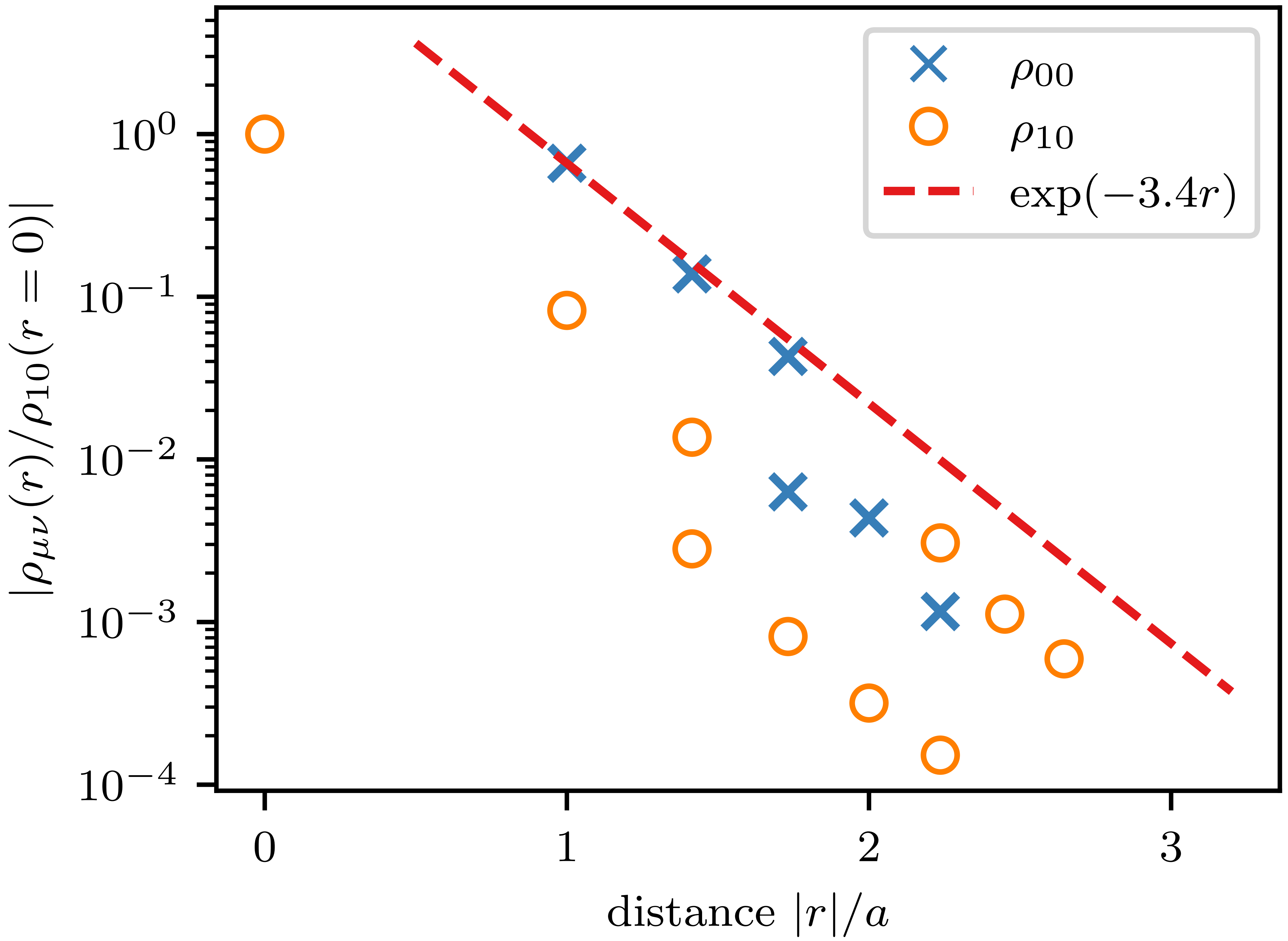
$\Rightarrow$ lattice artefacts expected to be substantially reduced:

$$~~\mathcal{O}(a^{2n})~~, \mathcal{O}(g^2 a^{2n}) \quad n = 1,2,\ldots$$

$\Rightarrow$ initiated a large activity, culminating in the discovery of **GW fermions**!

Classically perfect FP actions - locality

Is the FP action local? ⇒ Consider the couplings $A^{\mathrm{FP}} \propto \sum_k \tilde{\rho}_{\mu\nu}(k)\tilde{A}_\mu(k)\tilde{A}_\nu(-k)$



- perturbative calculation
- couplings fall off exponentially, as desired

Parametrization of the FP actions

Parametrization should be as local as possible, but still as expressive as possible.

- Wilson plaquette variable:

$$u_{\mu\nu} = \mathrm{ReTr}\left(1 - U^{pl}_{\mu\nu}\right)$$

from usual links U_μ, U_ν

- Smeared plaquette

$$w_{\mu\nu} = \mathrm{ReTr}\left(1 - W^{pl}_{\mu\nu}\right)$$

from asymmetrically smeared links

- FP action: $A^{FP}[U] = \sum_{\mu<\nu} f(u_{\mu\nu}, w_{\mu\nu})$ e.g. $f(u,w) = \sum_{k,l} p_{kl}\, u^k w^l$

- Asymmetrically smeared links:

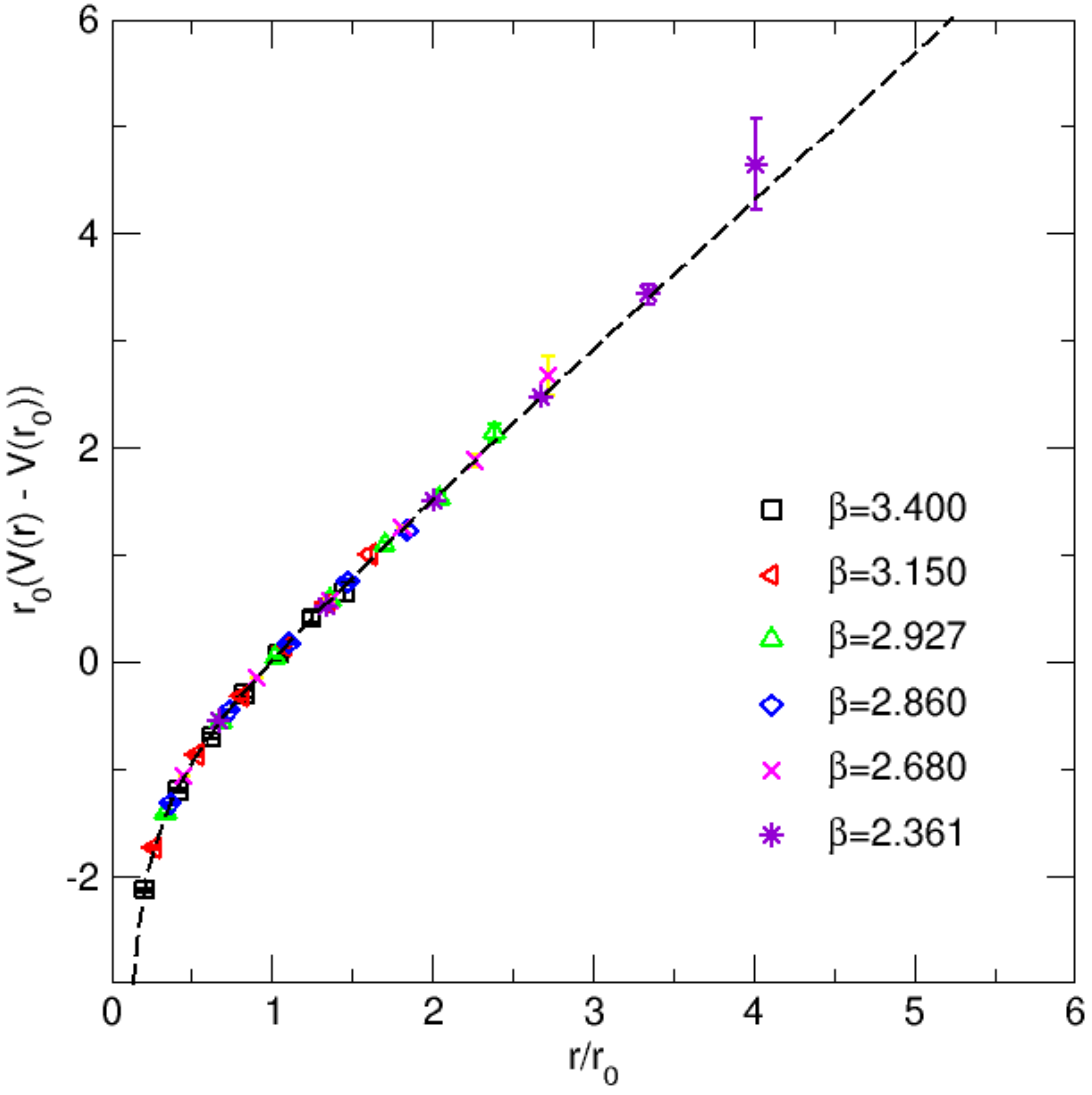
$$Q^S_\mu = \frac{1}{6}\sum_{\lambda\neq\mu} S^{(\lambda)}_\mu - U_\mu\,, \qquad Q^{(\nu)}_\mu = \frac{1}{4}\left(\sum_{\lambda\neq\mu,\nu} S^{(\lambda)}_\mu + \eta(x_\mu)\cdot S^{(\nu)}_\mu\right) - \left(1 + \frac{1}{2}\eta(x_\mu)\right) U_\mu\,,$$

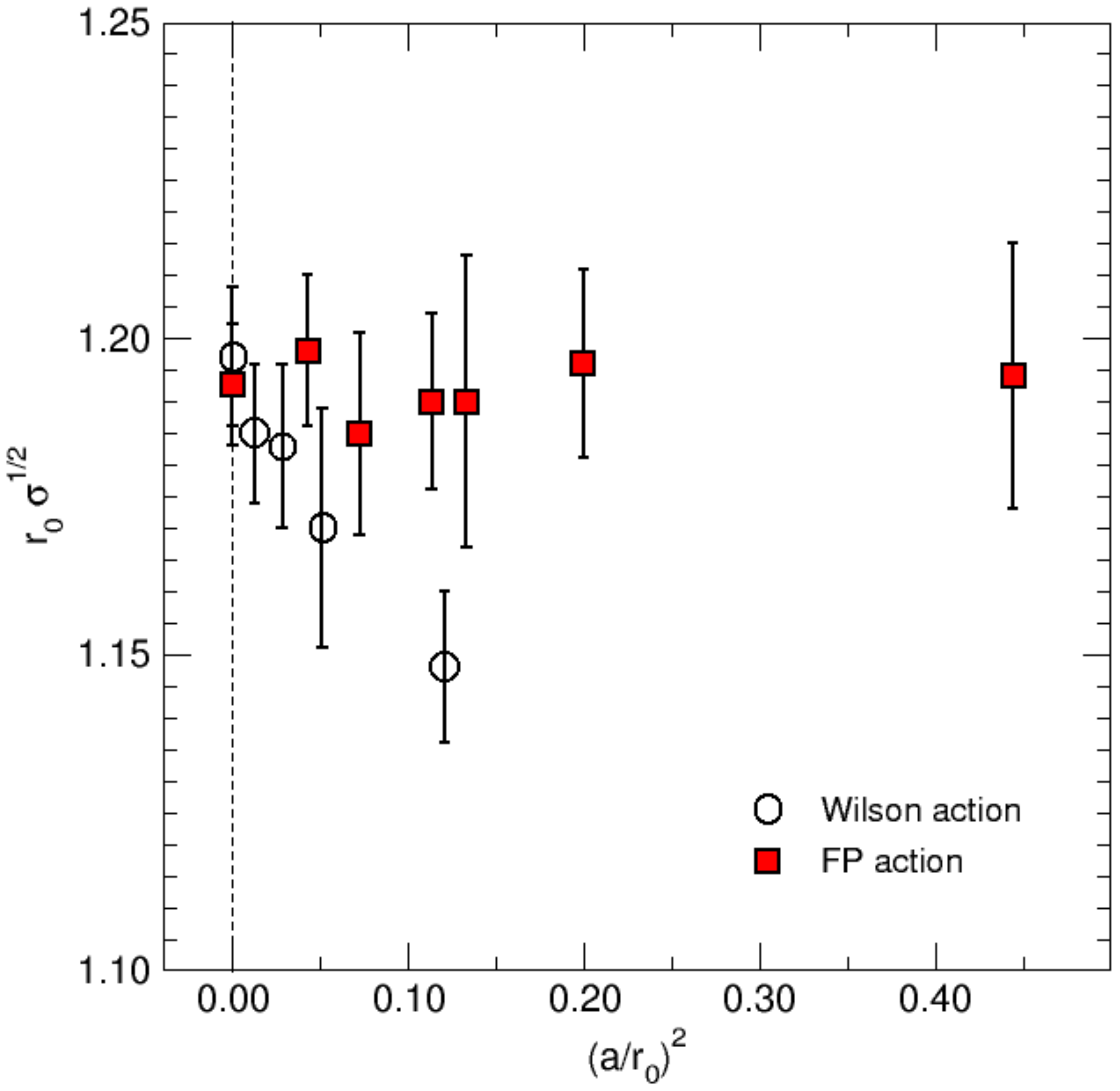
$$W^{(\nu)}_\mu = U_\mu + c_1(x_\mu)\cdot Q^{(\nu)}_\mu + c_2(x_\mu)\cdot Q^{(\nu)}_\mu U^\dagger_\mu Q^{(\nu)}_\mu + \ldots\,, \qquad x_\mu = \mathrm{ReTr}\left(Q^S_\mu \cdot U^\dagger_\mu\right)\,,$$

$$\eta(x) = \eta^{(0)} + \eta^{(1)}\cdot x + \eta^{(2)}\cdot x^2 + \ldots\,, \qquad c_i(x) = c_i^{(0)} + c_i^{(1)}\cdot x + c_i^{(2)}\cdot x^2 + \ldots$$

FP action in action

Static quark-antiquark potential, lattice spacings between $a = 0.33$ fm, $\cdots$, 0.10 fm :





[Niedermayer, Rüfenacht, UW, Nucl.Phys.B 597 (2001) 413, hep-lat/0007007]

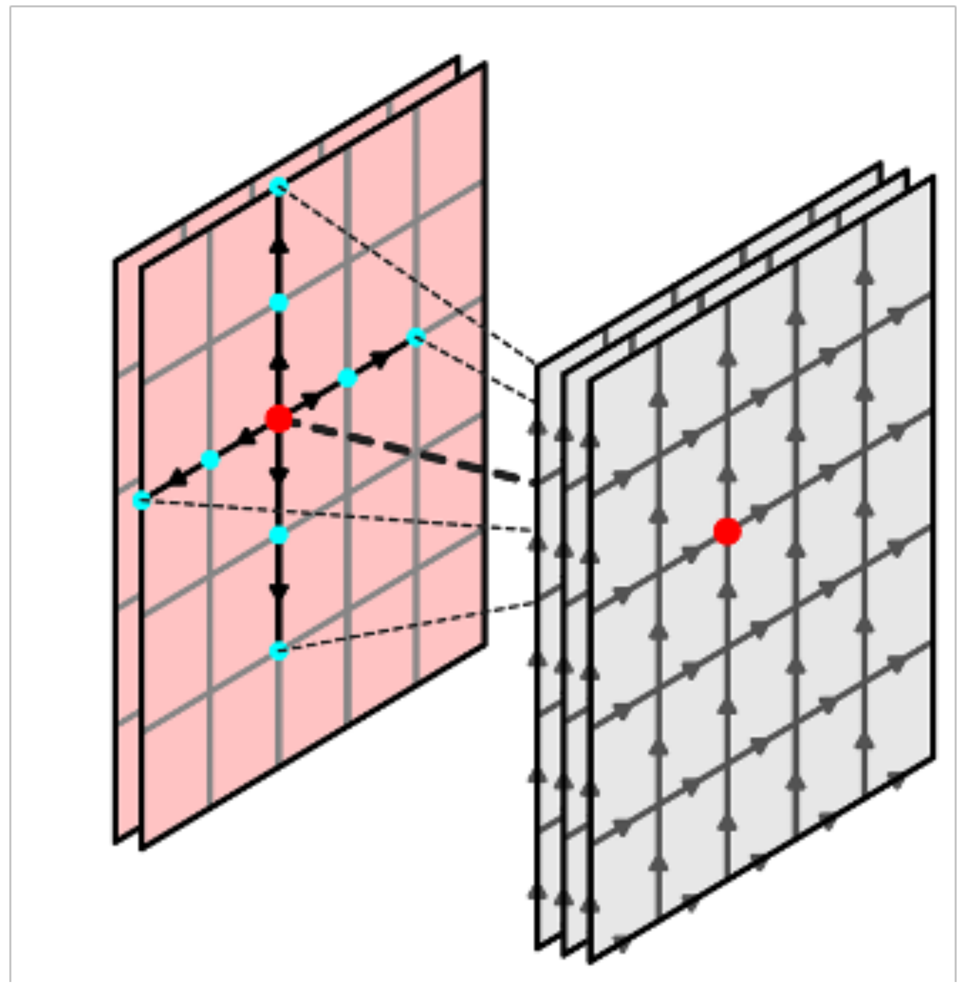
Part II:
Machine learning the FP action

Machine learning the FP action

ML architecture: Lattice gauge equivariant Convolutional Neural Network (L-CNN)

[Favoni, Ipp, Müller, Schuh, PRL 128 (2022) 3, 2012.12901]

