Machine learning a fixed point action (the sequel)

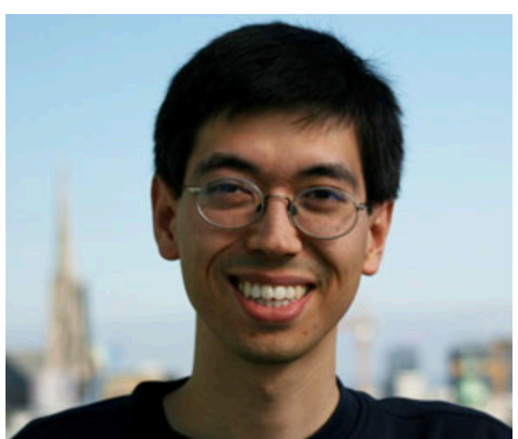
Kieran Holland
University of the Pacific
June 28 2023

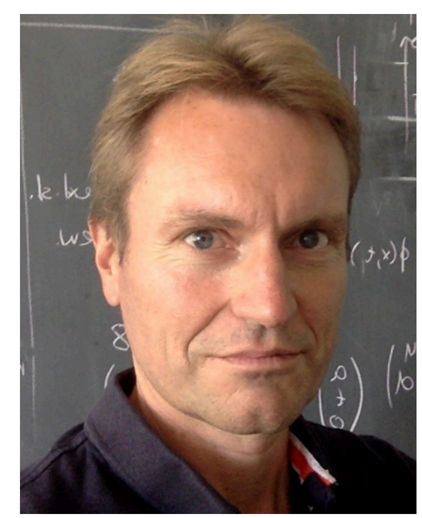
FONDAZIONE BRUNO KESSLER | ECT* EUROPEAN CENTRE FOR THEORETICAL STUDIES IN NUCLEAR PHYSICS AND RELATED AREAS

UNIVERSITY OF THE PACIFIC 1851

TU WIEN

u^b UNIVERSITÄT BERN





Andreas Ipp (Wien) David Müller (Wien) Urs Wenger (Bern)

Urs' talk: general properties of FP actions

my talk: applying ML to parametrize 4-d SU(3) gauge theory FP action

revisiting an old idea with new tools, motivated by difficulties pushing towards continuum limit (critical slowing down, topological freezing, …)

older work (90s): Anna & Peter Hasenfratz, Niedermayer, DeGrand, Wiese, Bietenholz, …

2 parts:
- what data can one generate for FP actions
- how to train an ML with this information

we know a lot about FP actions but not their explicit parametrization

looks like ideal match for ML !

connection to L-CNN work from Wien

(Lattice Gauge Equivariant - Convolutional Neural Networks)

RG step produces new action

$$\exp(-\beta' A'[V]) = \int dU \exp(-\beta(A[U] + T[U,V]))$$

RG blocking kernel - free to choose

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

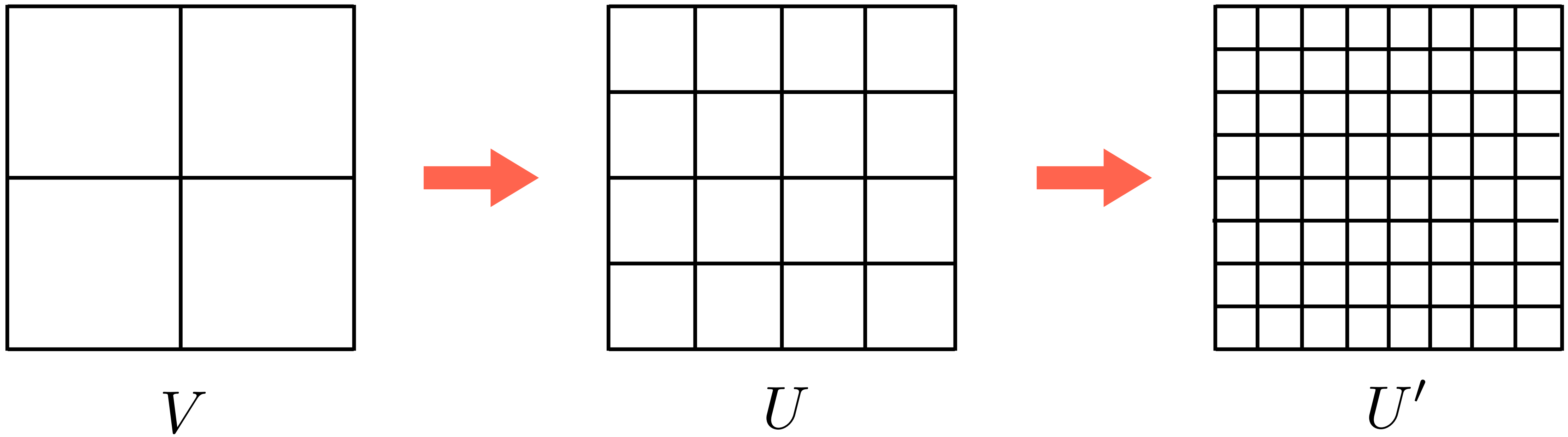
Fixed Point (FP) is saddle point eq in limit $\beta \to \infty$

$$A^{\mathrm{FP}}(V) = \min_{\{U\}}(A^{\mathrm{FP}}(U) + T(U,V))$$

with normalization term

$$\mathcal{N}^\infty_\mu = \max_{W \in \mathrm{SU}(N)} \{\mathrm{ReTr}(W Q^\dagger_\mu)\}$$

For given coarse configuration V, action value determined once minimizing fine configuration U is found



minimization is like reverse of RG step

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

$$= \min_{\{U'\}}\left(A^{\mathrm{FP}}(U') + T(U',U)\right)$$

in principle, iterate many times (Inception)

in practice, action density much smaller as lattice spacing decreases

can use a simpler action on RHS e.g. $A^{\mathrm{Wil}}(U)$

good choice for RHS determined by fluctuations of V

minimizing configuration has special properties

$$A^{\mathrm{FP}}(V) = \min_{\{U\}}(A^{\mathrm{FP}}(U) + T(U,V))$$

then

$$\frac{\delta A^{\mathrm{FP}}[V]}{\delta V^a_{x,\mu}} = \frac{\delta T[U,V]}{\delta V^a_{x,\mu}}$$

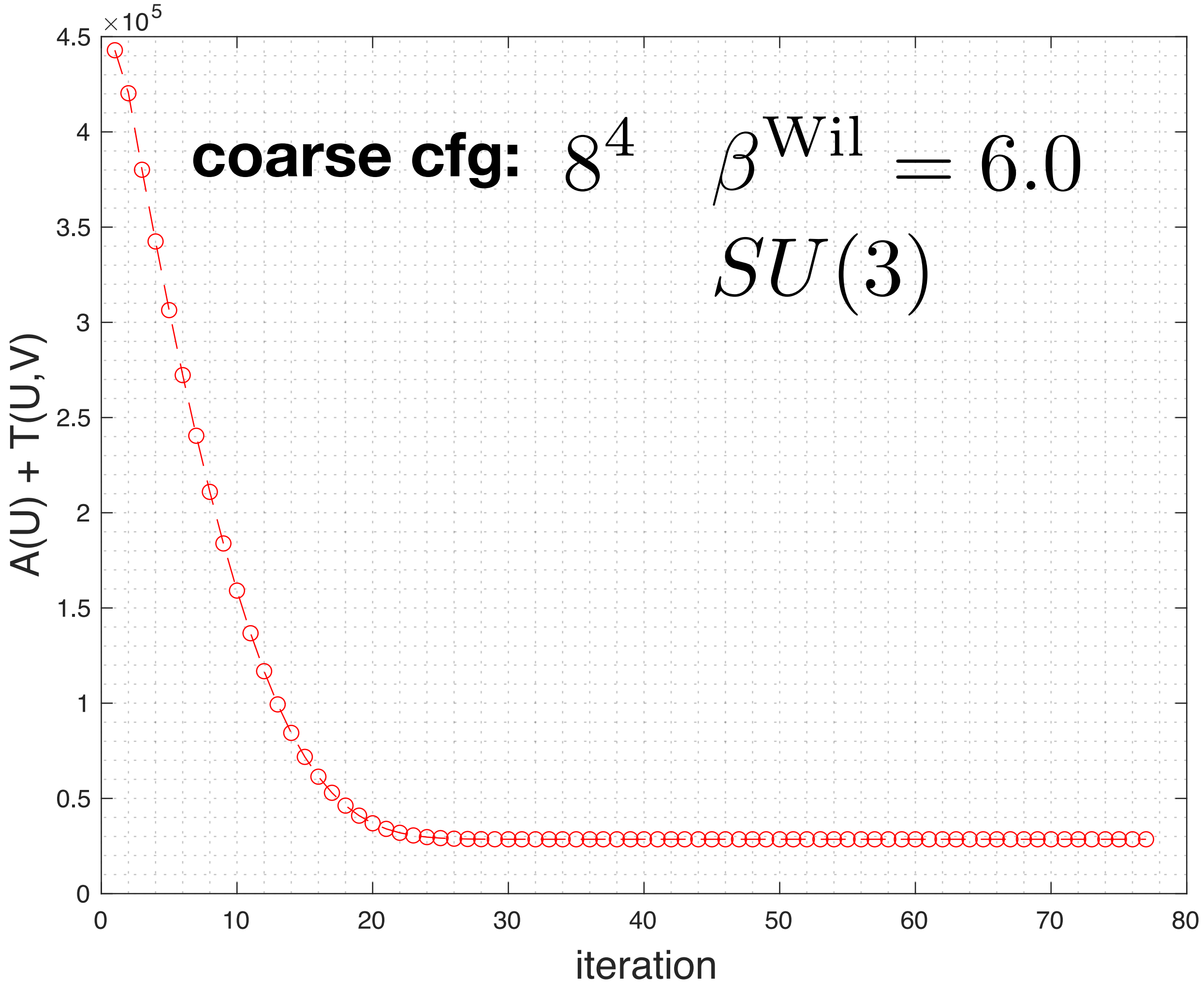
$$= -\kappa \ \mathrm{ReTr}(it^a V^a_{x,\mu} Q^\dagger_{x,\mu})$$ **not summed over** a

