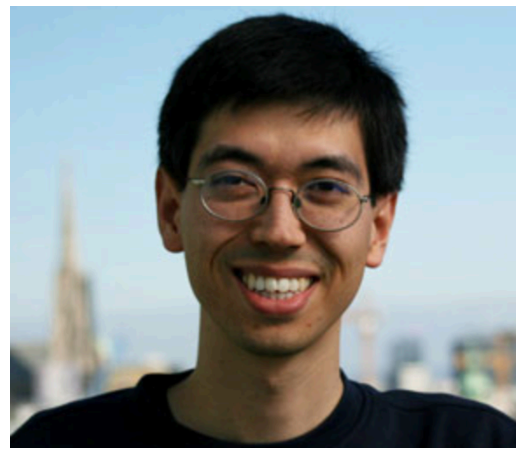
Machine learning a fixed point action

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

Urs Wenger

Albert Einstein Center for fundamental physics



in collaboration with Kieran Holland (University of Pacific), Andreas Ipp and David Müller (TU Wien)

40th Intl. lattice conference, 31 July 2023, Fermilab Chicago (IL), USA

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:

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

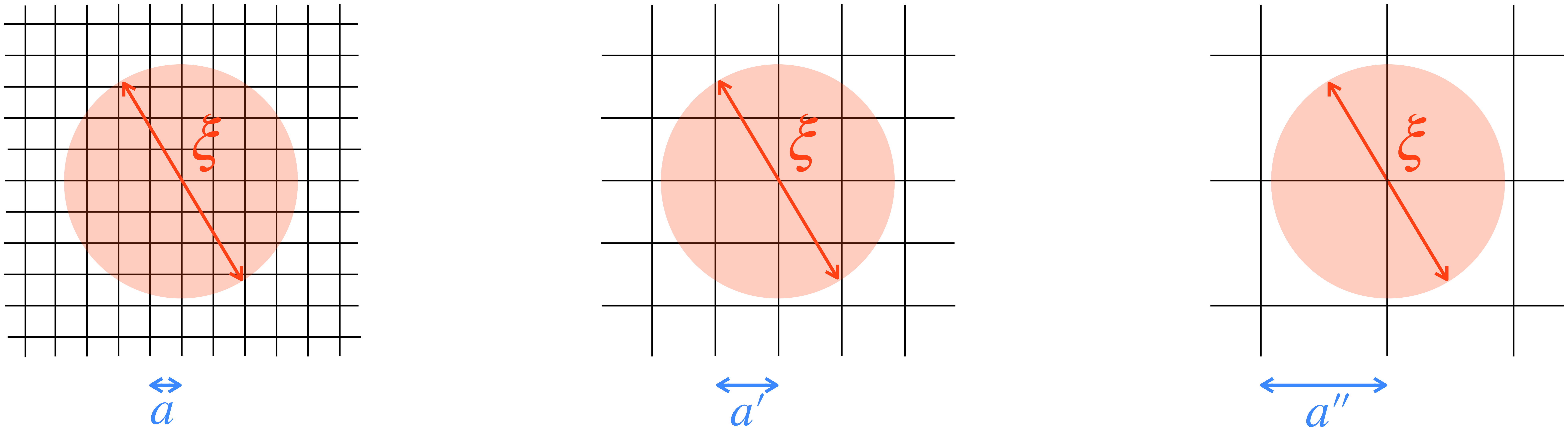
with a characteristic length scale ξ in units of the lattice spacing a:

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

Introduction

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

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



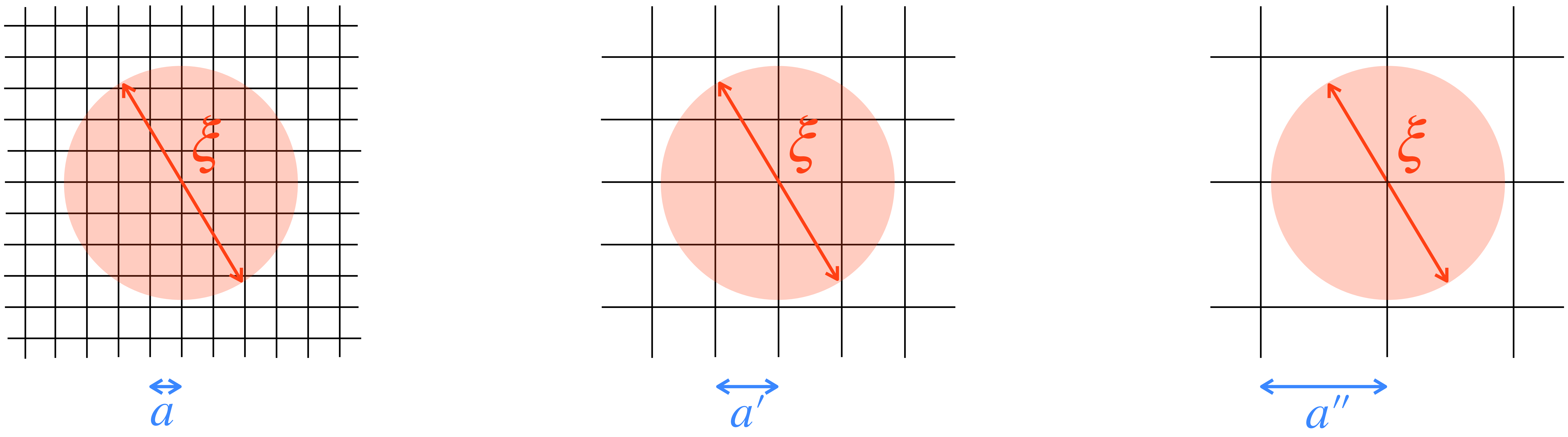
$$g \ll g' \ll g''$$

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

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?

Renormalization group transformation

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

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

⇒ provides solution for avoiding critical slowing down and lattice artefacts

Renormalization group transformation

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

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

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

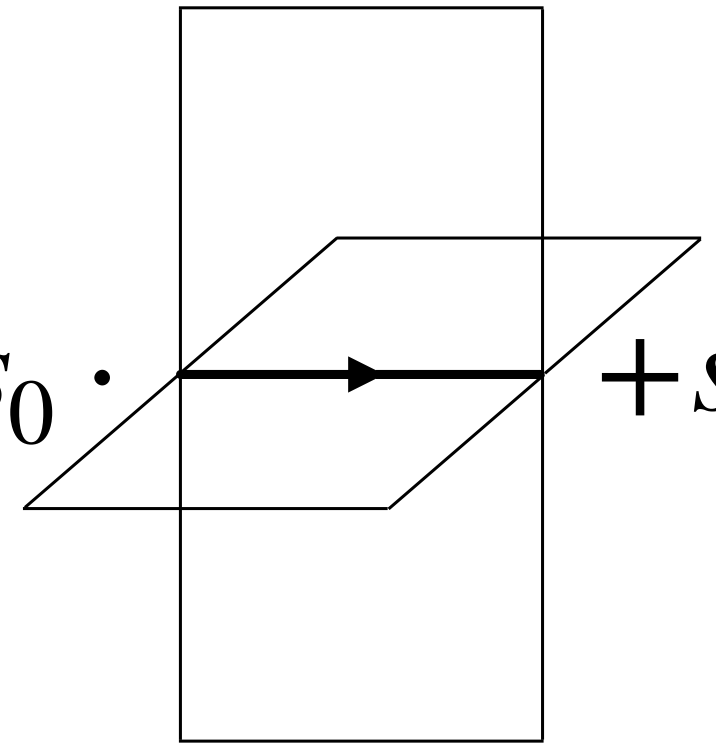
$$T[U,V] = -\frac{\kappa}{N_c}\sum_{x_B,\mu}\left\{\mathrm{ReTr}\left(V_\mu(x_B)\cdot Q_\mu^\dagger(x_B)\right) - \mathcal{N}_\mu^\beta\right\}$$