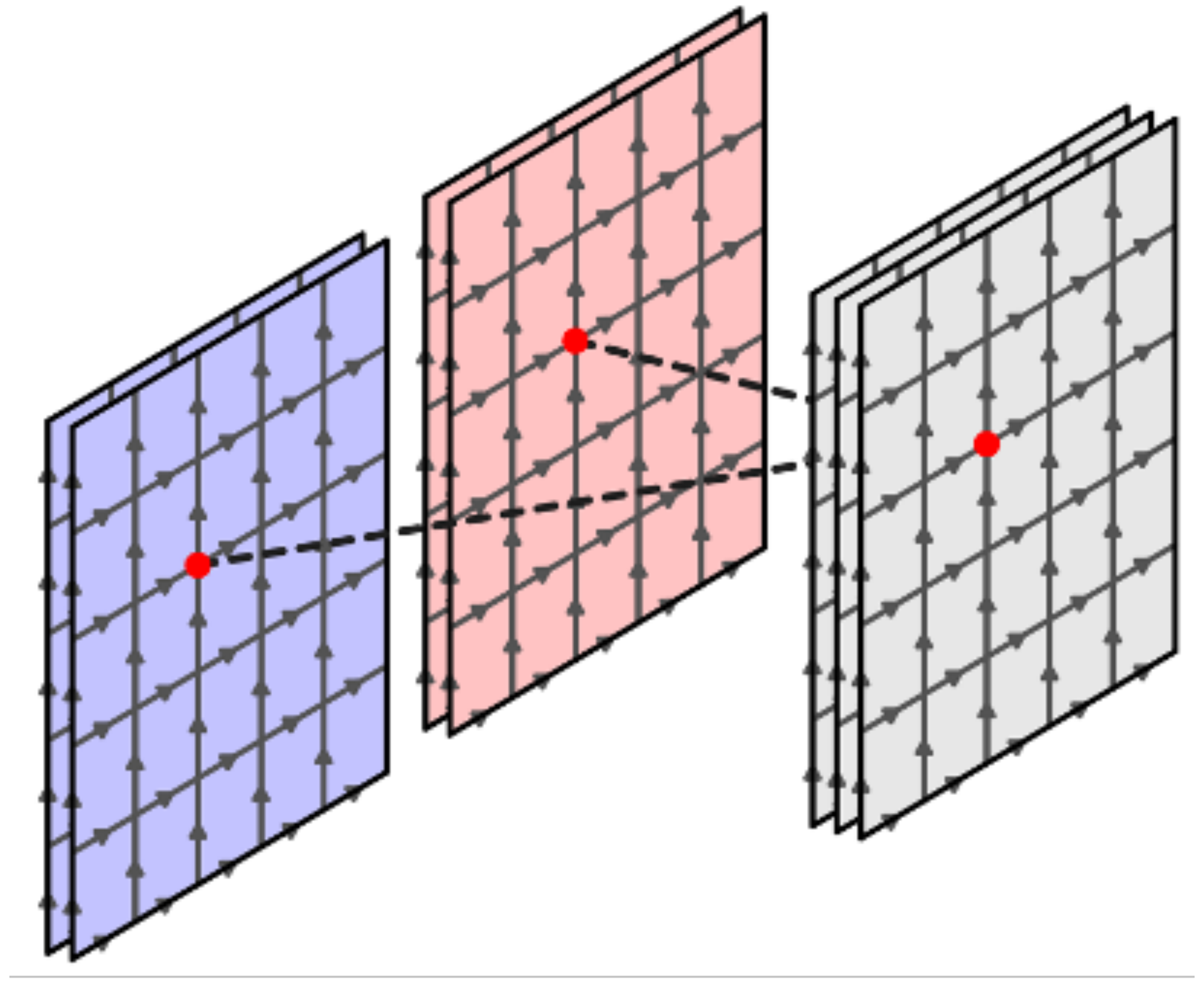
L-Conv:



$$(\mathcal{U},\mathcal{W}) \to (\mathcal{U},\mathcal{W}')$$

$$W'_{x+k\cdot\mu,j} = U_{x,k\cdot\mu} W_{x+k\cdot\mu,j} U^{\dagger}_{x,k\cdot\mu},$$

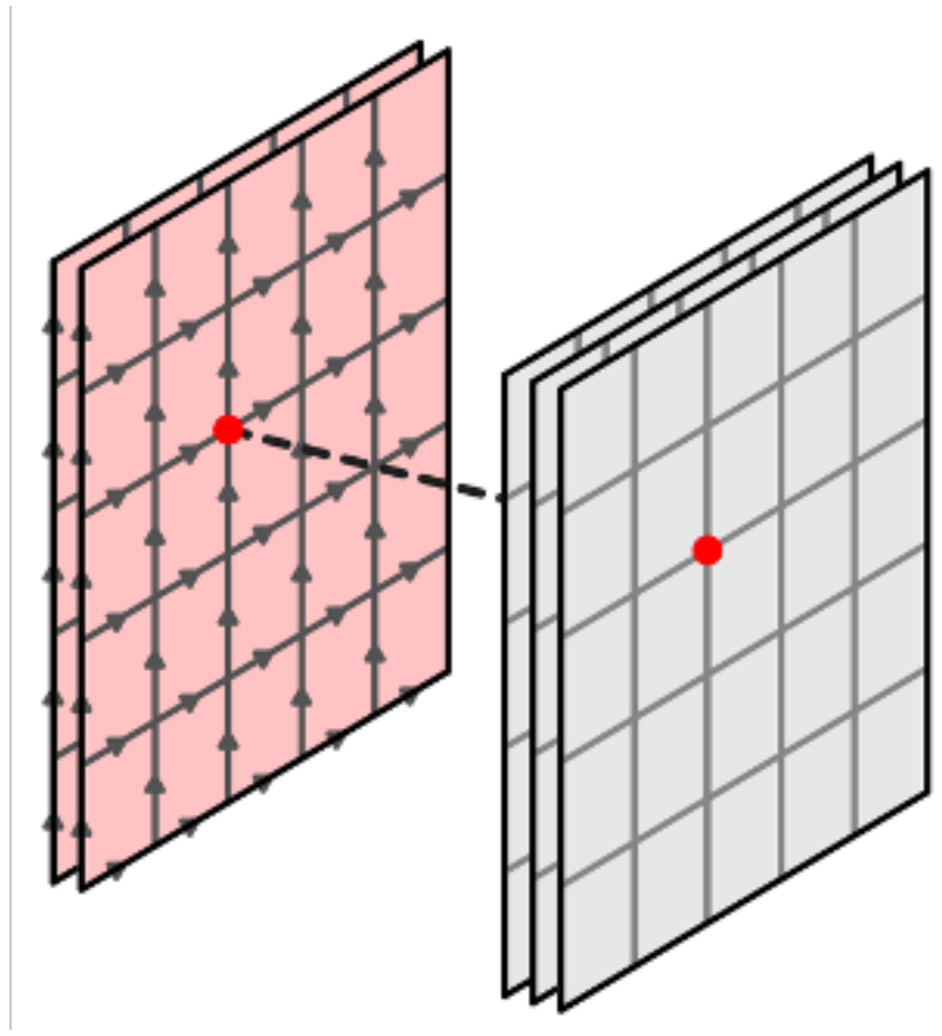
L-Bilin:



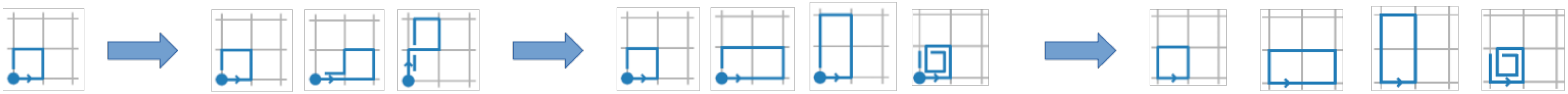
$$(\mathcal{U},\mathcal{W}) \times (\mathcal{U},\mathcal{W}') \to (\mathcal{U},\mathcal{W}'')$$

$$W_{x,i} \to \sum_{j,j',k} \alpha_{i,j,j',k} W_{x,j} W'_{x+k\cdot\mu,j'}.$$

Trace:

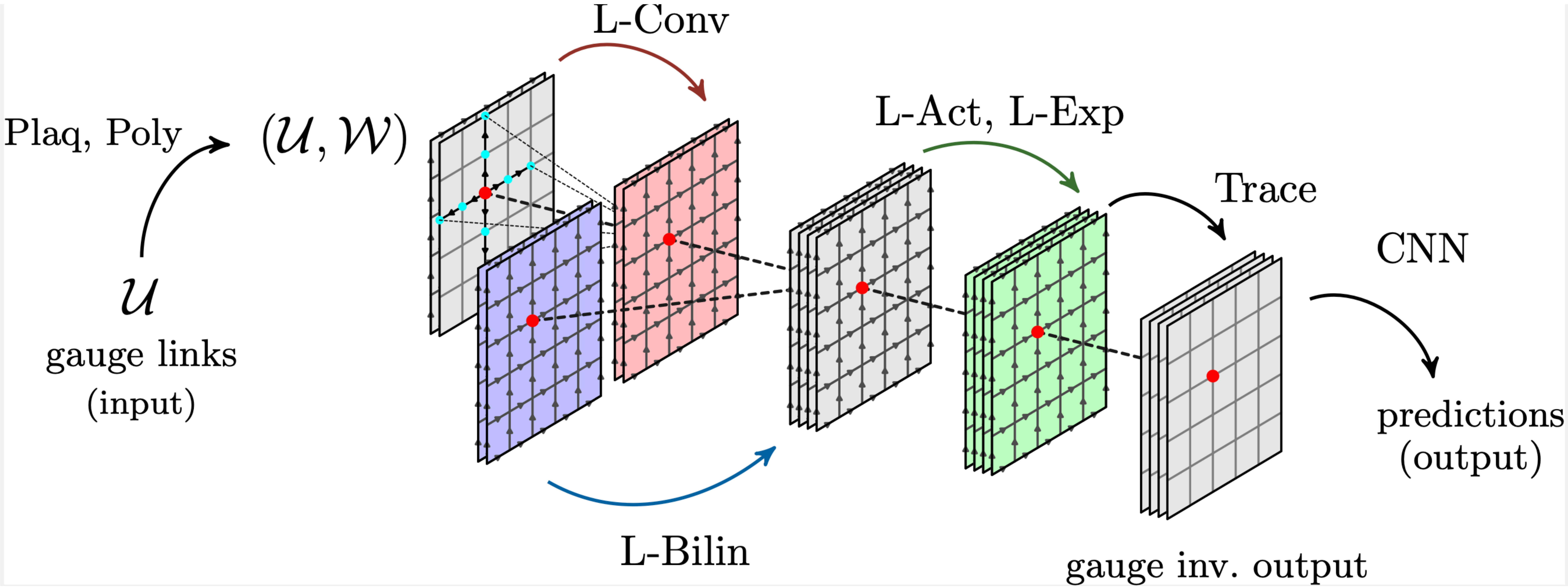


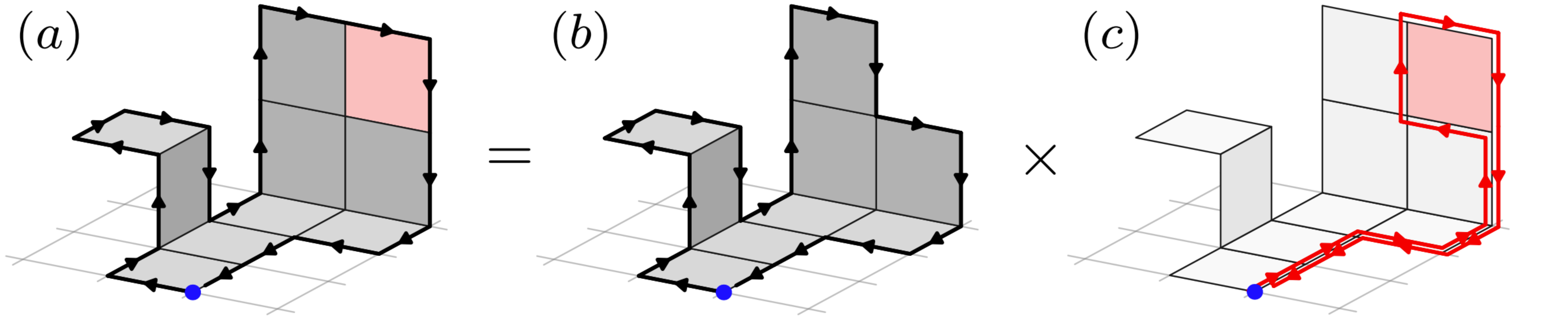
$$w_{\mathbf{x},i} = \mathrm{Tr}\, W_{\mathbf{x},i} \in \mathbb{C}$$



Machine learning the FP action

ML architecture: Lattice gauge equivariant Convolutional Neural Network (L-CNN)





Machine learning the FP action: FP data

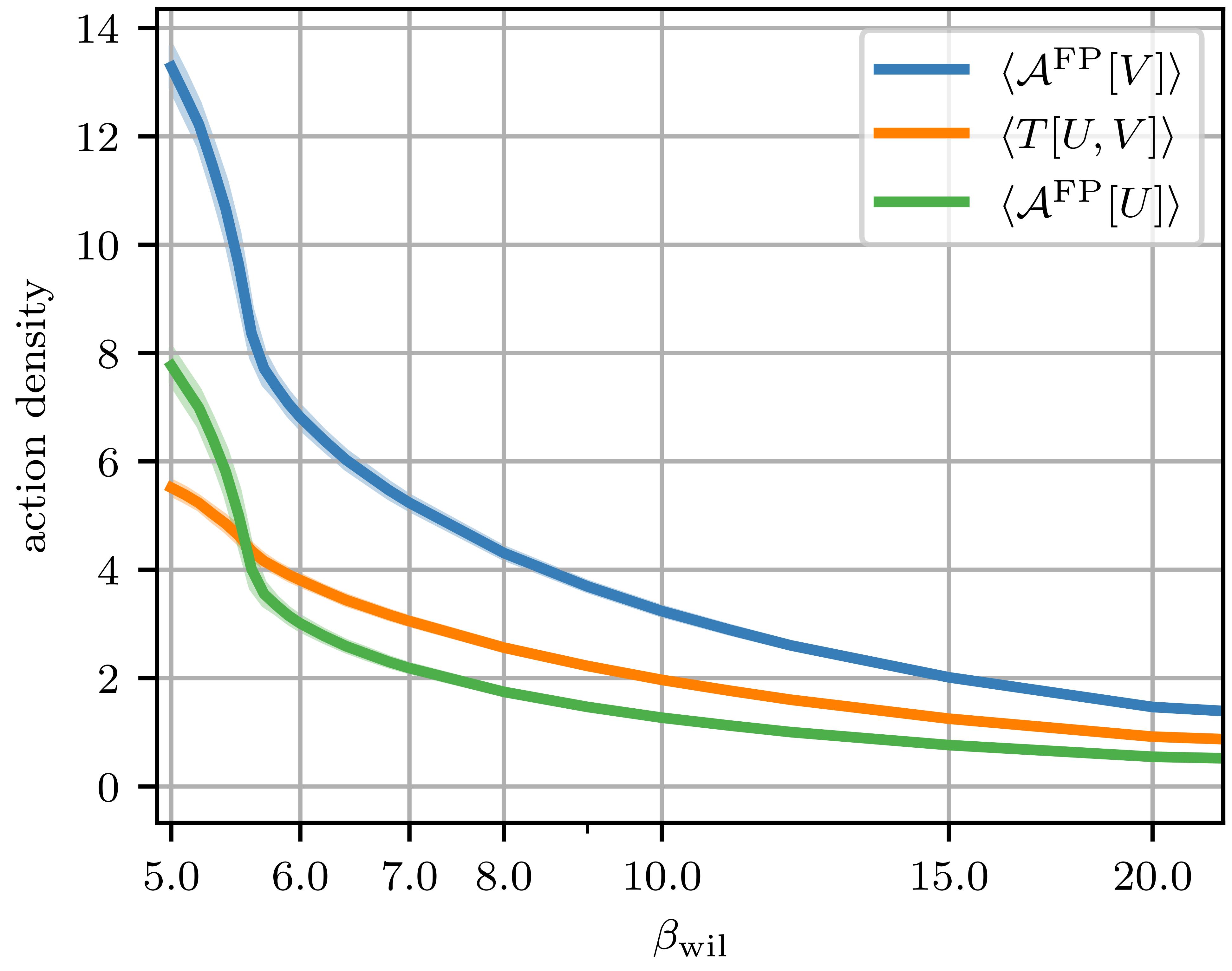
For given coarse V the FP action value determined by the minimizing confs. $U, U', \ldots$:

$$A^{\mathrm{FP}}[V] = \min_{\{U\}} \{A^{\mathrm{FP}}[U] + T[U,V]\} = \min_{\{U',U\}} \{A^{\mathrm{FP}}[U'] + T[U',U] + T[U,V]\}$$

Instead use approximate FP action values $A^{\mathrm{FP}}[U]$ in the first iteration

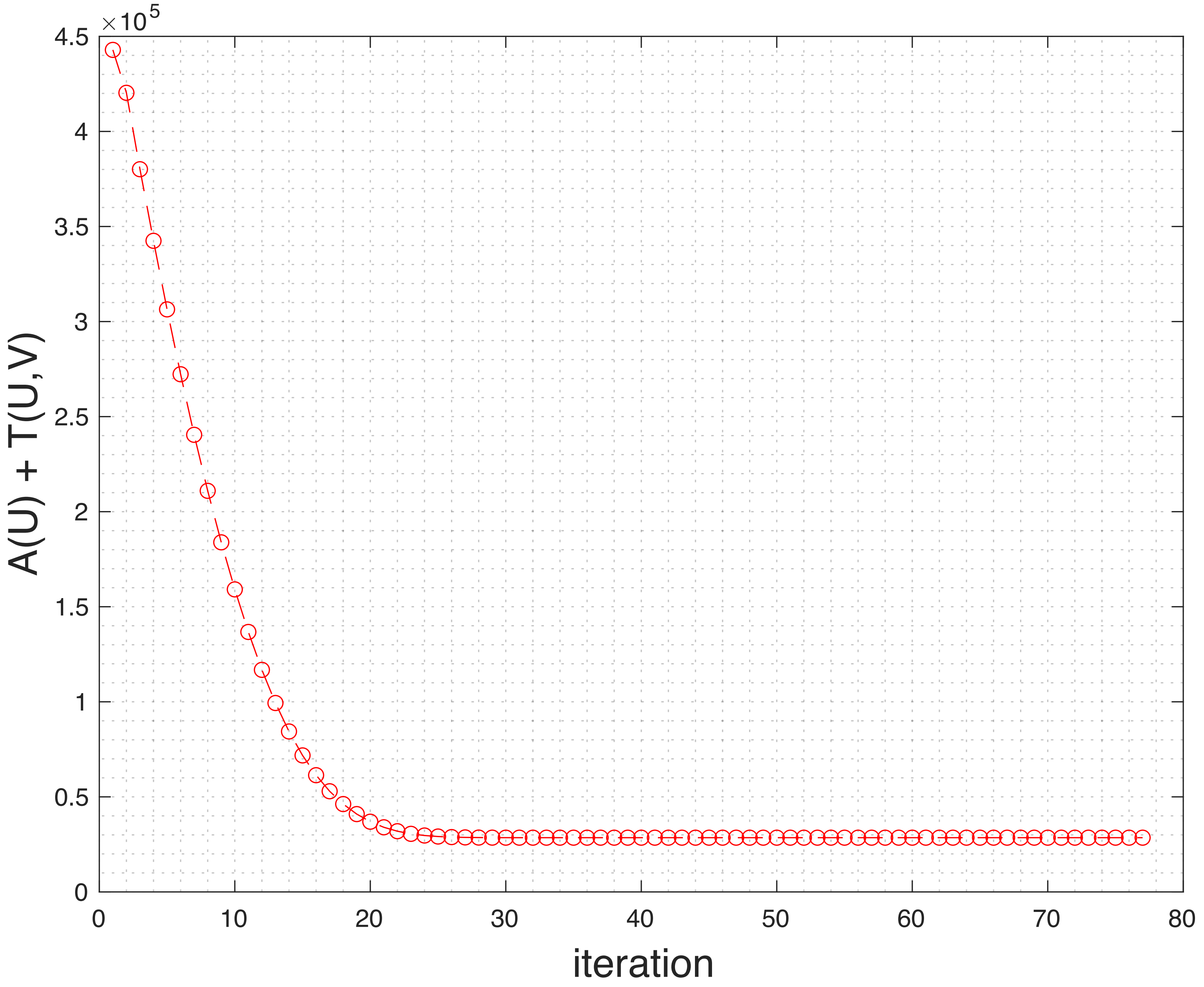
Minimising the RHS is crucial for generating the training data:

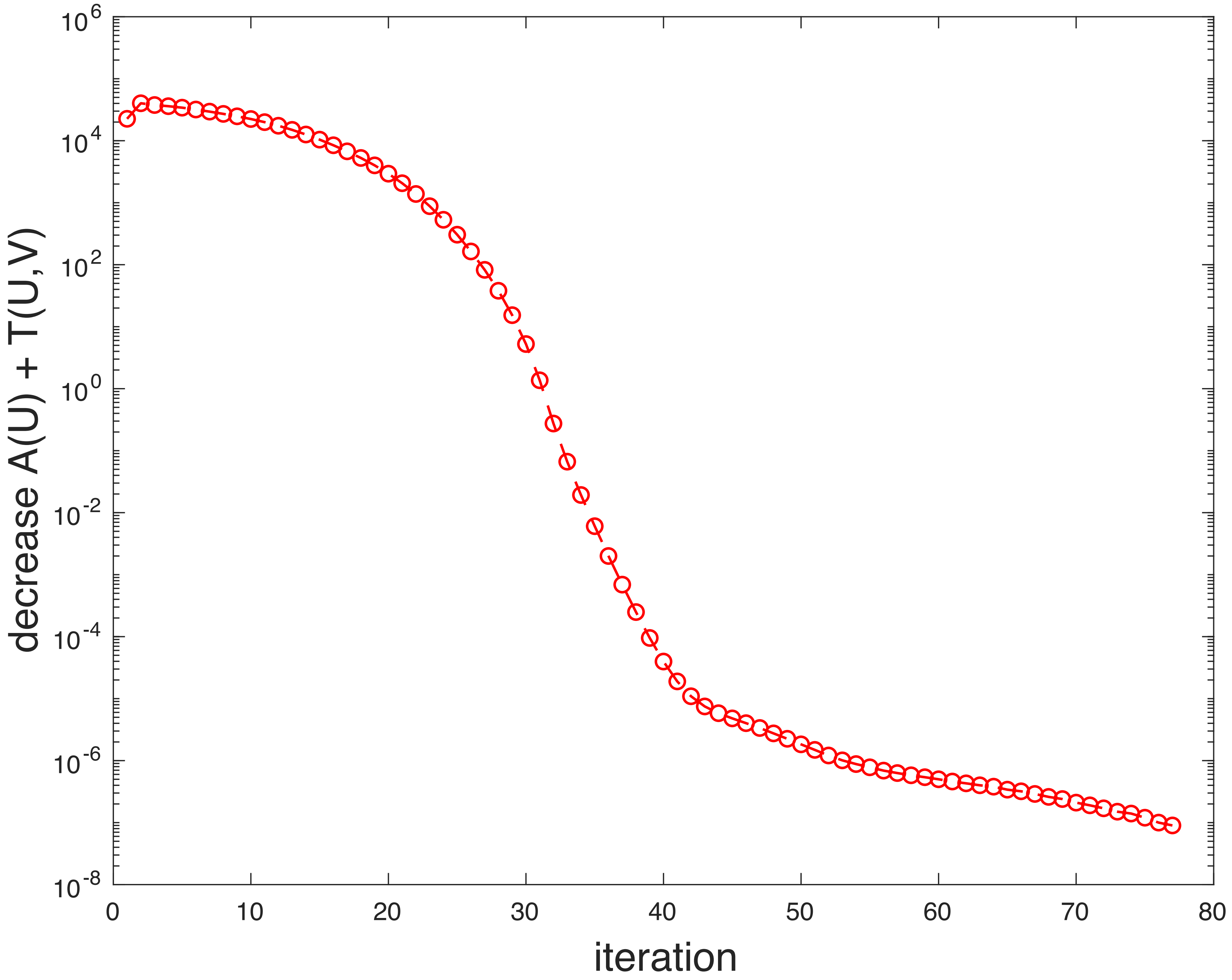
- generate ensembles of coarse gauge configurations for a range of field fluctuations
- find minimizing fine configuration



Machine learning the FP action: FP data

Minimisation evolution of fine configuration U for fixed coarse configuration V

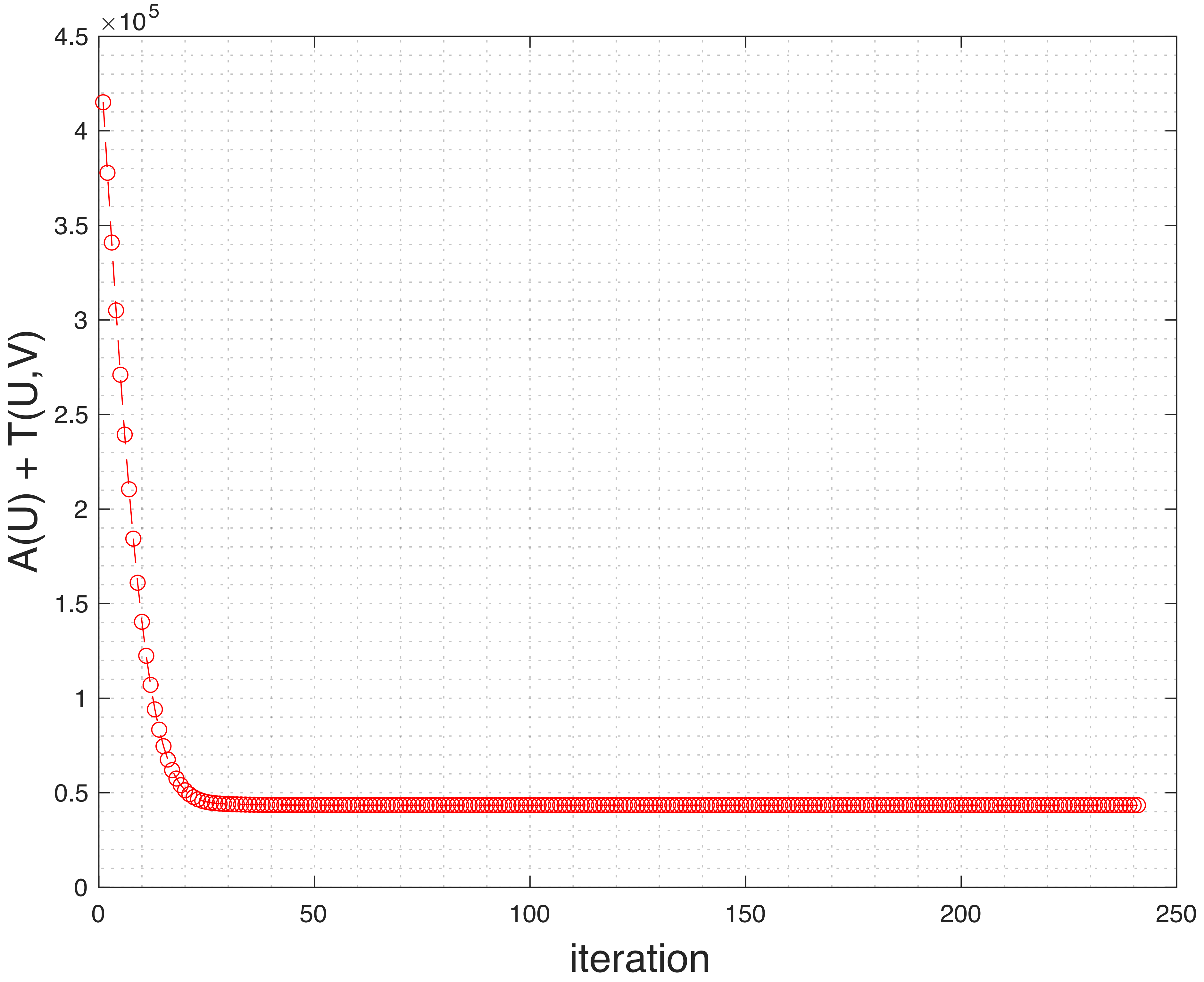


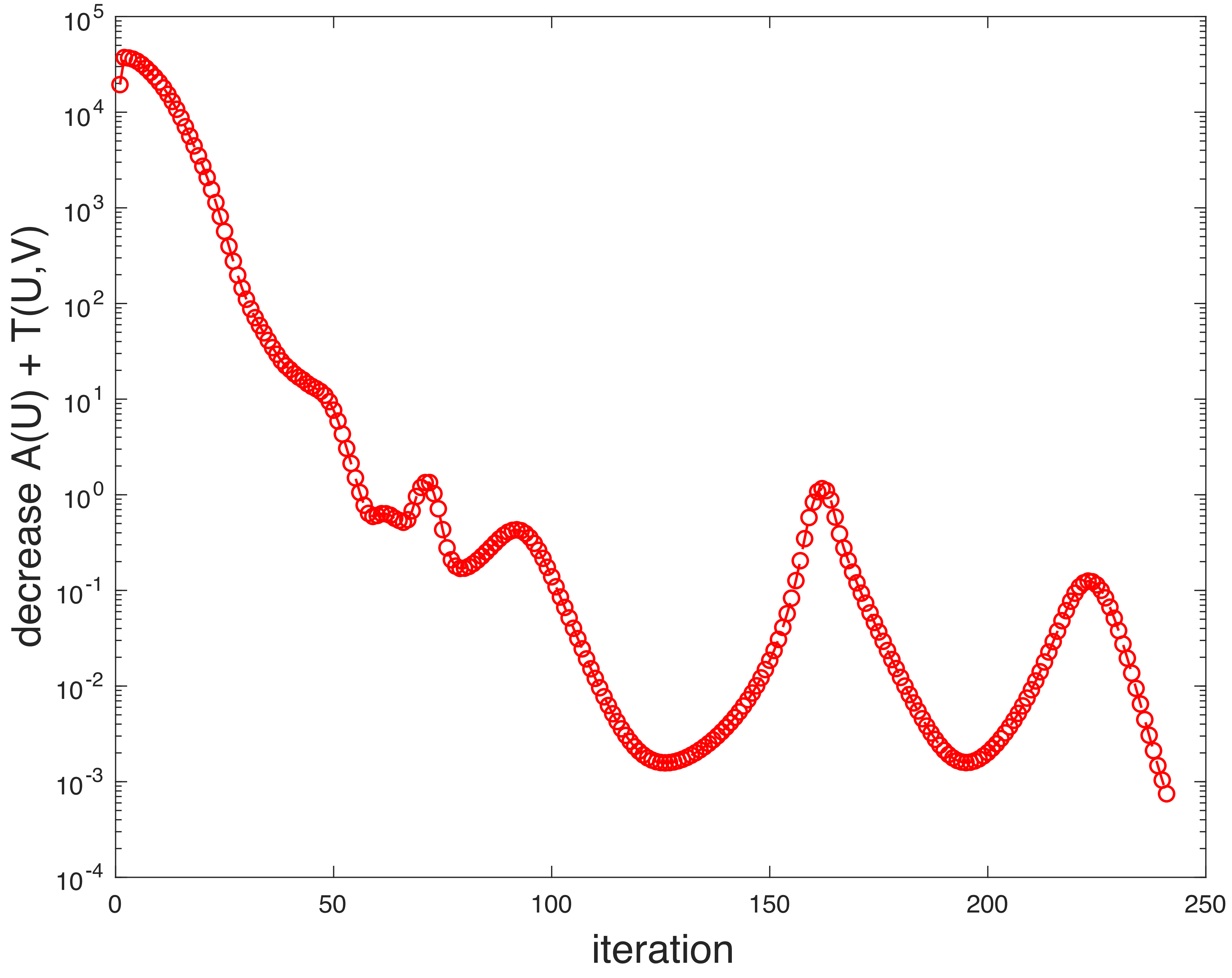


Coarse configuration: SU(3), $V = 8^4, \beta^{\text{Wil}} = 6.0$ $\Rightarrow$ lattice spacing $a \simeq 0.10$ fm

Machine learning the FP action: FP data

Minimisation evolution of fine configuration U for fixed coarse configuration V

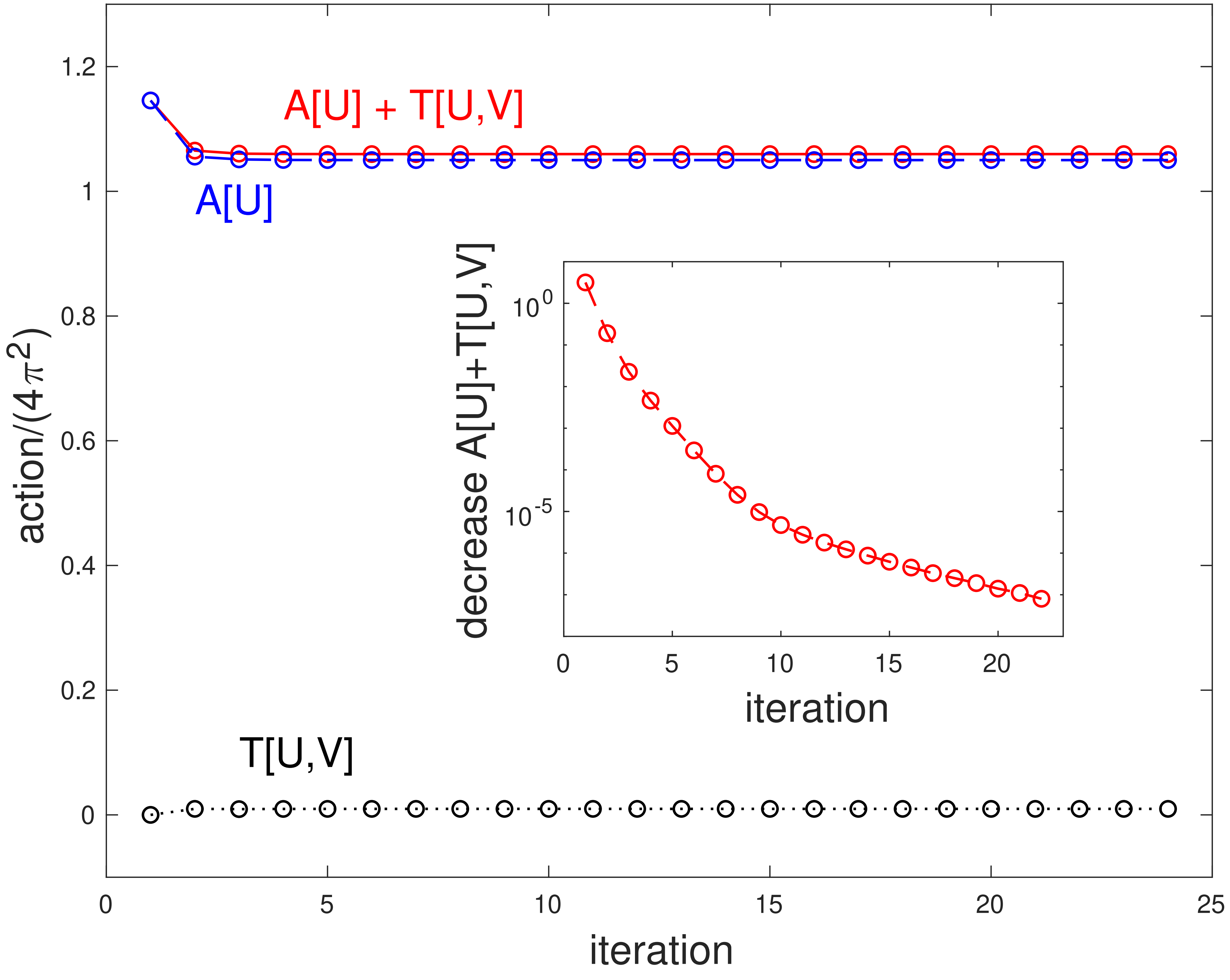




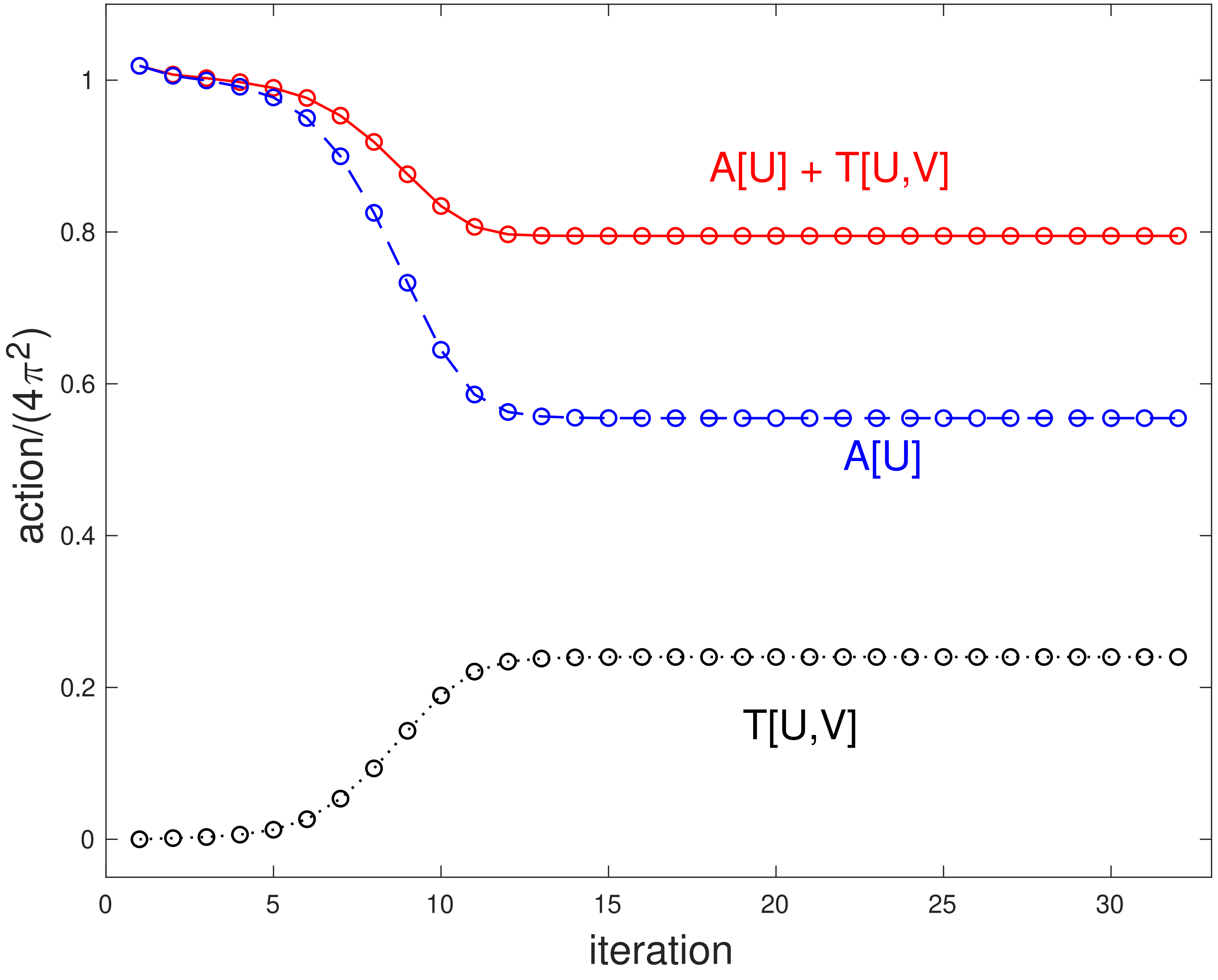
Coarse configuration: SU(3), $V = 8^4, \beta^{\mathrm{Wil}} = 5.4$ $\Rightarrow$ lattice spacing $a \simeq 0.25$ fm