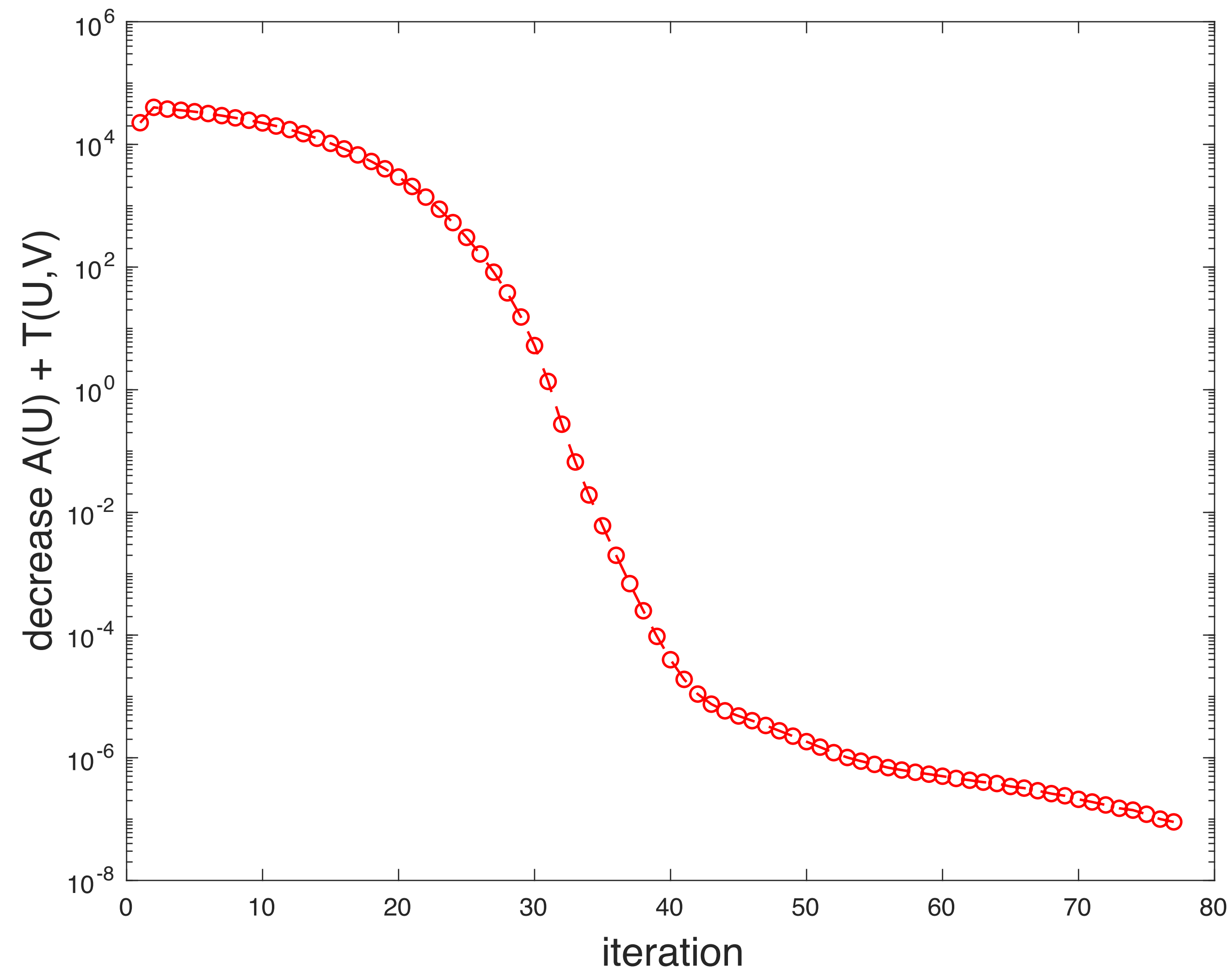
because other term in variation vanishes:

$$\frac{\delta}{\delta U}(\min_{\{U\}}(A^{\mathrm{FP}}[U] + T[U,V])) \cdot \frac{\delta U}{\delta V} = 0$$

derivatives have lots of information - use later for training

$$D^{\mathrm{FP}}_{x,\mu} = \sum_a t^a \frac{\delta A^{\mathrm{FP}}[V]}{\delta V^a_{x,\mu}}$$





minimization evolution of fine configuration U for fixed coarse configuration V

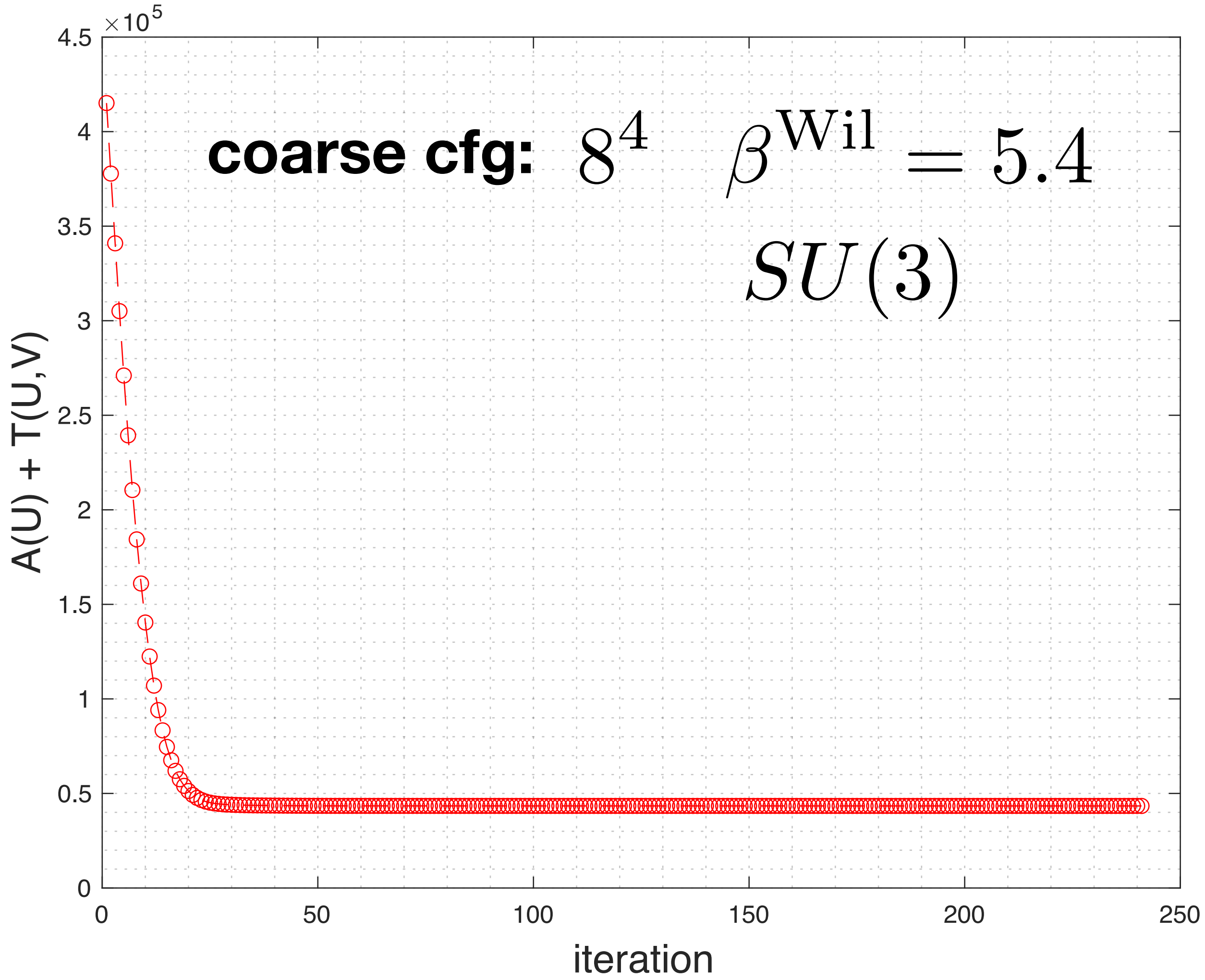
coarse cfg: Monte Carlo gauge configuration: lattice spacing ~ 0.1fm smoother

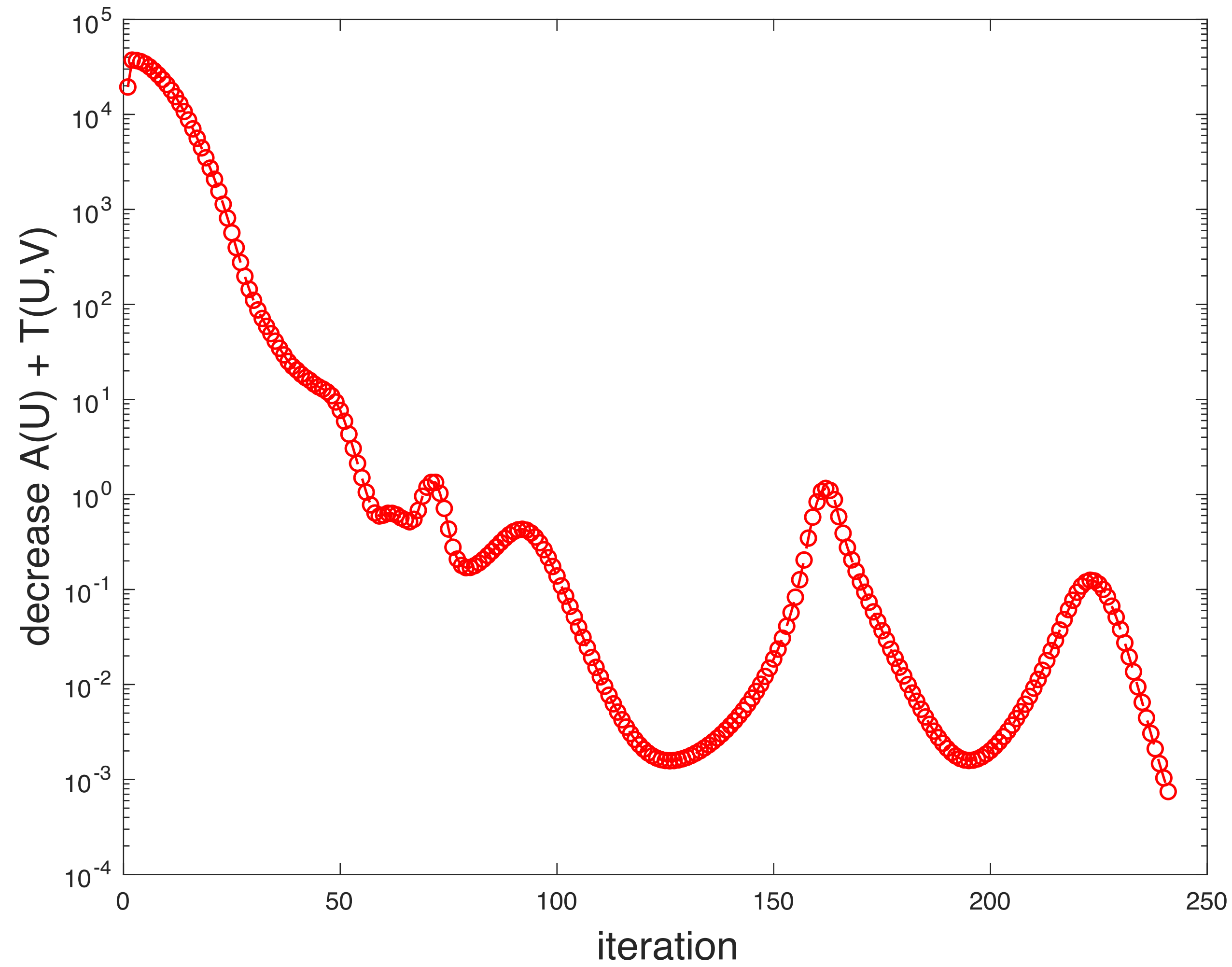
once chosen accuracy level reached, have not only action value $A^{\mathrm{FP}}(V)$

but also many derivatives $$\frac{\delta A^{\mathrm{FP}}(V)}{\delta V^a_{x,\mu}}$$