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

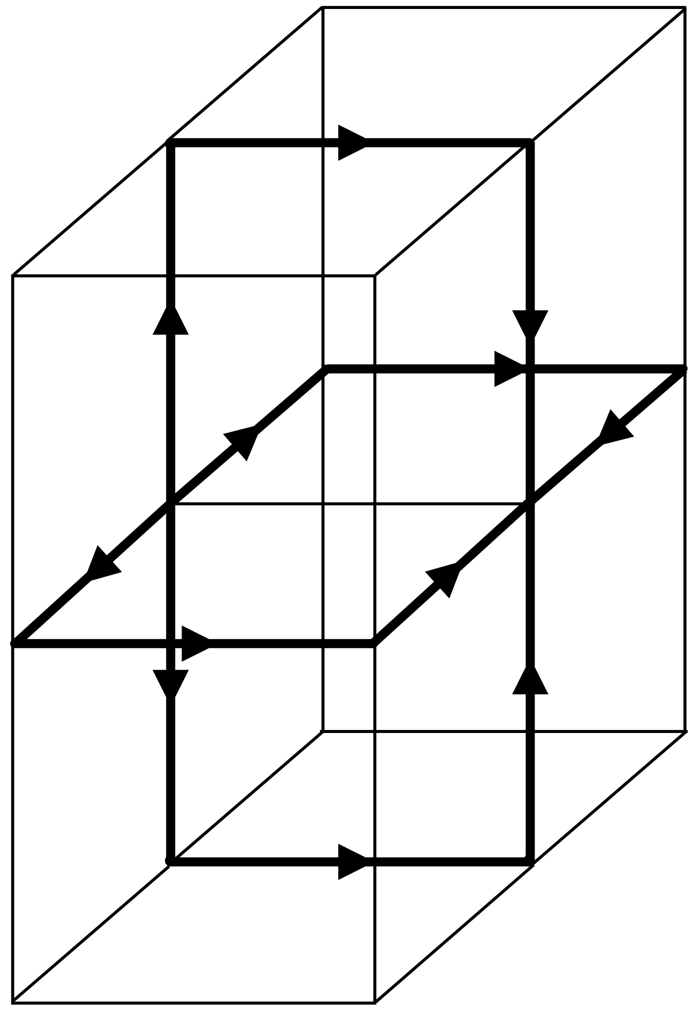
RGT blocking kernel

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

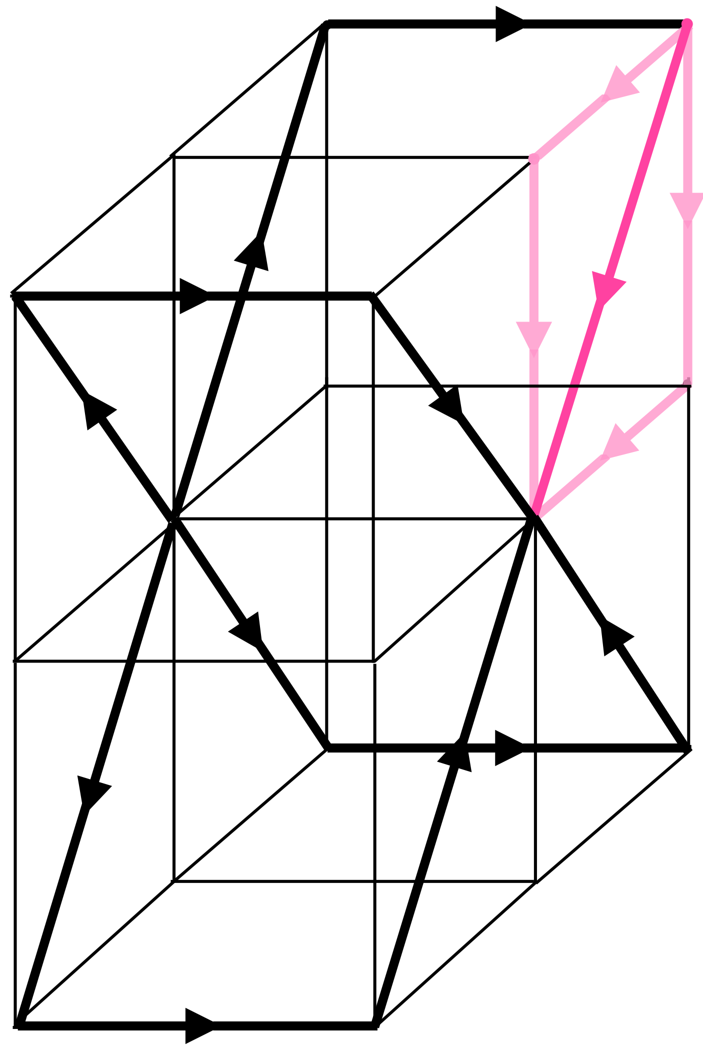
$S_\mu^{\text{smeared}}[U] = s_0 \cdot$



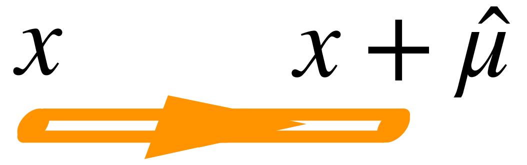
$+s_{pl}$



$+s_d$



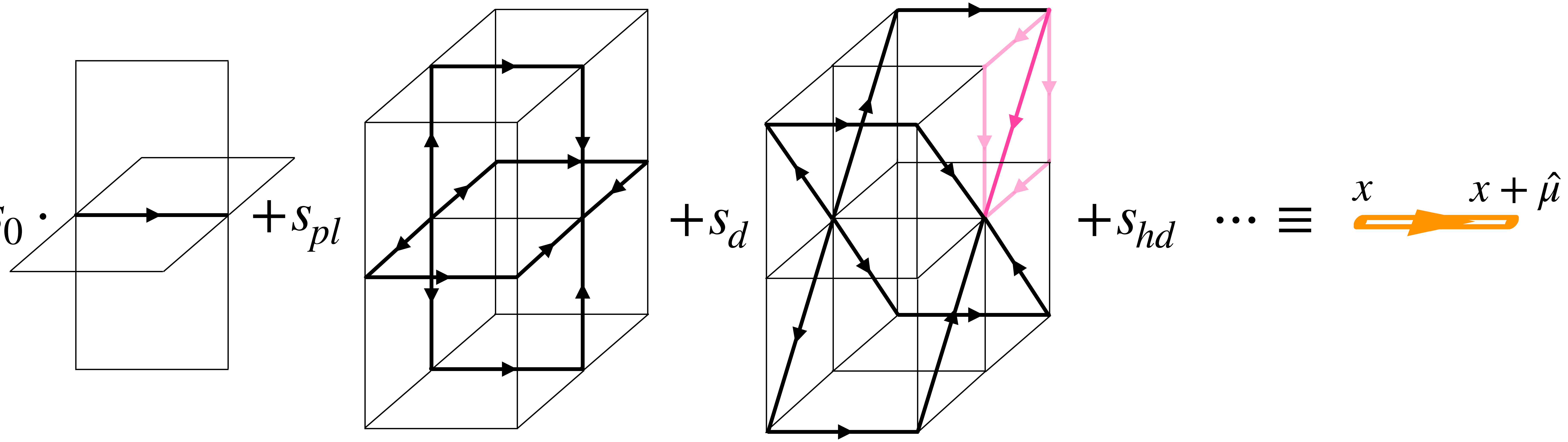
$+s_{hd} \cdots \equiv$



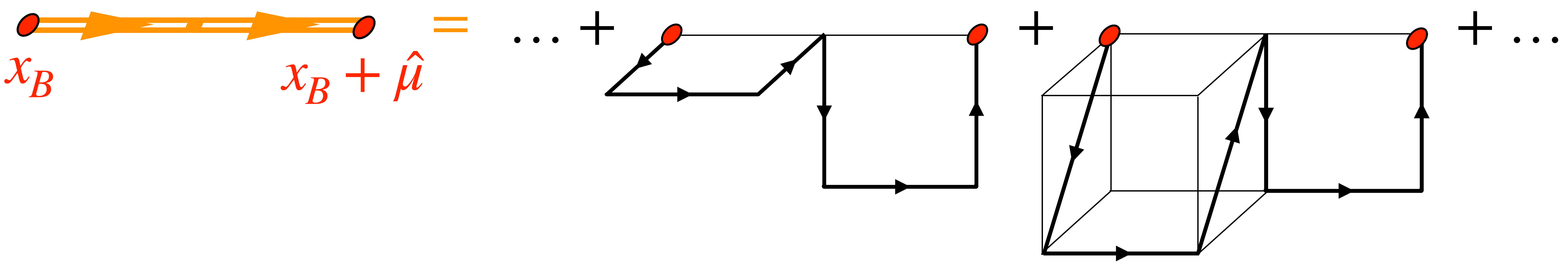
RGT blocking kernel

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

$$S_\mu^{\text{smeared}}[U] = s_0 \cdot$$



$$Q_\mu(x_B) =$$

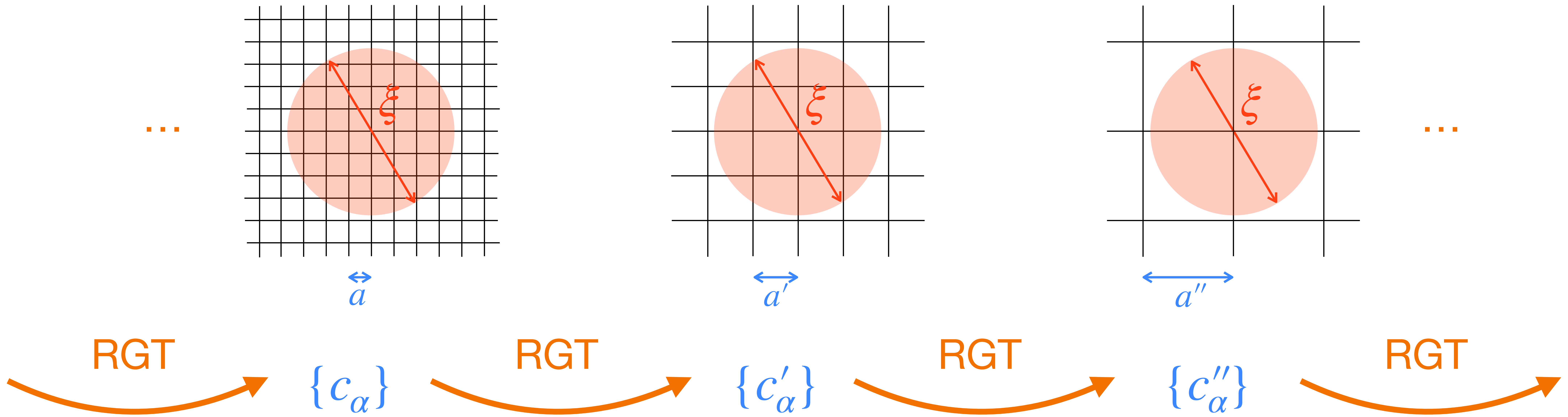


Renormalization group transformation

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

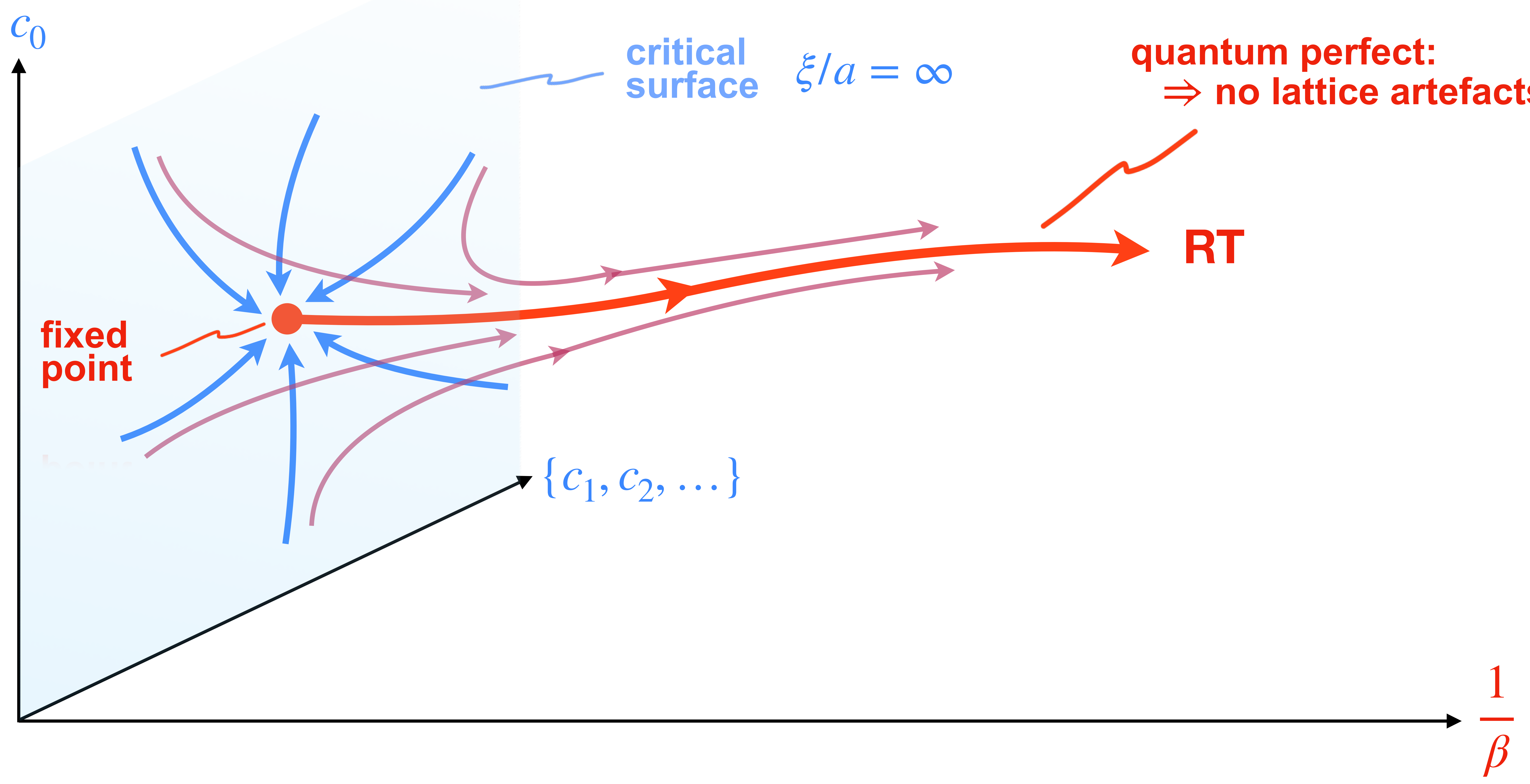
$$\exp\left\{-\beta' A'[V]\right\} = \int \mathscr{D}U \exp\left\{-\beta\left(A[U] + T[U,V]\right)\right\}$$