Machine learning the FP action: FP data

Minimisation of instanton configuration with ρ/a' on 16^4 blocked to 8^4:



$\rho/a' = 3.0$



$\rho/a' = 1.5$

Machine learning the FP action: FP data

Use the exact FP action values for training, plus the derivatives of the FP action:

$$\frac{\delta A^{\mathrm{FP}}[V]}{\delta V^a_{x,\mu}} = \frac{\delta T[U,V]}{\delta V^a_{x,\mu}} = -\kappa\,\mathrm{Re\ Tr}(it^a\, V_{x,\mu} Q^\dagger_{x,\mu}) \qquad Q^\dagger_{x,\mu} = Q^\dagger_{x,\mu}[U]$$

$\Rightarrow$ yields 4 x 8 x Volume (link) (color) (position) data per configuration

Gauge invariance of A^{FP} yields conserved local quantity via Noether's theorem:

$$D^{FP}_{x,\mu} = \sum_a t^a \frac{\delta A^{FP}[V]}{\delta V^a_{x,\mu}} \quad \Longrightarrow \quad \sum_\mu \mathscr{D}^B_\mu D^{FP}_{x,\mu}[V] = 0$$

$\Rightarrow$ consistency check satisfied up to the accuracy in minimization

- FP action values
- FP action derivatives

$\Bigg\} \Rightarrow$ data set for supervised ML

Machine learning the FP action: Architecture

FP action parameterised by

$$A^{\text{L-CNN}}[V] = \sum_x A_x^{\text{pre}}[V] \sum_{n=0}^{\infty} b^{(n)} \left(N_x[V] - N_x[1]\right)^n$$

local output of L-CNN

with

$$A_x^{\text{pre}}[V] = \frac{1}{N_c} \sum_C \sum_{m=1}^{M} c^{(m)}[\text{ReTr}(a - U_{x,C})]^m$$

for example Wil, tlSym, …

Note: $N_x[V \to 1] - N_x[1] \simeq O(a^2)$

In practice we use:

$$A^{\text{L-CNN}}[V] = \sum_x A_x^{\text{pre}}[V] \exp\left(N_x[V] - N_x[1]\right)$$

Machine learning the FP action: Loss functions

ML loss function from two weighted contributions:

$$L_1 = \frac{1}{L^4 N_{cfg}} \sum_{i=1}^{N_{cfg}} \left| A^{\text{FP}}[V_i] - A^{\text{L-CNN}}[V_i] \right|$$

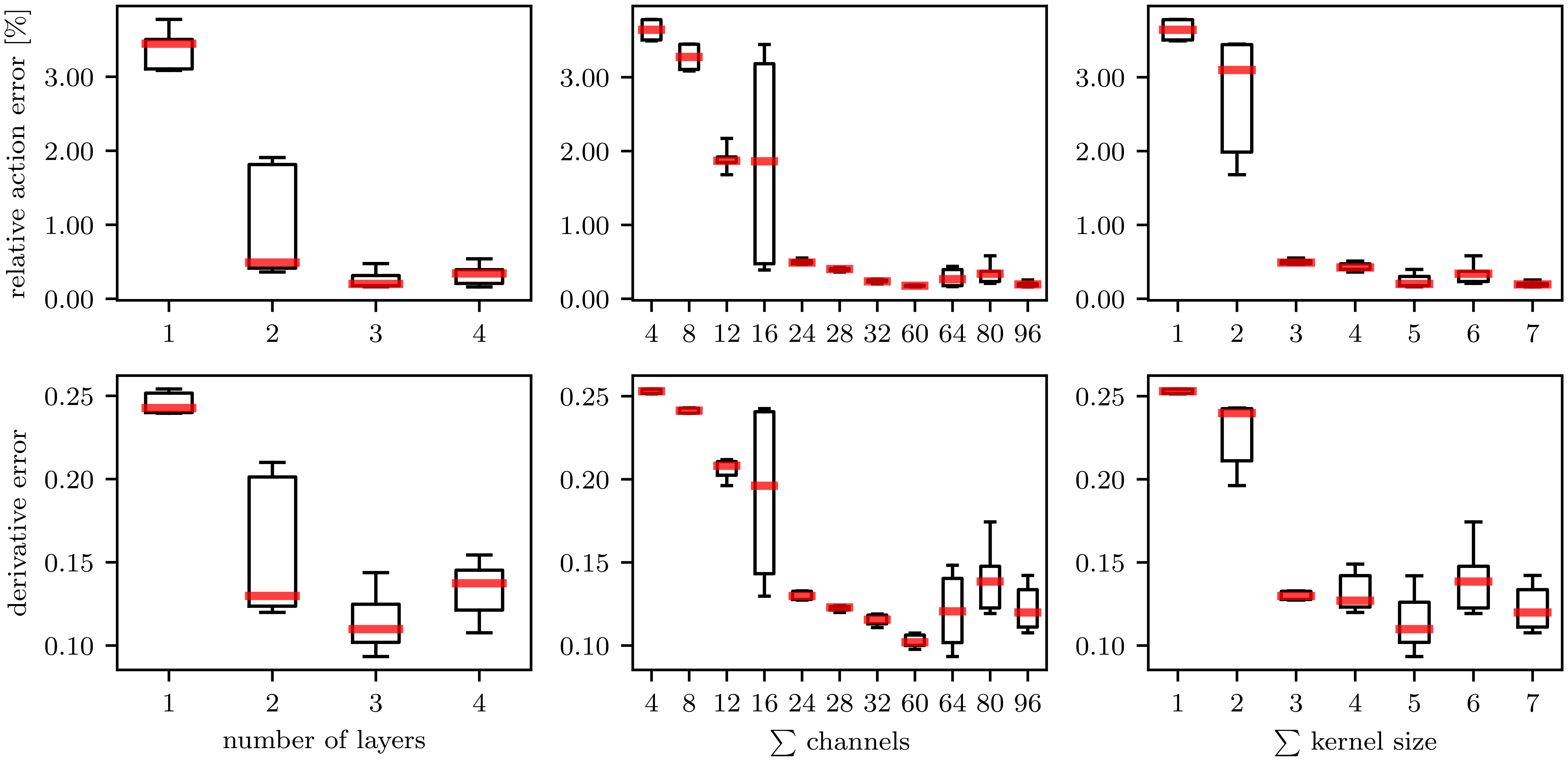
$$L_2 = \frac{1}{32 L^4 N_{cfg}} \sum_{i=1}^{N_{cfg}} \sum_{x,\mu} \text{Tr} \left[\left(D^{\text{FP}}_{x,\mu}[V_i] - D^{\text{L-CNN}}_{x,\mu}[V_i] \right)^2 \right]$$

$$L = w_1 L_1 + w_2 L_2$$

Technical point: derivatives in L-CNN are given through back propagation
(since they are derivatives of part of loss function w.r.t. input)

Machine learning the FP action: L-CNN

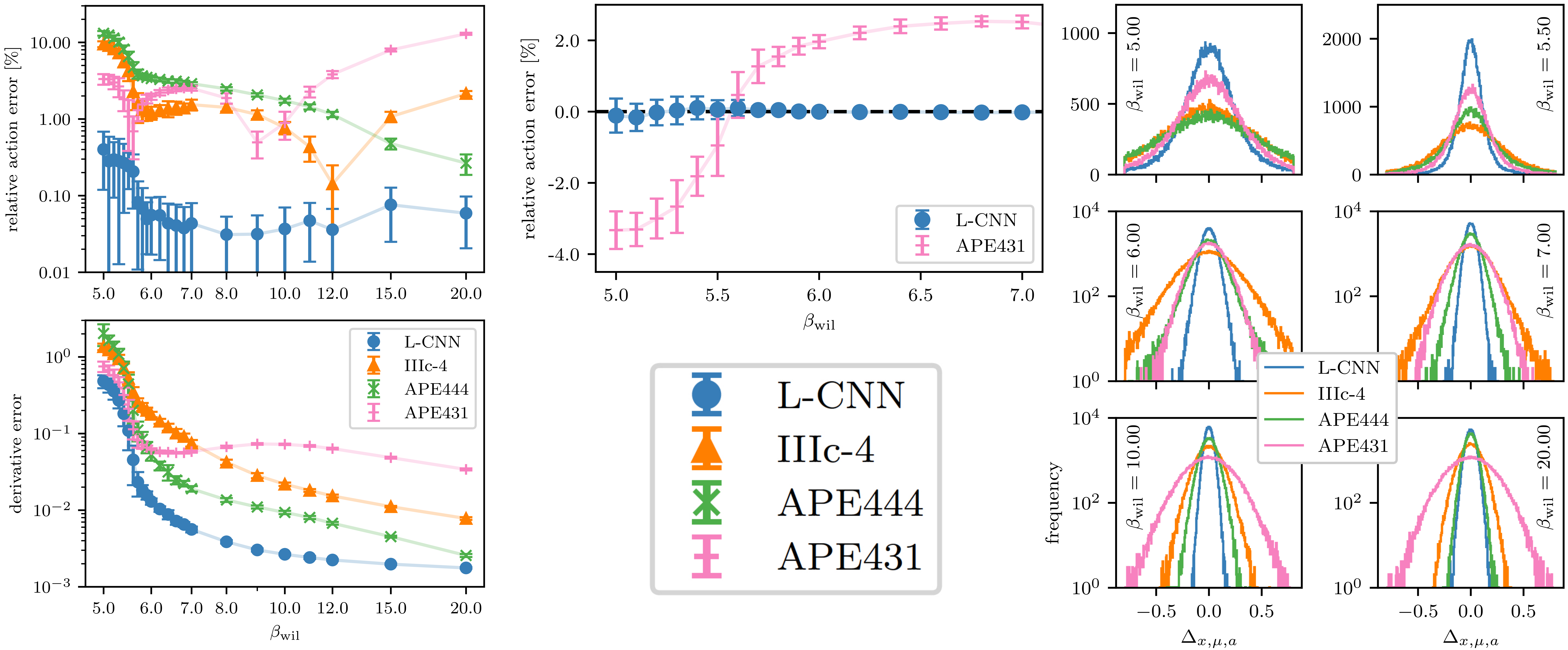
Architecture search:



⇒ 3 L-Bilin layers, kernel sizes {2,2,1}, output channels {12,24,24} with 443k parameters

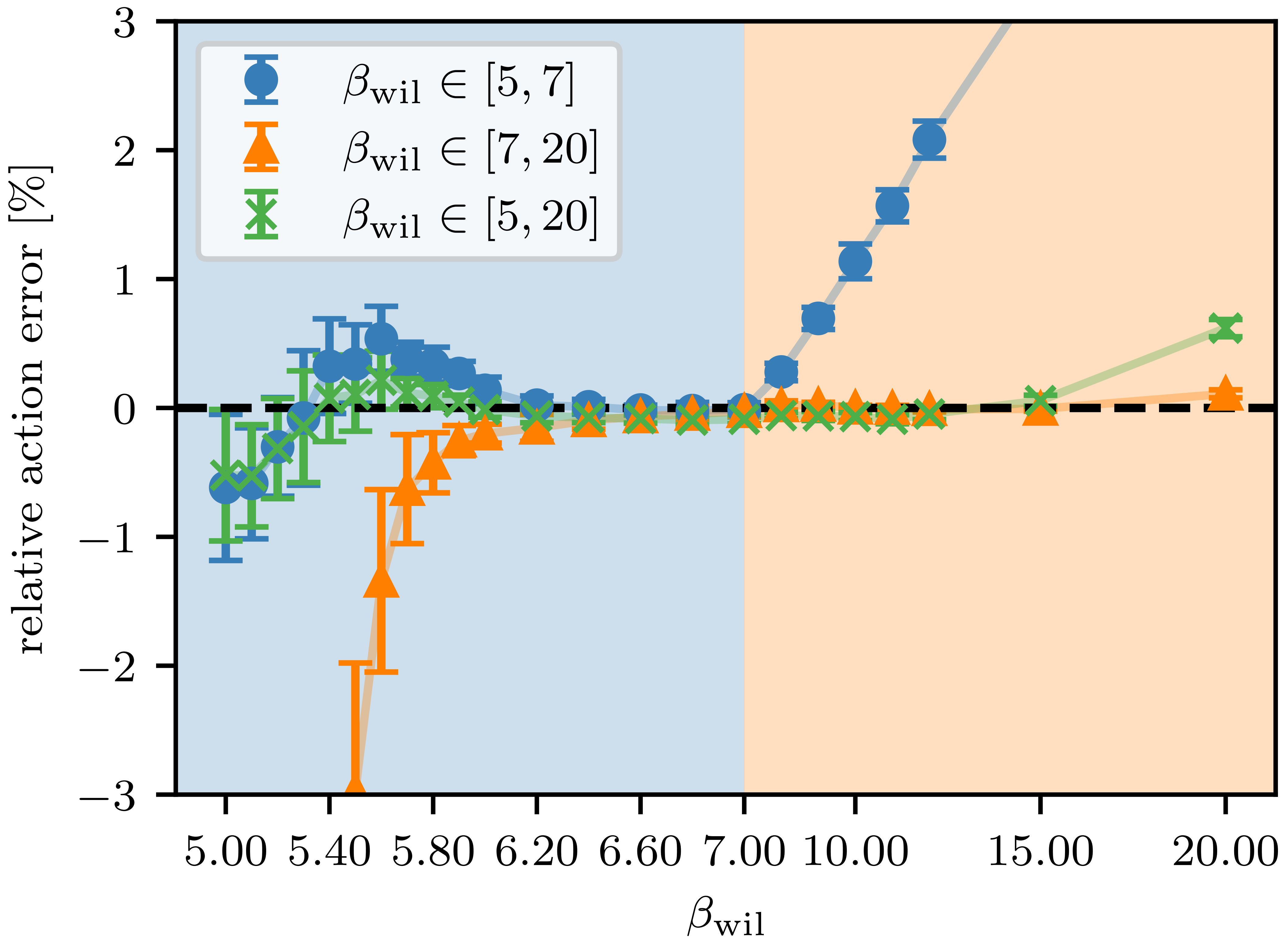
Machine learning the FP action: Results

Superiority of L-CNN over old parameterizations of FP action:



Machine learning the FP action: Results

Restricted training ranges:



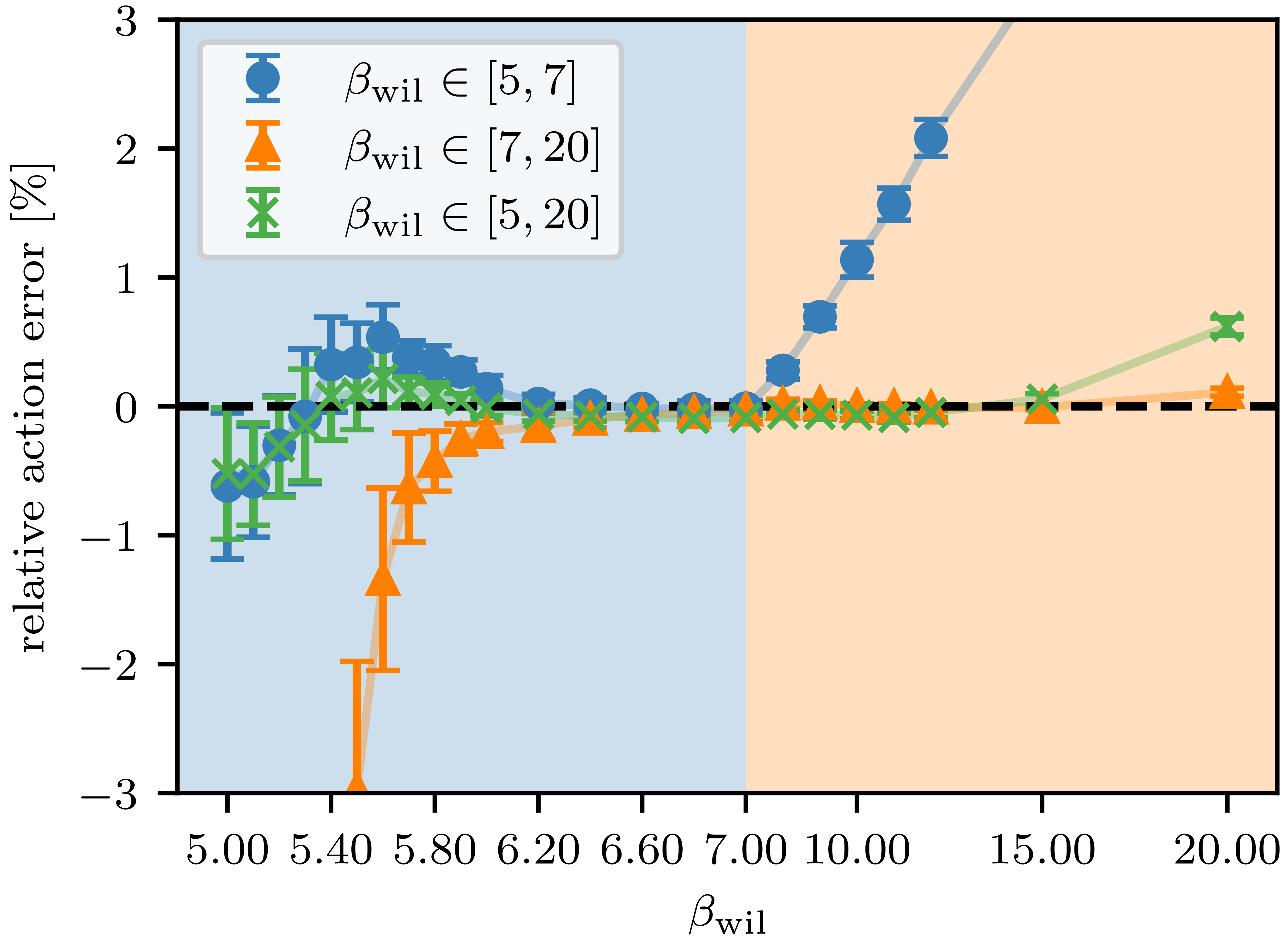
Transfer learning:

	relative error (test data)		
finetuned model	4^4	6^4	8^4
4^4	**0.178 %**	0.201 %	0.181 %
6^4	0.185 %	**0.196 %**	0.177 %
8^4	0.191 %	0.202 %	**0.176 %**