4 x 8 x Volume

(link) (color) (position)





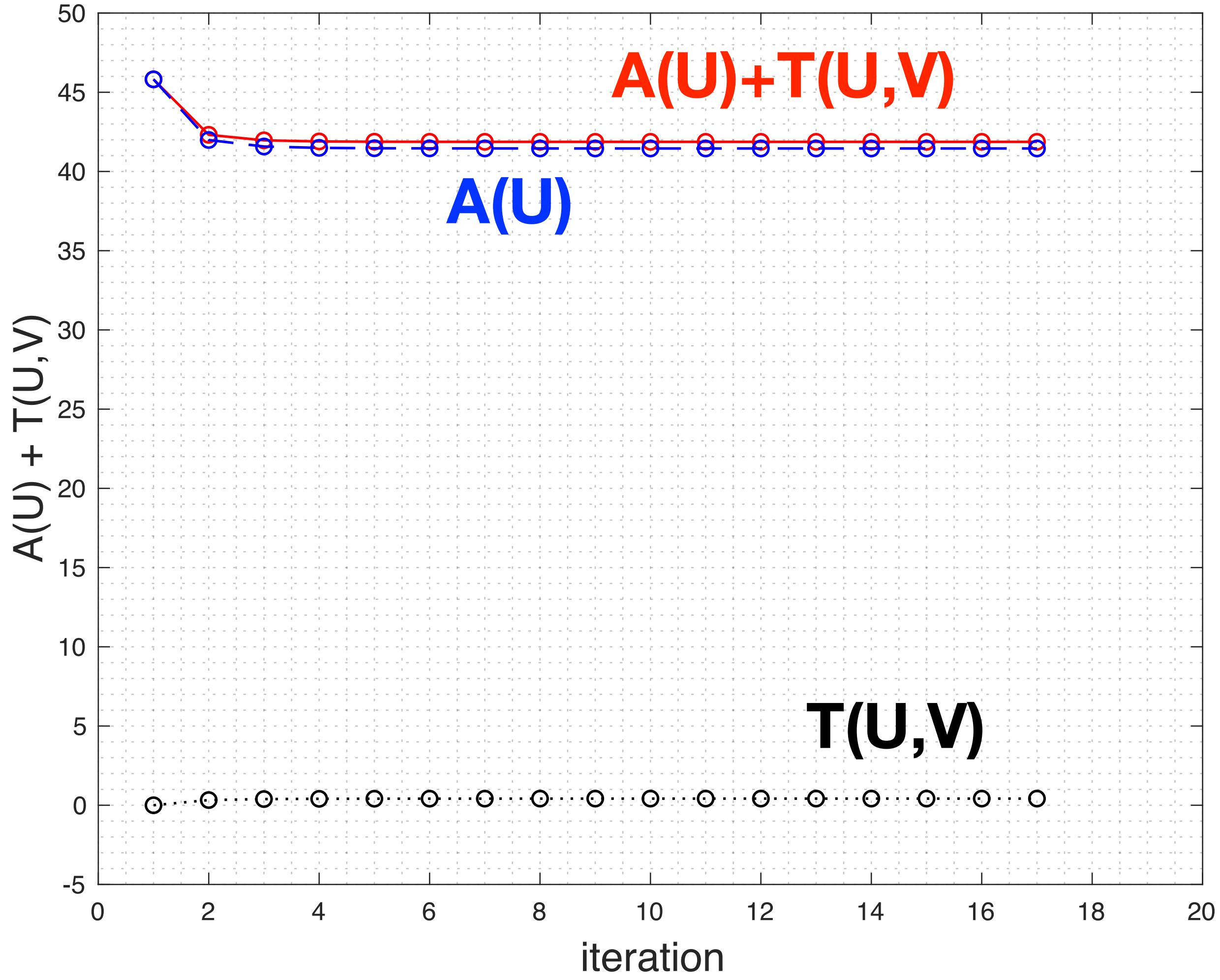
minimization evolution of U for coarse V at larger lattice spacing

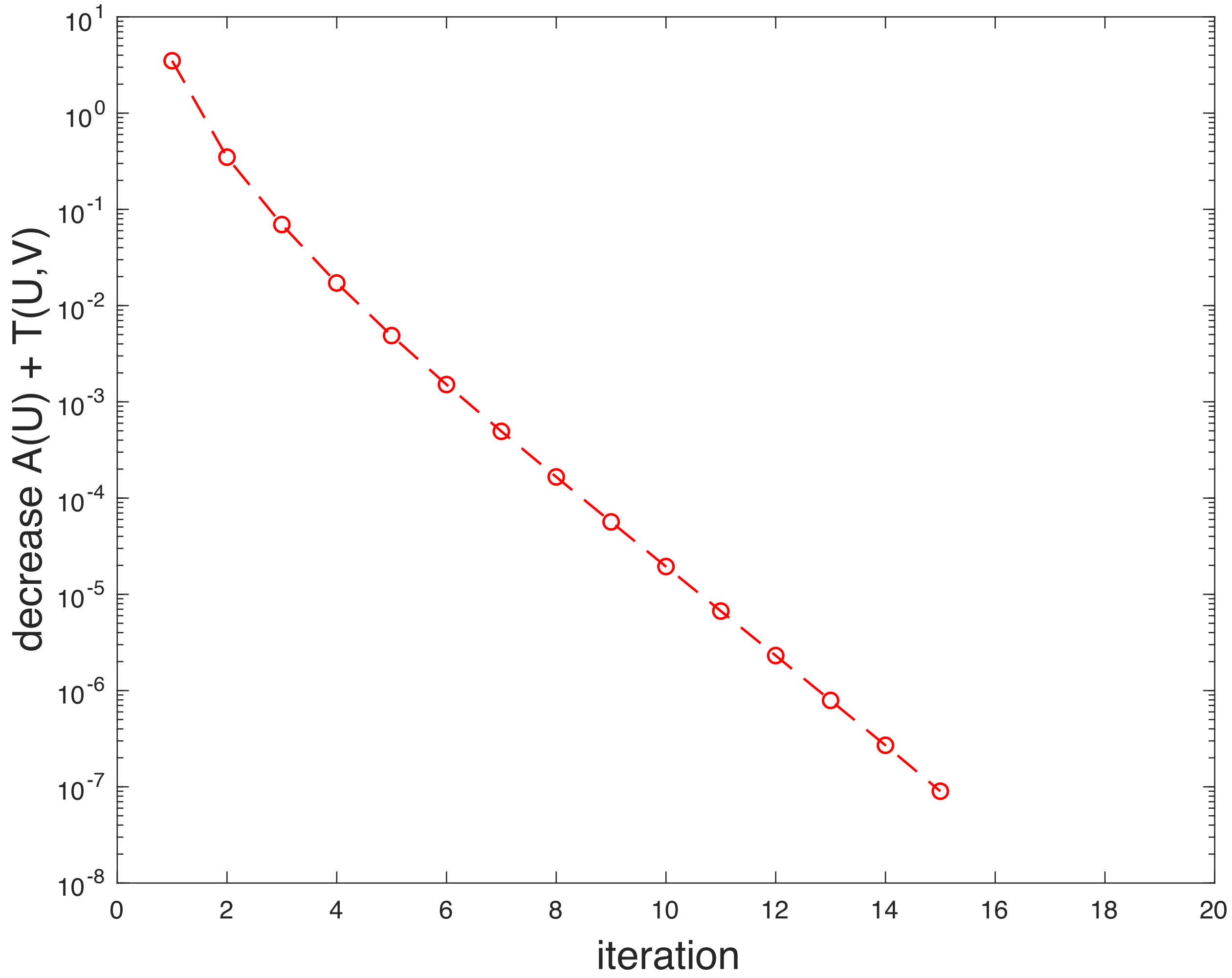
Monte Carlo configuration: lattice spacing ~ 0.25 fm   rougher

minimization takes somewhat longer to converge

generate ensembles of coarse lattice gauge configurations with a range of fluctuations, find minimizing fine configurations for each one, along with corresponding derivatives