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



Renormalization group transformation

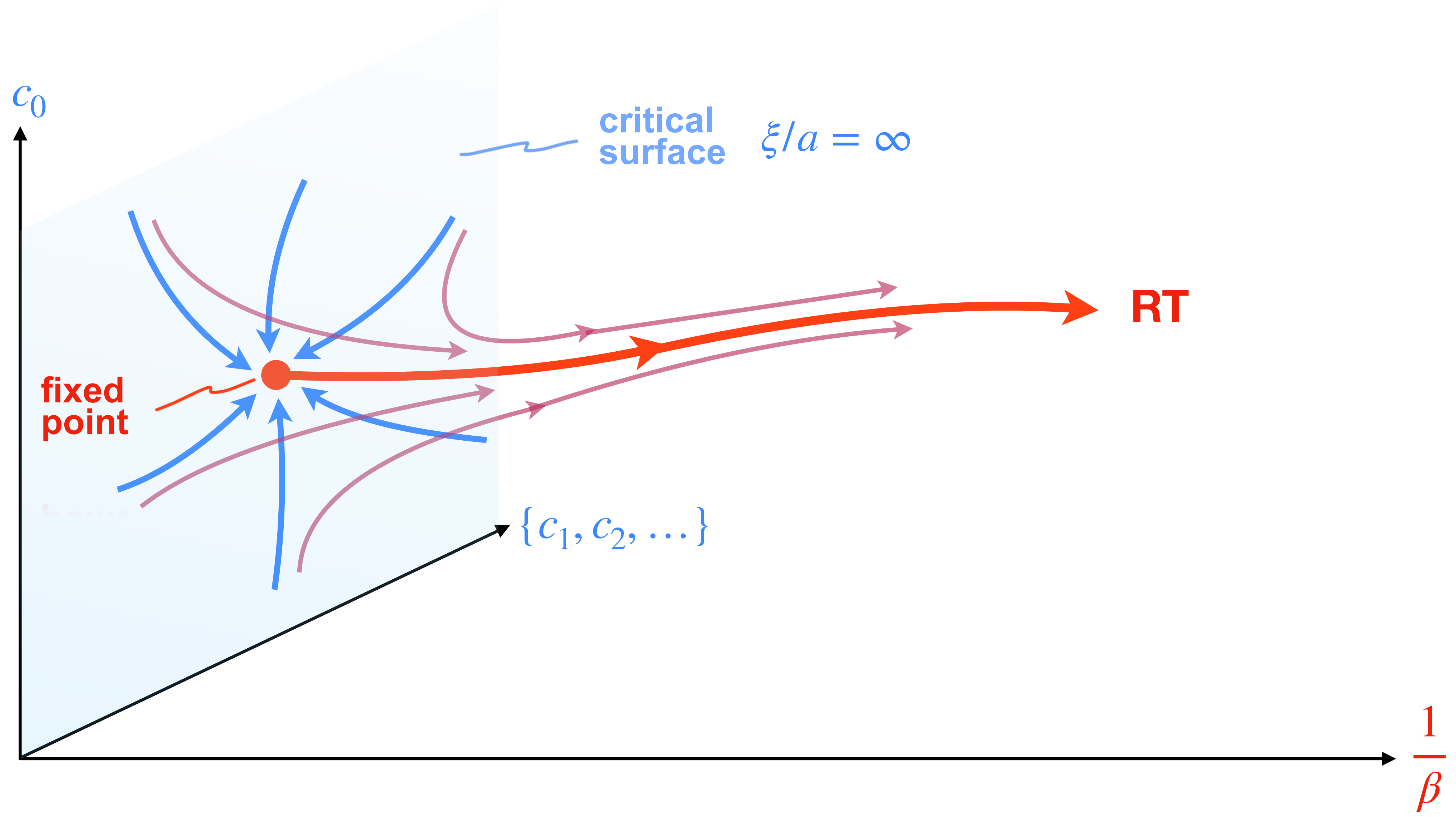
The effective action $\beta A[V]$ is described by infinitely many couplings $\{c_\alpha\}$:



$\Rightarrow$ 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 $\{c_\alpha^{\mathrm{FP}}\}$?

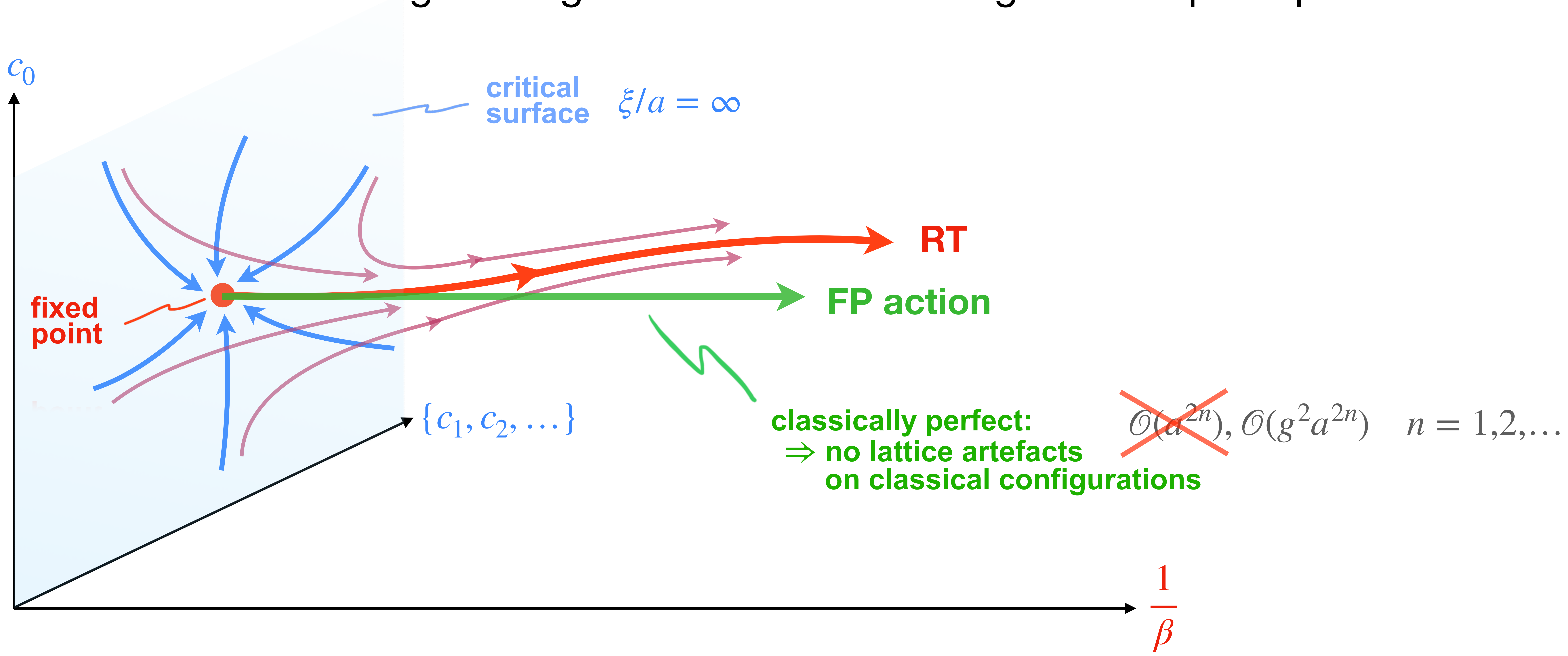
P. Hasenfratz, F. Niedermayer [Nucl.Phys.B 414 (1994) 785, hep-lat/9308004]