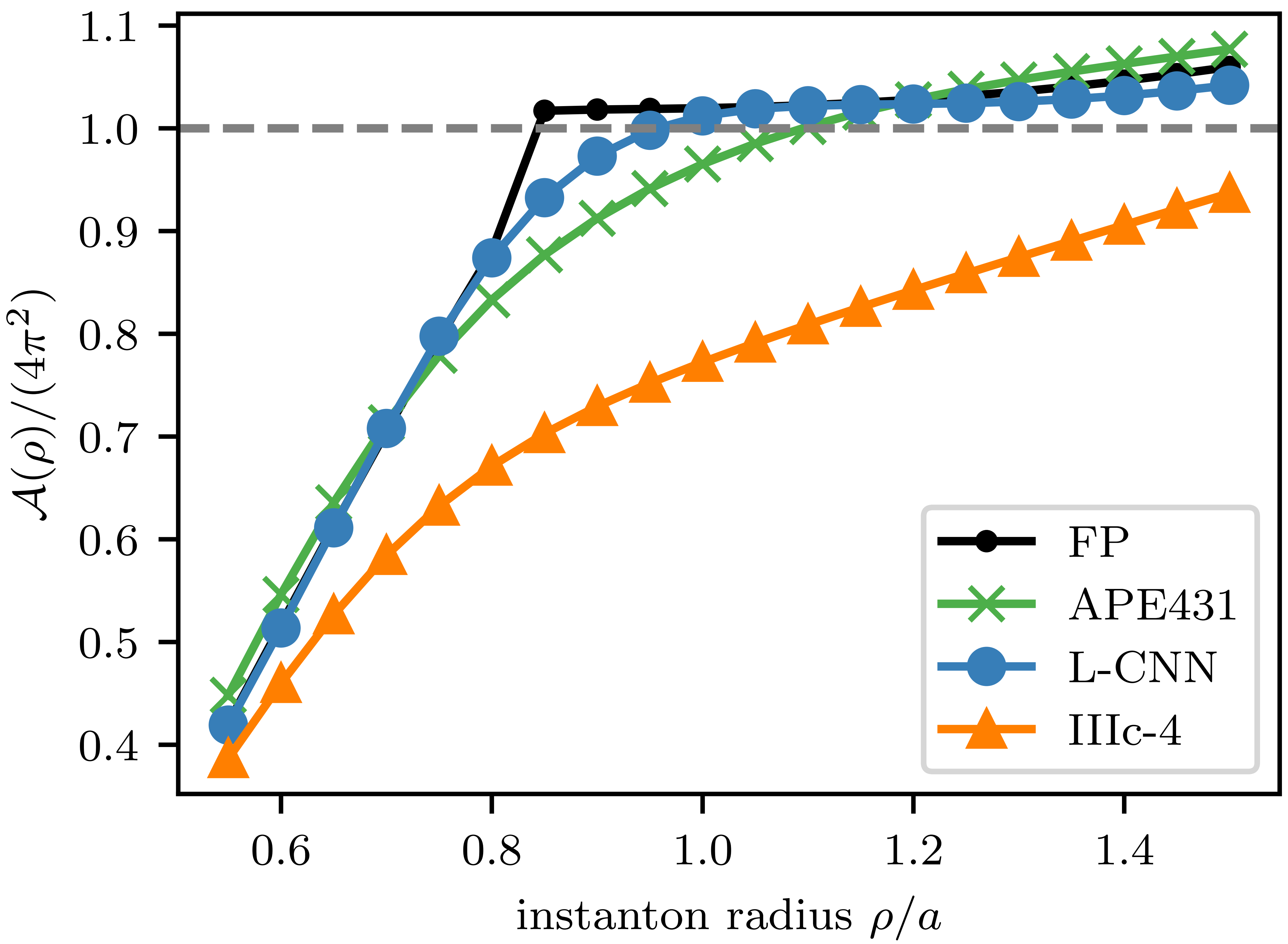
	derivative error (test data)		
finetuned model	4^4	6^4	8^4
4^4	$\mathbf{7.63 \times 10^{-2}}$	8.19×10^{-2}	8.22×10^{-2}
6^4	$\mathbf{7.39 \times 10^{-2}}$	7.93×10^{-2}	7.96×10^{-2}
8^4	$\mathbf{7.36 \times 10^{-2}}$	7.91×10^{-2}	7.93×10^{-2}

Machine learning the FP action: Results

Restricted training ranges:



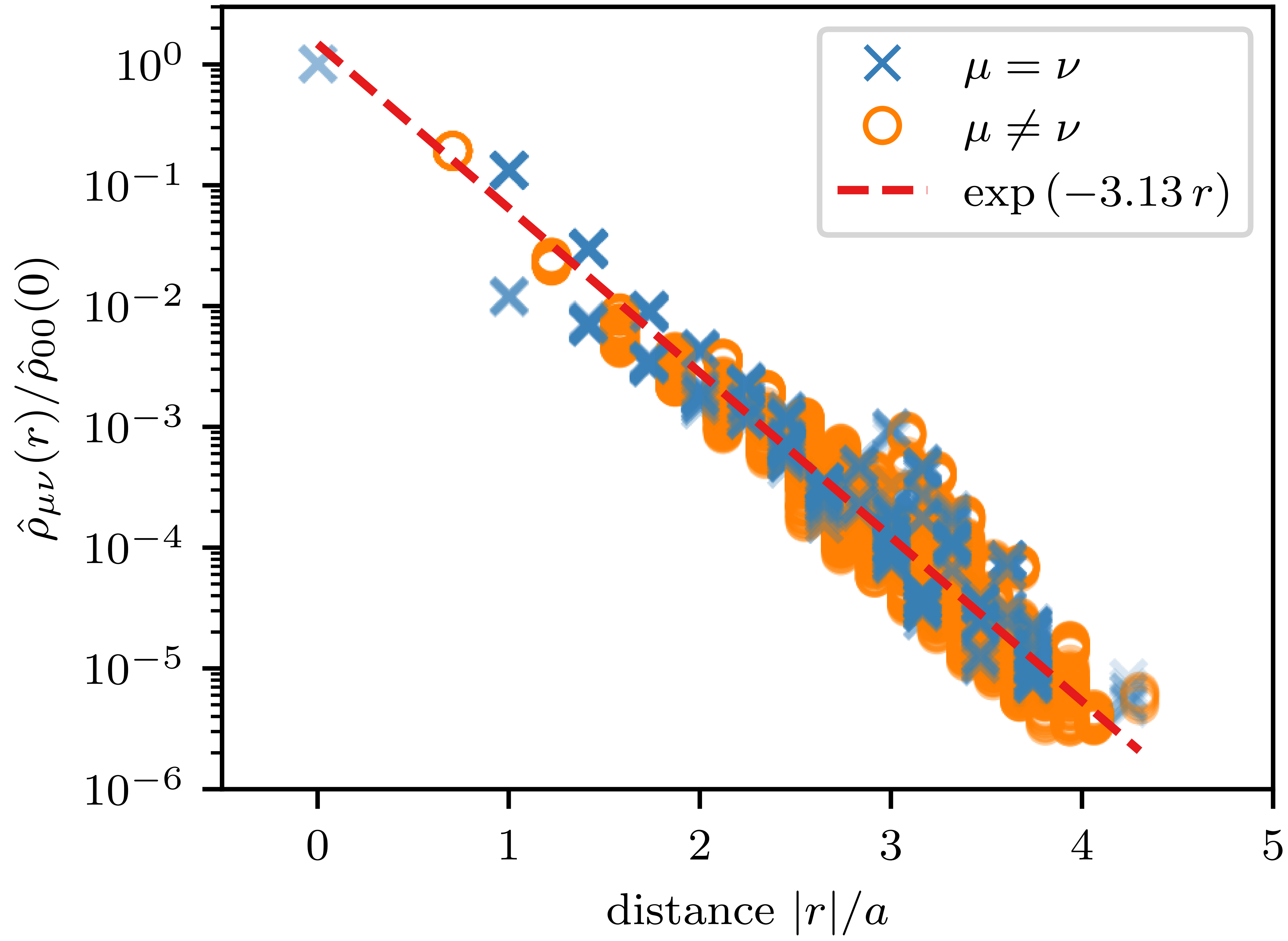
Finetuning on instantons:



⇒ new L-CNN parametrization is indeed very flexible and accurate

Machine learning the FP action: Locality

Locality of L-CNN trained FP action:



$$\hat{\rho}_{\mu\nu}(r) = \frac{1}{\sqrt{N_c^2 - 1}} \sqrt{\sum_{a,b} D_{\mu\nu}^{ab}(x,y) D_{\mu\nu}^{ab}(x,y)}$$

where $D_{\mu\nu}^{ab}(x,y) = \dfrac{\delta^2 A}{\delta V_{x,\mu}^{a} \delta V_{y,\nu}^{b}}$

- couplings fall off exponentially, as desired
- even on coarse configurations

Machine learning the FP action: Symmetries

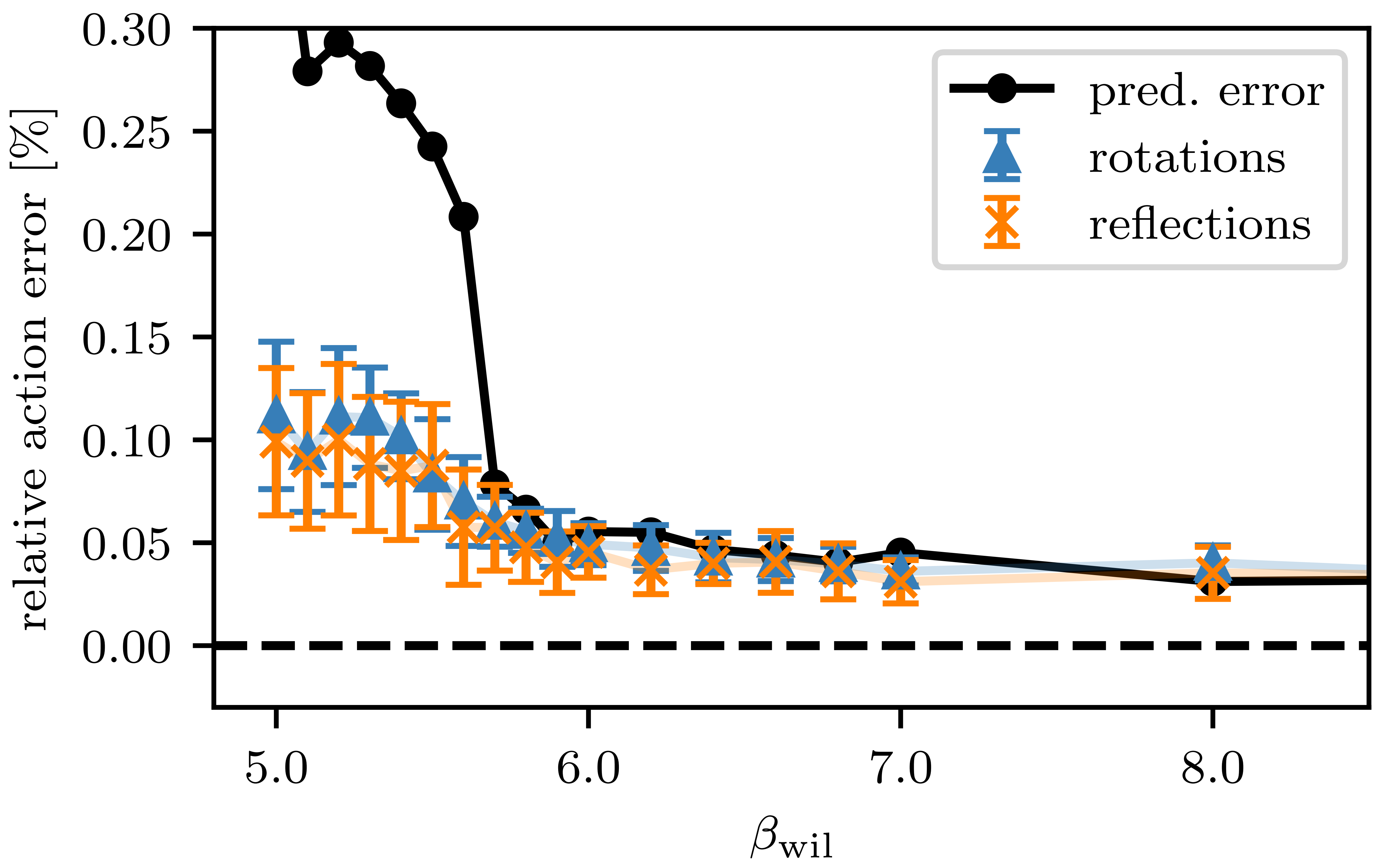
Test of lattice symmetries:

translations: $\Rightarrow A^{\text{L-CNN}}[U'_{(\text{shift})}] = A^{\text{L-CNN}}[U]$ by construction

rotations: $U \rightarrow U' = U_{(\text{rot})}$

reflections: $U \rightarrow U' = U_{(\text{refl})}$

a priori not present, but learned!



FP action with L-CNN:

So far, two questions were addressed:

- can the FP action be parametrised sufficiently well? ✓
- is the FP action sufficiently local for truncations to work? ✓

Could provide a solution to critical slowing down and topological freezing…

- how good are scaling properties of L-CNN FP action?

Availability of derivatives is the stepping stone for further developments:

- HMC, Langevin, GF (all based on derivatives)
- apply exact RGT step(s)

Part III:
Classically perfect gradient flow

Gradient flow

Use renormalized GF couplings as scaling quantities:

$$\frac{dA_\mu(t)}{dt} = D_\nu G_{\nu\mu} \qquad \langle t^2 E(t)\rangle = \frac{3(N^2-1)g^2}{128\pi^2}\left(1 + O(g^2)\right), \qquad E = \frac{1}{4}G_{\mu\nu}G_{\mu\nu}$$

On the lattice, artifacts are introduced through discretization of S^g, S^f, S^e:

$$\langle t^2 E(t)\rangle_a = \frac{3(N^2-1)g_0^2}{128\pi^2}\left(C(a^2/t) + O(g_0^2)\right)$$

Gradient flow

Use renormalized GF couplings as scaling quantities:

$$\frac{dA_\mu(t)}{dt} = D_\nu G_{\nu\mu} \qquad \langle t^2 E(t)\rangle = \frac{3(N^2-1)g^2}{128\pi^2}\left(1 + O(g^2)\right), \qquad E = \frac{1}{4}G_{\mu\nu}G_{\mu\nu}$$

On the lattice, artifacts are introduced through discretization of S^g, S^f, S^e:

$$\langle t^2 E(t)\rangle_a = \frac{3(N^2-1)g_0^2}{128\pi^2}\left(C(a^2/t) + O(g_0^2)\right)$$

⇒ turns out that GF with FP action is classically perfect!