the data set for ML training



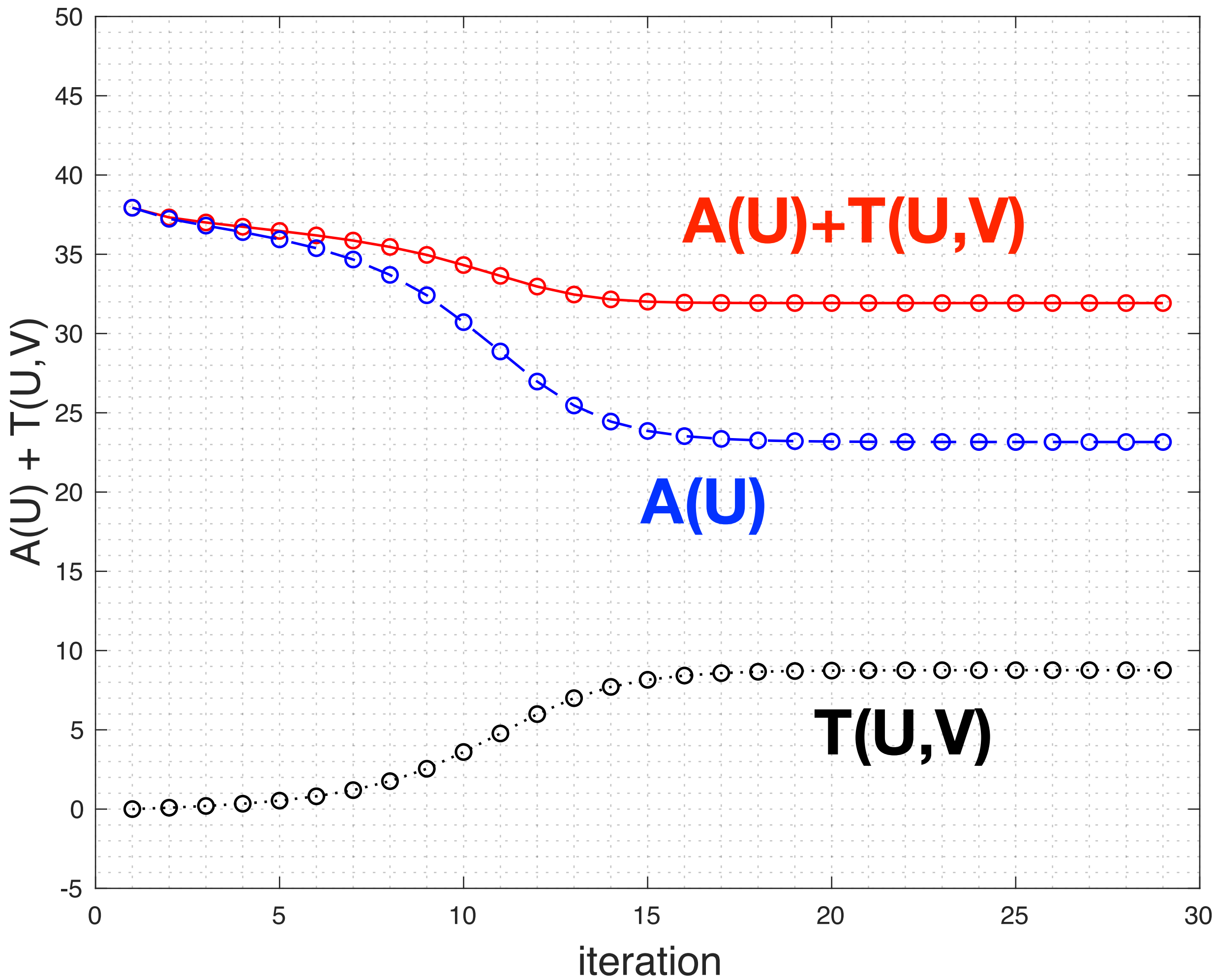


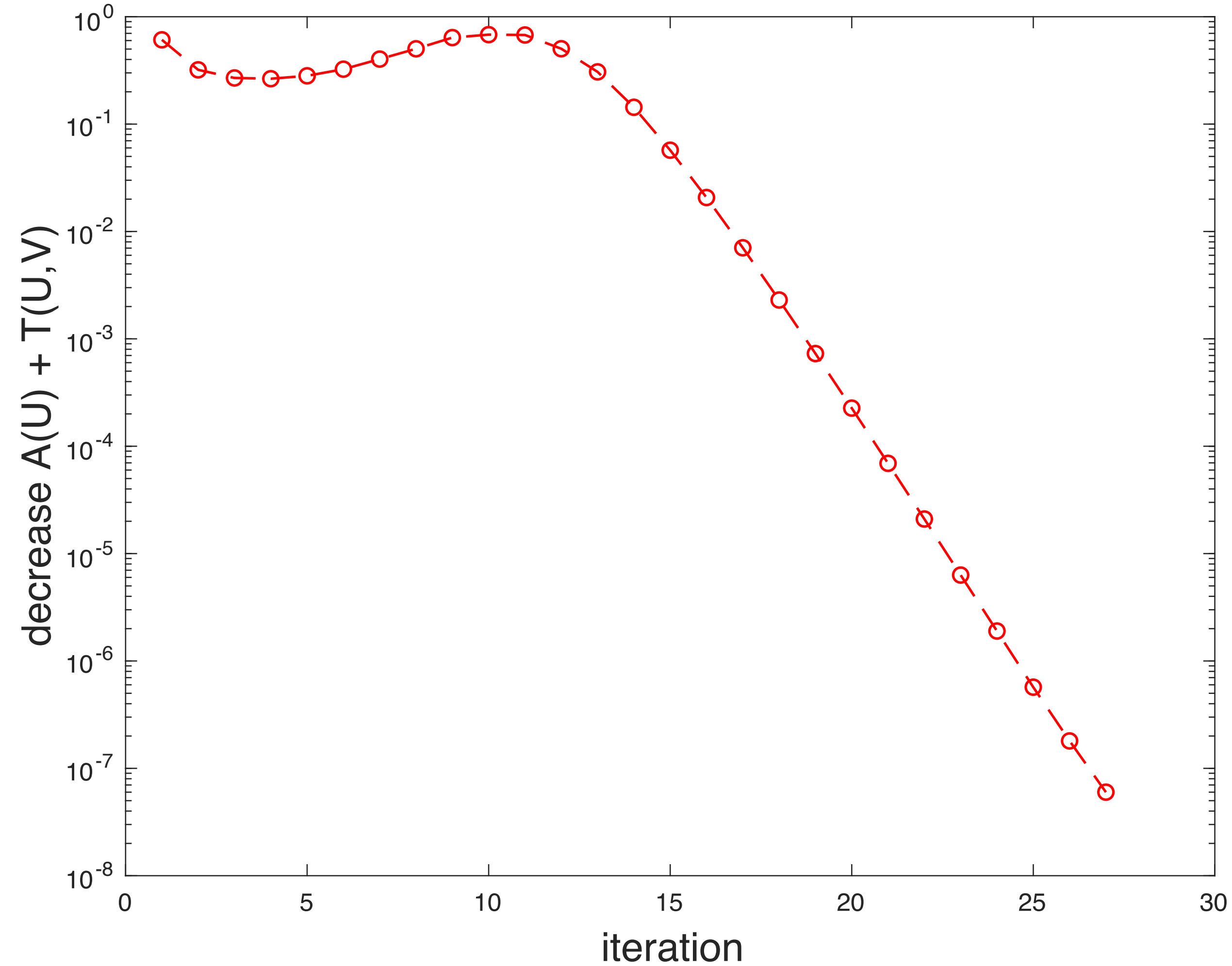
fine instanton configuration, volume 16^4 **radius before blocking** $r/a = 3.0$

coarse cfg: 8^4 **now find minimizing fine cfg U**

action almost unchanged, blocking kernel T(U,V) close to zero

A(U) + T(U,V) close to $4\pi^2$



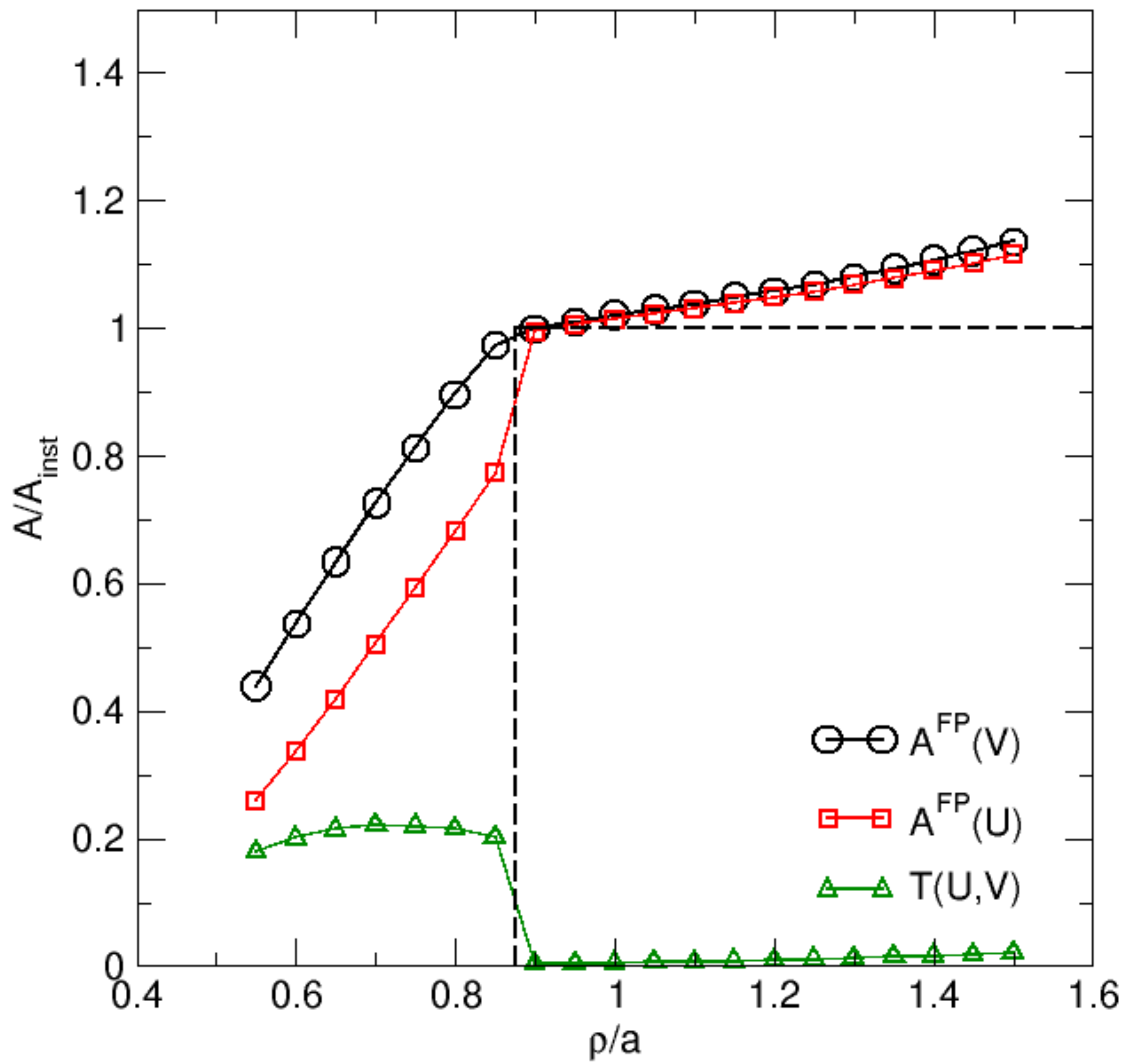


instanton configuration, volume 16^4 **radius before blocking** $r/a = 1.5$

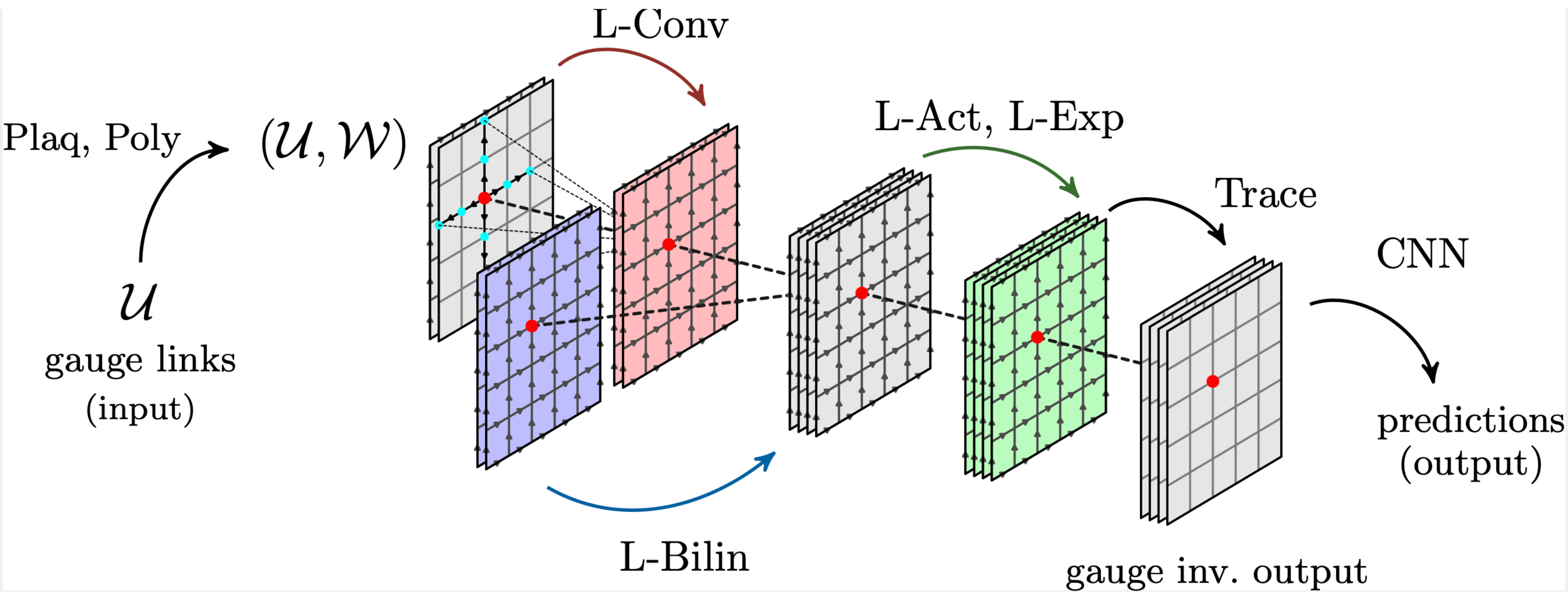
instanton "falls through" the lattice during RG blocking:

kernel T(U,V) much larger

total action below continuum instanton value



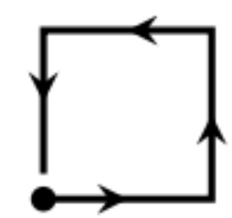
repeat for many instanton sizes, lost when radius too small