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

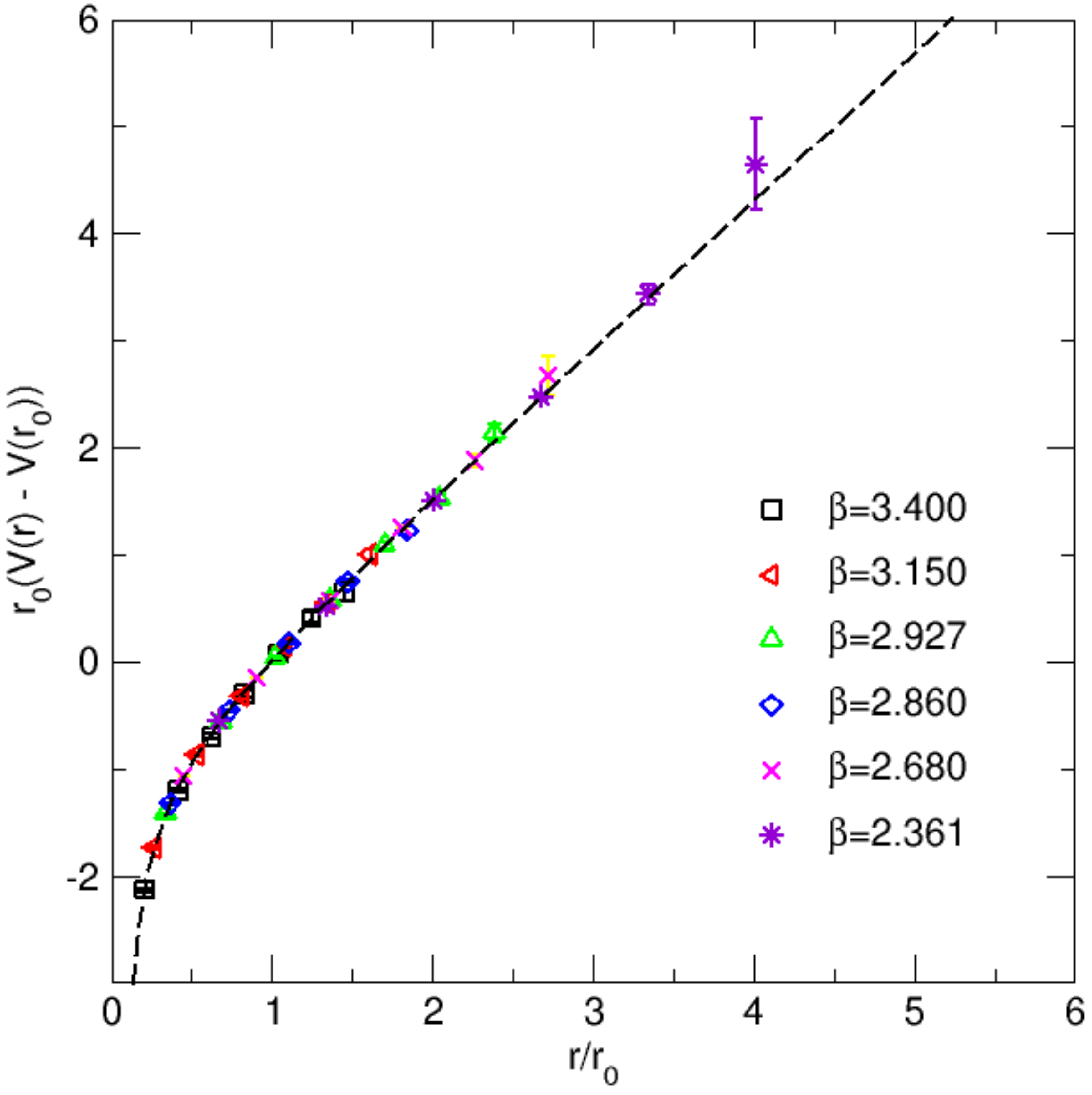
The FP action values for rough configurations defined through an inception procedure:

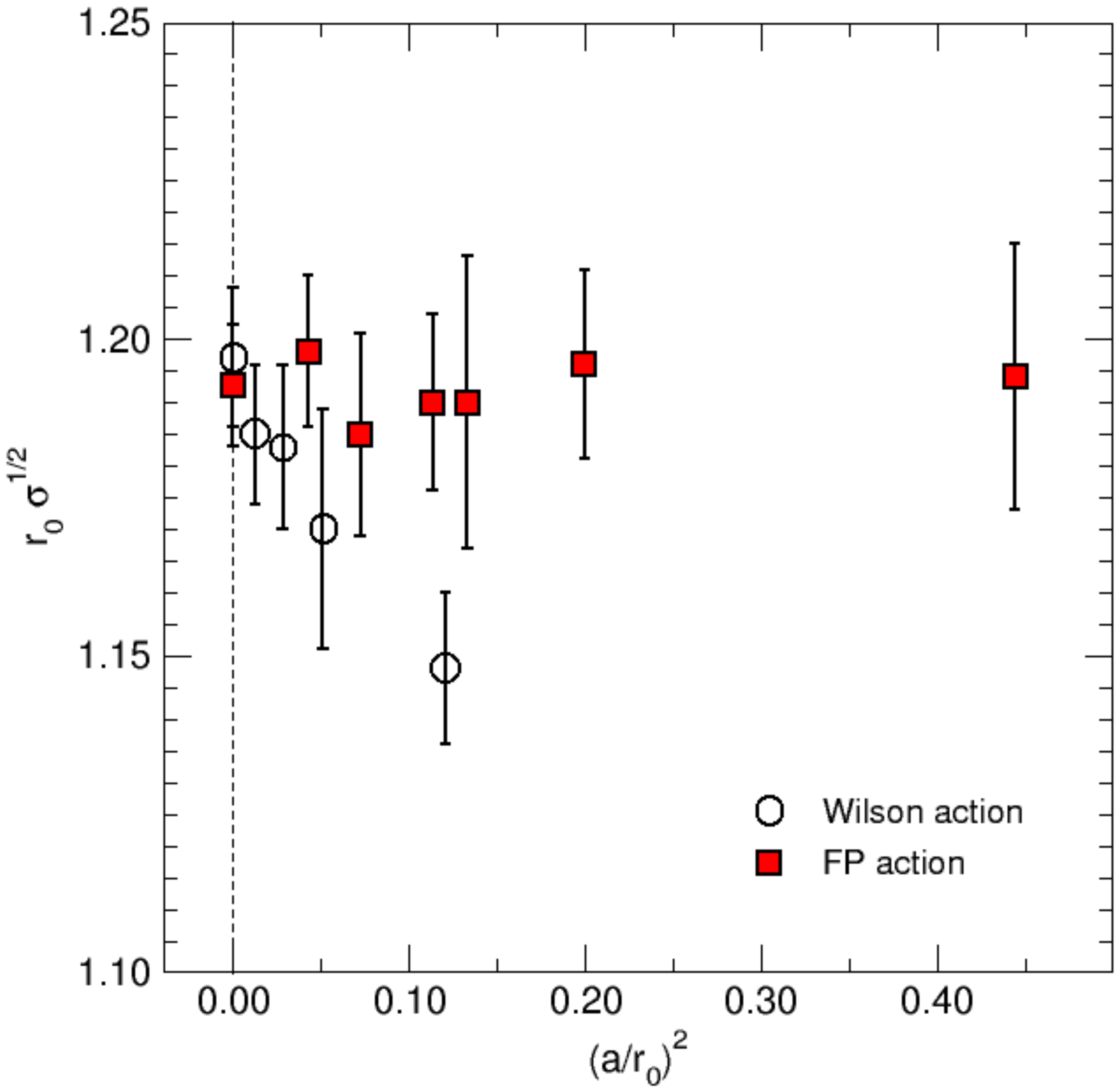


$$A^{\mathrm{FP}}[V] = \min_{\{U\}} \{A^{\mathrm{FP}}[U] + T[U, V]\} = \min_{\{U',U\}} \{A^{\mathrm{FP}}[U'] + T[U', U] + T[U, V]\}$$

FP action in action

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



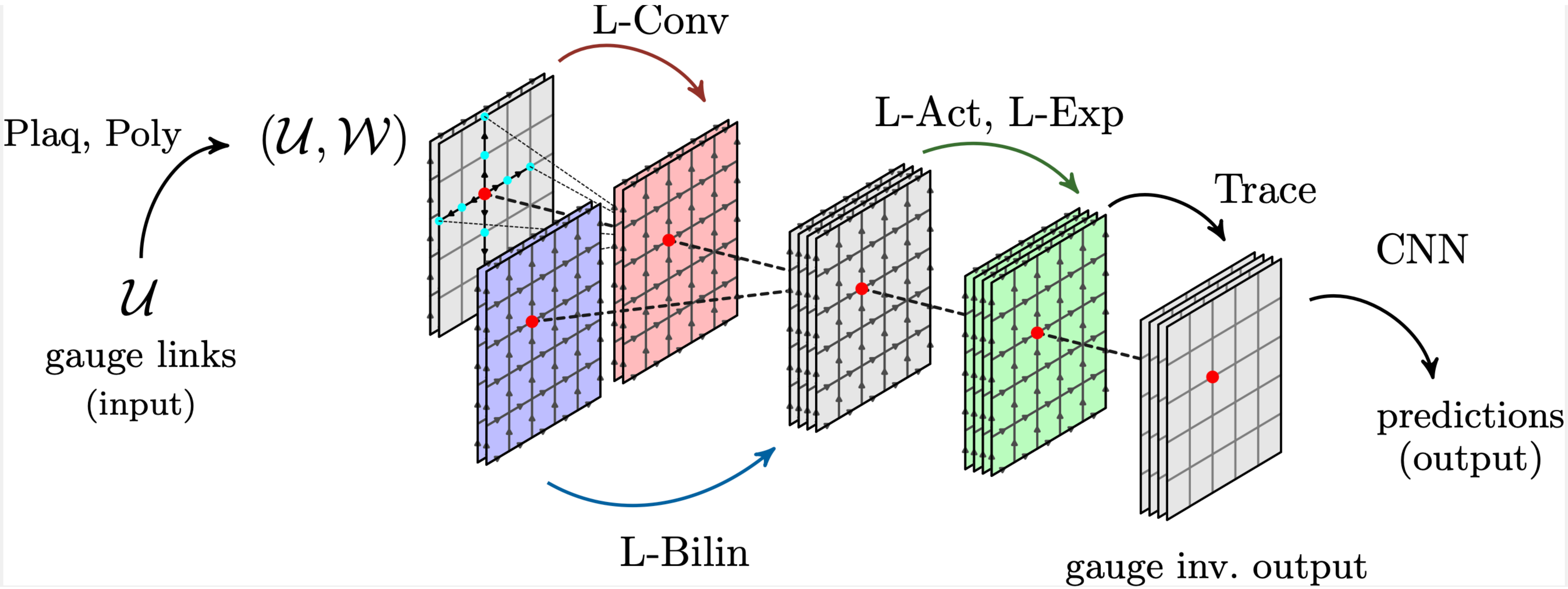


[Niedermayer, Rüfenacht, UW, Nucl.Phys.B 597 (2001) 413, hep-lat/0007007]

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:

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

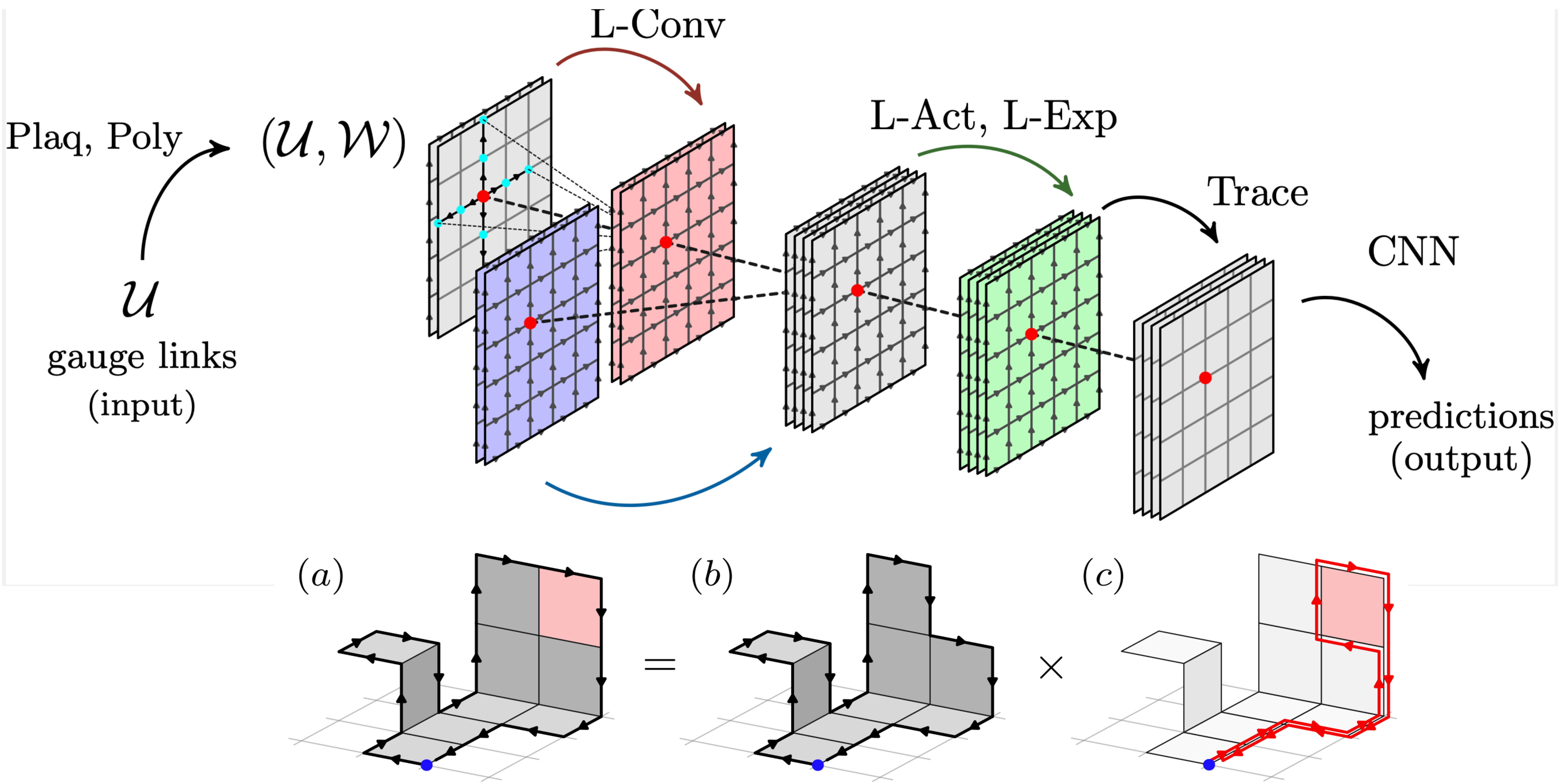
L-Bilin:

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

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]



Machine learning the FP action

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

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

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

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

$\Rightarrow$ data set for supervised ML

Machine learning the FP action

ML loss function from two weighted contribution:

$$\frac{1}{N_{\text{cfg}}} \sum_i |A^{\text{FP}}(V_i) - A^{\text{pred}}(V_i)| \qquad \frac{1}{32L^4}\frac{1}{N_{\text{cfg}}} \sum_{i,x,\mu} |\text{Tr}(D^{\text{FP}}_{x,\mu}(V_i) - D^{\text{pred}}_{x,\mu}(V_i))^2|$$

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

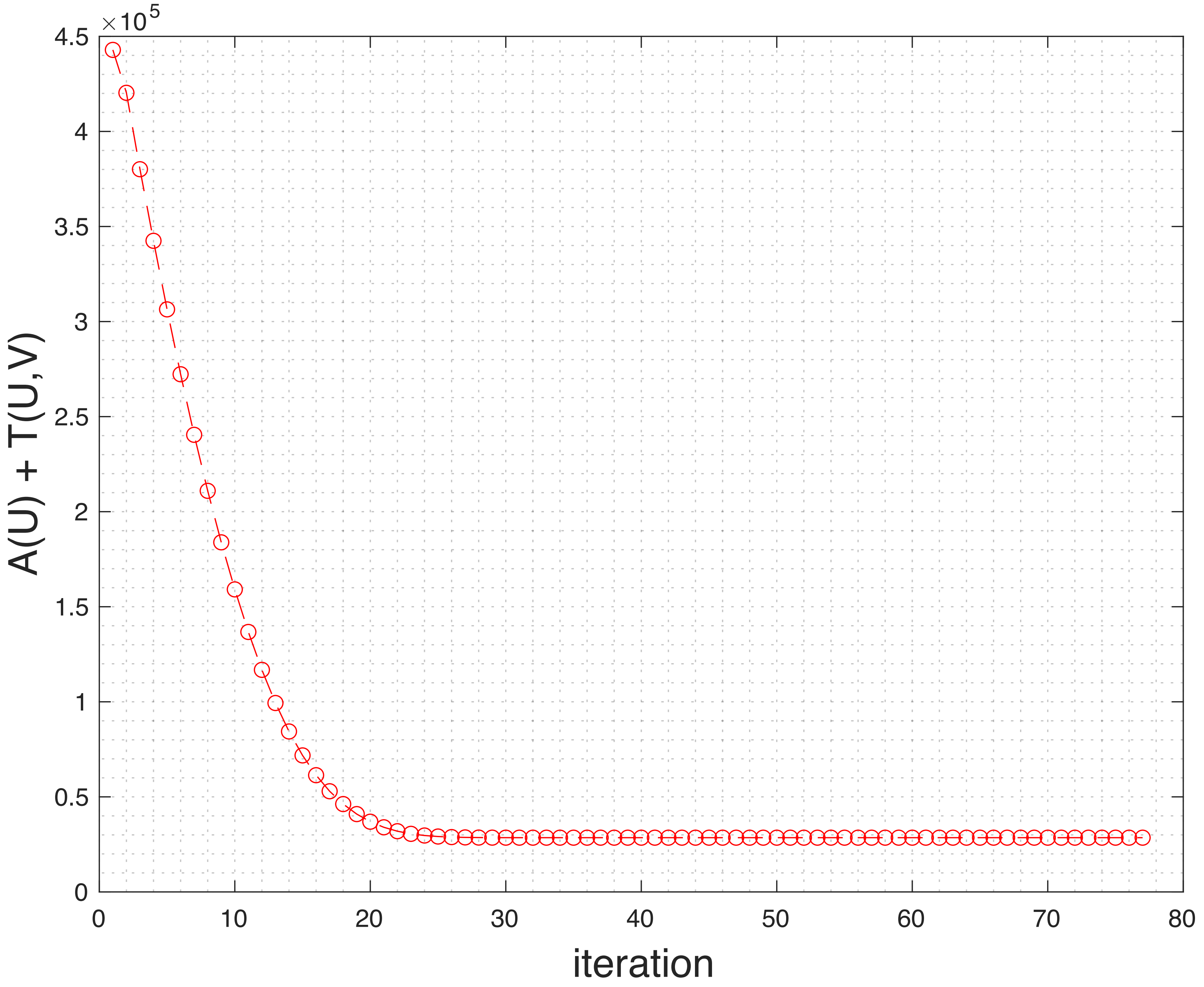
FP action parameterised by

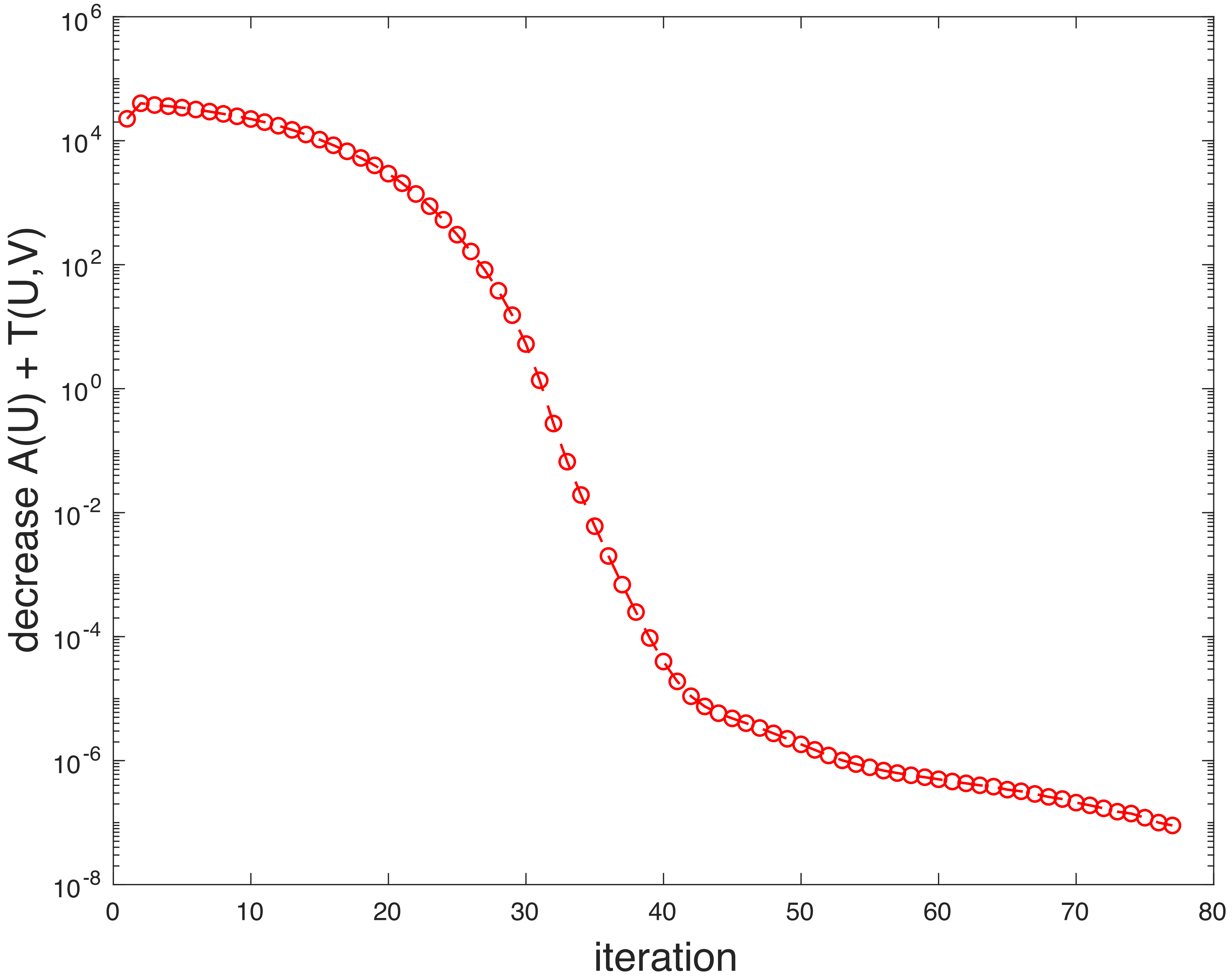
$$A^{\text{pred}}[V](x) = A^{\text{std.}}[V](x)\ \exp\{(N[V](x) - N[1])^m\}$$

std. = Wil, tlSym, ...