Gradient flow at tree level

At tree level, the flow equation reads: [Fodor, Holland, et al., *JHEP* 09 (2014) 018, 1406.0827]

$$\frac{dA_\mu(p,t)}{dt} = -\left(S^f_{\mu\nu}(p) + \mathcal{G}_{\mu\nu}\right) A_\nu(p,t)$$

with the solution

$$A_\mu(p,t) = \left[\exp\left\{-t\left(S^f(p) + \mathcal{G}\right)\right\}\right]_{\mu\nu} A_\nu(p,0)$$

leading to

$$\langle t^2 E(t)\rangle_a = \frac{(N^2-1)}{2} g_0^2 t^2 \int_{-\pi/a}^{\pi/a} \frac{d^4p}{(2\pi)^4} \mathrm{Tr}\left[e^{-t(S^f+\mathcal{G})}(S^g+\mathcal{G})^{-1} e^{-t(S^f+\mathcal{G})} \cdot S^e\right]$$

and with $S^f = S^g = S^e$:

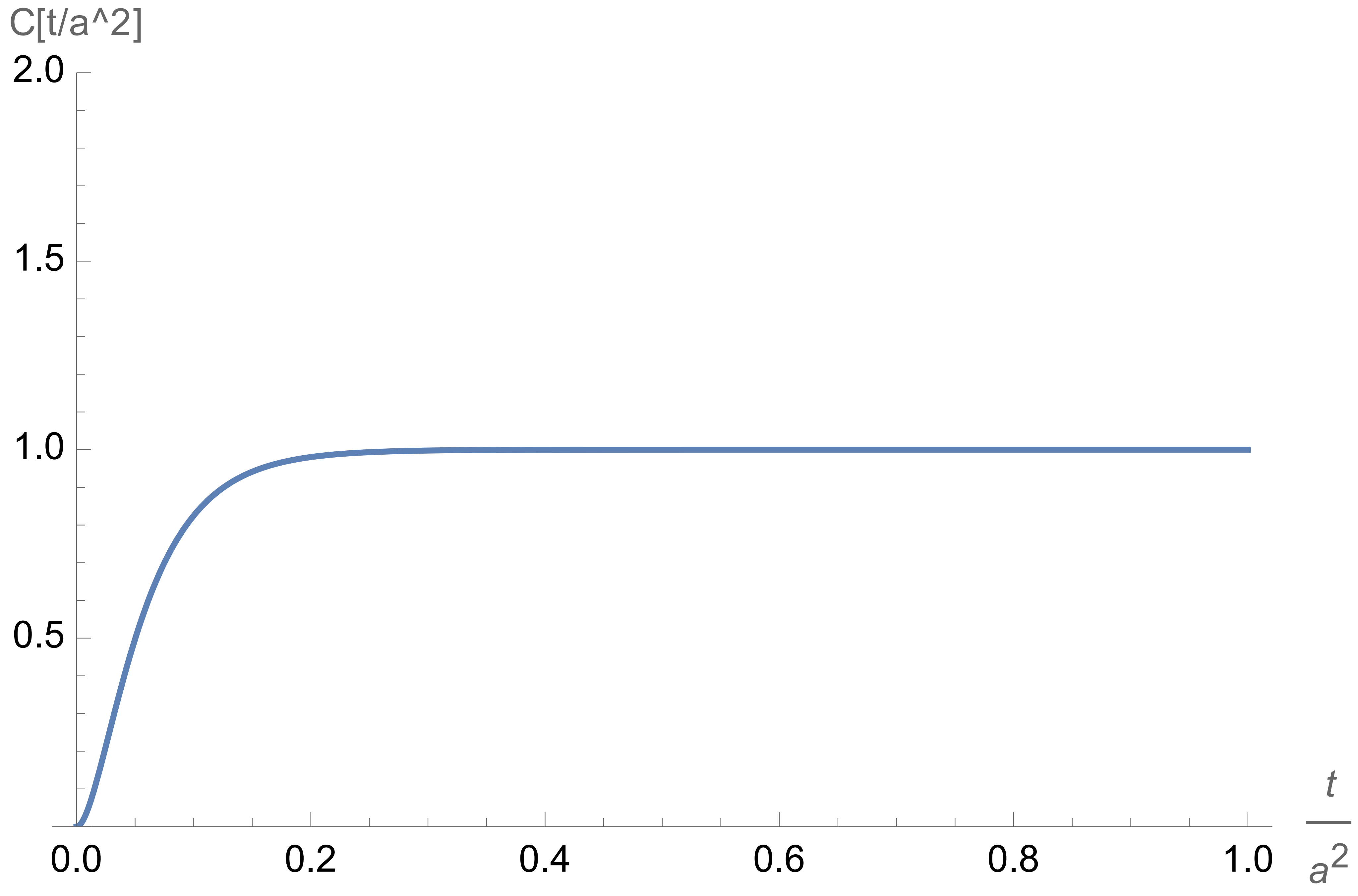
$$C(a^2/t) = \frac{64\pi^2}{3} t^2 \int_{-\pi/a}^{\pi/a} \frac{d^4p}{(2\pi)^4} \mathrm{Tr}\left[e^{-2t(S^f+\mathcal{G})}\right]$$

Gradient flow at tree level

Choose $S^f = S^g = S^e$

$S^{cutoff} \hat{=} \delta_{\mu\nu} p^2 - p_\mu p_\nu$:

$$\Rightarrow C(a^2/t) = \frac{64}{\pi^2} \left(\frac{t}{a^2}\right)^2 \left(\int_{-\pi/a}^{\pi/a} \frac{dp}{2\pi} e^{-2p^2 t} \right)^4$$

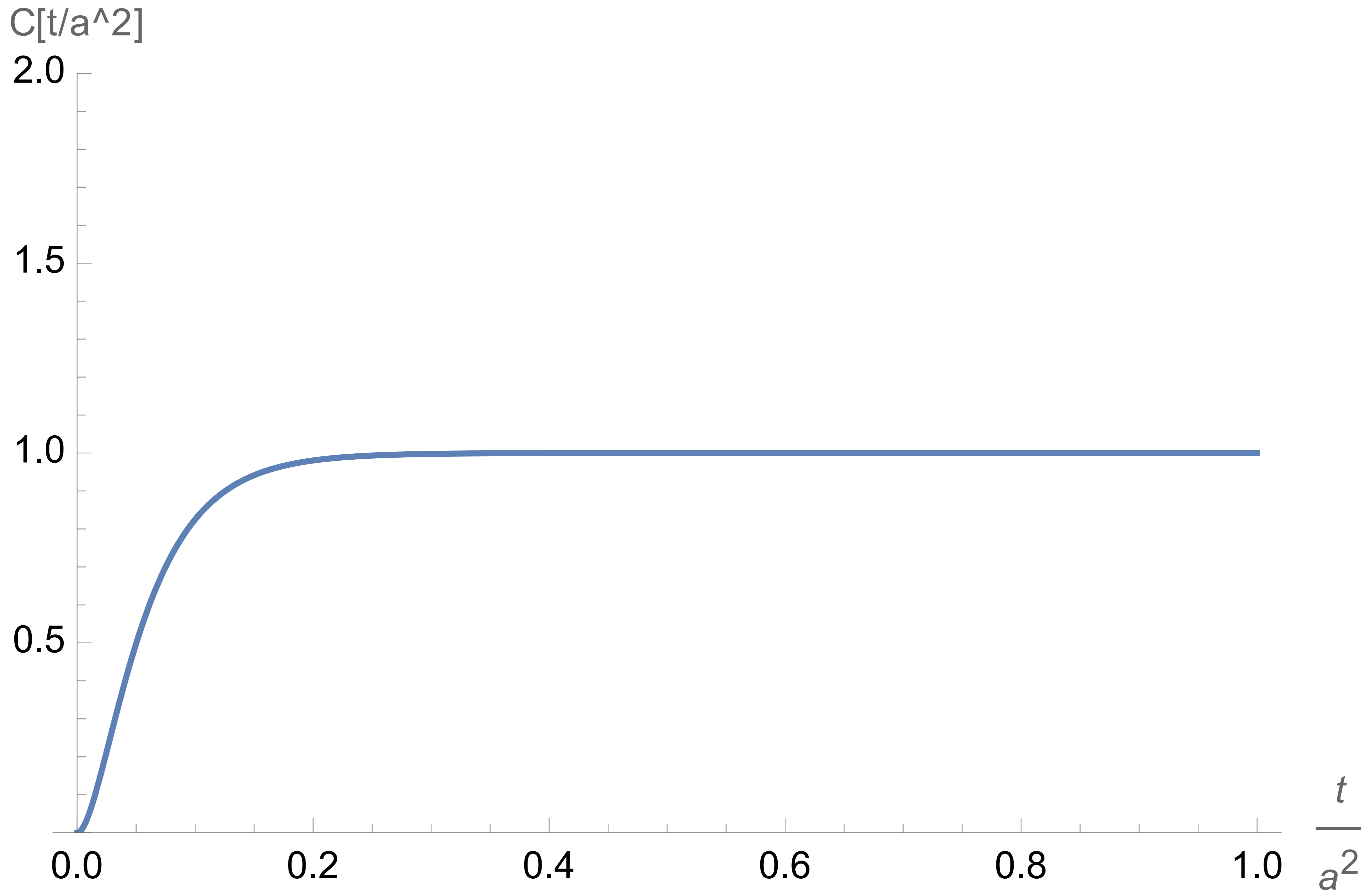


Gradient flow at tree level

Choose $S^f = S^g = S^e$

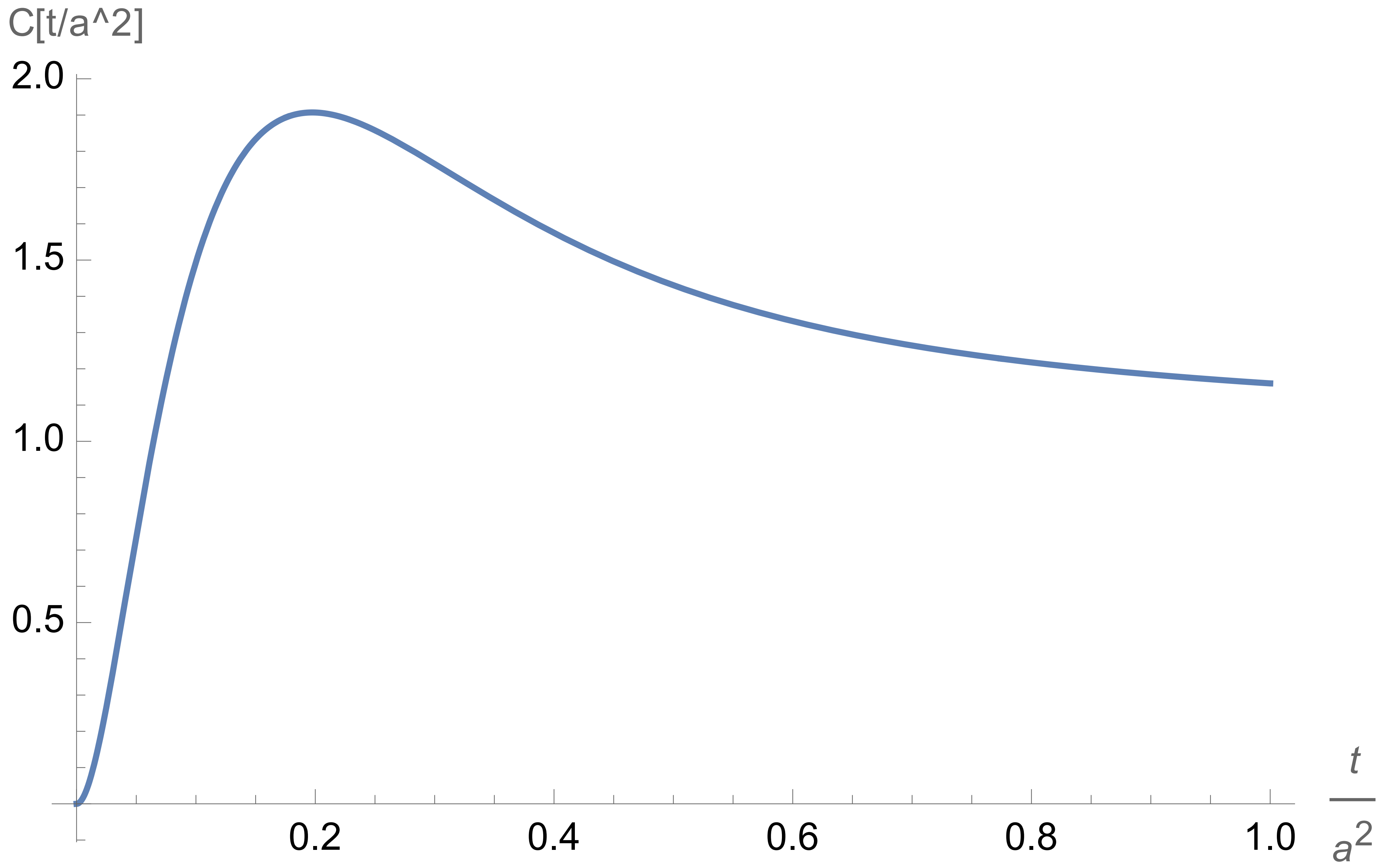
$S^{cutoff} \hat{=} \delta_{\mu\nu} p^2 - p_\mu p_\nu$:

$$\Rightarrow C(a^2/t) = \frac{64}{\pi^2}\left(\frac{t}{a^2}\right)^2 \left(\int_{-\pi/a}^{\pi/a} \frac{dp}{2\pi} e^{-2p^2 t}\right)^4$$



$S^{Wilson} \hat{=} \delta_{\mu\nu} \hat{p}^2 - \hat{p}_\mu \hat{p}_\nu$, $\quad \hat{p}_\mu = 2/a \sin(ap_\mu/2)$:

$$\Rightarrow C(a^2/t) = \frac{64}{\pi^2}\left(\frac{t}{a^2}\right)^2 \left(e^{-4t/a^2} I_0\left(4\,t/a^2\right)\right)^4$$



Gradient flow with FP action

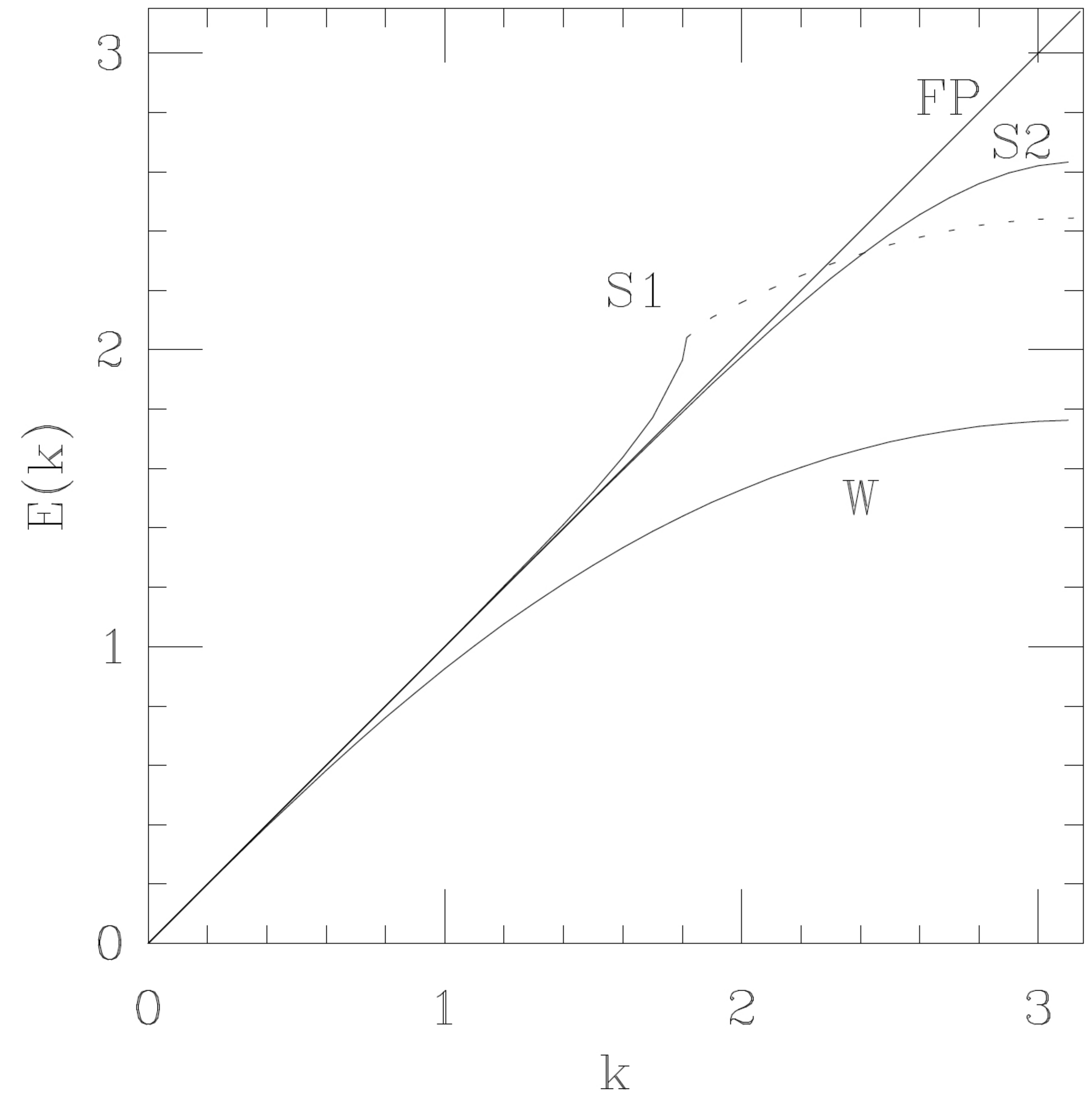
Choose $S^f = S^g = S^e = S^{FP}$

Iterating the FP equation in the quadratic approximation:

$$D'_{\mu\nu}(p) = \frac{1}{16}\sum_{l=0}^{1}\left[\omega\left(\frac{p+2\pi l}{2}\right) D\left(\frac{p+2\pi l}{2}\right)\omega^\dagger\left(\frac{p+2\pi l}{2}\right)\right]_{\mu\nu} + \frac{1}{\kappa}\delta_{\mu\nu}$$

After n iterations:

$$D^{(n)}_{\mu\nu}(p) \sim \left[\Omega^{(n)}\left(\frac{p+2\pi l}{2^n}\right)\Omega^{(n)\dagger}\left(\frac{p+2\pi l}{2^n}\right)\right]_{\mu\nu}\frac{1}{(p+2\pi l)^2}$$



The poles determine the dispersion relation:

$\forall p$ sums over l generate tower of poles $\Rightarrow$ full relativistic spectrum recovered

$$\Rightarrow C^{FP}(a^2/t) = \frac{64}{\pi^2}\left(\frac{t}{a^2}\right)^2\left(\int_{-\infty}^{\infty}\frac{dp}{2\pi}e^{-2p^2 t}\right)^4 = 1$$