ML architecture: L-CNN (also in talks by Andreas, Daniel, Matteo)

input: gauge links and untraced plaquettes

$$U_{x,\mu} \qquad U_{x,\mu\nu} = U_{x,\mu} U_{x+\mu,\nu} U^{\dagger}_{x+\nu,\mu} U^{\dagger}_{x,\nu} =$$



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

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

what kind of loss function?

$$\frac{1}{N_{\mathrm{cfg}}} \sum_{i} |A^{\mathrm{FP}}(V_i) - A^{\mathrm{pred}}(V_i)|$$

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

weight the two contributions

technical point:

derivatives in L-CNN given by back propagation, since they are derivatives of part of loss function w.r.t. input

training example

L-CNN model with 3 layers

12 channels in each layer

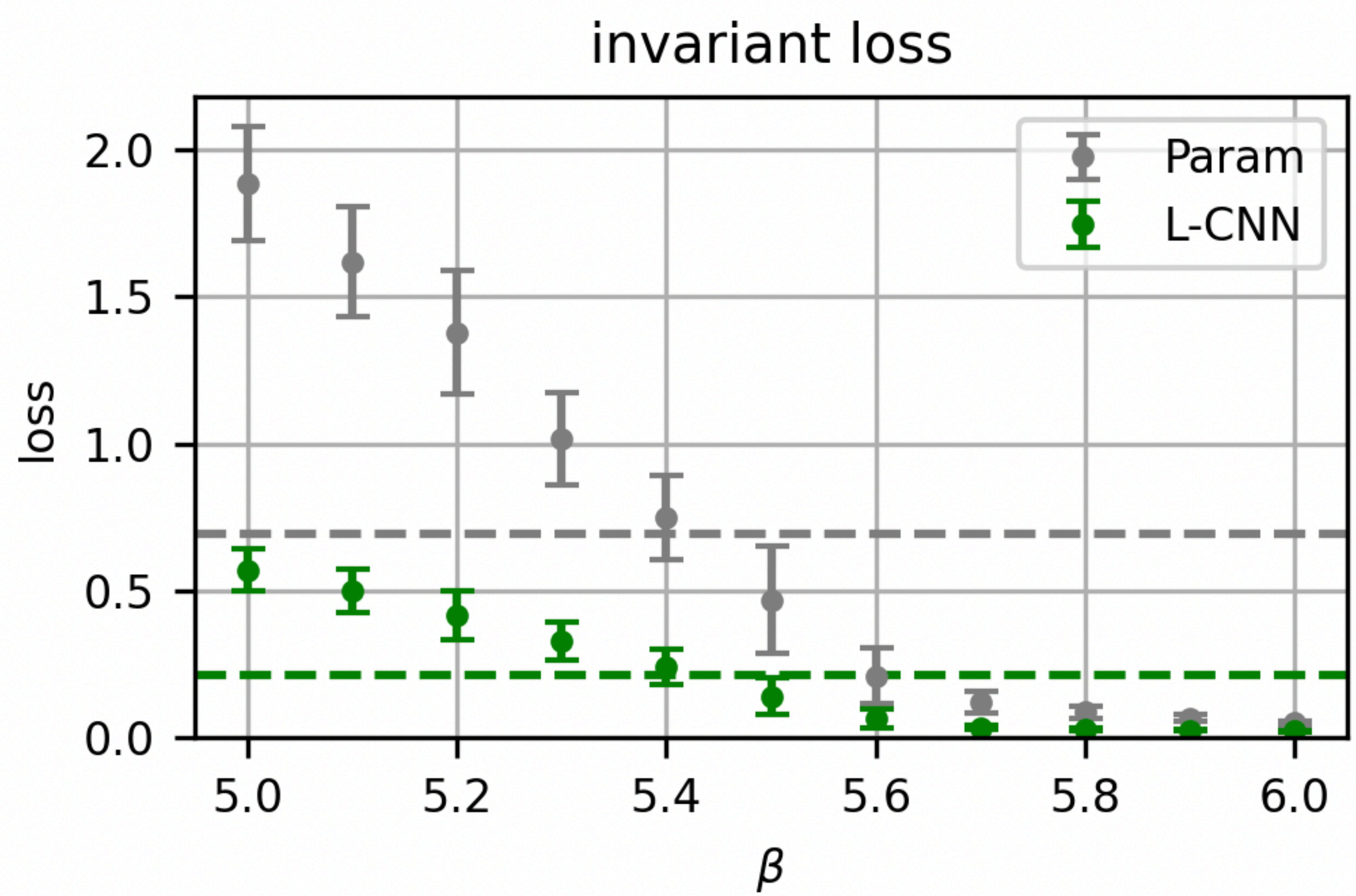
parallel transport ± 1 in 1st two layers