Machine learning the FP action

Minimisation evolution of fine configuration U for fixed coarse configuration V

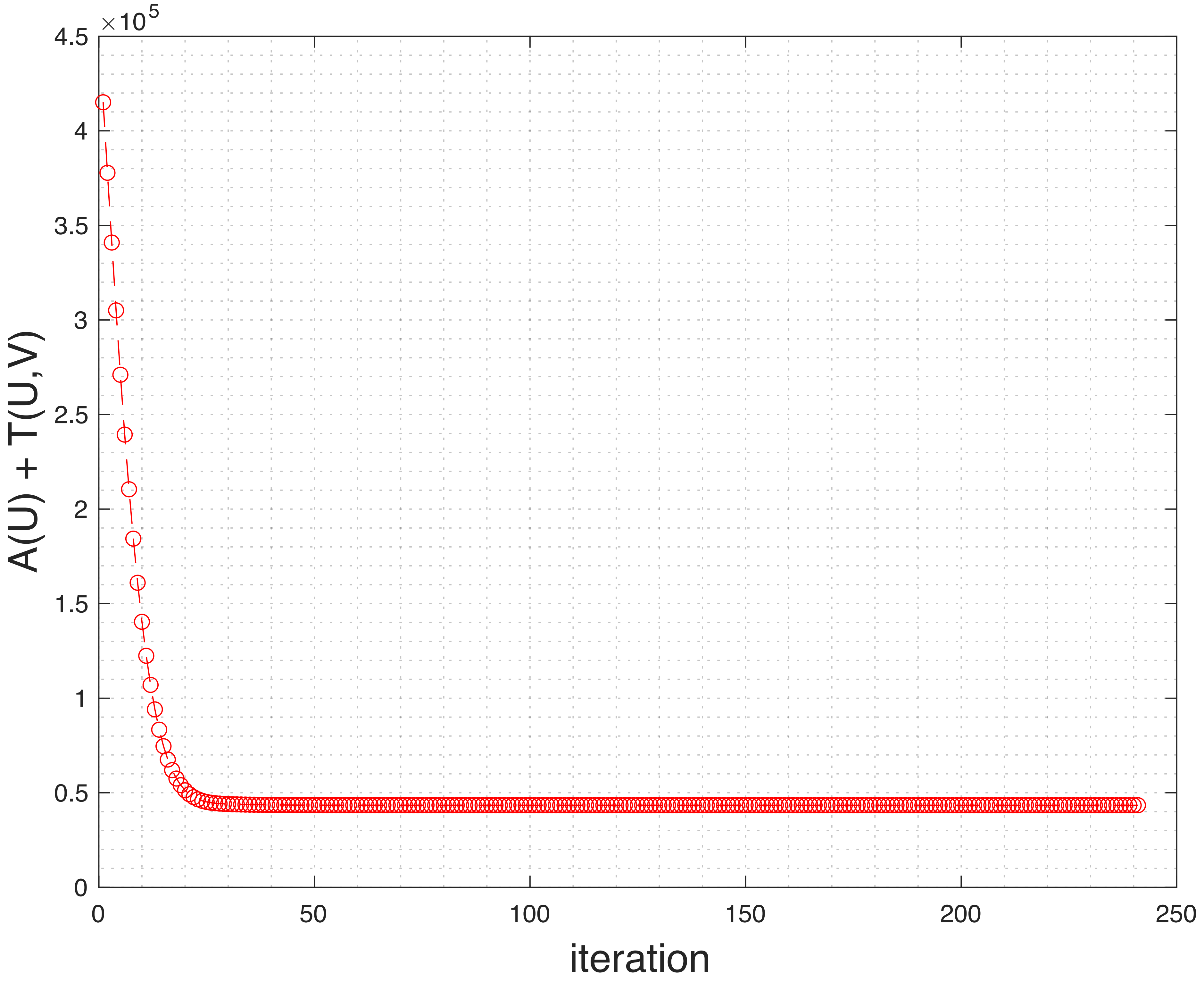


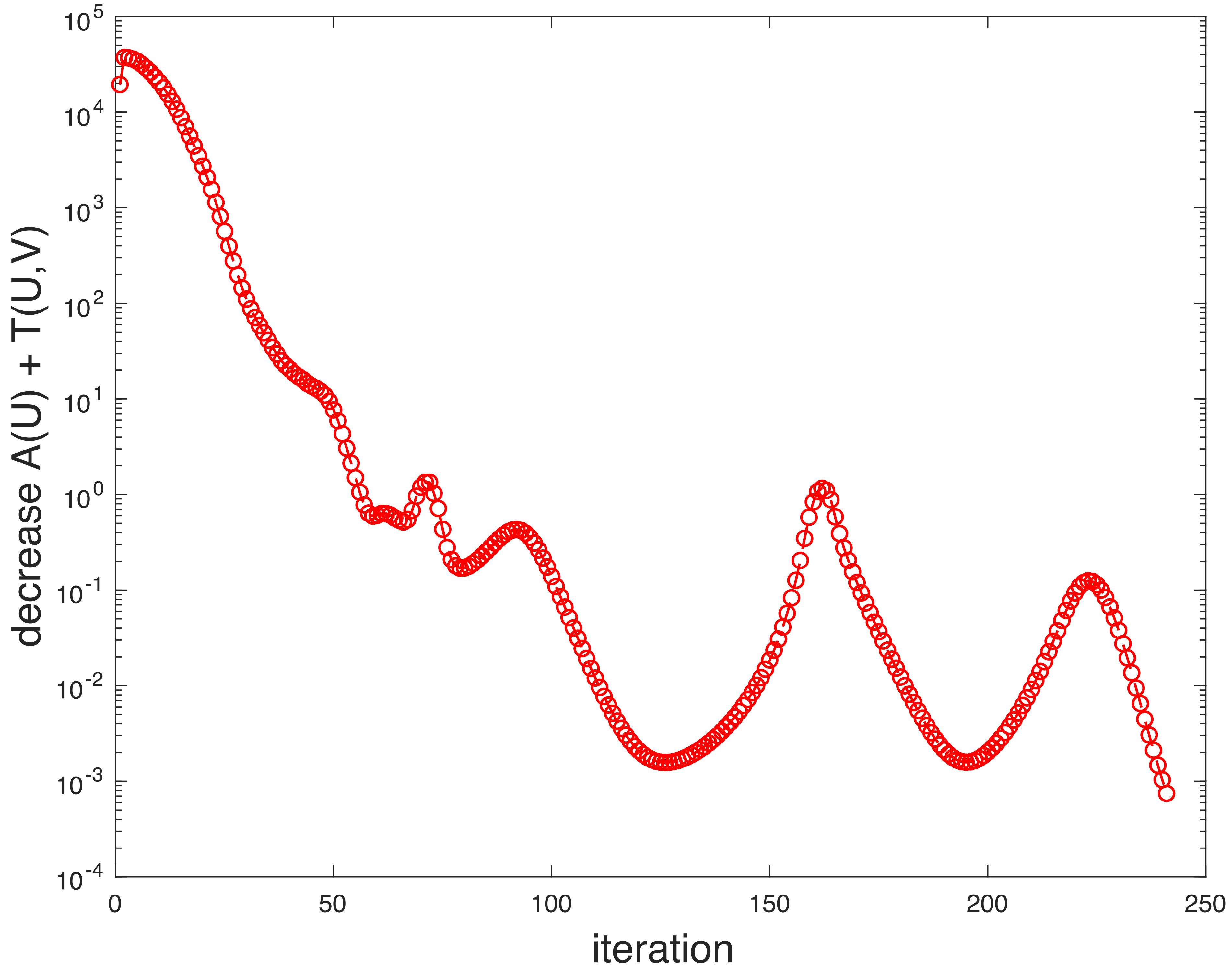


Coarse configuration: SU(3), $V = 8^4, \beta^{\text{Wil}} = 6.0$ $\Rightarrow$ lattice spacing $a \simeq 0.10$ fm

Machine learning the FP action

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

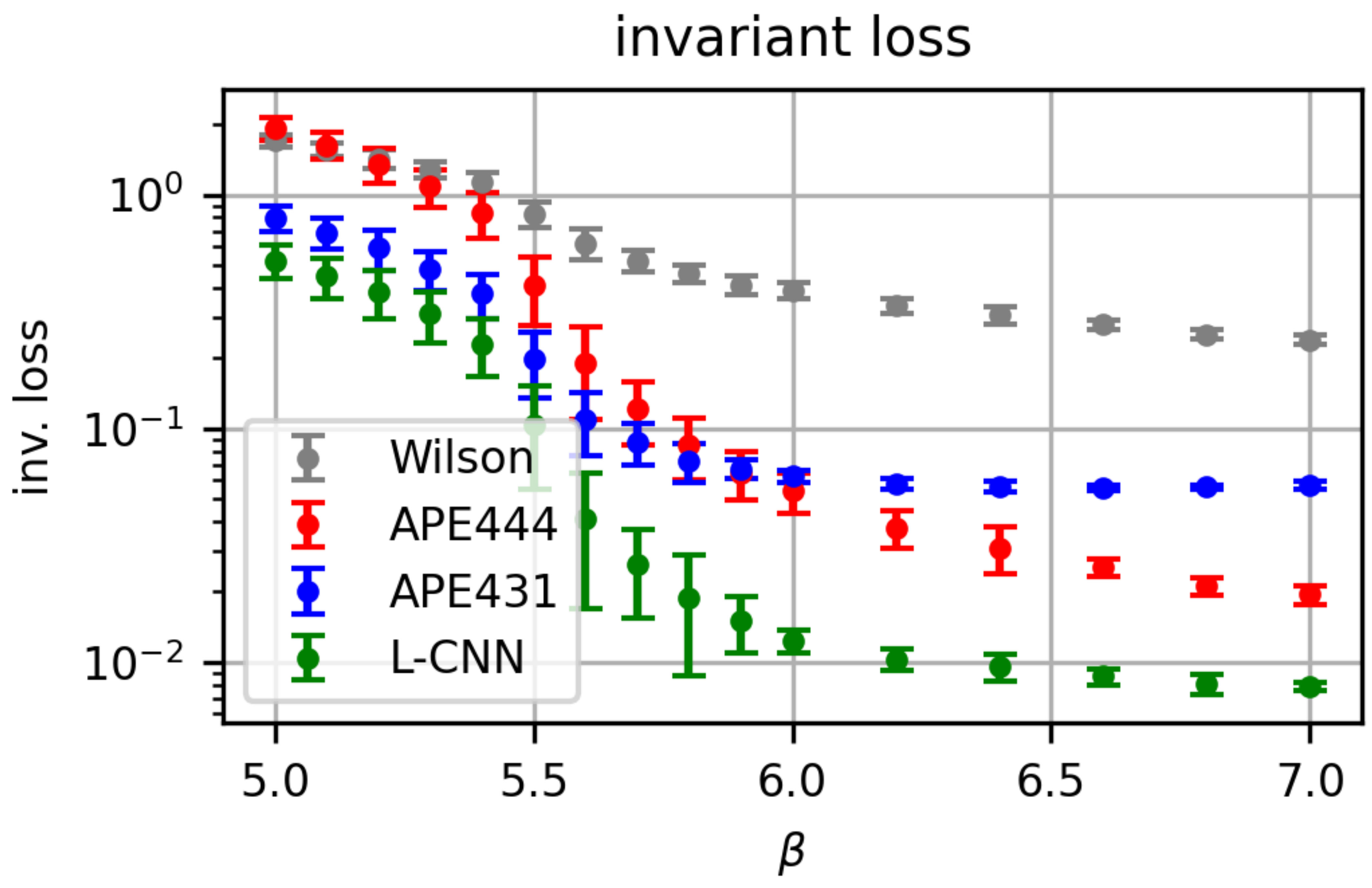
Machine learning the FP action

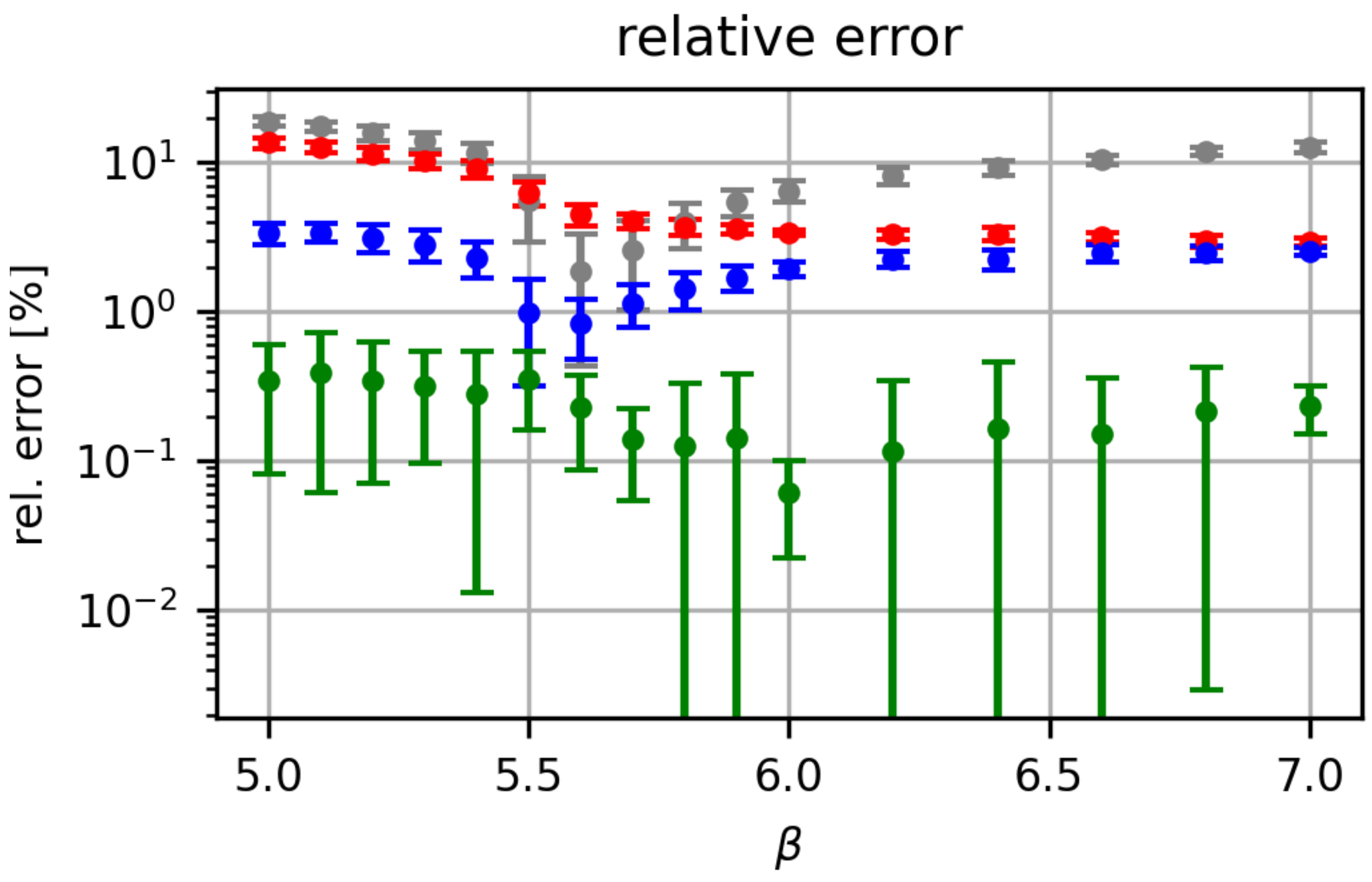
Training example: L-CNN model with

- 3 layers with 12, 24, 24 channels each
- parallel transport in ± 1 in first 2 layers
- local in 3rd layer

	L-CNN	APE431	APE444	Wilson
L1 (A/V)	0.02148	0.19690	0.62189	0.90898
rel. err.	0.226%	2.1965%	6.1356%	9.7577%
inv. loss (DA)	0.13480	0.23799	0.49264	0.73533

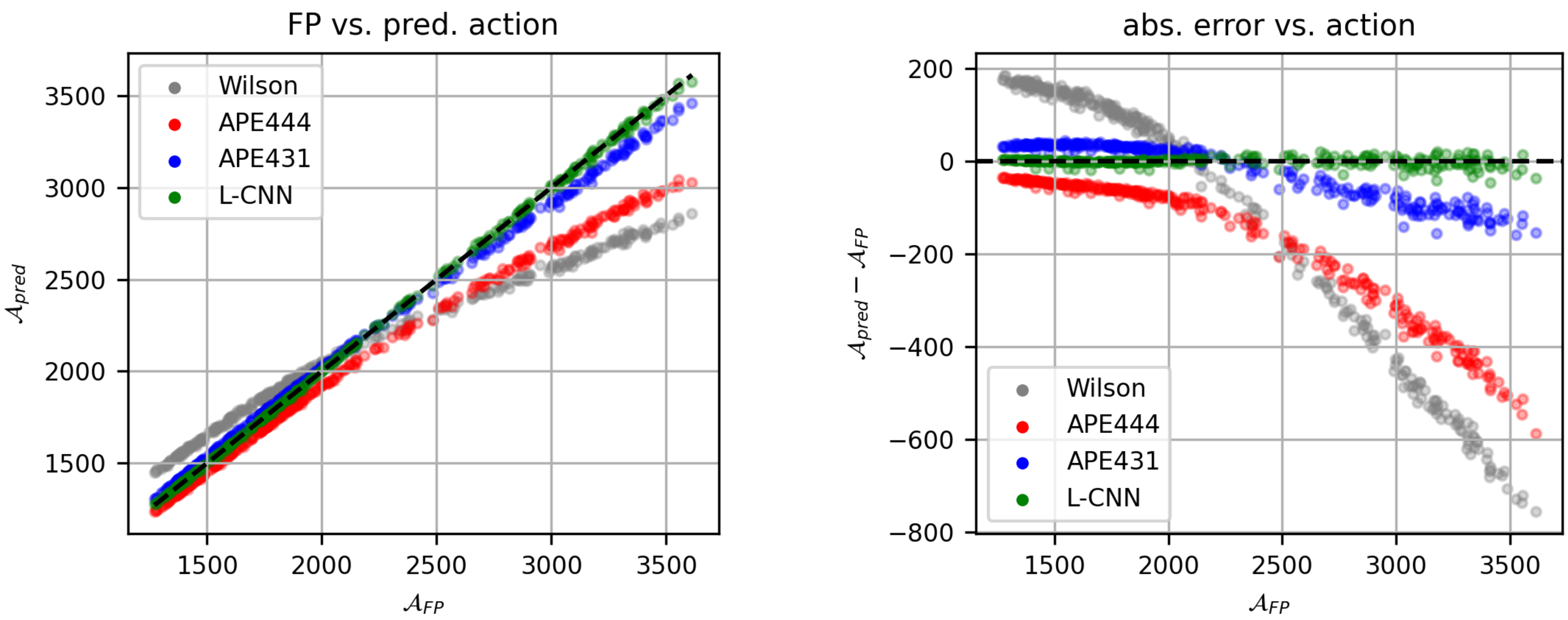
Older parametrizations of FP action as baselines:





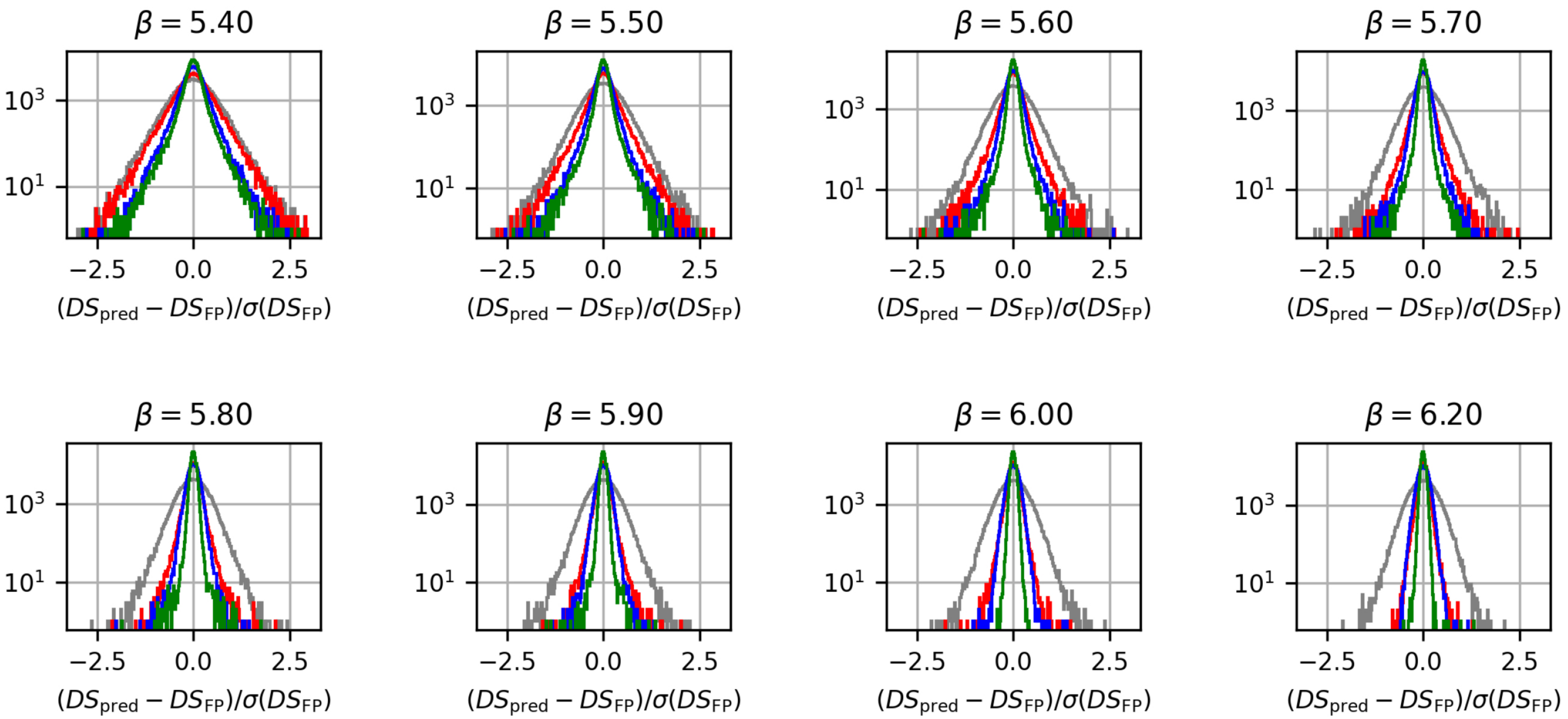
Machine learning the FP action

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



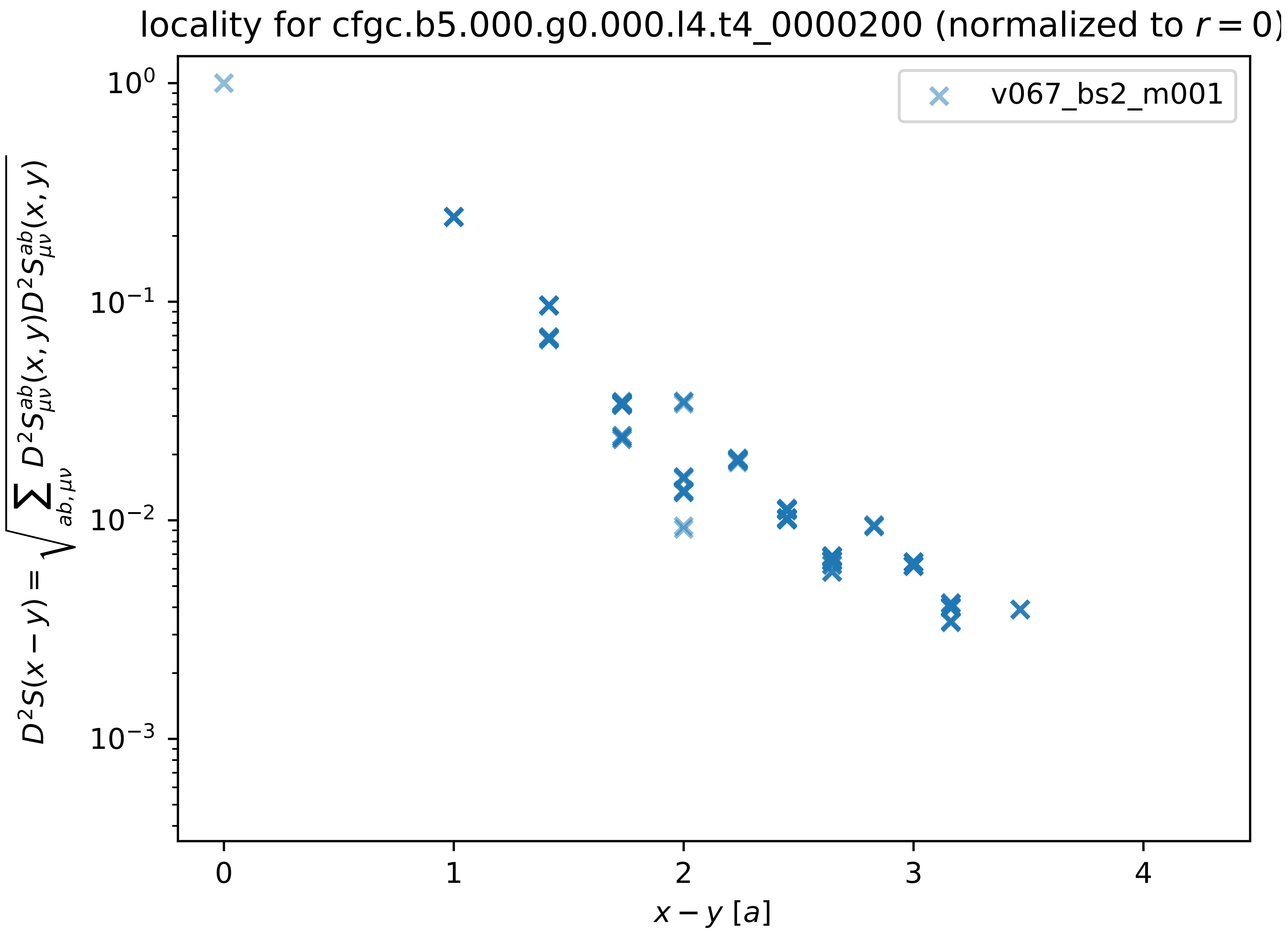
Machine learning the FP action

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



Machine learning the FP action

Locality of L-CNN trained FP action $\dfrac{\delta^2 A^{\mathrm{pred}}(V)}{\delta V^a_{x,\mu} \delta V^b_{y,\nu}}$:



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

FP action with L-CNN

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 (based on derivatives)
- apply exact RGT step(s)

Backup slides

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:

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

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

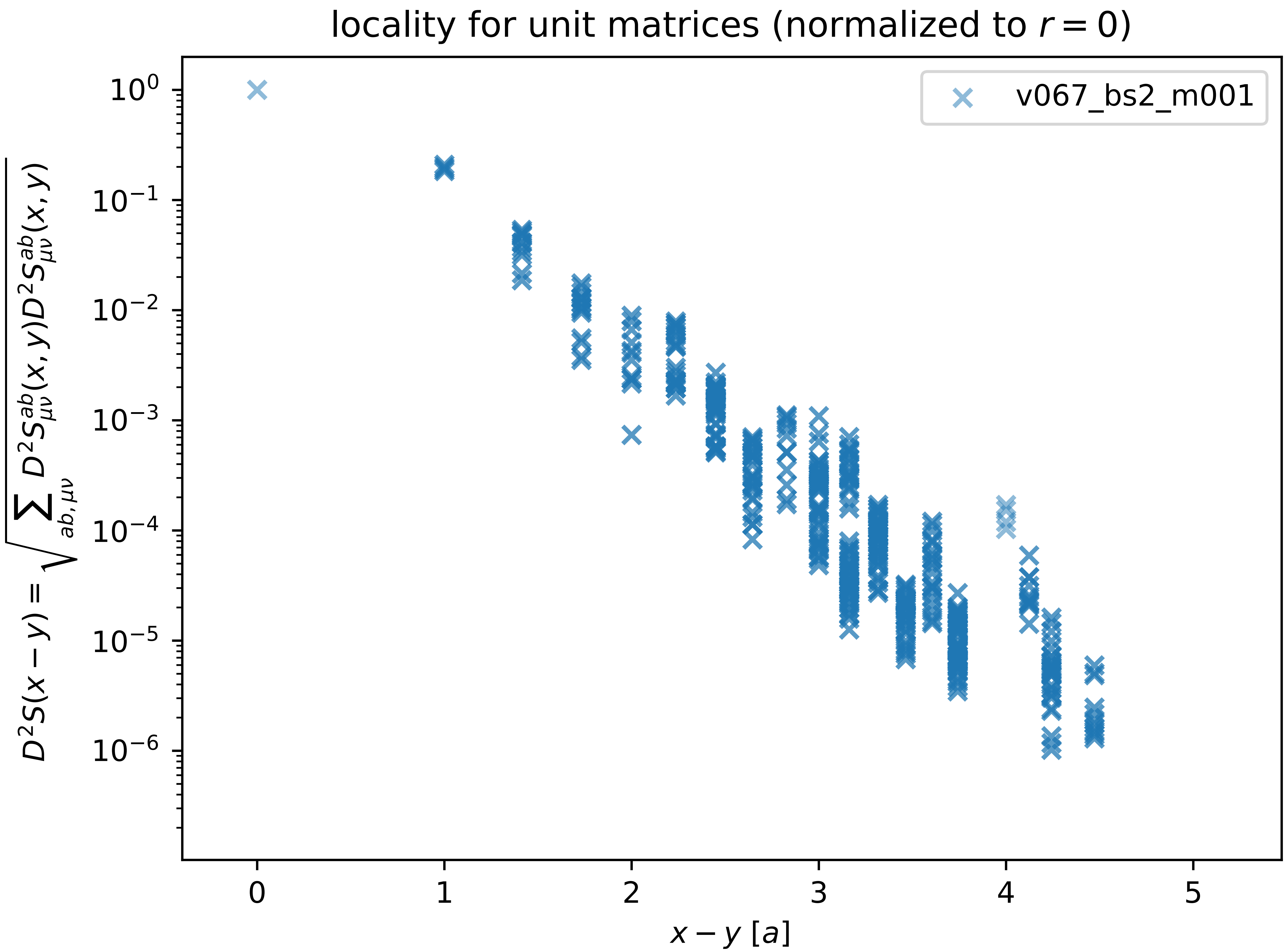
- FP action: $A^{FP}[U] = \sum_{\mu<\nu} f(u_{\mu\nu}, w_{\mu\nu})$ e.g. $f(u,w) = \sum_{k,l} p_{kl}\, u^k w^l$

- Asymmetrically smeared links:

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

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

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



locality of L-CNN trained action $\dfrac{\delta^2 A^{\mathrm{pred}}(V)}{\delta V^a_{x,\mu}\,\delta V^b_{y,\nu}}$ **exponential fall off, as desired**