Part IV:
HMC results and GF data

Hybrid Monte Carlo (HMC)

$$H(p,U) = p^2/2 + \mathcal{A}(U)$$

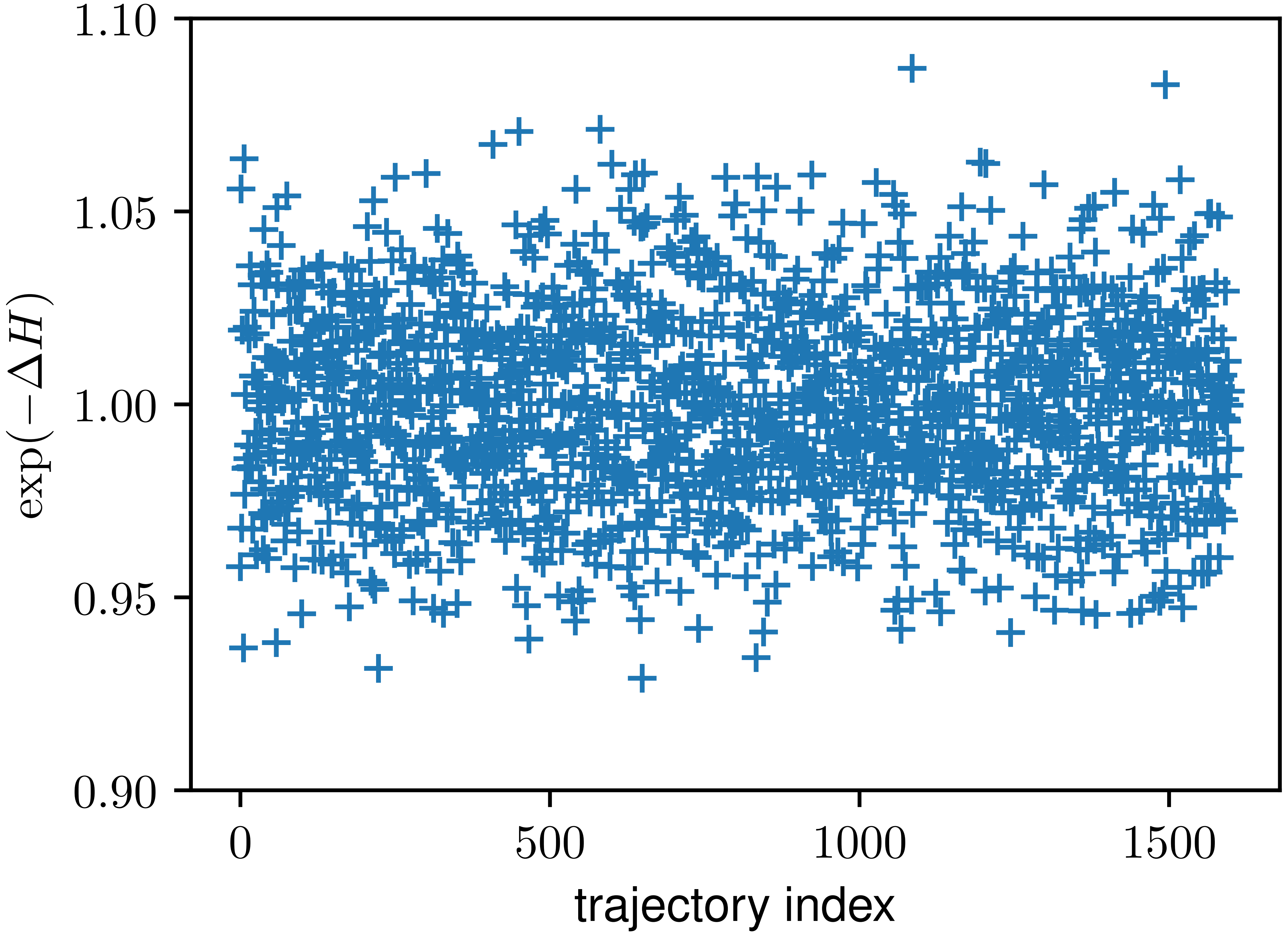
add momentum p

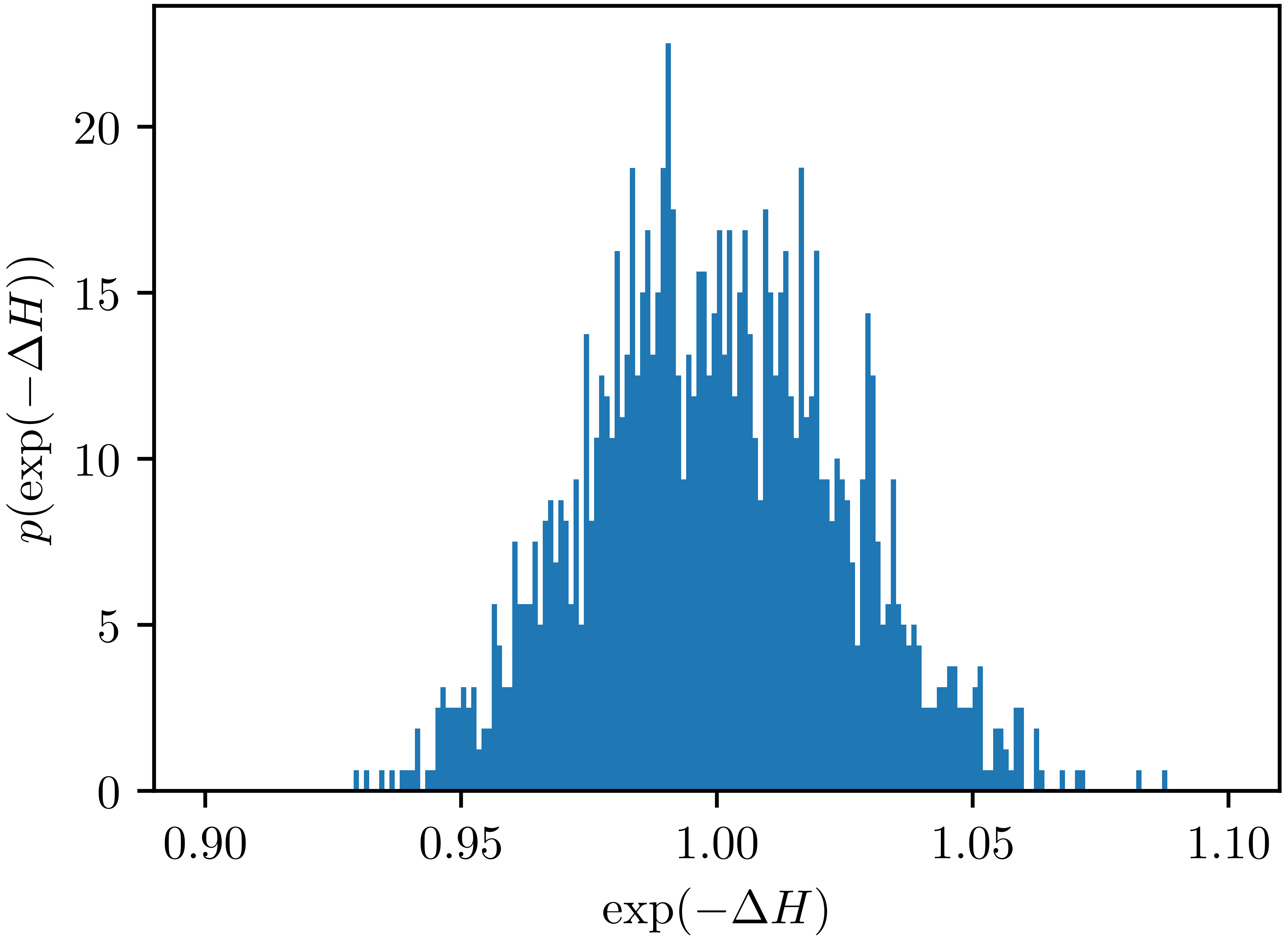
$$\frac{dU}{dt} = \frac{\partial H}{\partial p} = p$$

$$\frac{dp}{dt} = -\frac{\partial H}{\partial U} = -\frac{\partial \mathcal{A}}{\partial U}$$

- sample momenta p
- integrate eqs of motion (leapfrog, Omelyan)
- correct for H non-conservation

$$\langle \exp(-\Delta H) \rangle = 1$$

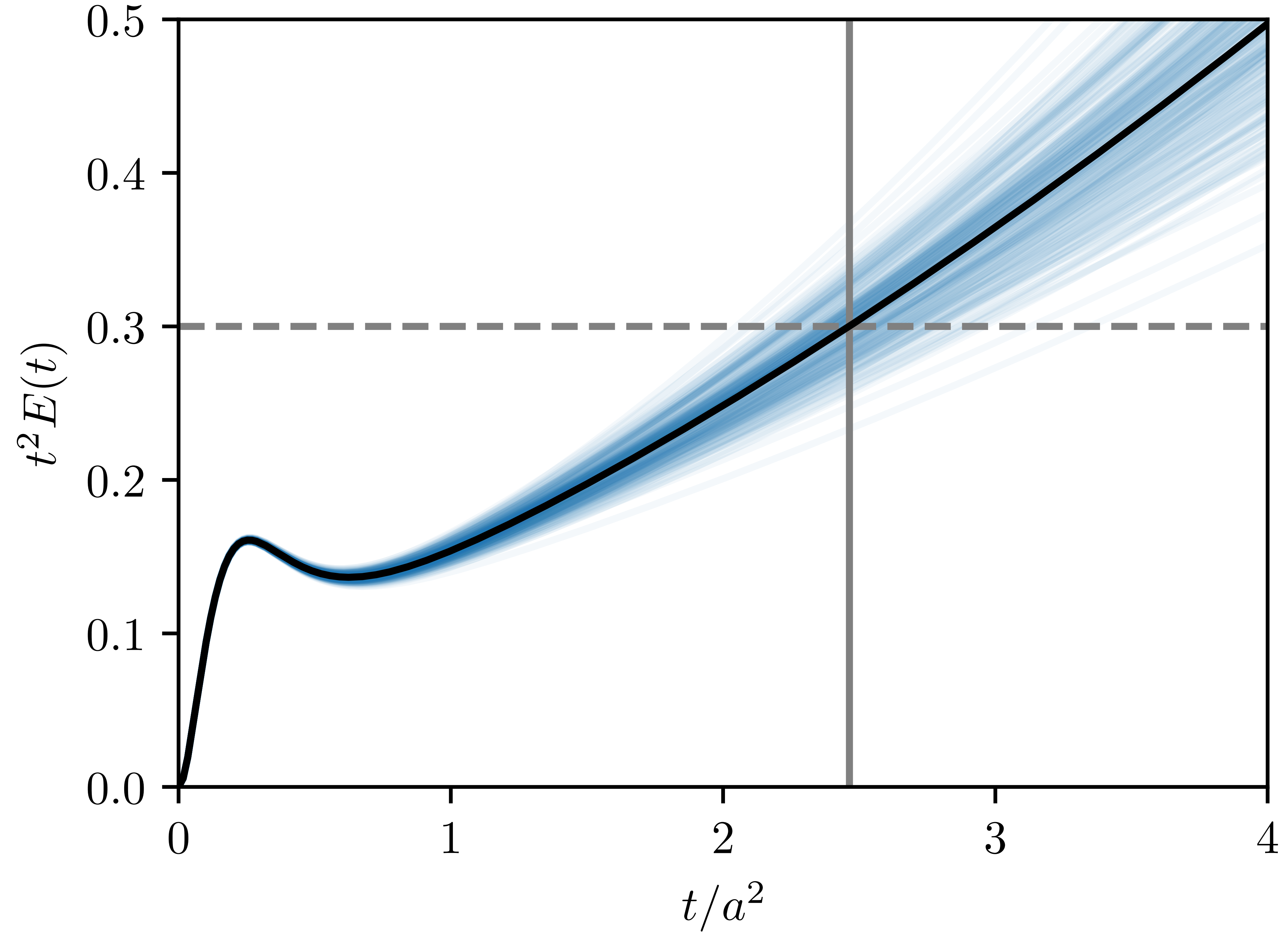


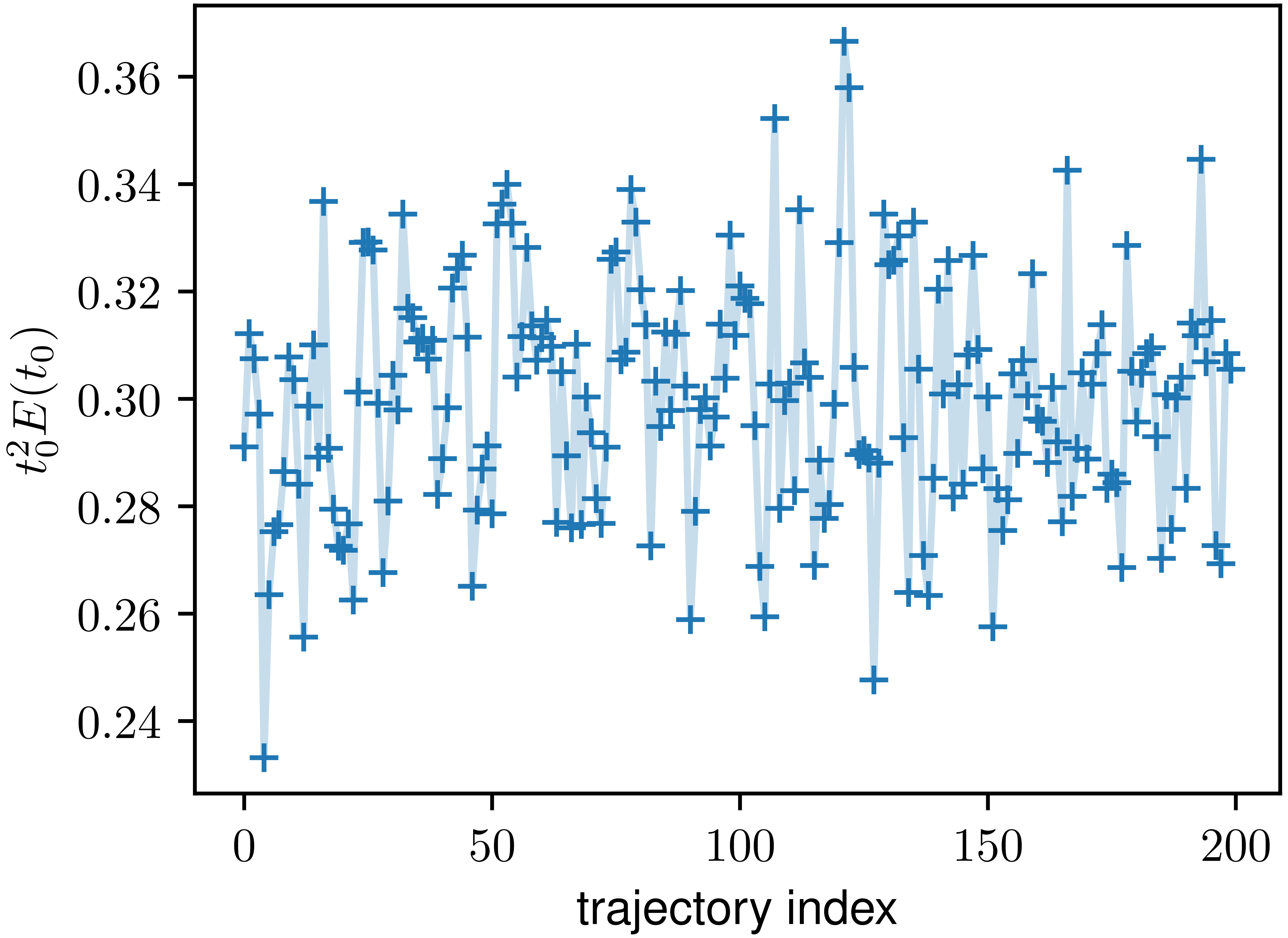


Gradient Flow (GF)

continuum $\frac{dA_\mu}{dt} = D_\nu G_{\nu\mu}$ $G_{\mu\nu} = \partial_\mu A_\nu - \partial_\nu A_\mu + [A_\mu, A_\nu]$ $A_\mu(t)$ $E = \frac{1}{4} G^a_{\mu\nu} G^a_{\mu\nu}$

scale definition $t^2 \langle E \rangle|_{t_0} = 0.3$ $t_0 = 0.167 \text{ fm}$ SU(3) gauge theory

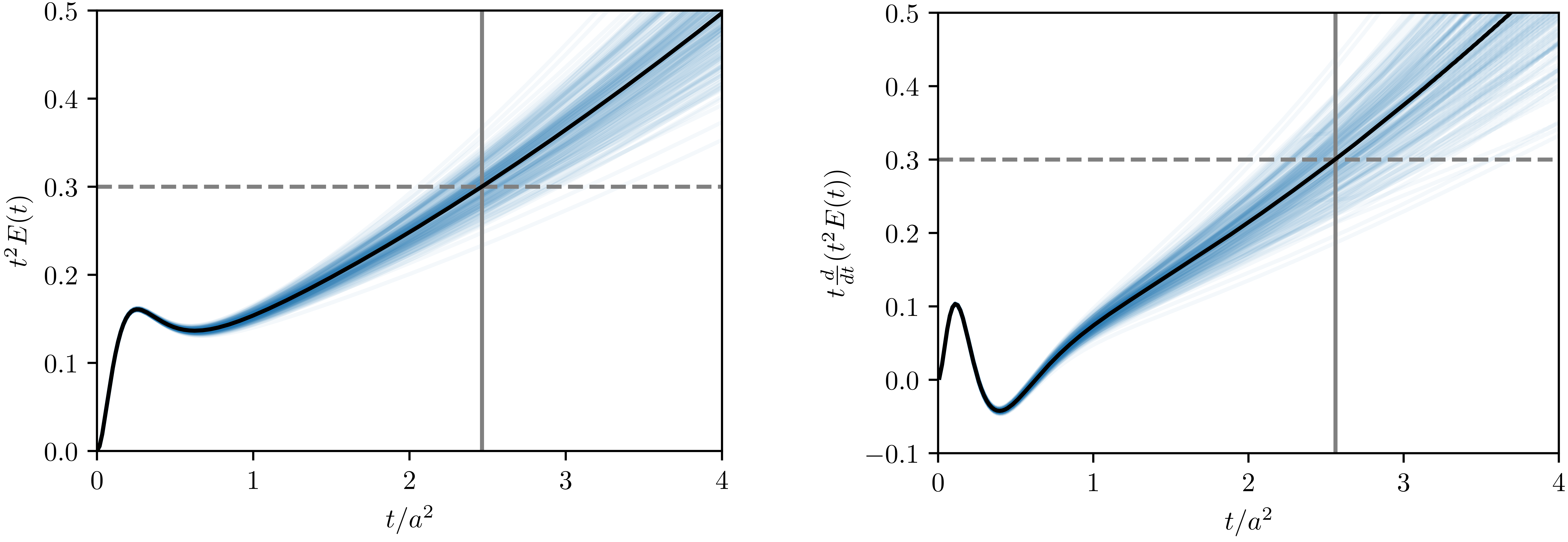




Gradient Flow (GF)

alternative scale determination w_0 $\qquad t\frac{d}{dt}\left(t^2\langle E\rangle\right)|_{t=w_0^2} = 0.3$

ratio t_0/w_0^2 $\qquad$ universal value in the continuum limit $a^2/t_0 \to 0$



Gradient Flow (GF)

how does FP action perform?