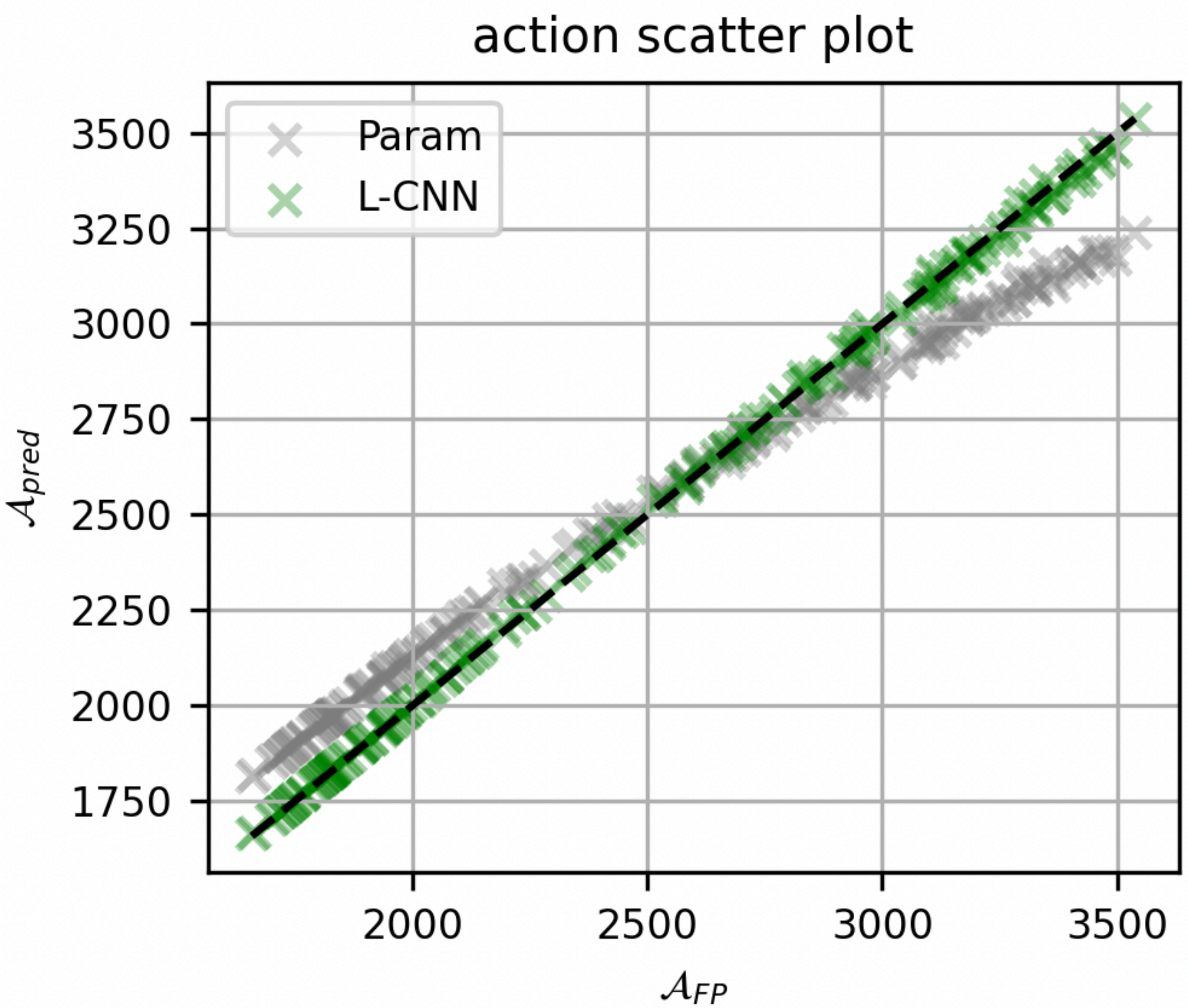
local in 3rd layer

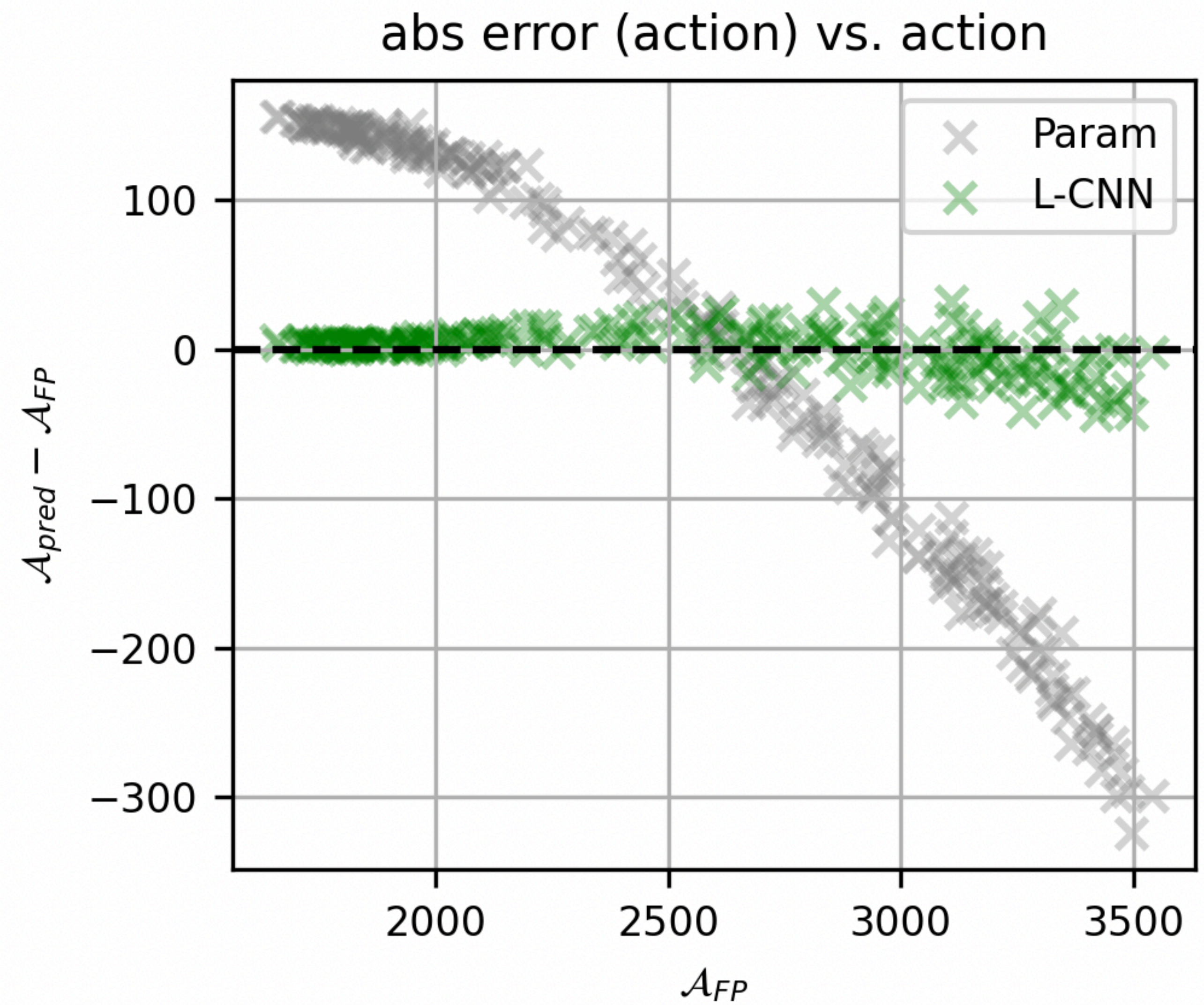
have older parametrization of FP action as baseline

L-CNN does much better

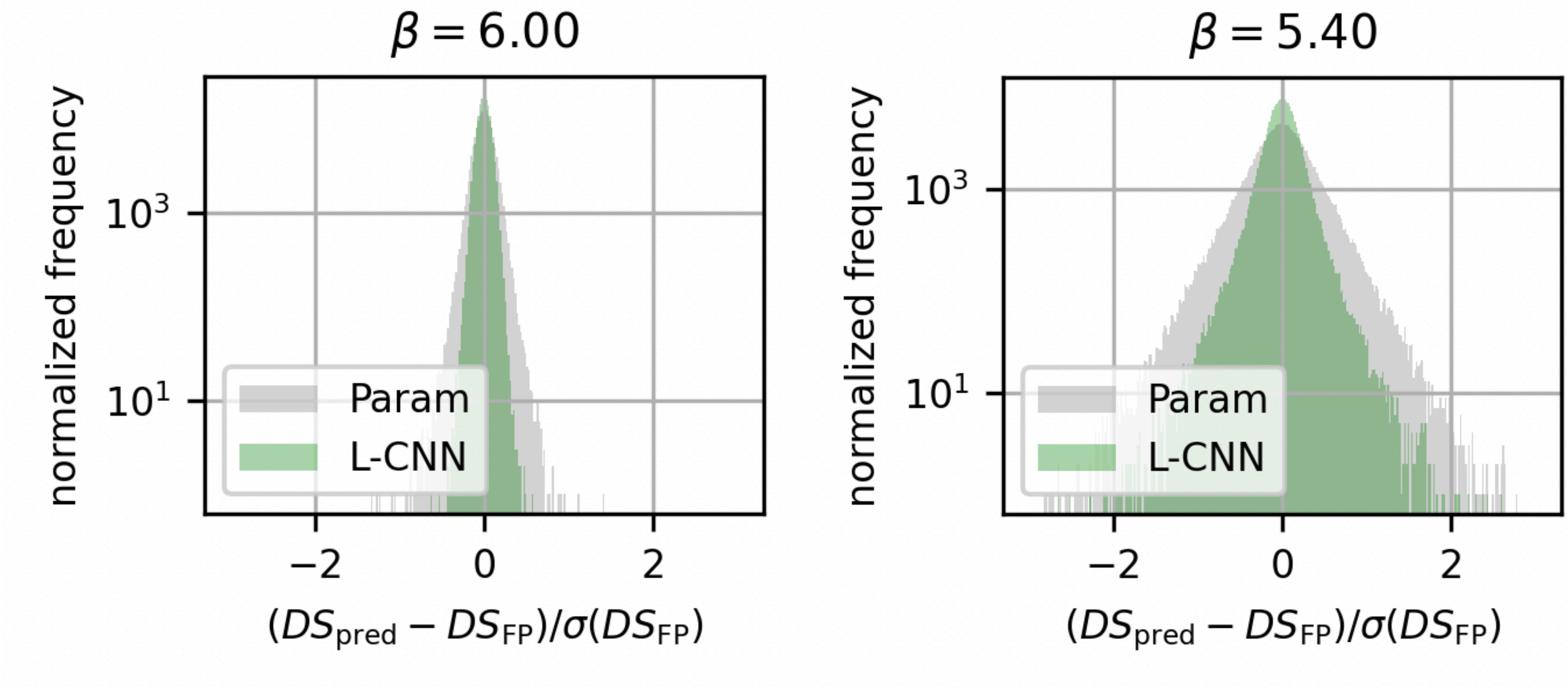
	L-CNN*	Param*	Param (bare)
L1 (A)	9.71447	130.31018	206.87717
L1 (DA)	0.10489	0.19214	-
rel. err (A)	0.360%	5.4557%	7.3985%
inv. loss (DA)	0.21537	0.69545	-





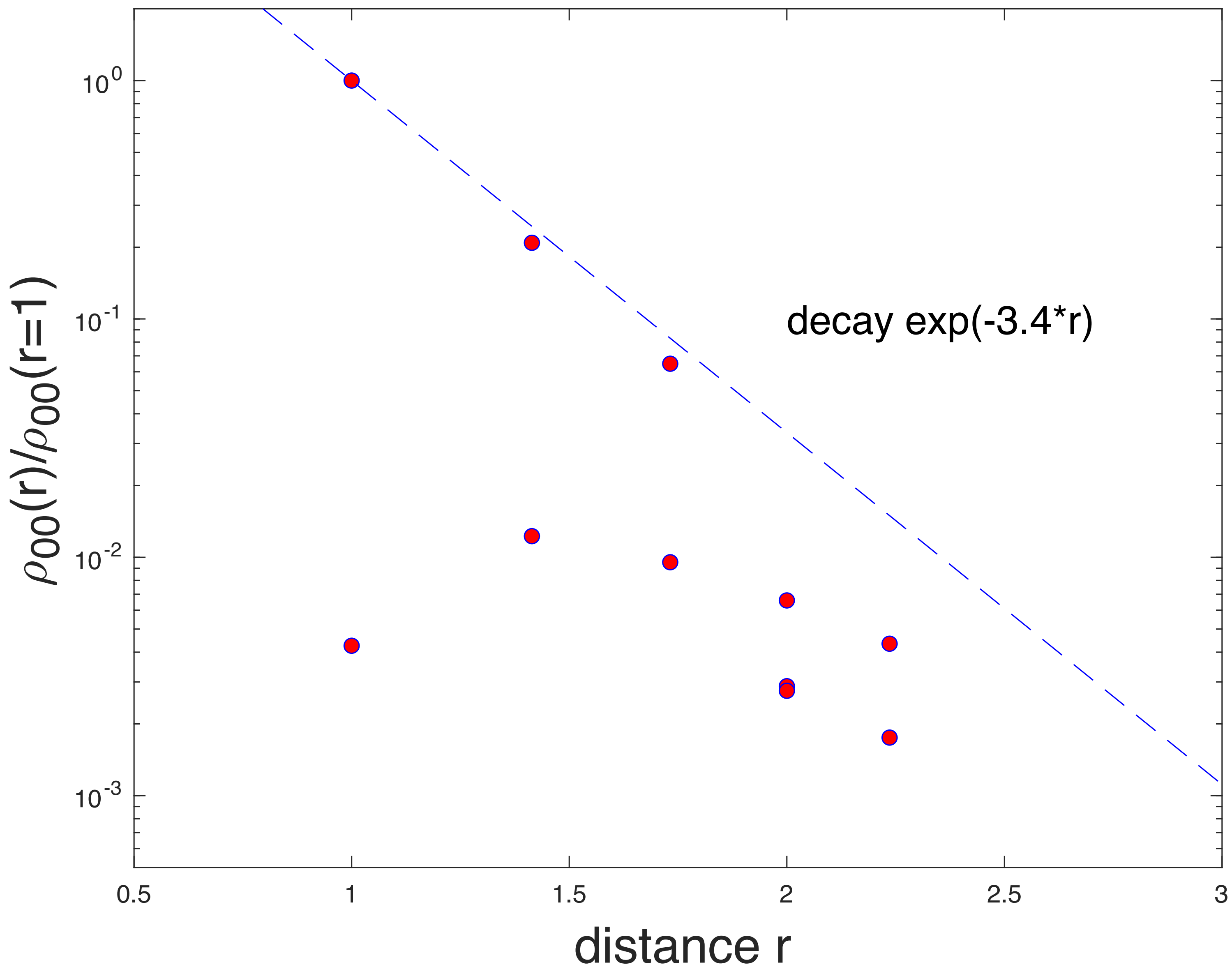


superiority of L-CNN over Param of FP action more visible in action differences



distributions of action derivatives are narrower for L-CNN than Parametrized

(note vertical axis is log scale)

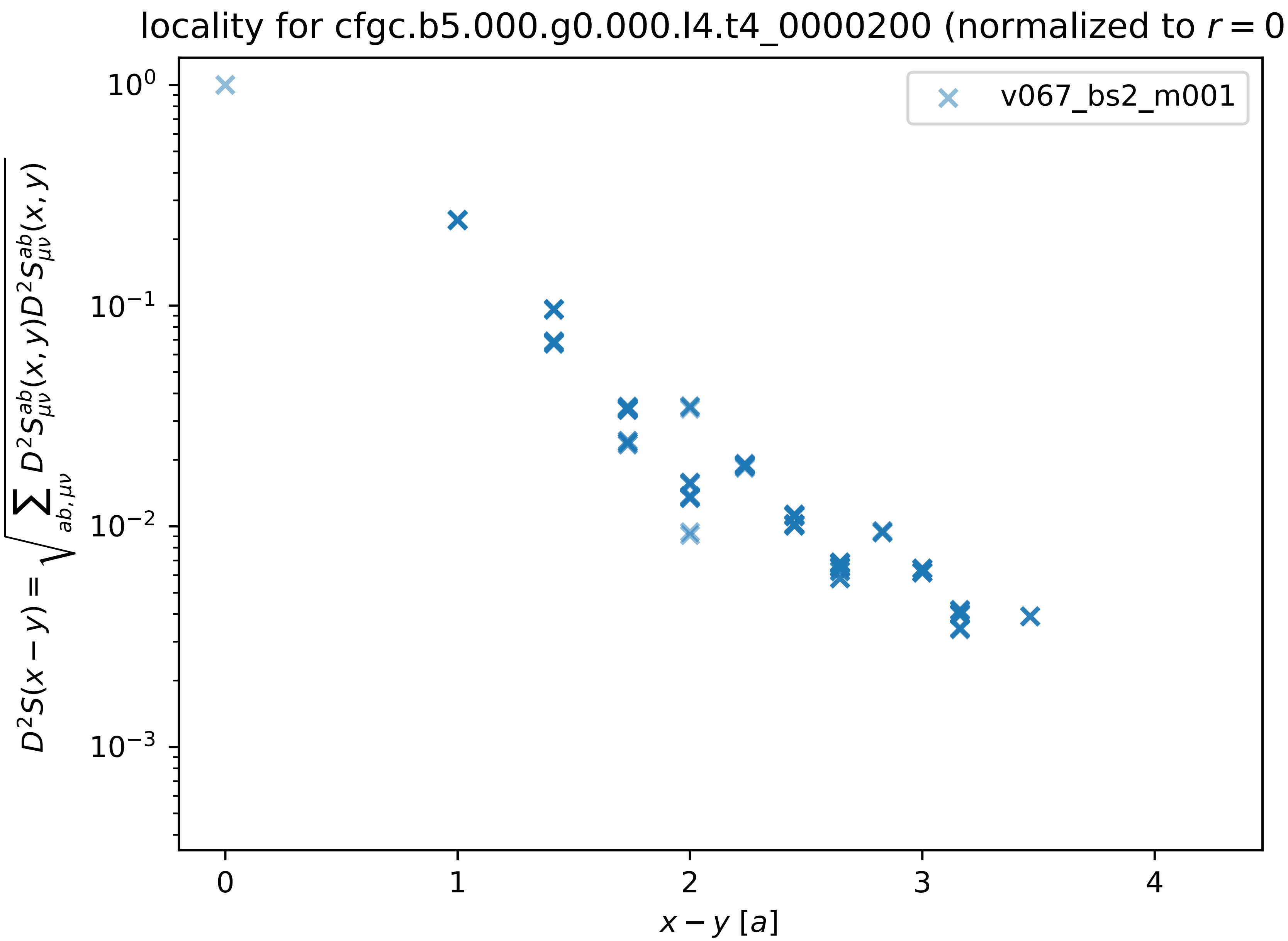


how local is the action?

perturbative expansion to optimize RG blocking choice T(U,V)

action = $$\sum_{x,r} A_\mu(x)\rho_{\mu\nu}(r)A_\nu(x+r)$$ **couplings** $\rho_{\mu\nu}(r)$

must fall off exponentially with distance to guarantee correct continuum limit



locality of L-CNN trained action $$\frac{\delta^2 A^{\text{pred}}(V)}{\delta V^a_{x,\mu} \delta V^b_{y,\nu}}$$

how much are links separated by |x-y| coupled

decreases exponentially, as desired

where we are:

L-CNN captures FP action properties well (better than before)

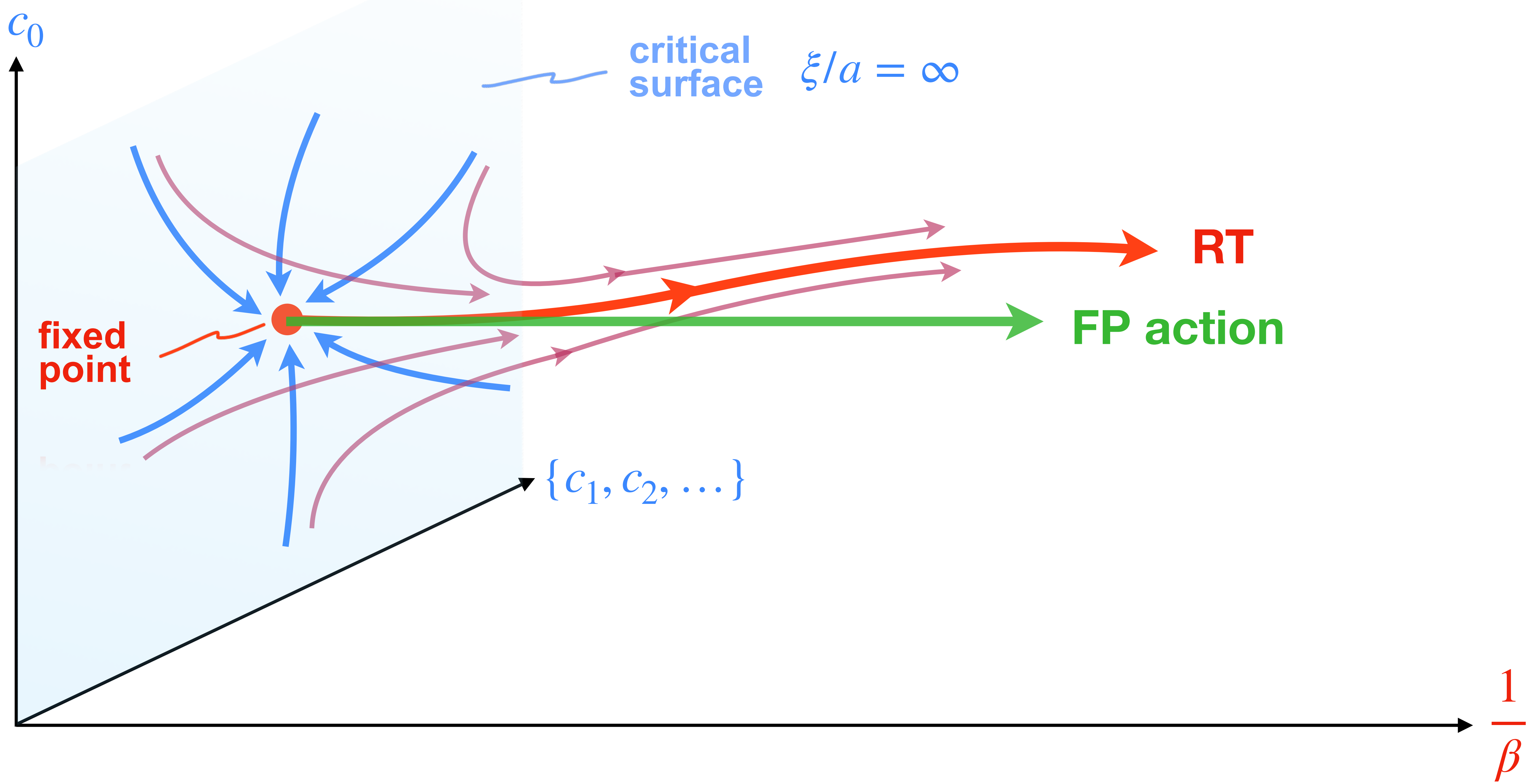
continuing to explore possible architectures

how good are scaling properties of L-CNN action i.e. lattice artifacts reduced?
— don't currently know, requires MC simulation, but older results are encouraging

availability of derivatives hints at algorithm ideas (HMC, …)

go beyond FP, follow actual RG trajectory?

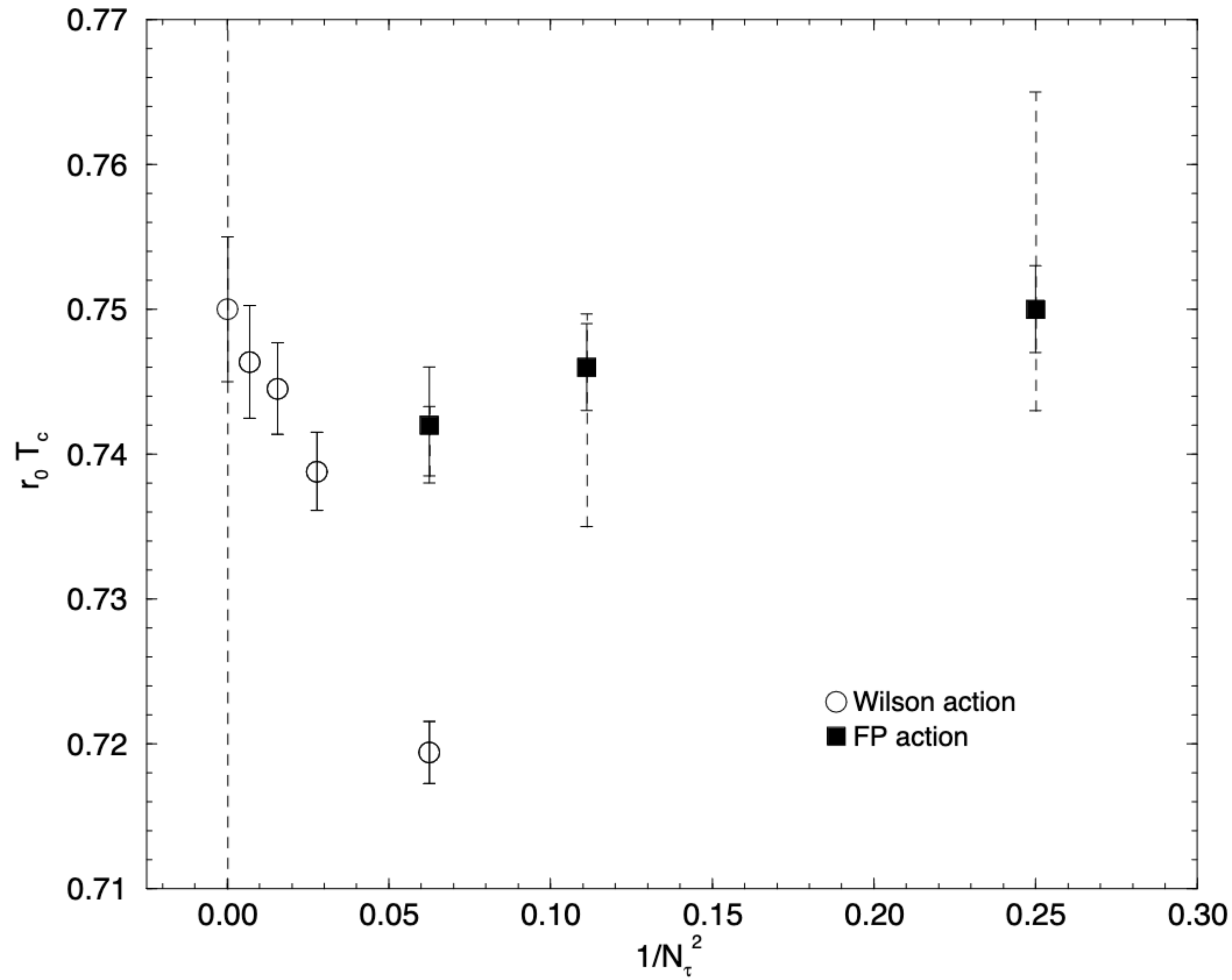
Extra slides



Fixed Point (FP) action close to the RG trajectory at weak coupling

classically perfect

reduced lattice artifacts in full quantum theory

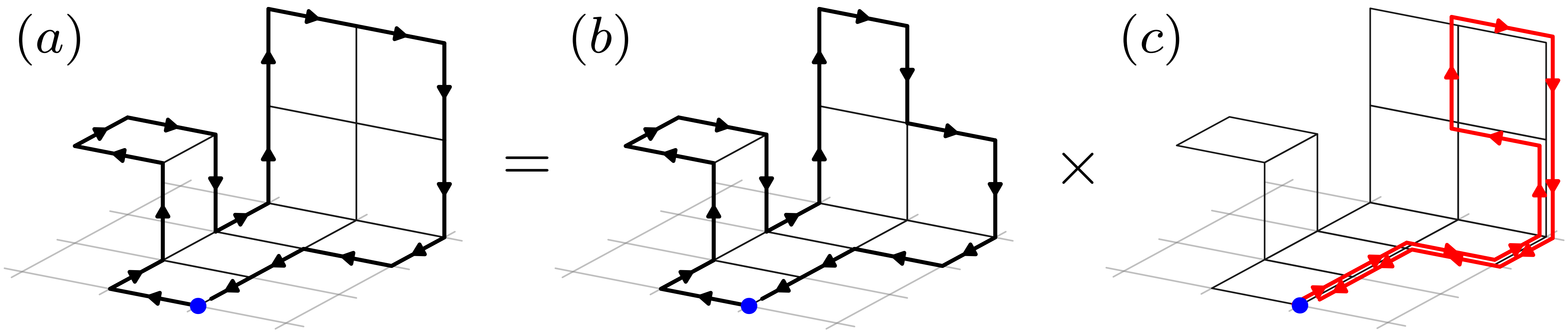


scaling test of previous parametrized FP action

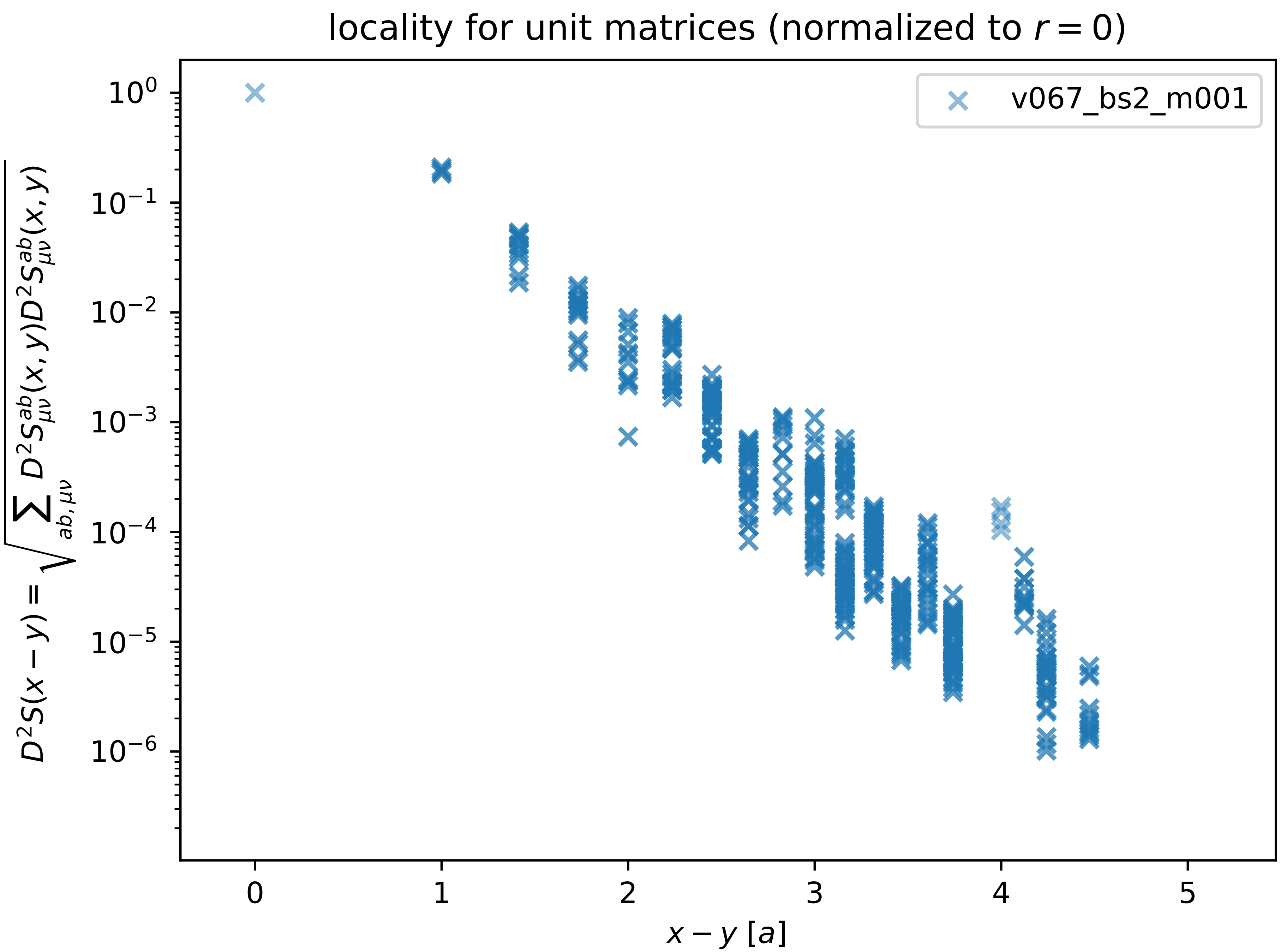
critical temperature for deconfinement transition T_c

Sommer parameter from quark potential r_0

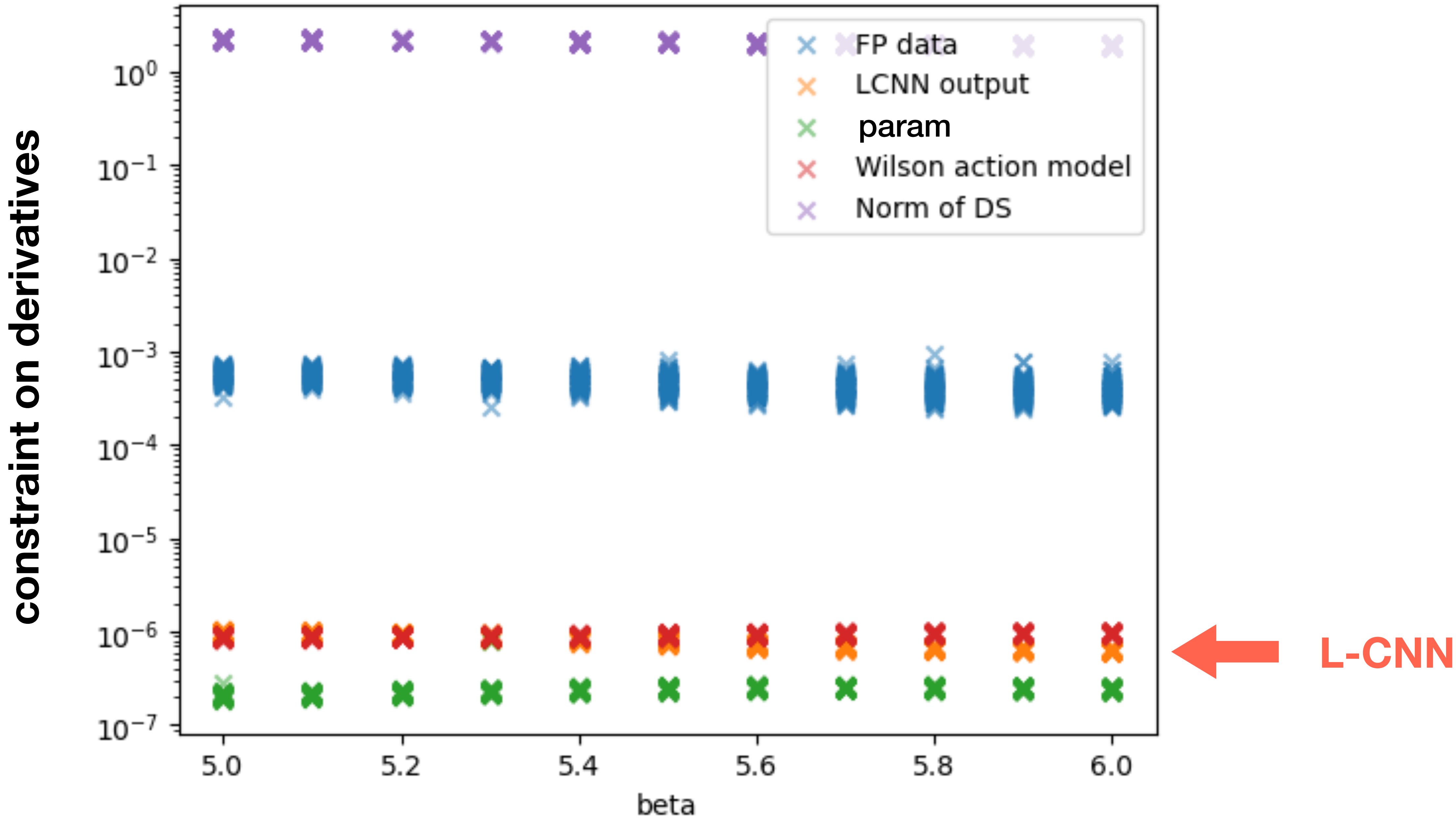
dimensionless combination $r_0 T_c$



proof than sufficiently deep L-CNN with enough layers can generate any arbitrary Wilson loop



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



constraint on action derivatives — not independent

$$\sum_\mu \mathcal{D}^B_\mu D_{x,\mu} S[U] = \sum_\mu \Big(D_{x,\mu} S[U] - U^\dagger_{x-\mu,\mu} (D_{x-\mu,\mu} S[U]) U_{x-\mu,\mu} \Big) = 0$$

satisfied to high accuracy by L-CNN