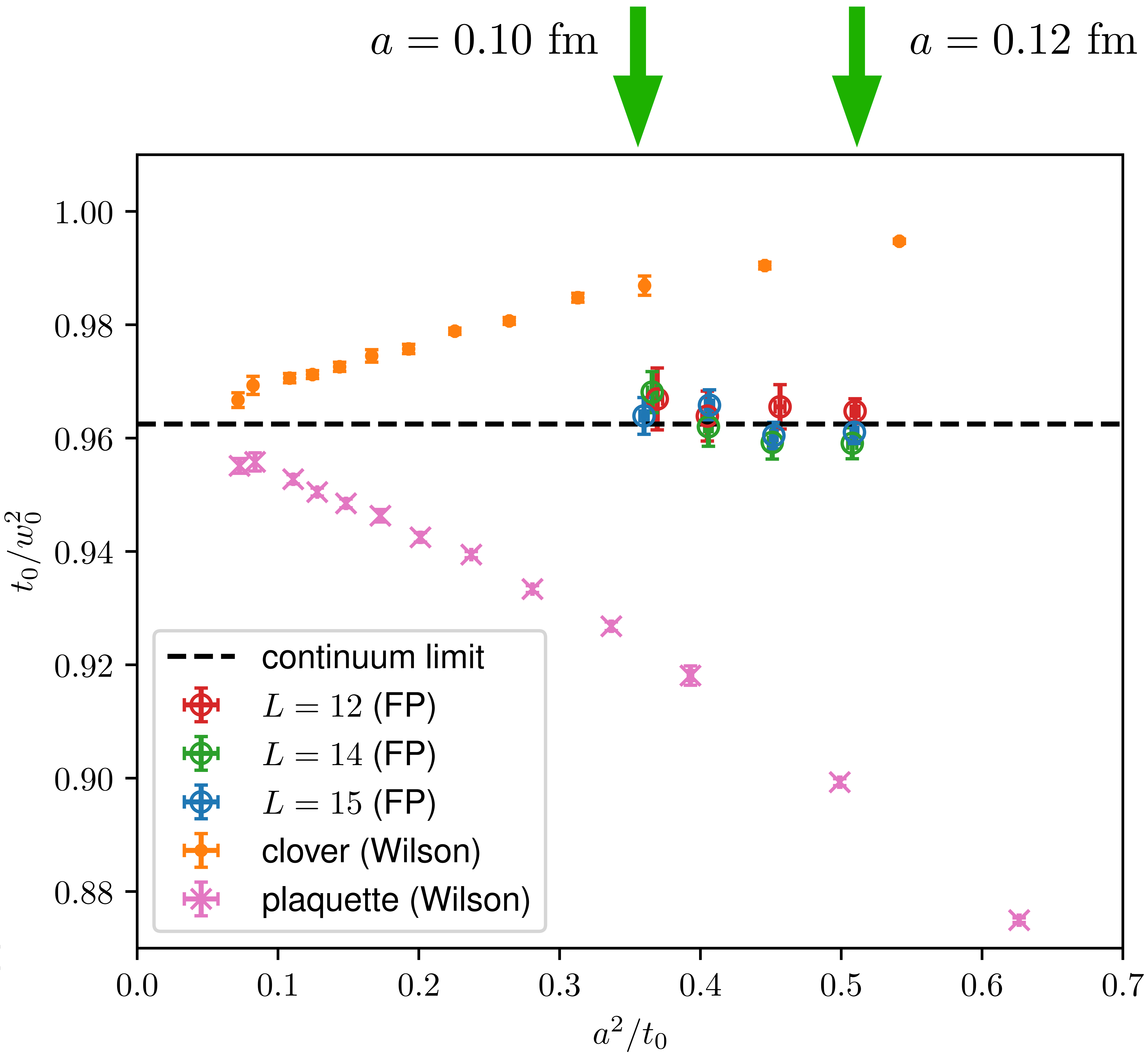
ongoing part of project

comparison with independent Wilson gauge action simulations

continuum limit $a^2/t_0 \to 0$

Wilson action: artifacts $\mathcal{O}(a^2)$

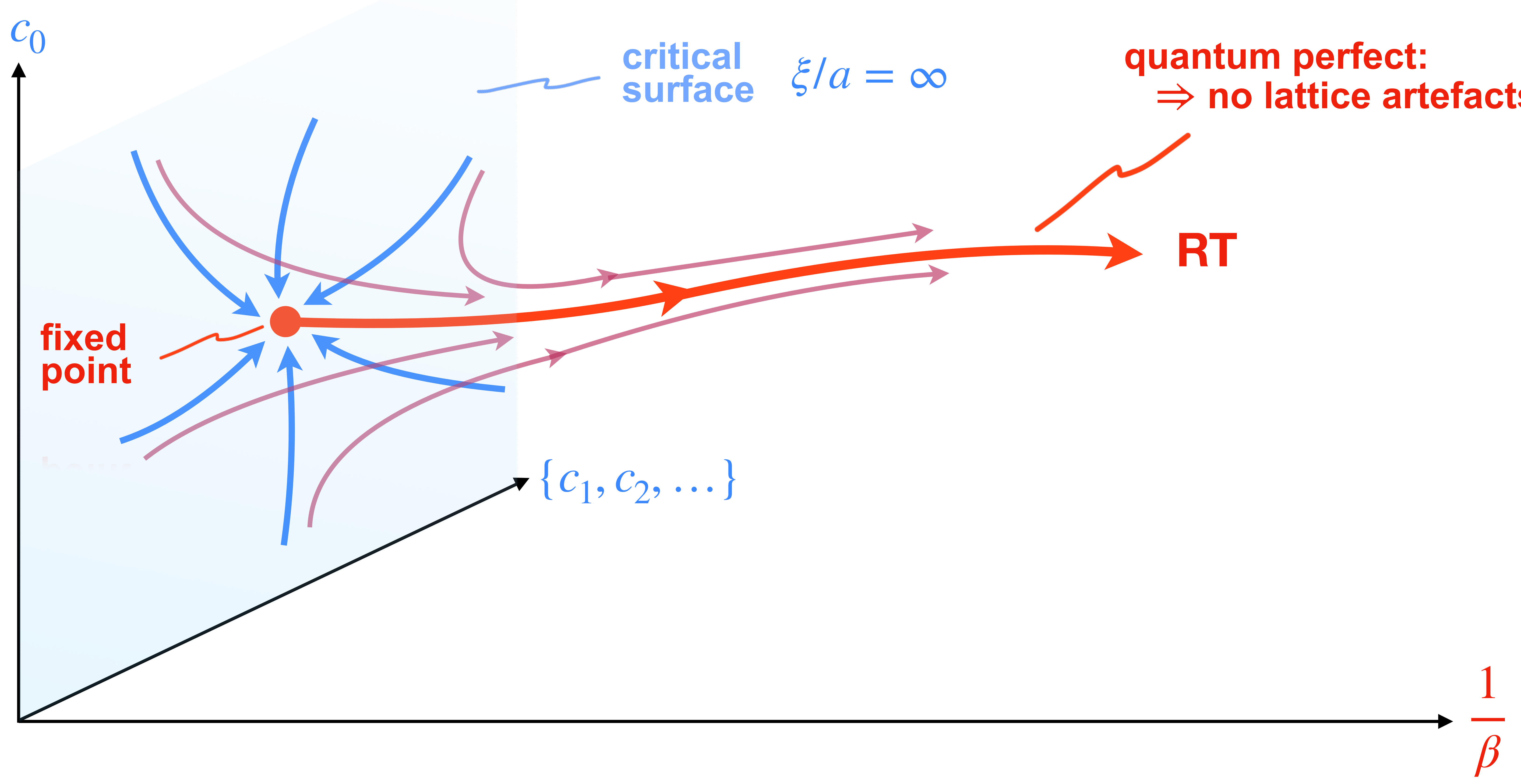
FP action: artifacts appear to be smaller, consistent with continuum limit at lattice spacing ~ 0.1 fm



Part V:
The quantum perfect RG action

The quantum perfect RG action

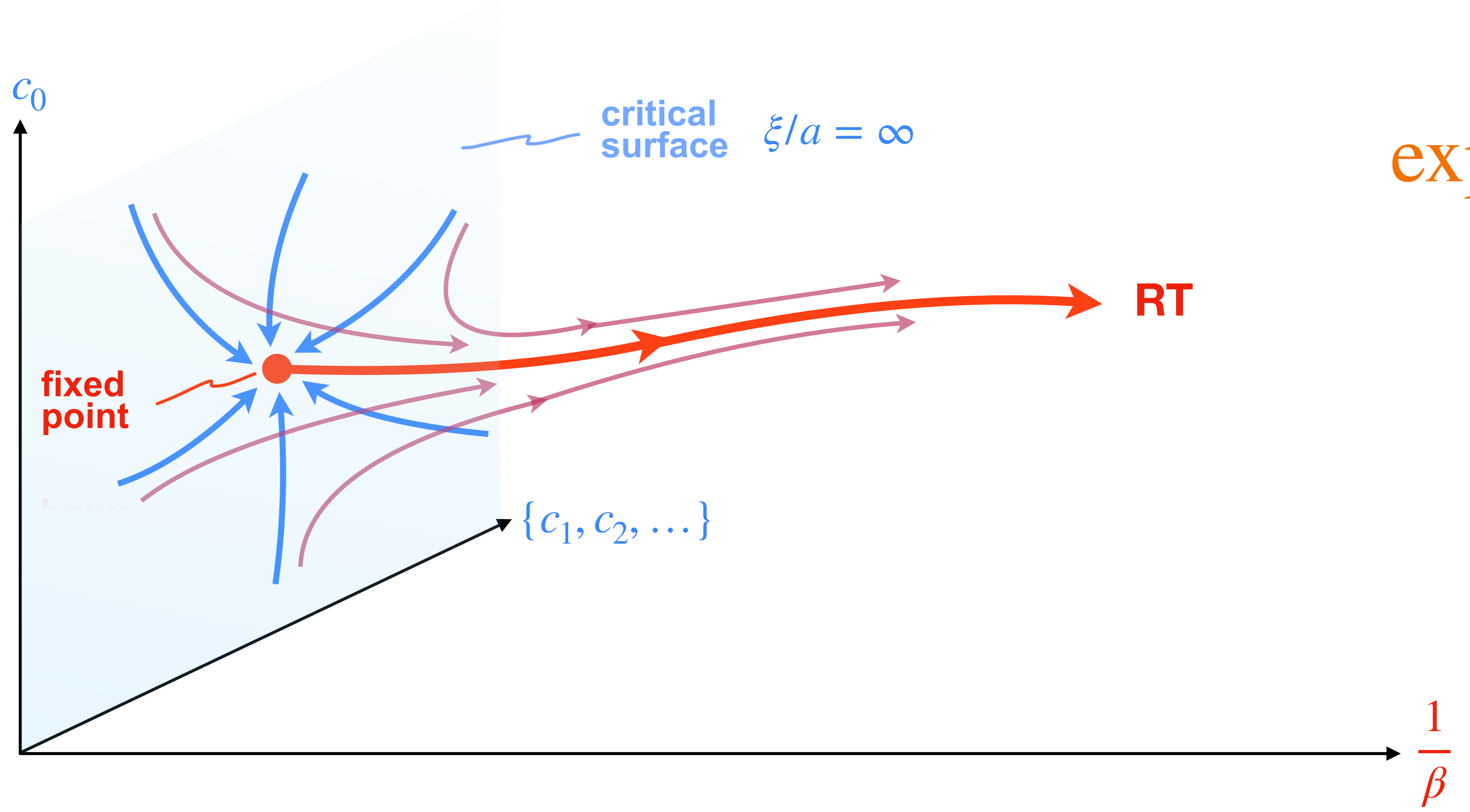
The effective action $\beta A_{RT}[V]$ is described by infinitely many couplings $\{c_\alpha\}$:



RT $\Rightarrow$ start from the fixed point and apply RGTs: $\{c_\alpha^*\}$ $\xrightarrow{\text{RGT}}$ $\{c_\alpha^{RT}\}$

The quantum perfect RG action

The effective action $\beta A_{RT}[V]$ is described by infinitely many couplings $\{c_\alpha^{RT}\}$:



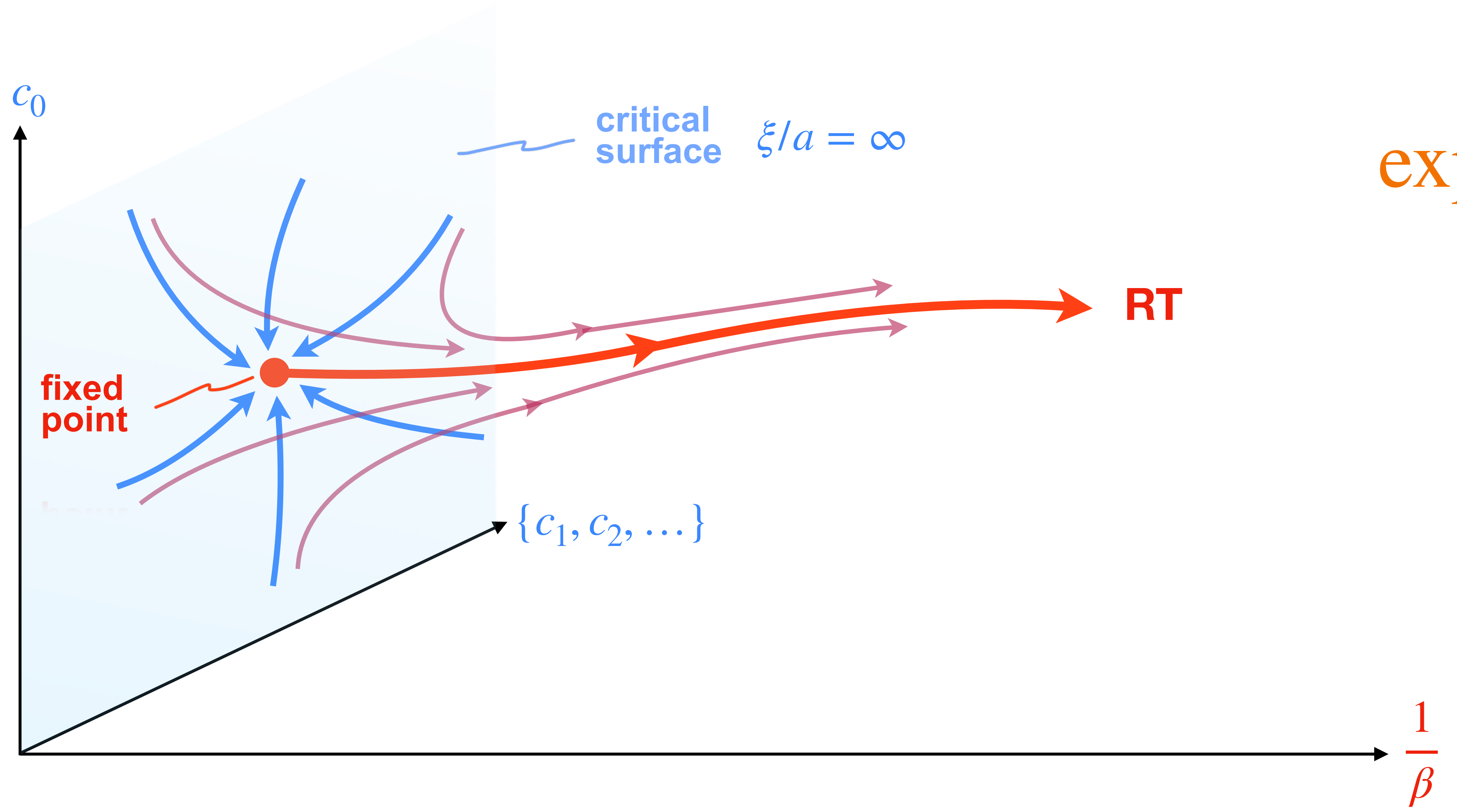
$$\exp\left\{-\beta' A'_{RT}[V]\right\} = \int \mathscr{D}U \exp\left\{-\left(\beta A_{RT}[U] + T[U,V]\right)\right\}$$

Two practical problems:

- how to parametrize RT, i.e., which set $\{c_\alpha^{RT}\}$?
- how to determine $\{c_\alpha^{RT}\}$?

The quantum perfect RG action

The effective action $\beta A_{RT}[V]$ is described by infinitely many couplings $\{c_\alpha^{RT}\}$:



$$\exp\left\{-\beta' A'_{RT}[V]\right\} = \int \mathcal{D}U \exp\left\{-\left(\beta A_{RT}[U] + T[U,V]\right)\right\}$$

Two practical problems:

- how to parametrize RT, i.e., which set $\{c_\alpha^{RT}\}$?
- how to determine $\{c_\alpha^{RT}\}$?

$\Rightarrow$ change of action from change of configuration $V \to V_\epsilon$ via MC simulation:

$$\exp\left\{-\beta'\left(A'_{RT}[V^\epsilon] - A'_{RT}[V]\right)\right\} = \left\langle \exp\left\{-\left(T[U,V^\epsilon] - T[U,V]\right)\right\}\right\rangle_V$$

The quantum perfect RG action

⇒ change of action from change of configuration $V \to V_\epsilon$:

$$\exp\left\{-\beta'\left(A'_{RT}[V_\epsilon] - A'_{RT}[V]\right)\right\} = \left\langle e^{-\left(T[U,V^\epsilon] - T[U,V]\right)}\right\rangle_V$$

with

$$\langle \mathcal{O} \rangle_V = \frac{\int \mathcal{D}U\, e^{-\left(\beta A_{RT}[U] + T[U,V]\right)} \cdot \mathcal{O}}{\int \mathcal{D}U\, e^{-\left(\beta A_{RT}[U] + T[U,V]\right)}}$$

Note: difference $T[U, V_\epsilon] - T[U, V]$ only depends on the blocked link $Q_\mu(x_B)[U]$:

$$T[U,V^\epsilon] - T[U,V] = -\frac{\kappa}{N_c}\sum_{x_B,\mu} \mathrm{ReTr}\left(\left\{V^\epsilon_\mu(x_B) - V_\mu(x_B)\right\} \cdot Q^\dagger_\mu(x_B)\right)$$

The quantum perfect RG action

$\Rightarrow$ change of action from change of configuration $V \to V_\epsilon$:

$$\exp\left\{-\beta'\left(A'_{RT}[V^\epsilon] - A'_{RT}[V]\right)\right\} = \left\langle e^{-\left(T[U,V^\epsilon] - T[U,V]\right)}\right\rangle_V$$

In practice:

1. generate $\{V\}$, e.g., with $\beta' A^{FP}$, at fine lattice spacing a'
2. $\forall V$ simulate U with βA^{FP} at finer lattice spacing $a = a'/2$
3. store $Q_\mu(x_B)[U]$ and calculate $\dfrac{\delta(\beta' A'_{RT}[V])}{\delta V^a_{x,\mu}}$ from $V \to V_\epsilon$

training data for L-CNN!

4. repeat with β'' at coarser a'' and simulate $\beta' A'_{RT}$ at $a' = a''/2$

Conclusions: RG from ML

New technologies enable old ideas: