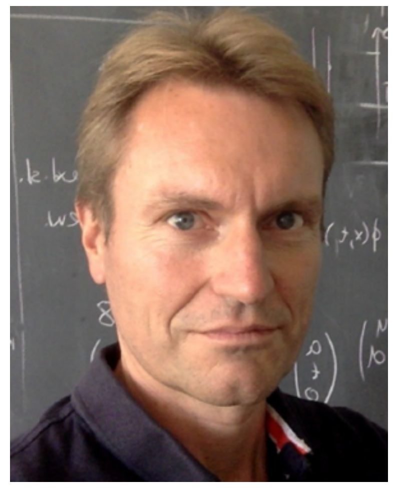
Lattice simulations with machine-learned classically-perfect fixed-point actions



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

e-Print: 2401.06481 [hep-lat]

July 26 2024, ML meets LFT, Swansea University, UK

Outline

- The continuum limit in lattice gauge theory
- Renormalization group transformation and fixed-point (FP) actions
- Learning a new FP action **2401.06481**
- Hybrid Monte-Carlo and FP gradient flow
- Preliminary results for flow observables **New!**

Lattice gauge theory and the continuum limit

Lattice gauge theory

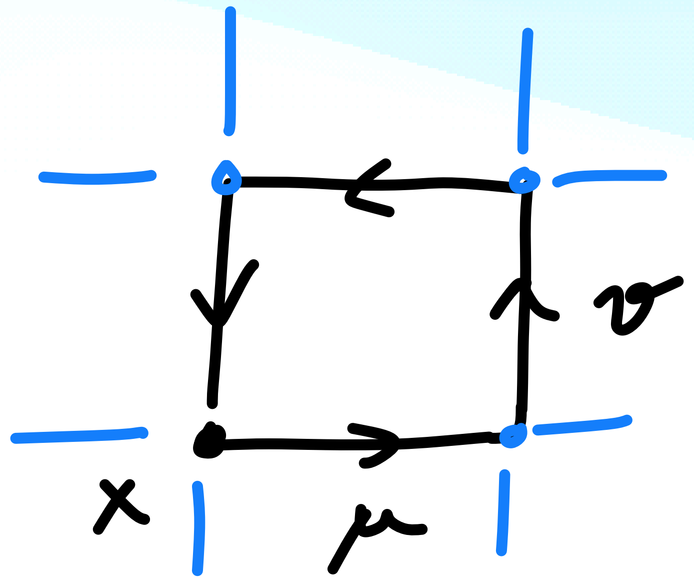
Yang-Mills action

$$S[A_\mu] = \frac{1}{4g^2}\int d^4x\, F^a_{\mu\nu}(x)F^a_{\mu\nu}(x)$$

discretization →

Wilson action

$$S[U] = \frac{\beta}{N_c}\sum_{x,\mu<\nu} \mathrm{ReTr}(1 - U_{x,\mu\nu}) = \beta\mathscr{A}[U]$$



Euclidean lattice path integral for observables $\mathscr{O}$

$$Z(\beta) = \int [\prod_{x,\mu}\mathscr{D}U_{x,\mu}]\exp\left[-\beta\mathscr{A}[U]\right]$$

$$\langle\mathscr{O}\rangle_\beta = \frac{1}{Z(\beta)}\int \mathscr{D}U\,\mathscr{O}[U]\,\exp\left[-\beta\mathscr{A}[U]\right]$$

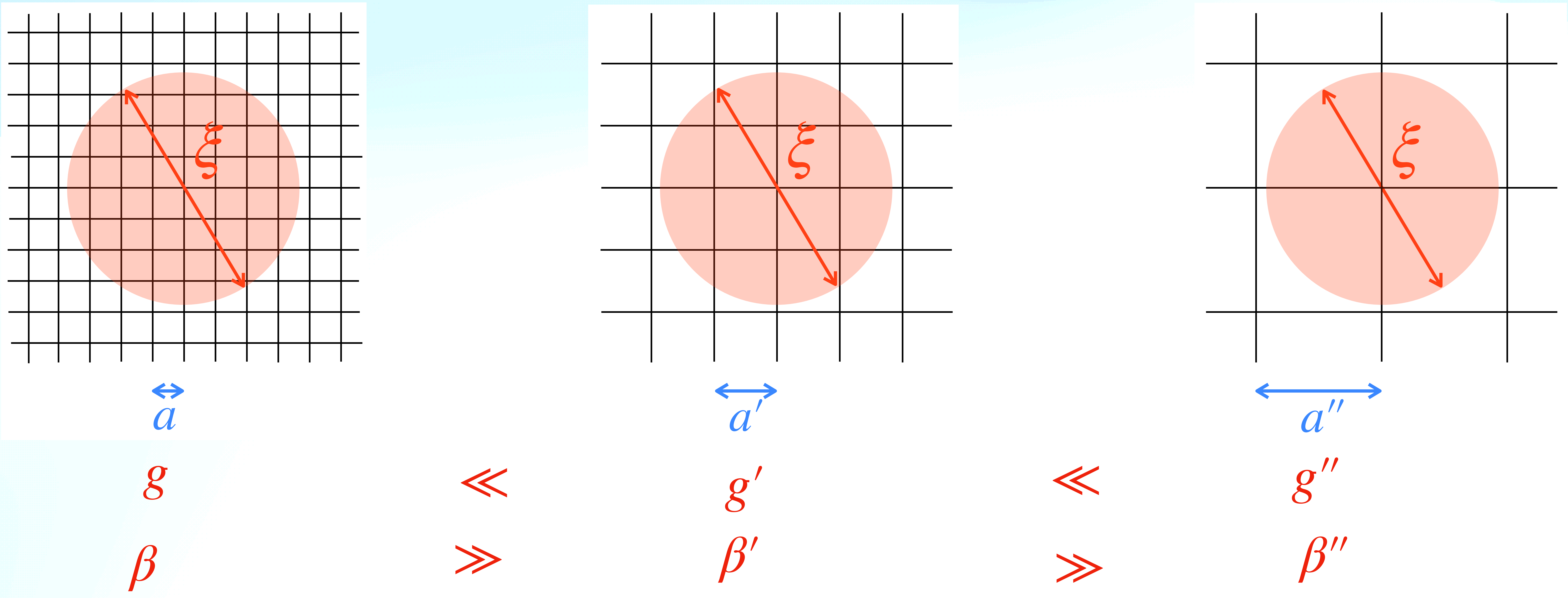
Tractable with Monte Carlo simulations

Renormalization: lattice spacing a is determined by coupling β ("scale setting")

The continuum limit

Extrapolate lattice spacing $a \to 0$

1. Infinite volume $L \to \infty$ (thermodynamic limit)
2. Vanishing gauge coupling $g \to 0$ $(\beta \to \infty)$



The continuum limit

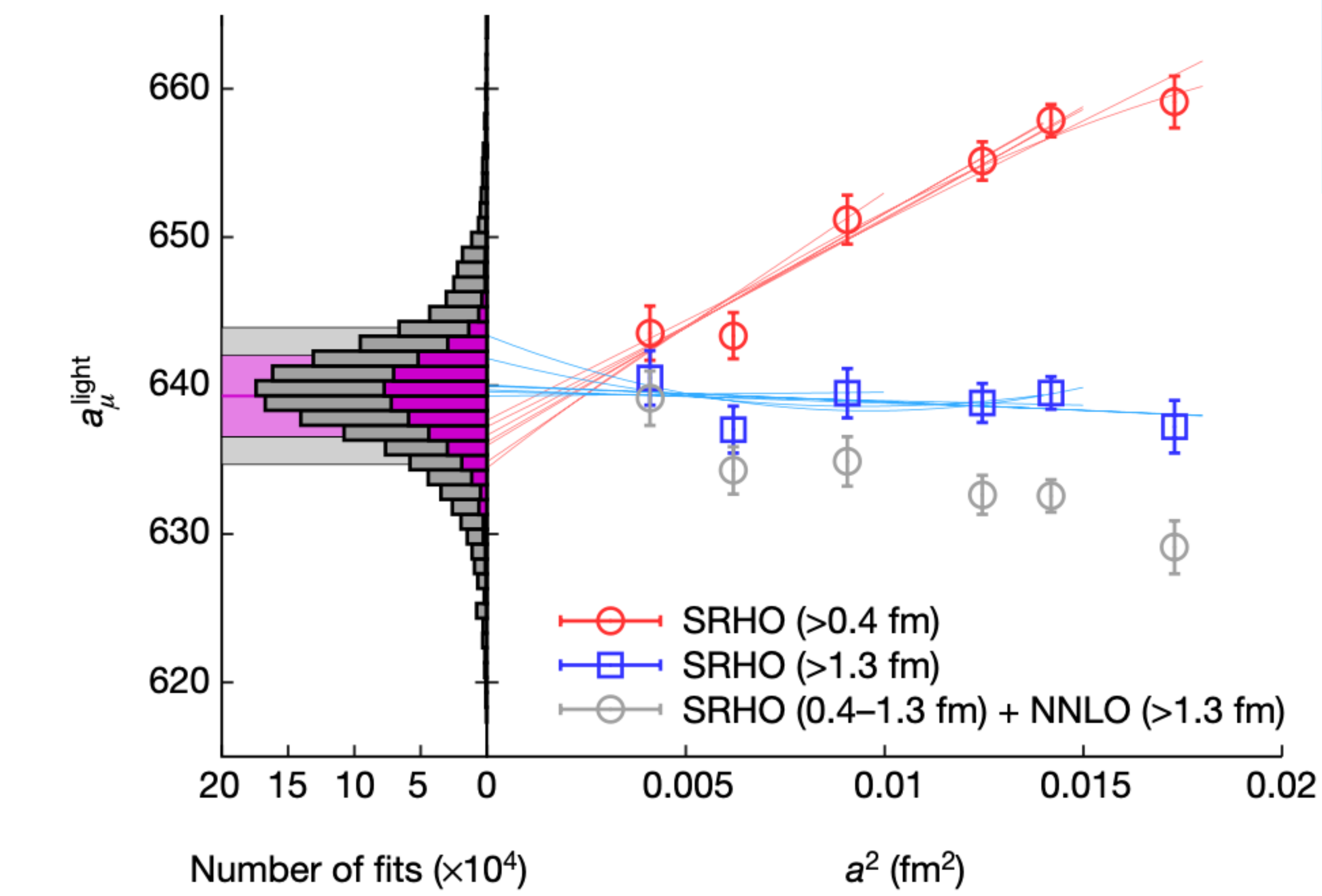
Extrapolate lattice spacing $a \to 0$

Article | Published: 07 April 2021

Leading hadronic contribution to the muon magnetic moment from lattice QCD

Sz. Borsanyi, Z. Fodor, J. N. Guenther, C. Hoelbling, S. D. Katz, L. Lellouch, T. Lippert, K. Miura, L. Parato, K. K. Szabo, F. Stokes, B. C. Toth, Cs. Torok & L. Varnhorst

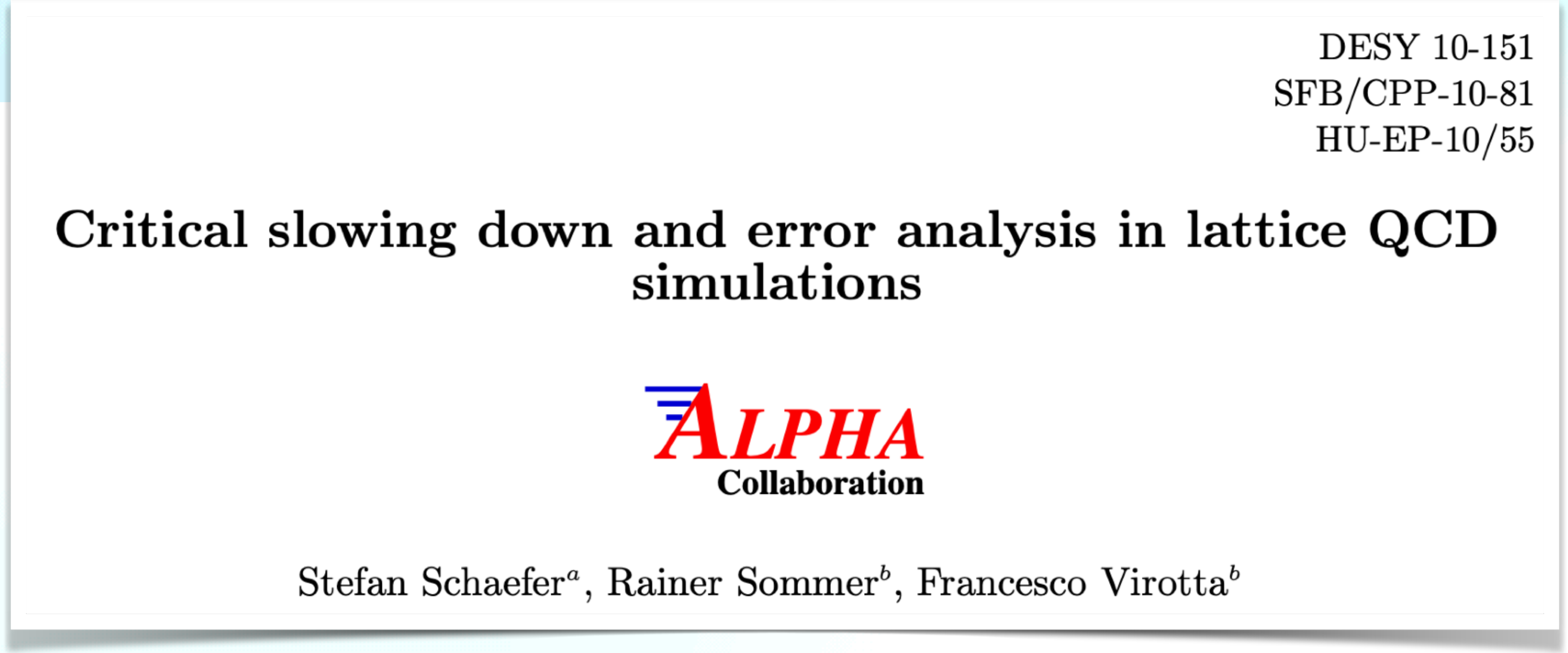
Nature **593**, 51–55 (2021) | Cite this article



The continuum limit

Critical slowing down

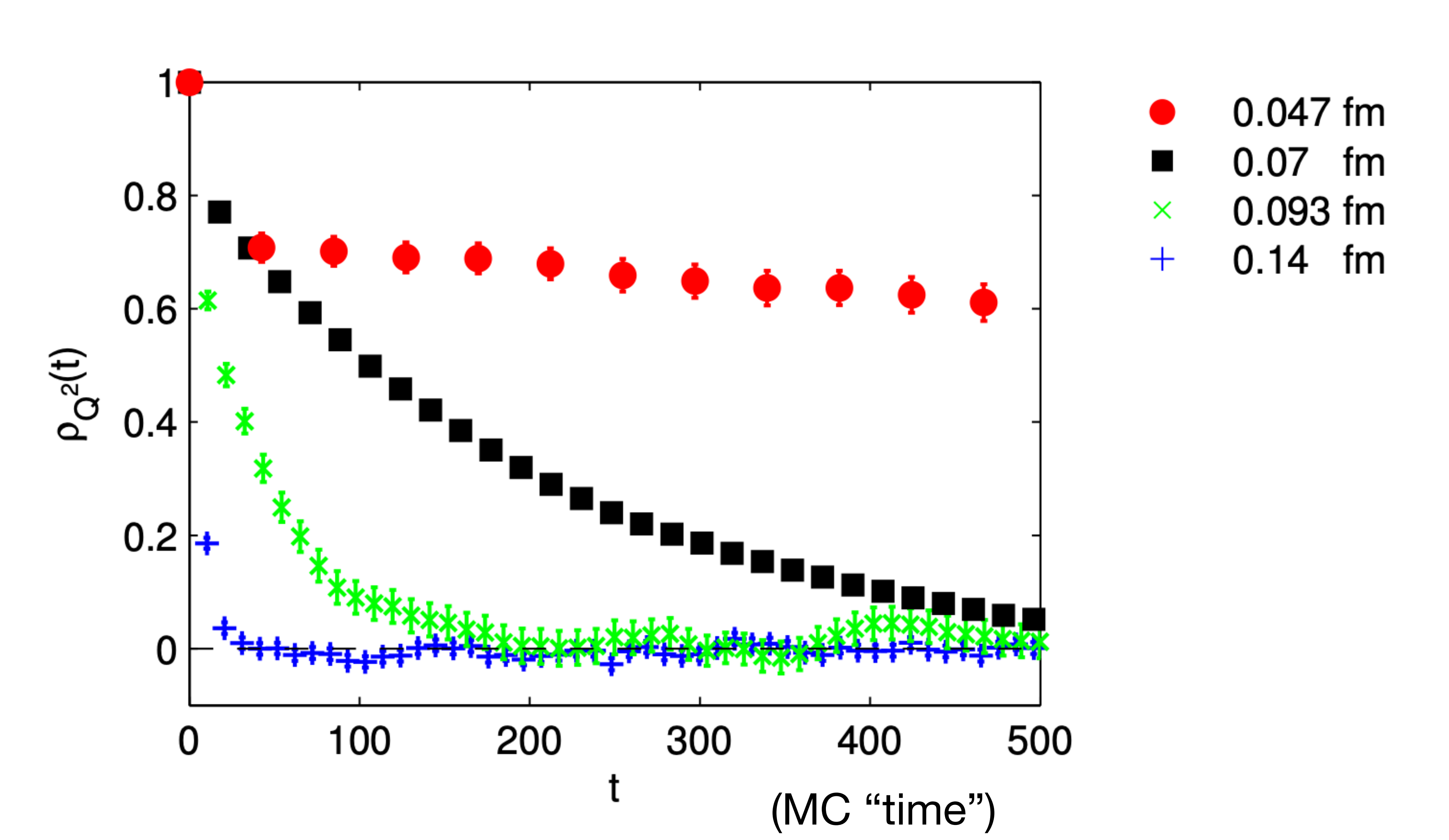
Autocorrelation time (ACT) grows with small lattice spacing a


DESY 10-151
SFB/CPP-10-81
HU-EP-10/55

Critical slowing down and error analysis in lattice QCD simulations

ALPHA Collaboration

Stefan Schaefer[a], Rainer Sommer[b], Francesco Virotta[b]



ACF for topological charge Q^2

Accurate simulations on coarse lattices?

Renormalization group transformation and fixed-point (FP) actions

Renormalization group transformation

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

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

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

Renormalization group transformation

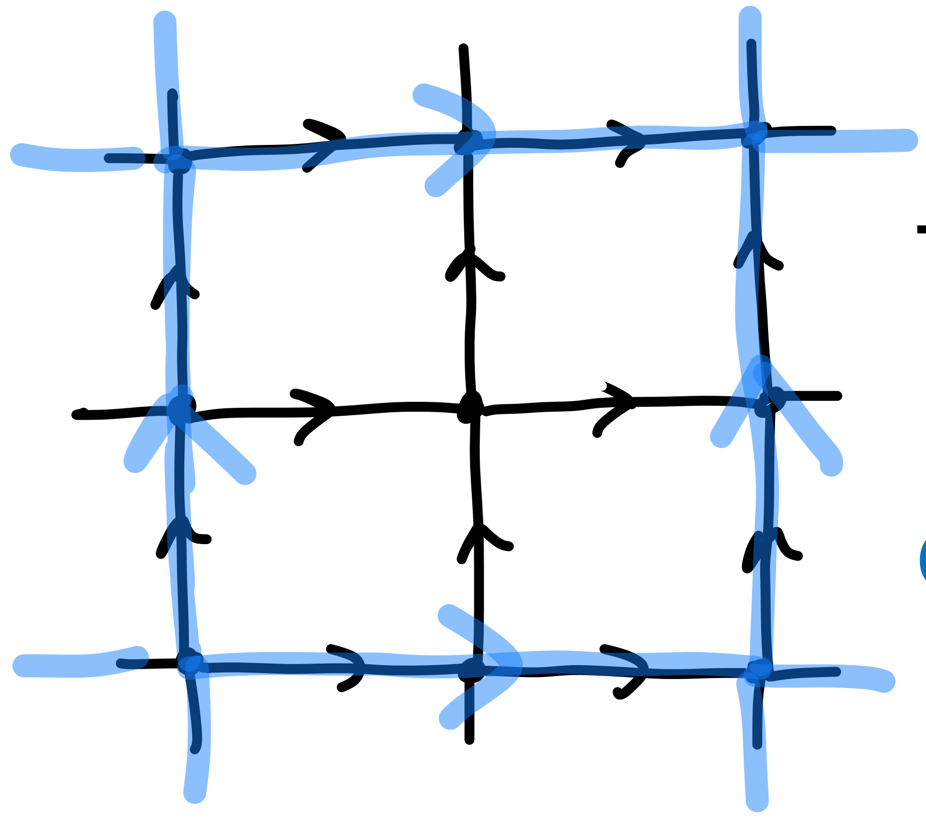
Introduce (coordinate space) renormalization group transformation (RGT):

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

where $\mathscr{T}[U,V]$ is a blocking kernel relating the fine links U to the coarse links V

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

blocked links $Q_{x_B,\mu}$



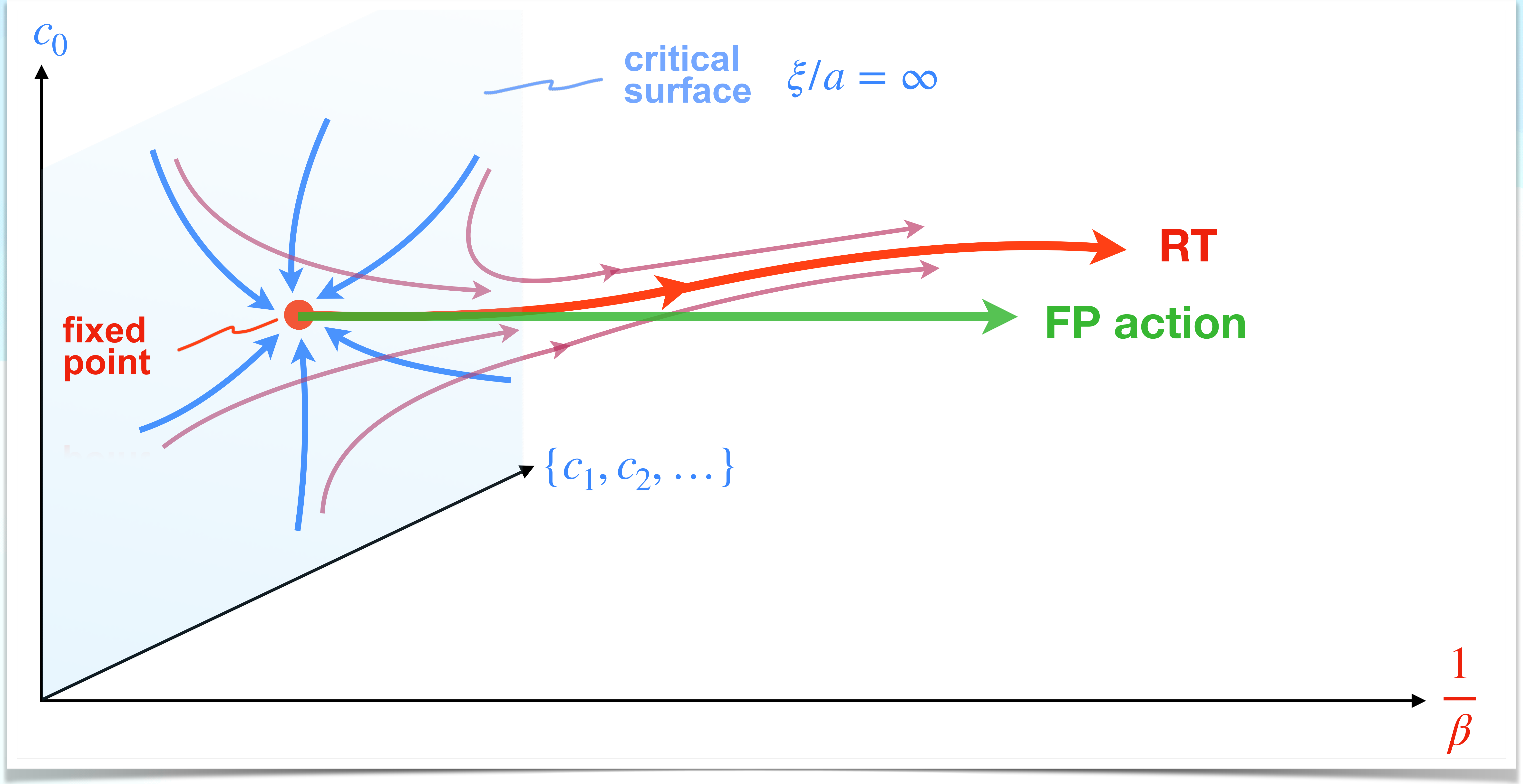
fine links $U_{x,\mu}$

coarse links $V_{x_B,\mu}$

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

Renormalization group transformation

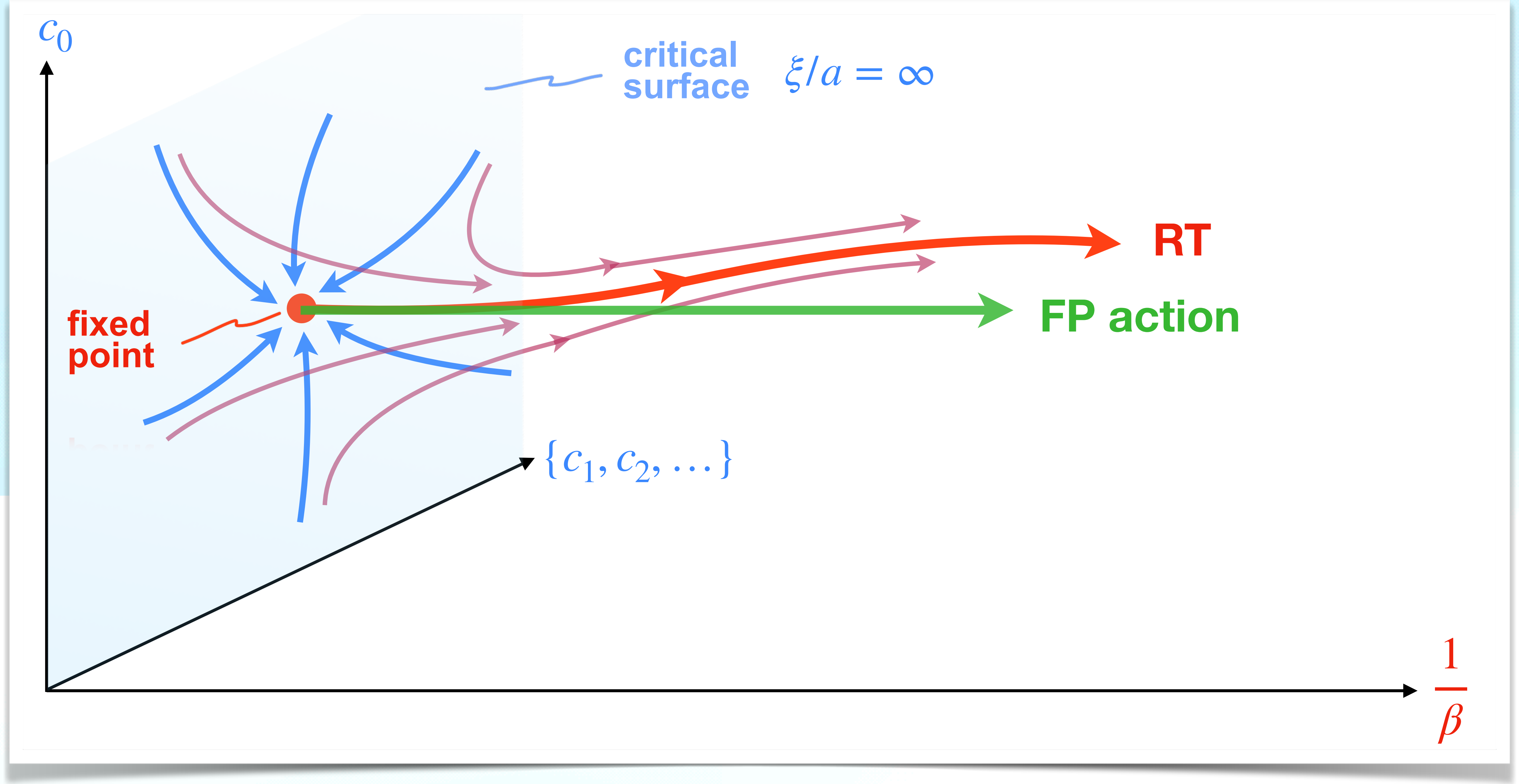
The effective action $\beta' \mathscr{A}'[V]$ is described by infinitely many couplings $\{c_\alpha\}$:



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

Renormalization group transformation

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



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

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:

$$\mathscr{A}^{\mathrm{FP}}[V] = \min_U \left(\mathscr{A}^{\mathrm{FP}}[U] + \mathscr{T}[U,V]\right)$$

The classically-perfect FP action

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

- There are no lattice artefacts on classical configurations (perfect action):

$$\frac{\delta \mathcal{A}^{\mathrm{FP}}[V]}{\delta V} = 0 \quad \Rightarrow \quad \left. \frac{\delta \mathcal{A}^{\mathrm{FP}}[U]}{\delta U} \right|_{U^\star} = 0$$

$$\mathcal{A}^{\mathrm{FP}}[V] = \mathcal{A}^{\mathrm{FP}}[U^\star]$$

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

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

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

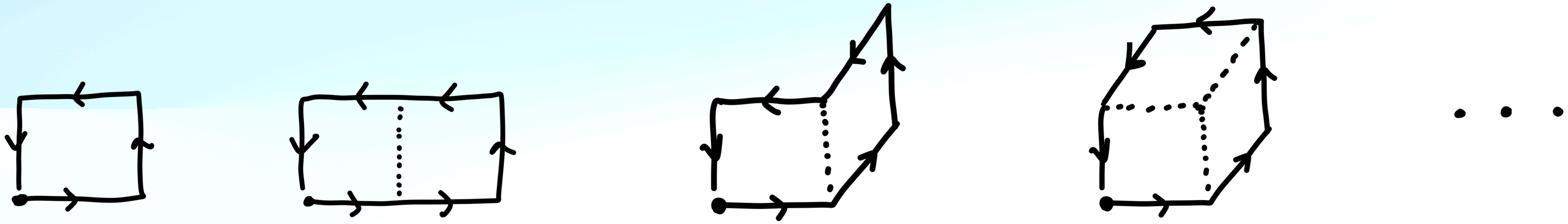
$$~~\mathcal{O}(a^{2n})~~, \mathcal{O}(g^2 a^{2n}) \quad n = 1,2,\ldots$$

For practical MC simulations: find approximations to $\mathcal{A}^{\mathrm{FP}}[V]$

The classically-perfect FP action

Ansatz: sum of powers of loops with free parameters $p_C^{(m)}$

$$\mathscr{A}^{\mathrm{FP}}[U] = \frac{1}{N_c} \sum_x \sum_C \sum_{m=1}^{M} p_C^{(m)} \left[\mathrm{ReTr}(1 - U_{x,C})\right]^m$$



Solve FP equation numerically:

$$\mathscr{A}^{\mathrm{FP}}[V] \approx \min_U \left(\mathscr{A}^{\mathrm{FP}}_{\mathrm{quad}}[U] + \mathscr{T}[U,V] \right)$$

⇒ Find set of parameters to approximate numerical FP data
(+ normalization and Symanzik conditions)

Improve approximations: iterate FP equation

$$\mathscr{A}^{\mathrm{FP}}_{(n+1)}[V] = \min_U \left(\mathscr{A}^{\mathrm{FP}}_{(n)}[U] + \mathscr{T}[U,V] \right)$$

The classically-perfect FP action

Selection of previous parametrizations for SU(3) gauge theory

- Quadratic, Type I, Type II [DeGrand, Hasenfratz, Hasenfratz, Niedermayer (1995)]
- Type IIIa, IIIb, IIIc [Blatter, Niedermayer (1996)]
- FP with APE-smearing [Niedermayer, Rüfenacht, Wenger (2000)]

  APE 444 (for smooth fields)

  APE 431 (for coarse fields)

  number of parameters $< \mathcal{O}(100)$

- Anisotropic FP [Rüfenacht, Wenger (2001)]
- **L-CNN FP** [Ipp, Holland, Müller, Wenger (2023)]

Learning a new FP action

Generating FP training data

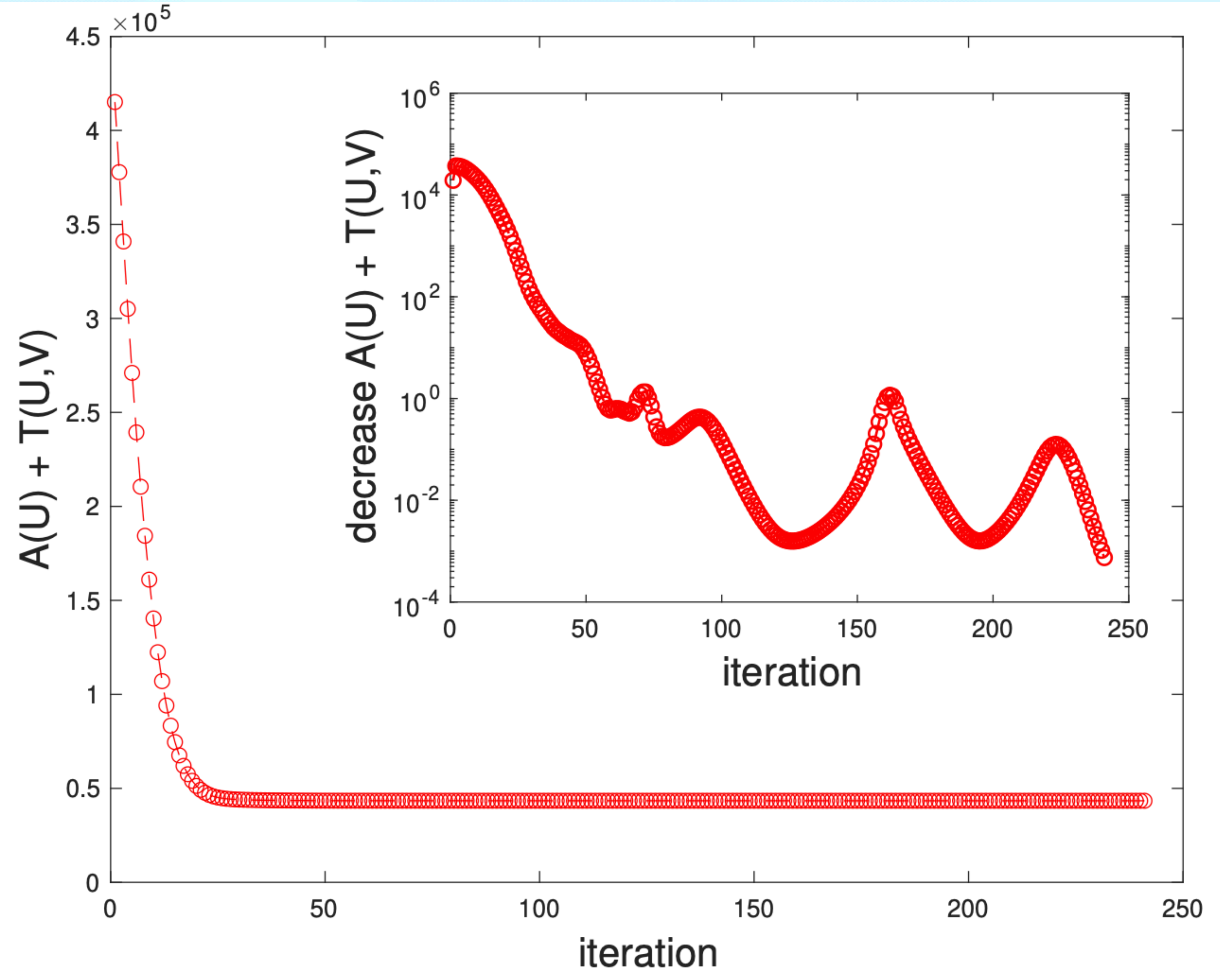
Selection of coarse configurations

200 on 4^4, 6^4, 8^4 with $\beta_w \in [5,100]$ (Wilson HMC)

For each coarse $V_{x,\mu}$ configuration

numerically minimize w.r.t. $U_{x,\mu}$

$$\mathscr{A}^{\mathrm{FP}}_{\mathrm{data}}[V] = \min_{U} \left(\mathscr{A}^{\mathrm{FP}}_{444}[U] + \mathscr{T}[U,V] \right)$$
$$= \mathscr{A}^{\mathrm{FP}}_{444}[U^\star] + \mathscr{T}[U^\star, V]$$



Local information: derivatives w.r.t. links

$$D^a_{x,\mu}\mathscr{A}^{\mathrm{FP}}_{\mathrm{data}}[V] = D^a_{x,\mu}\mathscr{T}[U^\star, V]$$

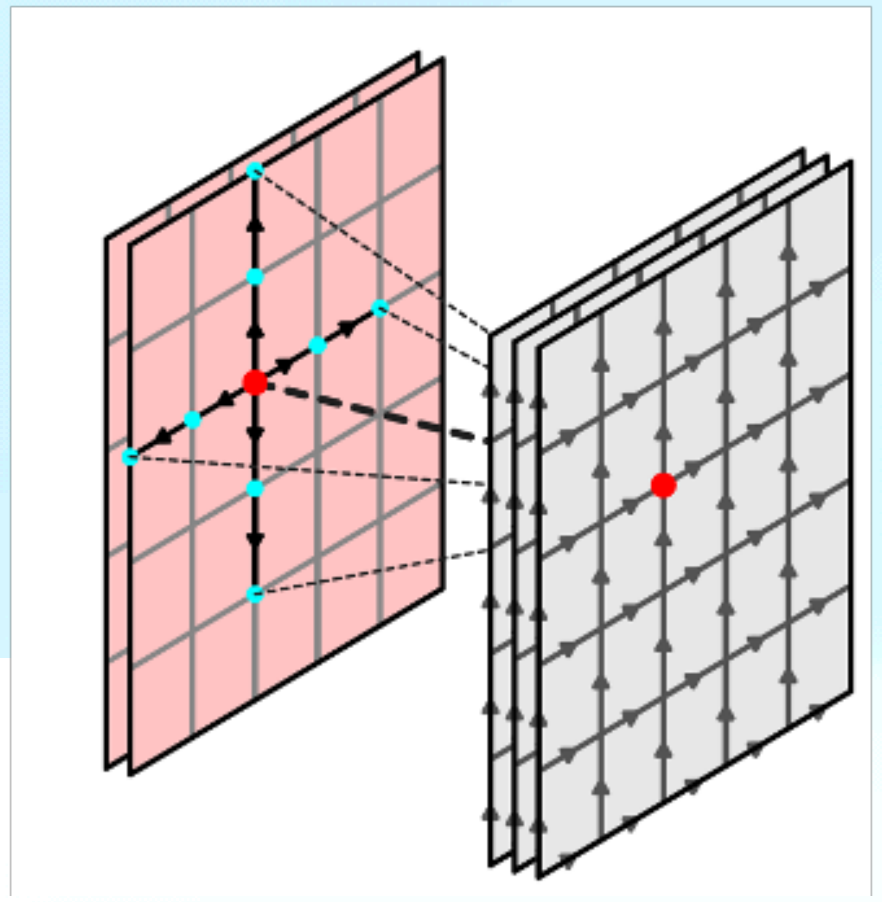
Note: $D^a_{x,\mu}\mathscr{A}[U] := \frac{d}{d\varepsilon}\mathscr{A}[e^{i\varepsilon T^a}U_{x,\mu}]|_{\varepsilon=0}$

Machine-learned FP action

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

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

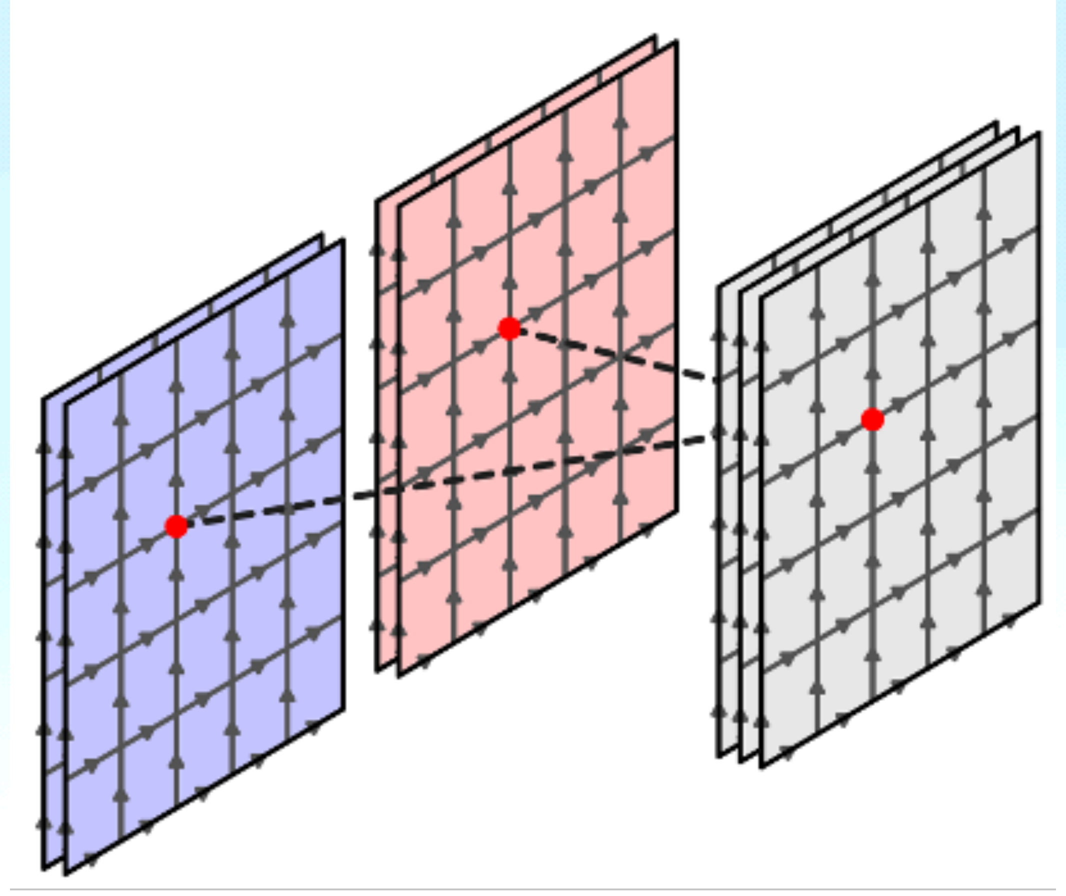
L-Conv:



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

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

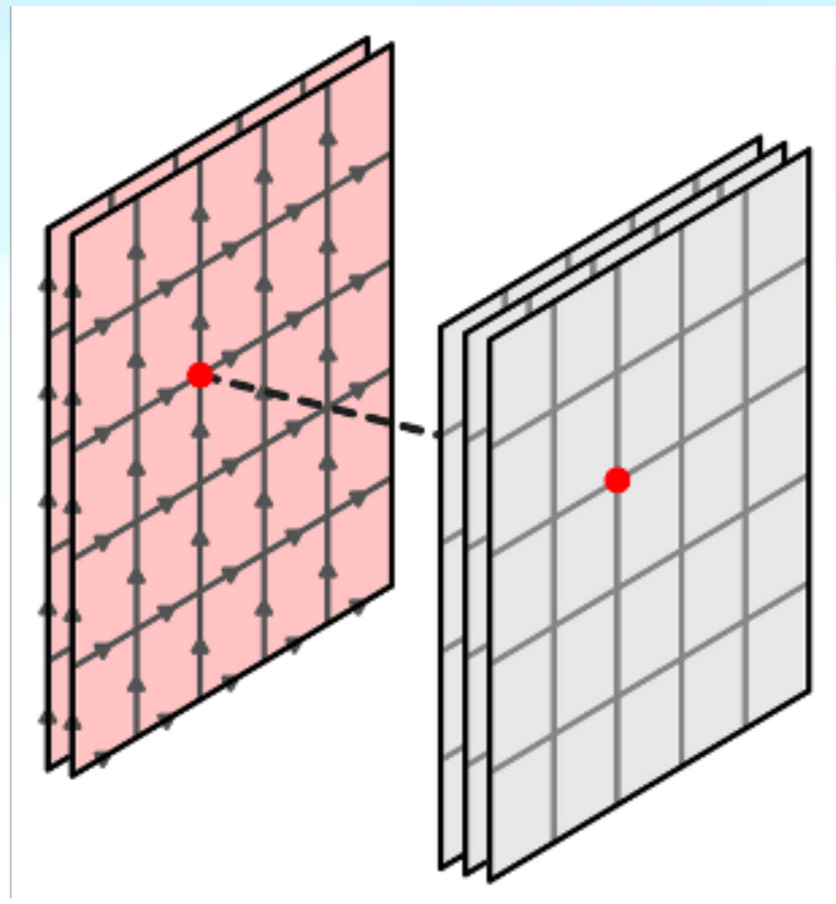
L-Bilin:



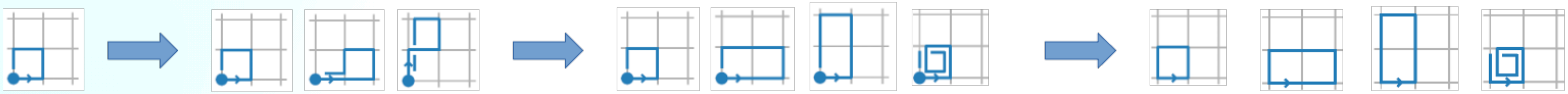
$$(\mathcal{U},\mathcal{W}) \times (\mathcal{U},\mathcal{W}') \to (\mathcal{U},\mathcal{W}'')$$

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

Trace:



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

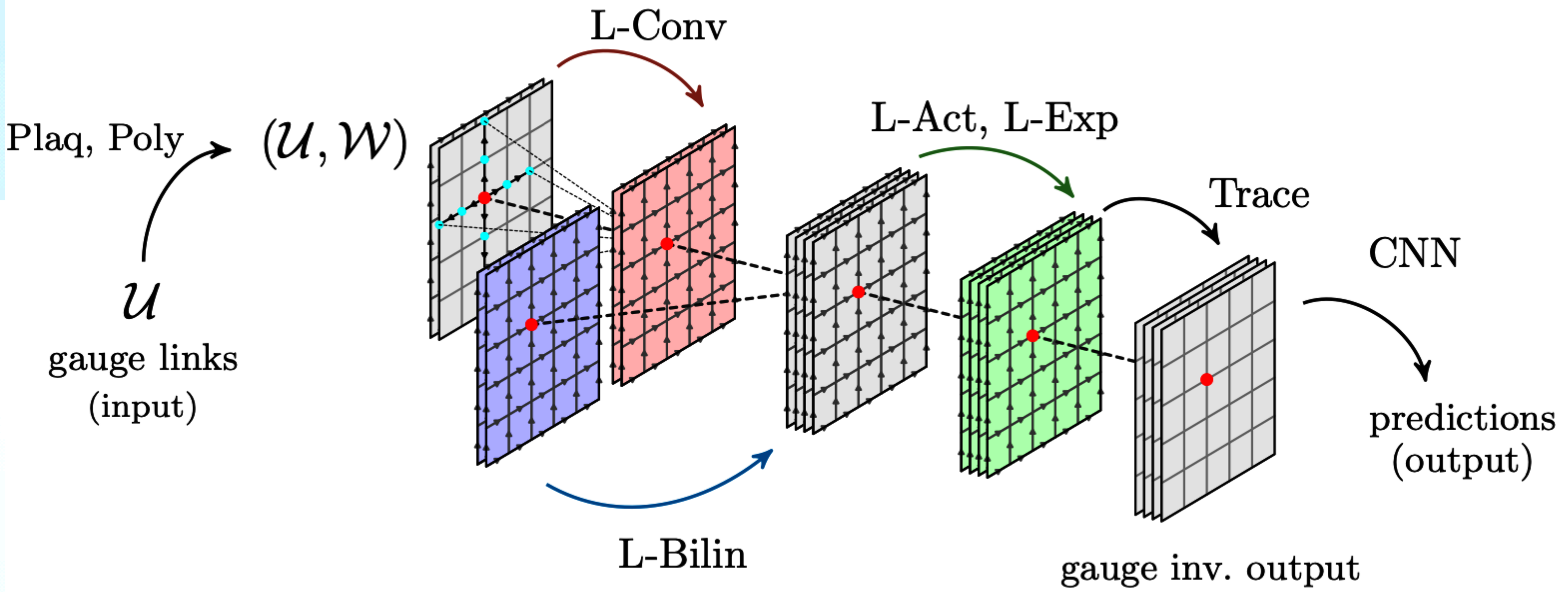


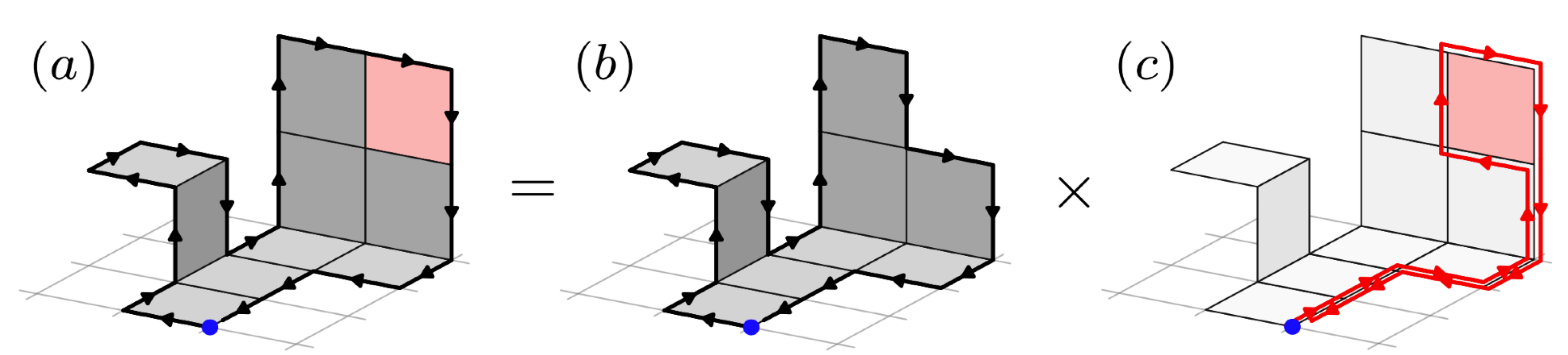
Machine-learned FP action

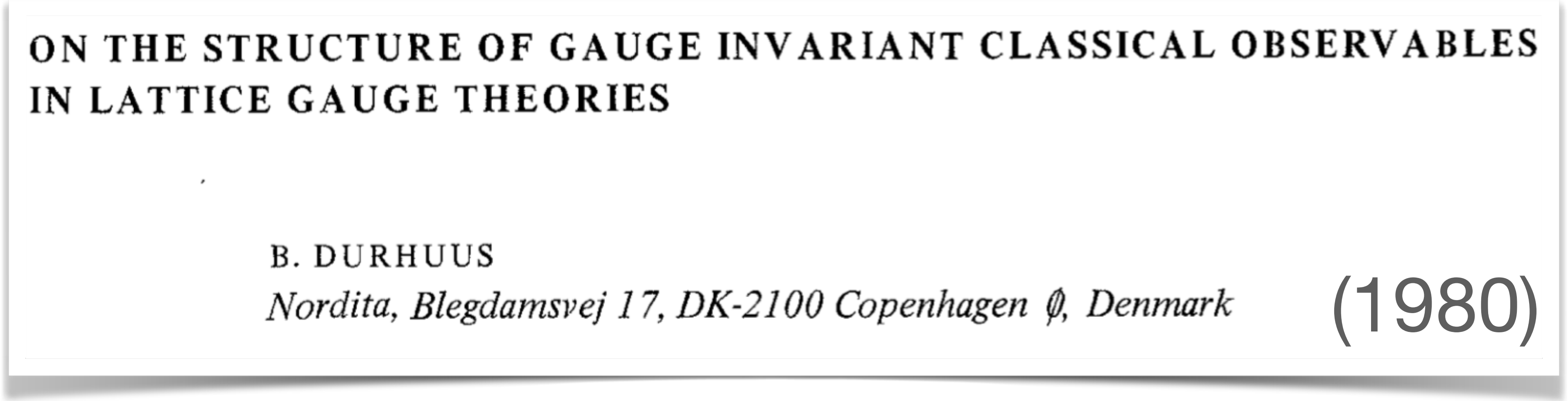
Architecture: Lattice gauge equivariant Convolutional Neural Network (L-CNN)

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

"Universal approximator" for gauge invariant functions on the lattice



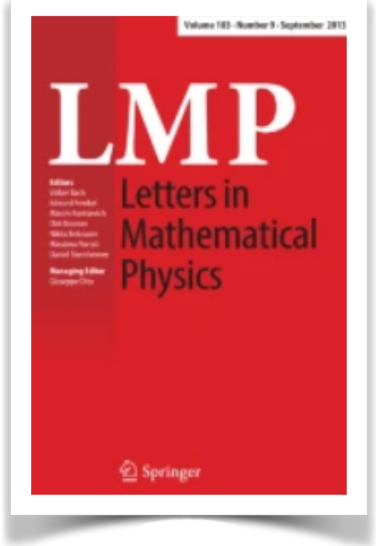



ON THE STRUCTURE OF GAUGE INVARIANT CLASSICAL OBSERVABLES IN LATTICE GAUGE THEORIES

B. DURHUUS

Nordita, Blegdamsvej 17, DK-2100 Copenhagen Ø, Denmark

(1980)



Machine-learned FP action

FP action parameterised by

$$\mathscr{A}^{\mathrm{L-CNN}}[U] = \sum_x \mathscr{A}_x^{\mathrm{pre}}[U] \sum_{n=0}^{\infty} b^{(n)} \left(N_x[U] - N_x[1]\right)^n$$

local output of L-CNN

with “prefactor” action

$$\mathscr{A}_x^{\mathrm{pre}}[U] = \frac{1}{N_c} \sum_C \sum_{m=1}^{M} p_C^{(m)} [\mathrm{ReTr}(1 - U_{x,C})]^m$$

e.g. Wilson, Symanzik,
a sum of more general loops

guaranteed continuum limit

In practice we use:

$$\mathscr{A}^{\mathrm{L-CNN}}[U] = \sum_x \mathscr{A}_x^{\mathrm{pre}}[U] \exp\left(N_x[U] - N_x[1]\right)$$

Machine-learned FP action

Loss function from two weighted contributions:

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

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

$$L = w_1 L_1 + w_2 L_2$$

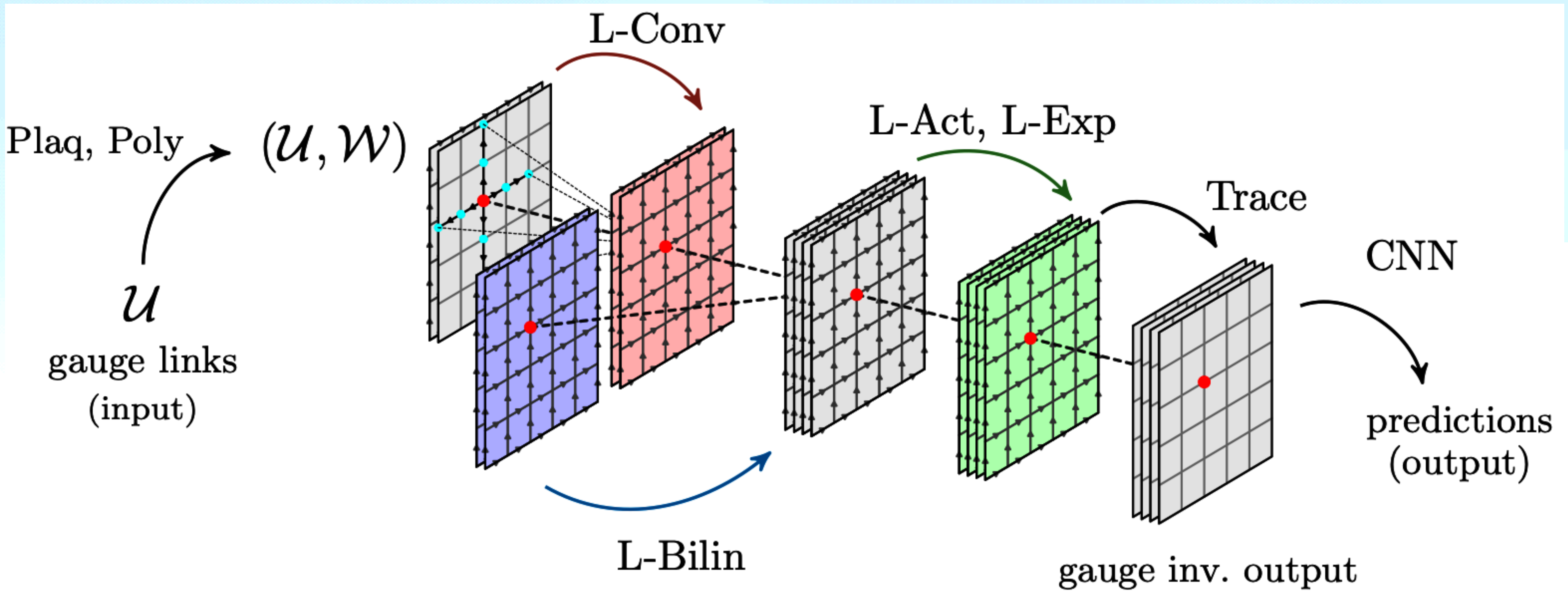
$$D_{x,\mu}[U] := T^a \frac{d}{d\varepsilon} \mathscr{A}[e^{i\varepsilon T^a} U_{x,\mu}]|_{\varepsilon=0}$$

Technical point: derivatives in L-CNN are given through backpropagation

Machine-learned FP action

Architecture search

layers	kernel sizes	channels	parameters
1	1	4	9.61K
	2	8	170K
	2	16	340K
2	1, 1	4, 8	10.3K
	2, 1	8, 16	174K
	2, 2	16, 12	454K
3	2, 1, 1	4, 4, 8	85.8K
	2, 2, 1	8, 8, 16	194K
	2, 2, 1	12, 24, 24	443K
	2, 2, 1	16, 16, 32	527K
4	2, 1, 1, 1	8, 8, 16, 32	212K
	2, 2, 1, 1	16, 16, 16, 32	544K
	2, 2, 2, 1	16, 24, 24, 32	1.15M

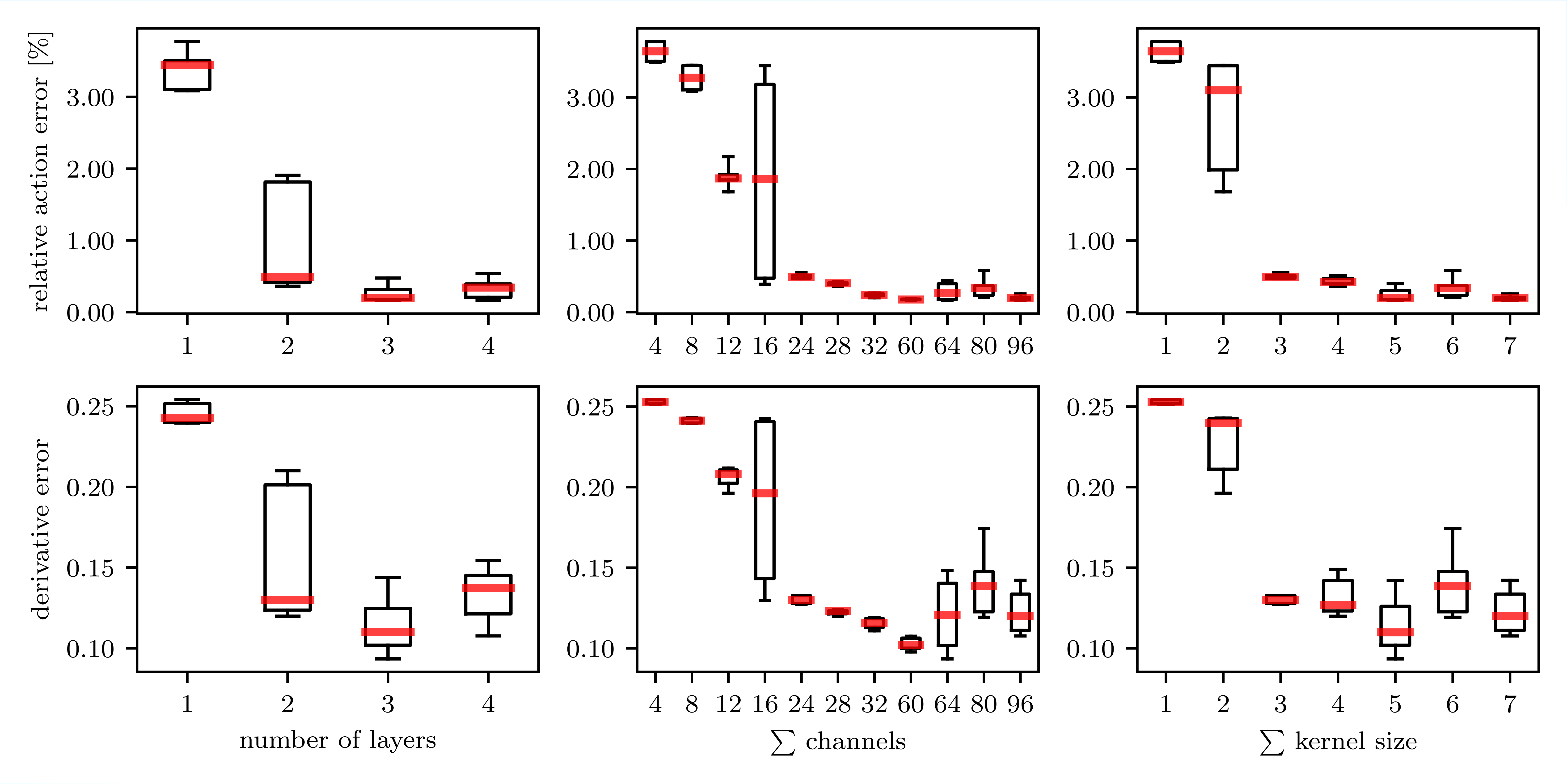


400 - 1000 training epochs on 4^4

five random initializations

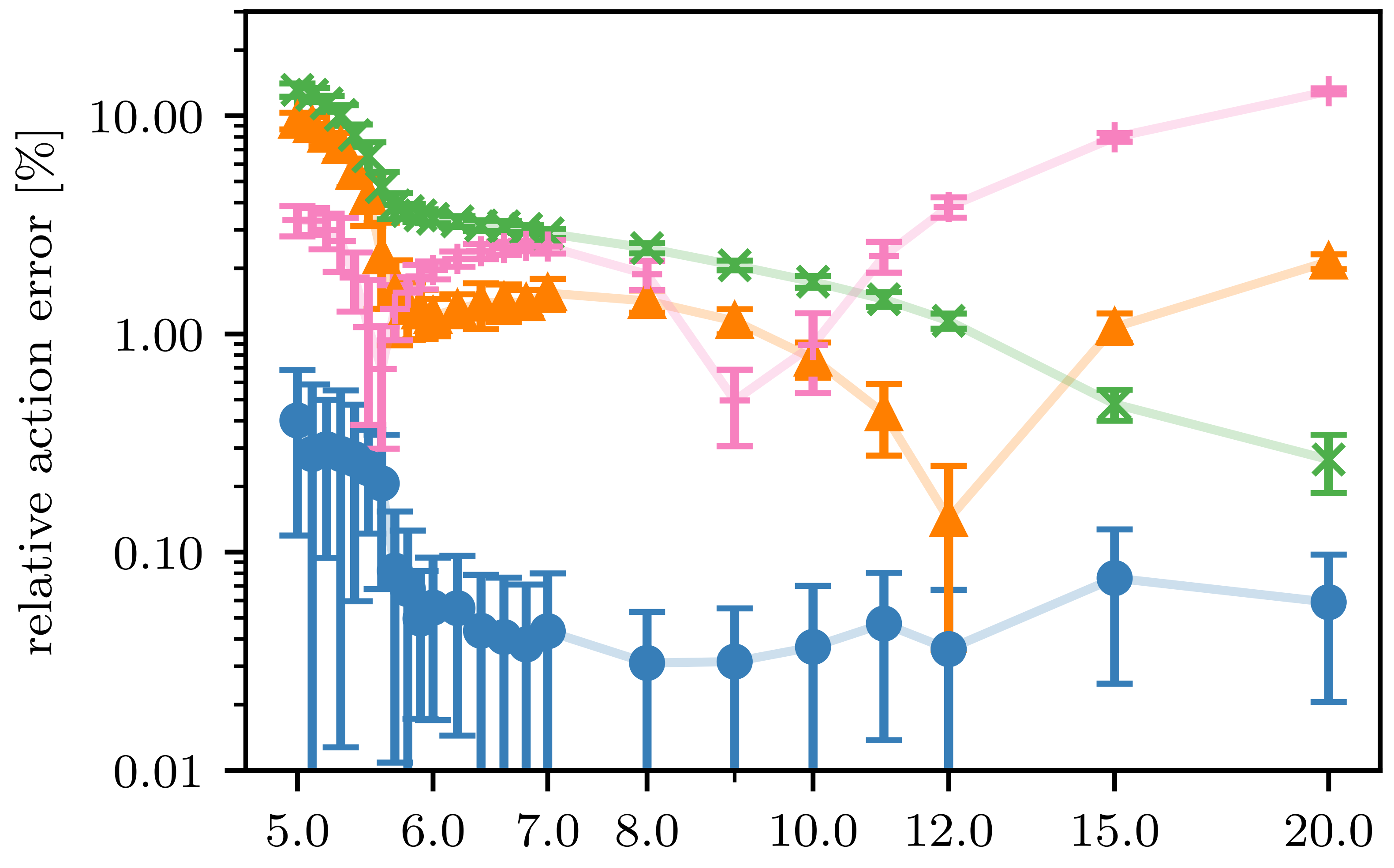
Machine-learned FP action

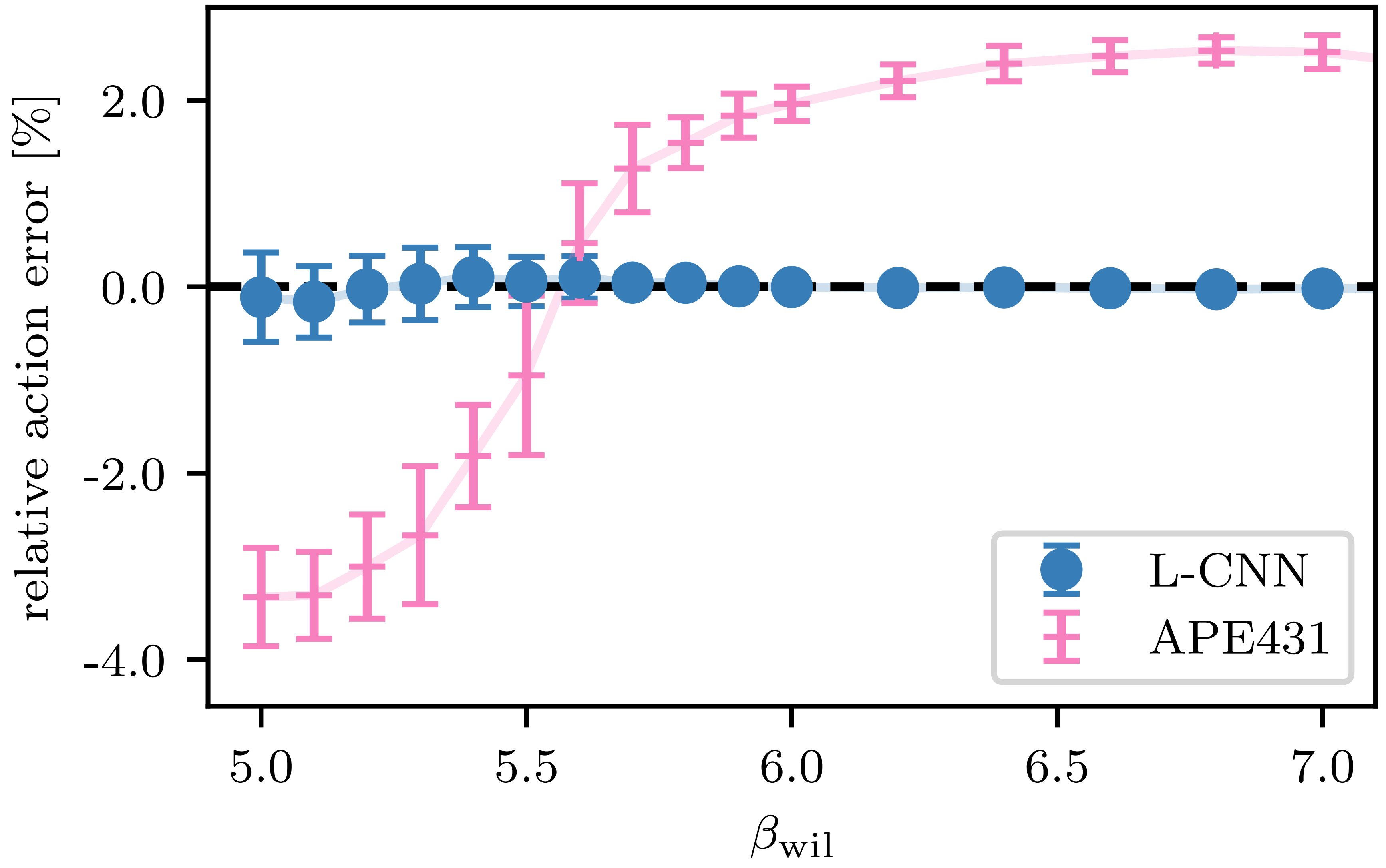
Architecture search   3 layers, kernel sizes (2,2,1), ~60 channels, $\mathcal{O}(10^5)$ parameters

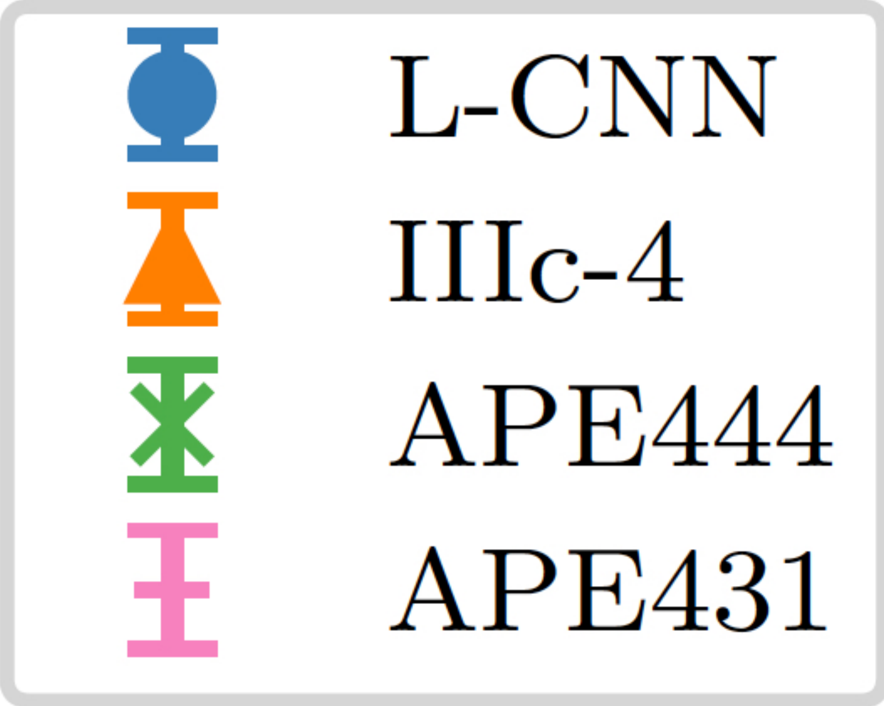


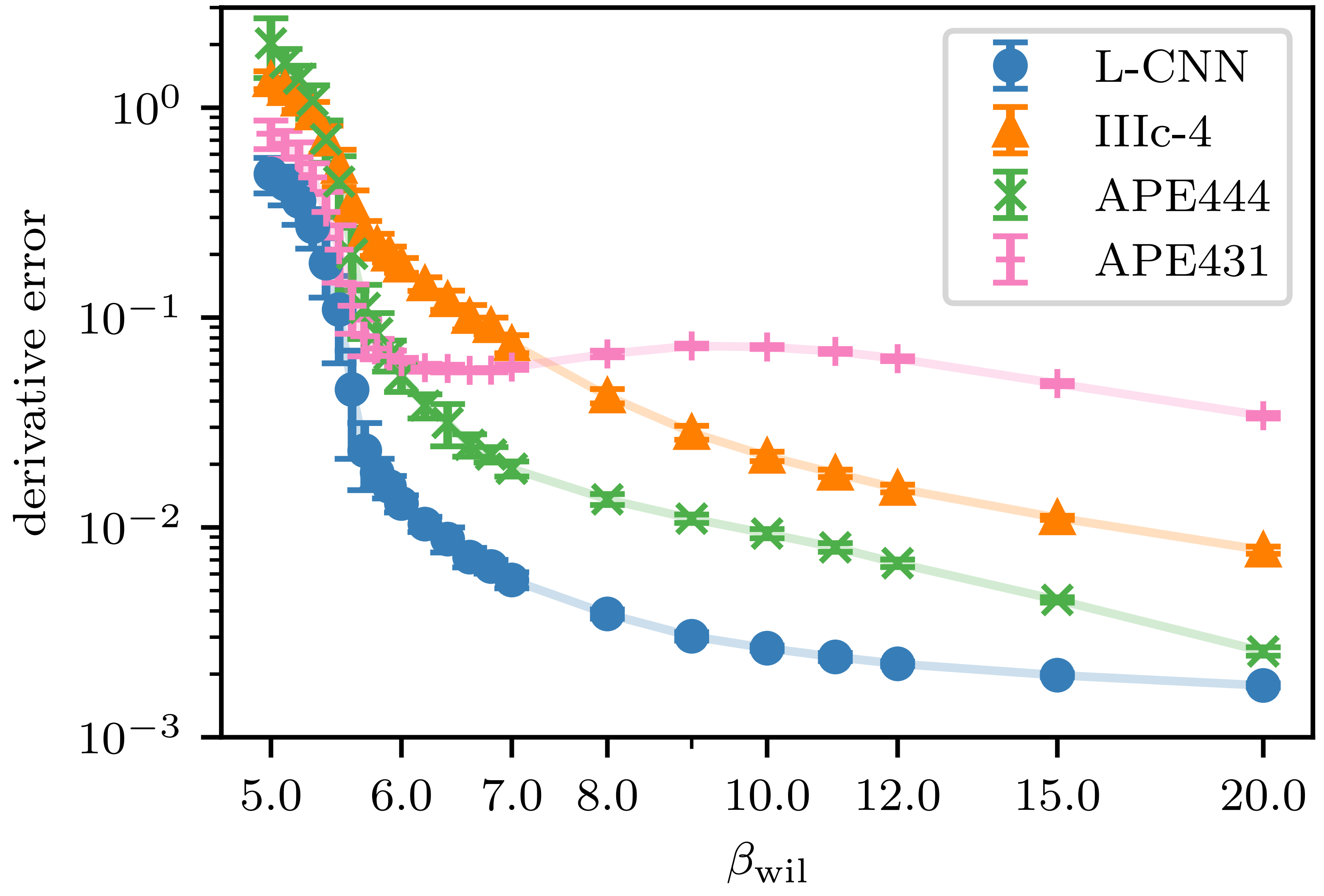
+ further training with instantons, different lattice sizes, …

Machine-learned FP action









⇒ beats all previous parametrizations

⇒ works on fine and coarse configurations

HMC and FP gradient flow

Hybrid Monte Carlo (HMC)

Euclidean lattice path integral for observables $\mathcal{O}$

$$Z(\beta) = \int [\prod_{x,\mu} \mathcal{D}U_{x,\mu}] \exp\left[-\beta \mathcal{A}[U]\right] \qquad \langle \mathcal{O} \rangle_\beta = \frac{1}{Z(\beta)} \int \mathcal{D}U\, \mathcal{O}[U] \exp\left[-\beta \mathcal{A}[U]\right]$$

$$\int \mathcal{D}U e^{-S[U]} \propto \int \mathcal{D}U \int \mathcal{D}P e^{-\frac{1}{2}P^2 - S[U]} = \int \mathcal{D}U \int \mathcal{D}P e^{-H[P,U]}$$

Hamiltonian

$$H[P,U] = \sum_{x,\mu,a} \mathrm{Tr}[P_{x,\mu} P_{x,\mu}] + \beta \mathcal{A}[U]$$

Hamiltonian EOM

$$\dot{P}_{x,\mu}(t) = \beta\, D_{x,\mu} \mathcal{A}[U(t)]$$

$$\dot{U}_{x,\mu}(t) = -\, i P_{x,\mu}(t)\, U_{x,\mu}(t)$$

(Again: derivatives via backpropagation)

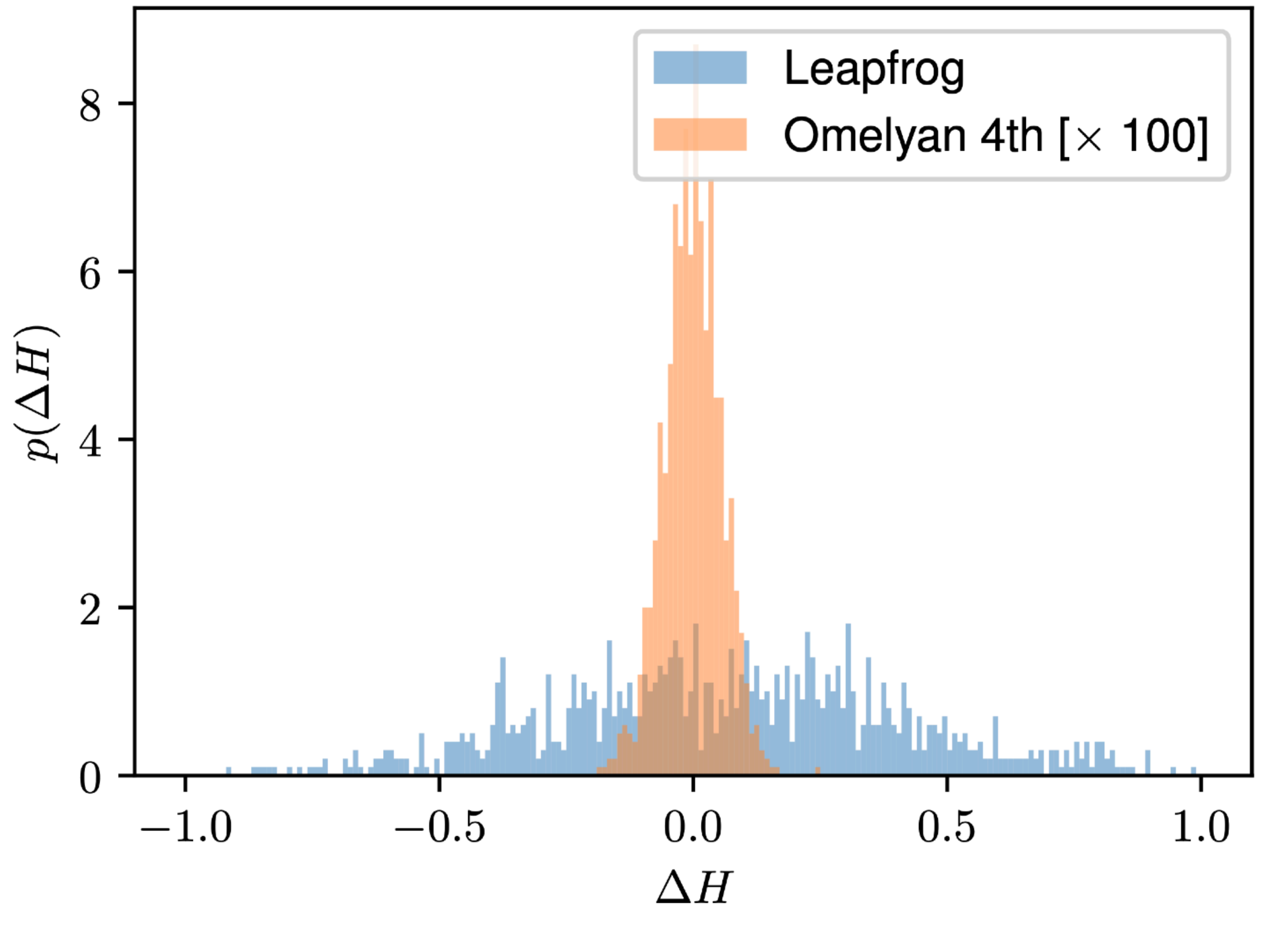
Hybrid Monte Carlo (HMC)

Euclidean lattice path integral for observables $\mathcal{O}$

$$Z(\beta) = \int [\prod_{x,\mu} \mathcal{D}U_{x,\mu}] \exp\left[-\beta \mathcal{A}[U]\right] \qquad \langle \mathcal{O} \rangle_\beta = \frac{1}{Z(\beta)} \int \mathcal{D}U \, \mathcal{O}[U] \, \exp\left[-\beta \mathcal{A}[U]\right]$$

HMC algorithm

1. Pick random momenta $P_{x,\mu} \sim \exp(-\mathrm{Tr} P_{x,\mu}^2)$
2. Solve Hamiltonian EOM from $t = 0$ to τ (LF/Omelyan)
3. Metropolis accept-reject $\Delta H = H(\tau) - H(0) \approx 0$
4. Repeat



Classically-perfect FP gradient flow

Ensemble generation

$$U_{x,\mu} \sim \exp(-\beta \mathscr{A}_g[U])$$

Gradient flow

$$\dot{U}_{x,\mu}(t_f) = -\, iD_{x,\mu}\mathscr{A}_f[U]\, U_{x,\mu}(t_f)$$

(adaptive RK23 solver)

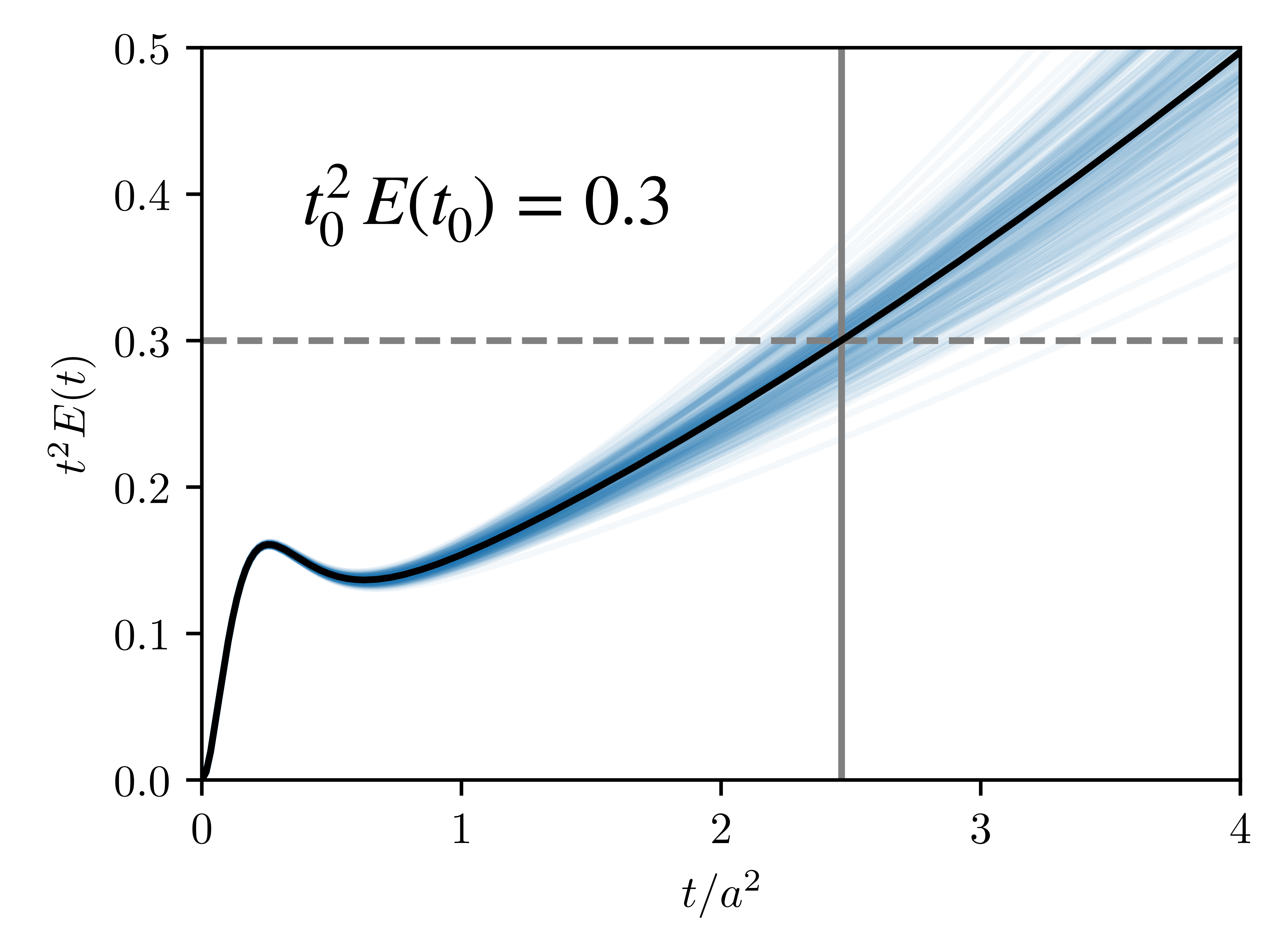
Energy density

$$E(t_f) = \frac{1}{L^4}\langle \mathscr{A}_e[U(t_f)] \rangle$$

FP flow $\mathscr{A}_g = \mathscr{A}_f = \mathscr{A}_e = \mathscr{A}^{\mathrm{FP}}$

GF with FP action is classically-perfect!

Lüschers flow scale $\sqrt{t_0} \approx 0.167\,\mathrm{fm}$



High-precision scale setting

Classically-perfect FP gradient flow

Ensemble generation

$$U_{x,\mu} \sim \exp(-\beta \mathscr{A}_g[U])$$

Gradient flow

$$\dot{U}_{x,\mu}(t_f) = -\, iD_{x,\mu}\mathscr{A}_f[U]\, U_{x,\mu}(t_f)$$

(adaptive RK23 solver)

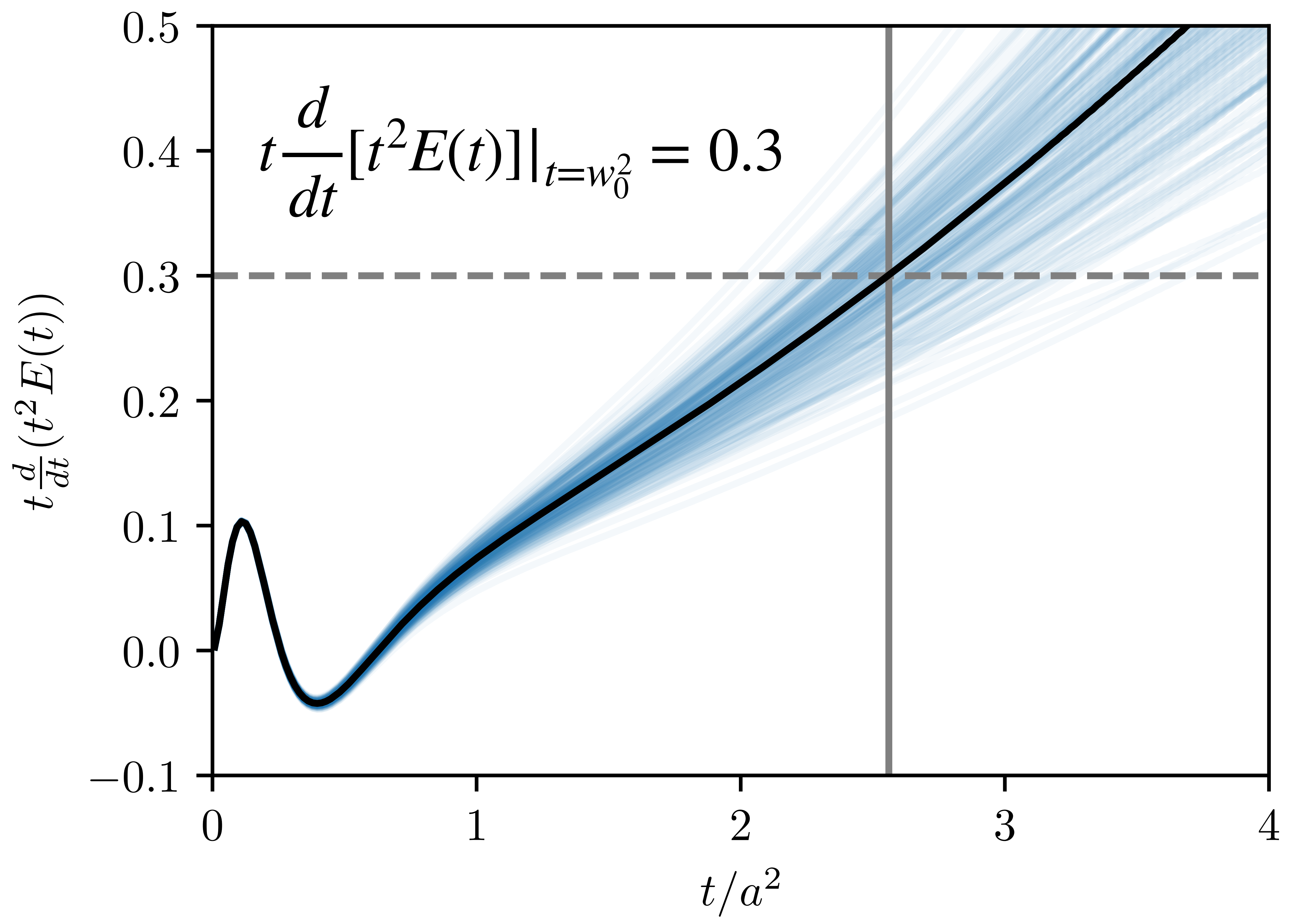
Energy density

$$E(t_f) = \frac{1}{L^4}\langle \mathscr{A}_e[U(t_f)]\rangle$$

FP flow $\mathscr{A}_g = \mathscr{A}_f = \mathscr{A}_e = \mathscr{A}^{\mathrm{FP}}$

GF with FP action is classically-perfect!

Wuppertal scale $w_0 \approx 0.176\,\mathrm{fm}$



High-precision scale setting

Preliminary results

Ratio of gradient flow scales

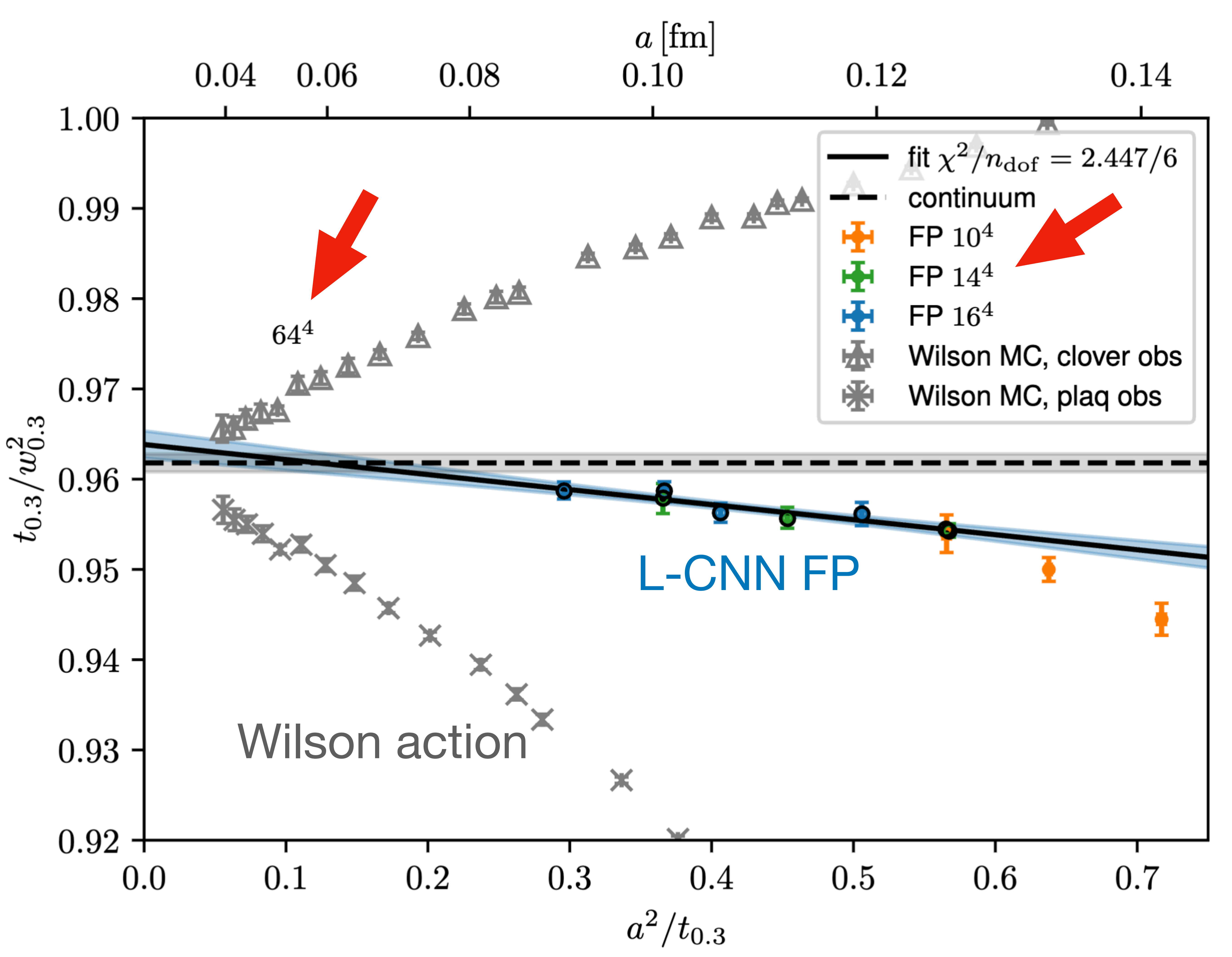
Lüschers flow time t_c

$$t^2 E(t)\Big|_{t_c} = c$$

Wuppertal scale w_c

$$t\frac{d}{dt}[t^2 E(t)]\Big|_{t=w_c^2} = c$$

L-CNN FP shows scaling on very small (coarse) lattices!



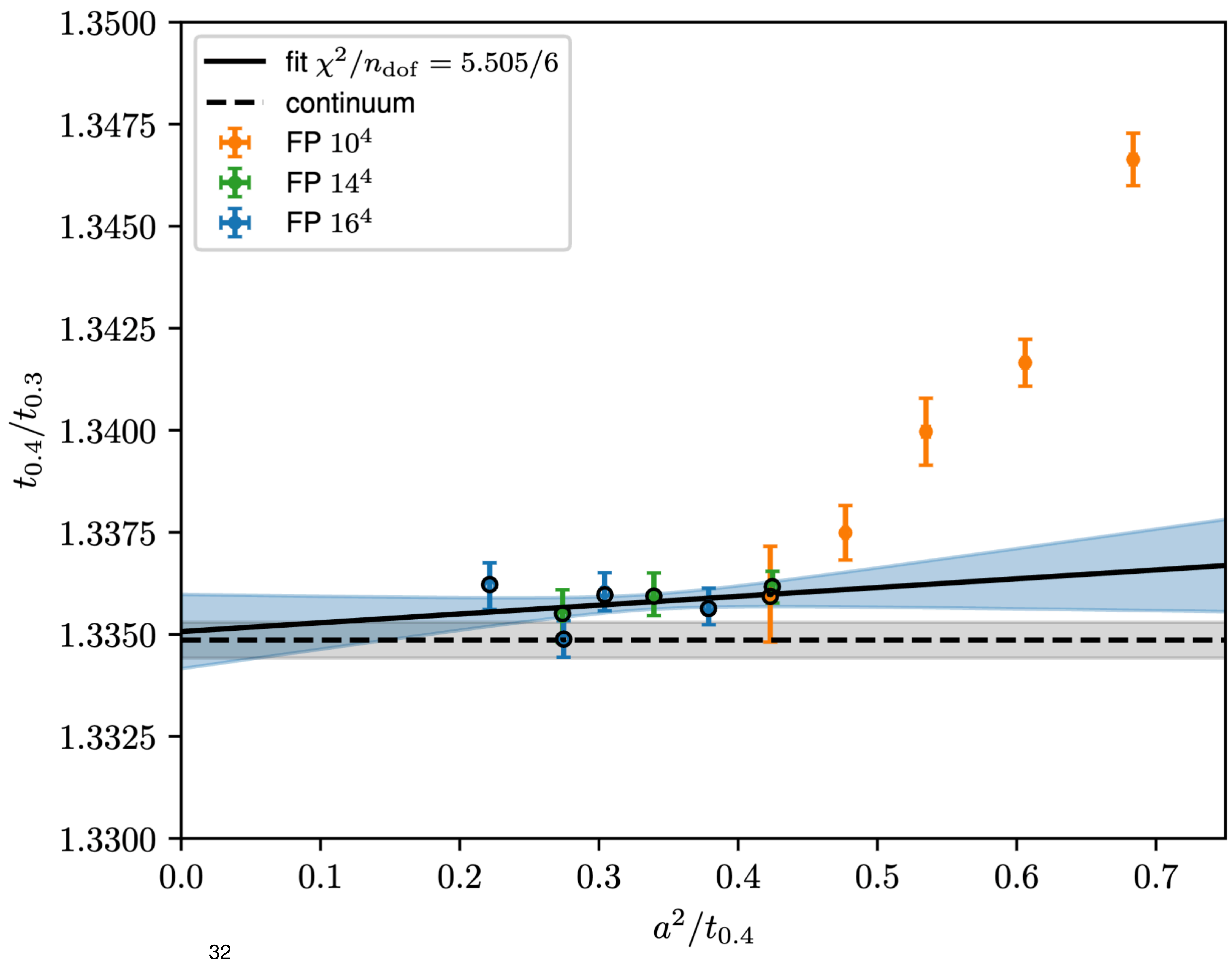
More flow time ratios

Lüschers flow time t_c

$$t^2 E(t)\Big|_{t_c} = c$$

Dimensionless ratios $t_{c'}/t_c$

Note: $t_{0.4}$ also called t_2



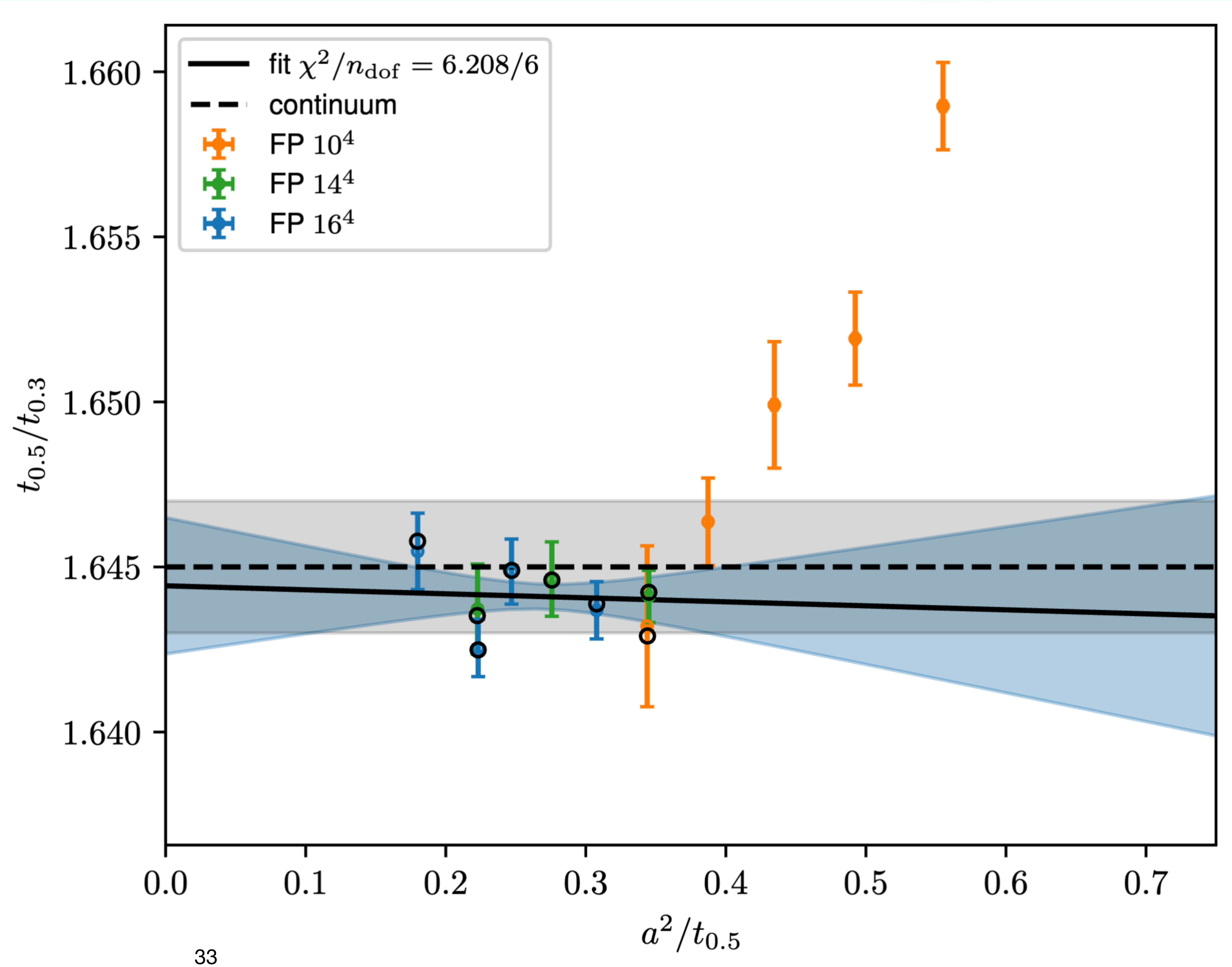
More flow time ratios

Lüschers flow time t_c

$$t^2 E(t)\Big|_{t_c} = c$$

Dimensionless ratios $t_{c'}/t_c$

Note: $t_{0.5}$ also called t_1



β-function

Renormalized coupling through gradient flow

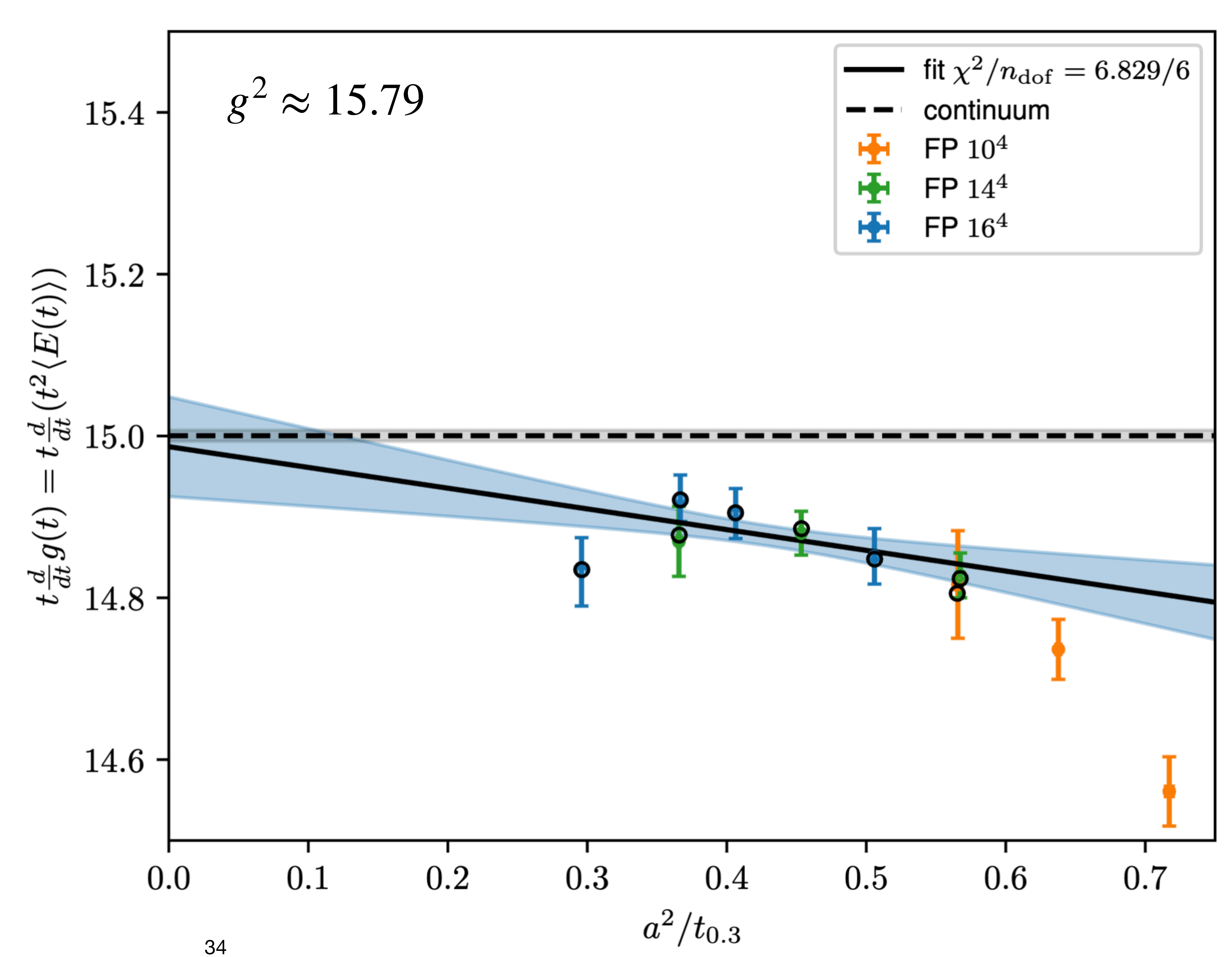
$$t^2 E(t) = \frac{3}{16\pi^2} g^2(t)$$

Beta function

$$\beta(g^2(t)) = t\frac{d}{dt} g^2(t)$$

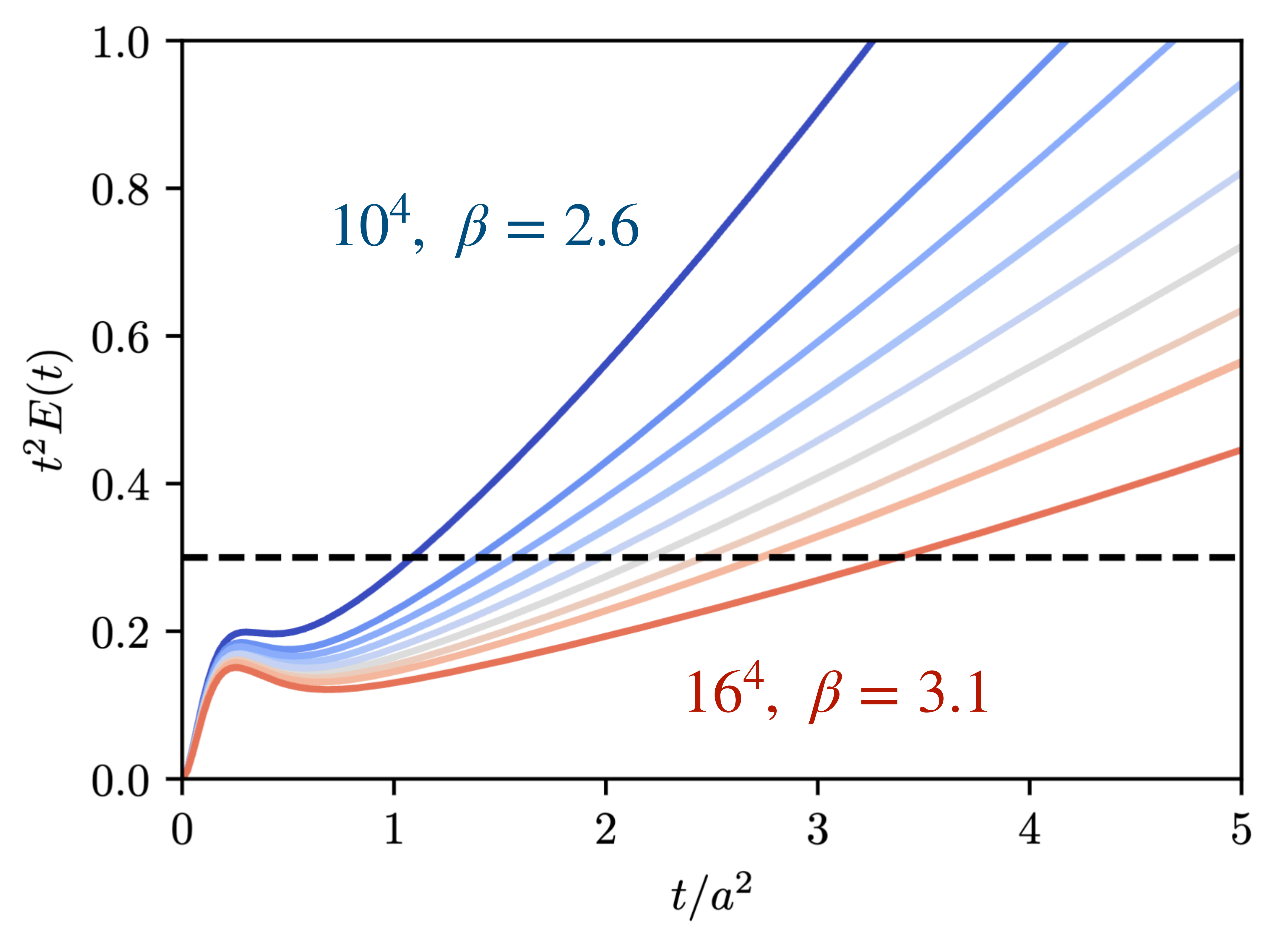
Flow time $t_{0.3}$ sets $g^2 \approx 15.79$

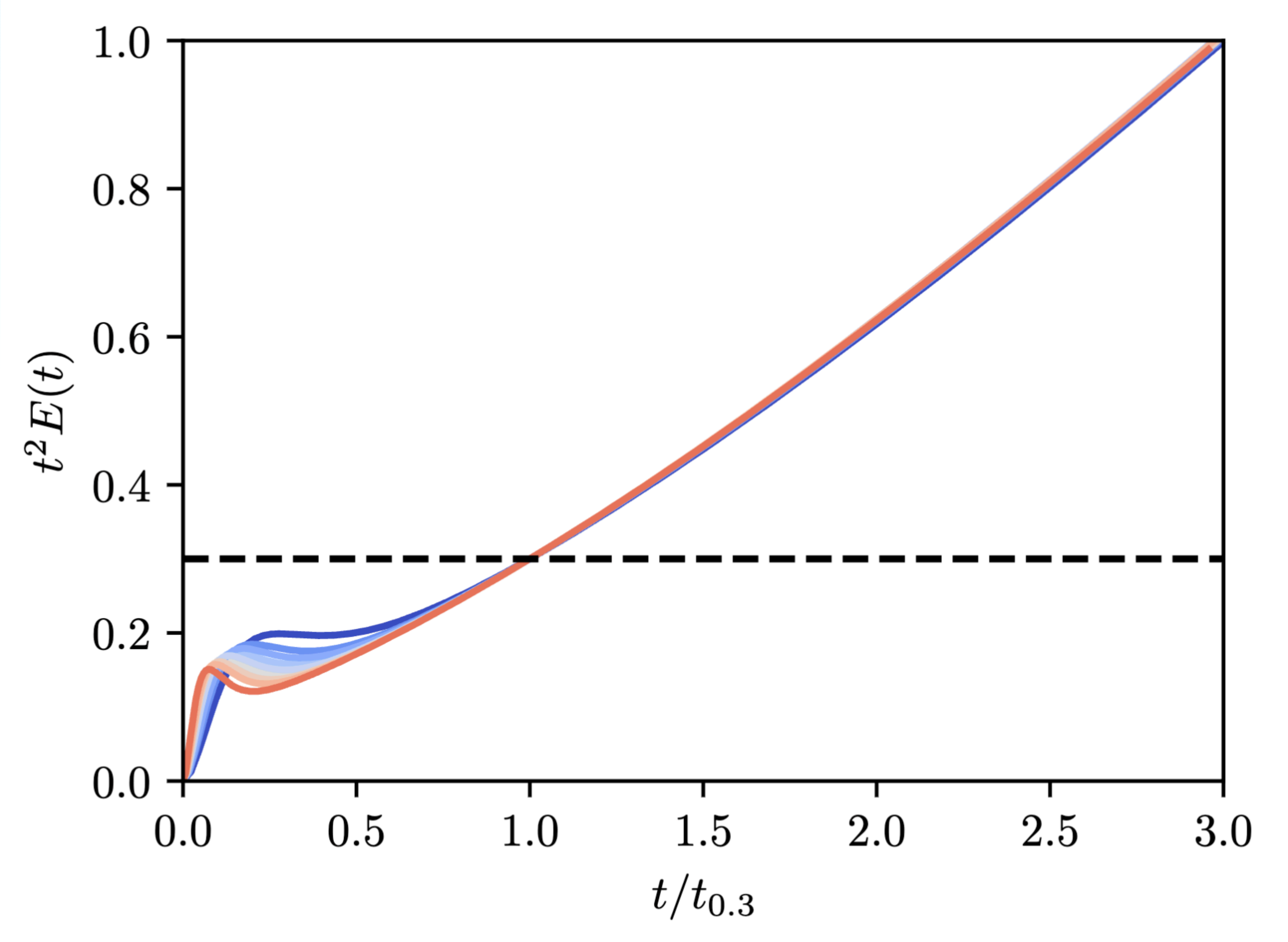
see e.g. Wong et al [2301.06611]



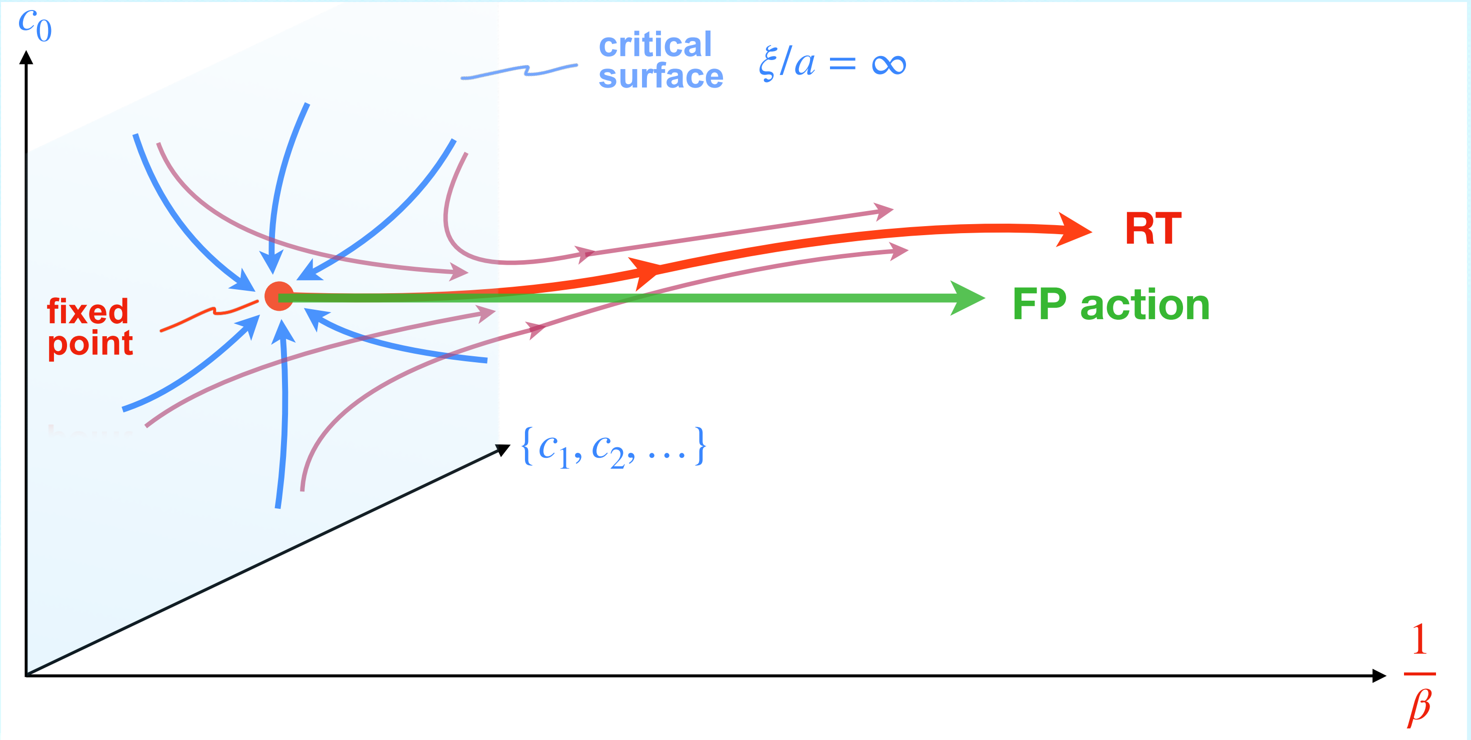
Universal flow trajectores

Rescale flow trajectories at different β with $t_{0.3}(\beta)$



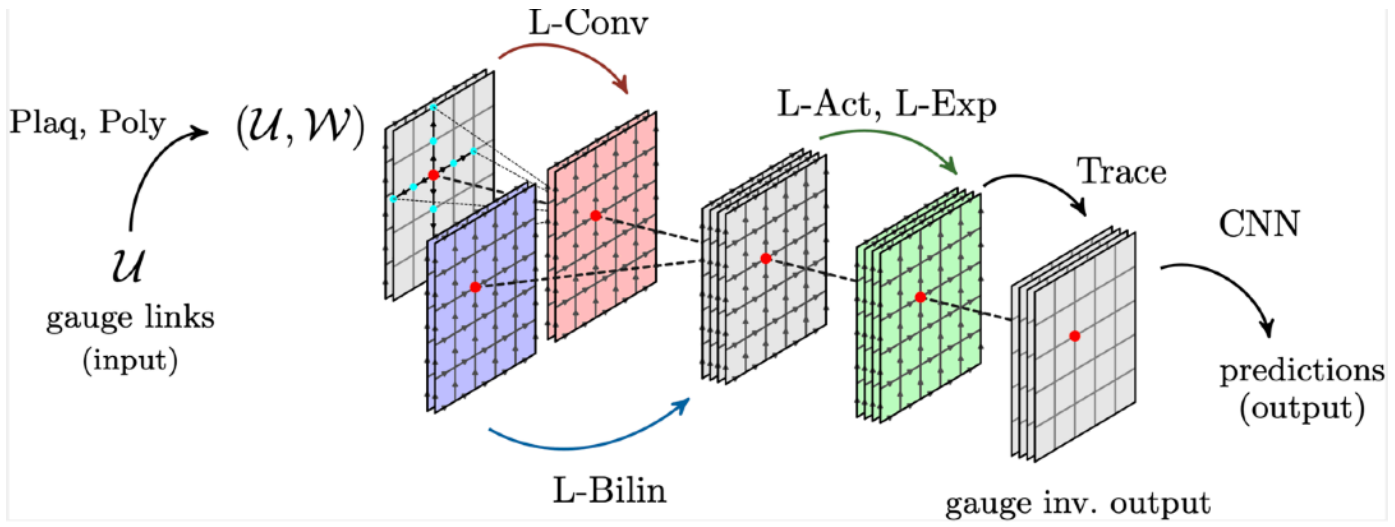


Conclusions and outlook



- Highly improved FP parametrization using L-CNNs
- HMC and gradient flow with FP action
- Gradient flow observables show correct scaling on coarse lattices

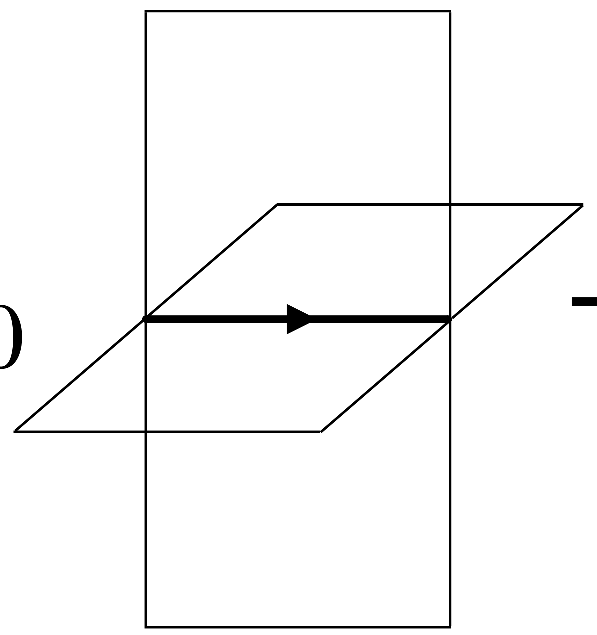
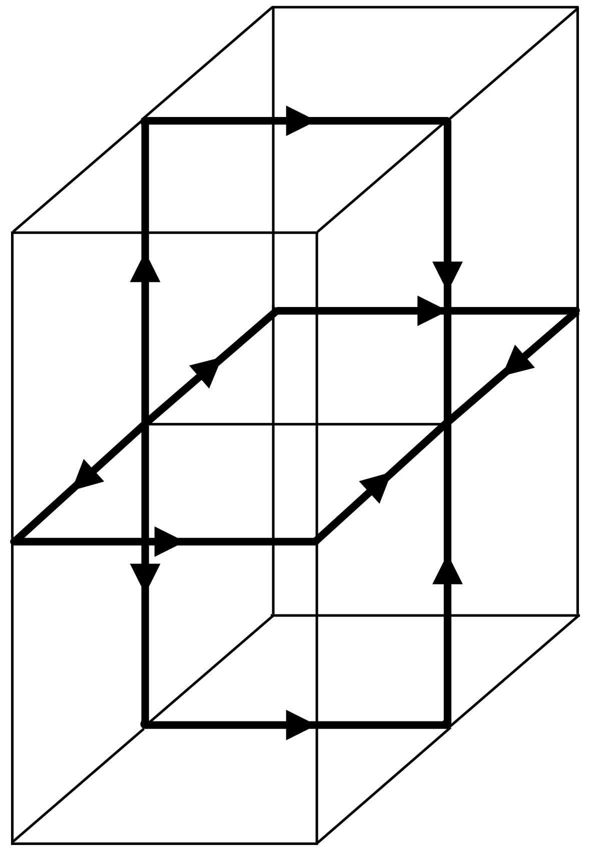
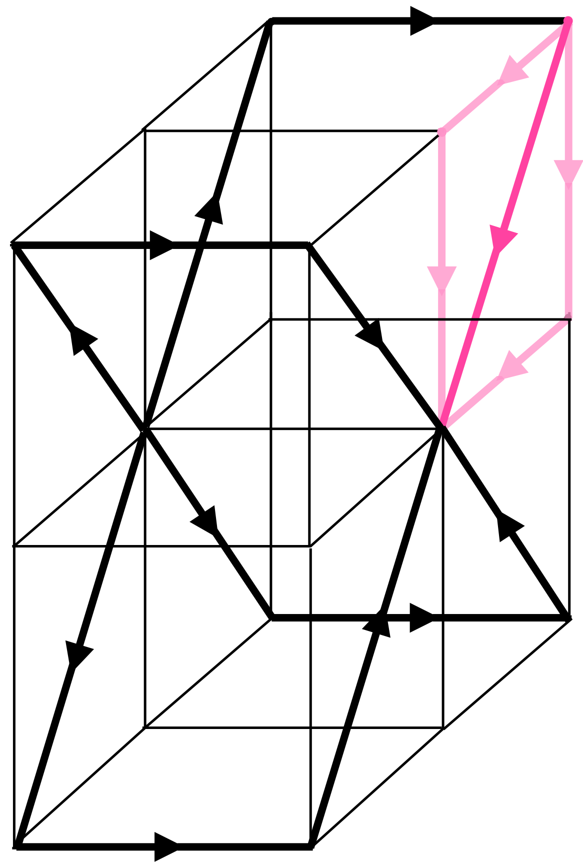
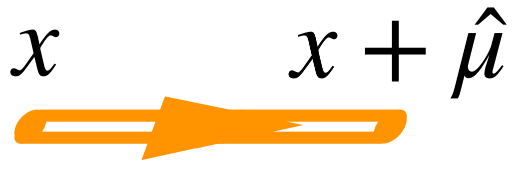
- Other observables (critical temperature, glueballs, …)
- Simulations with a quantum perfect action

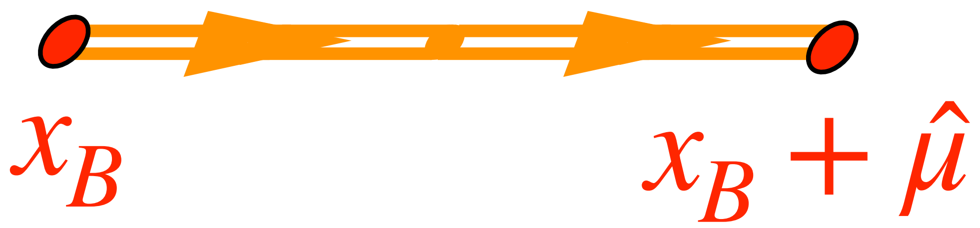


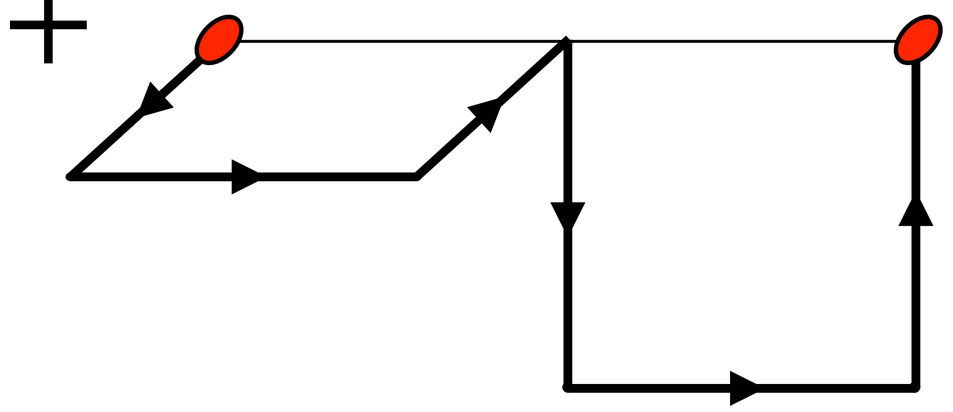
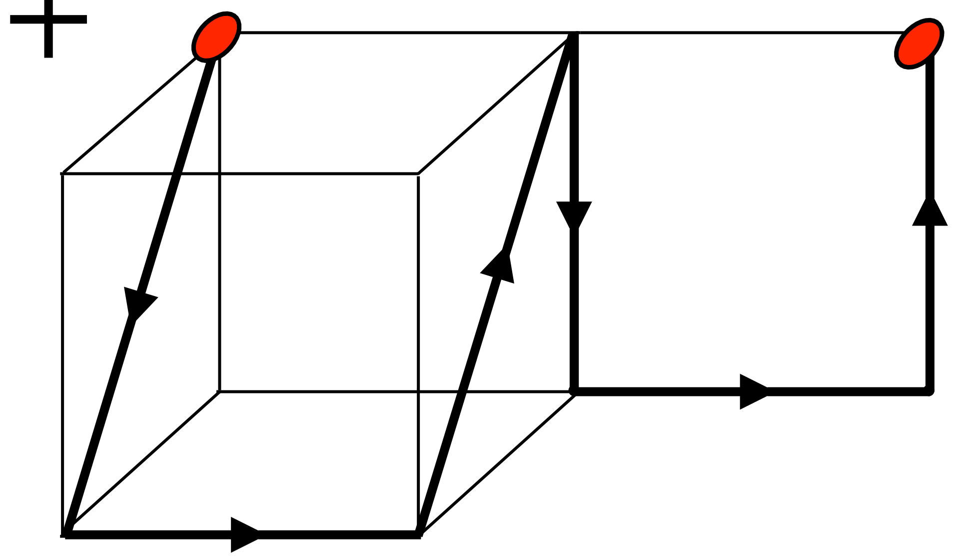
Appendix

RGT blocking kernel

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

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

$Q_\mu(x_B) =$ 

$= \ldots +$  $+$  $+ \ldots$

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:

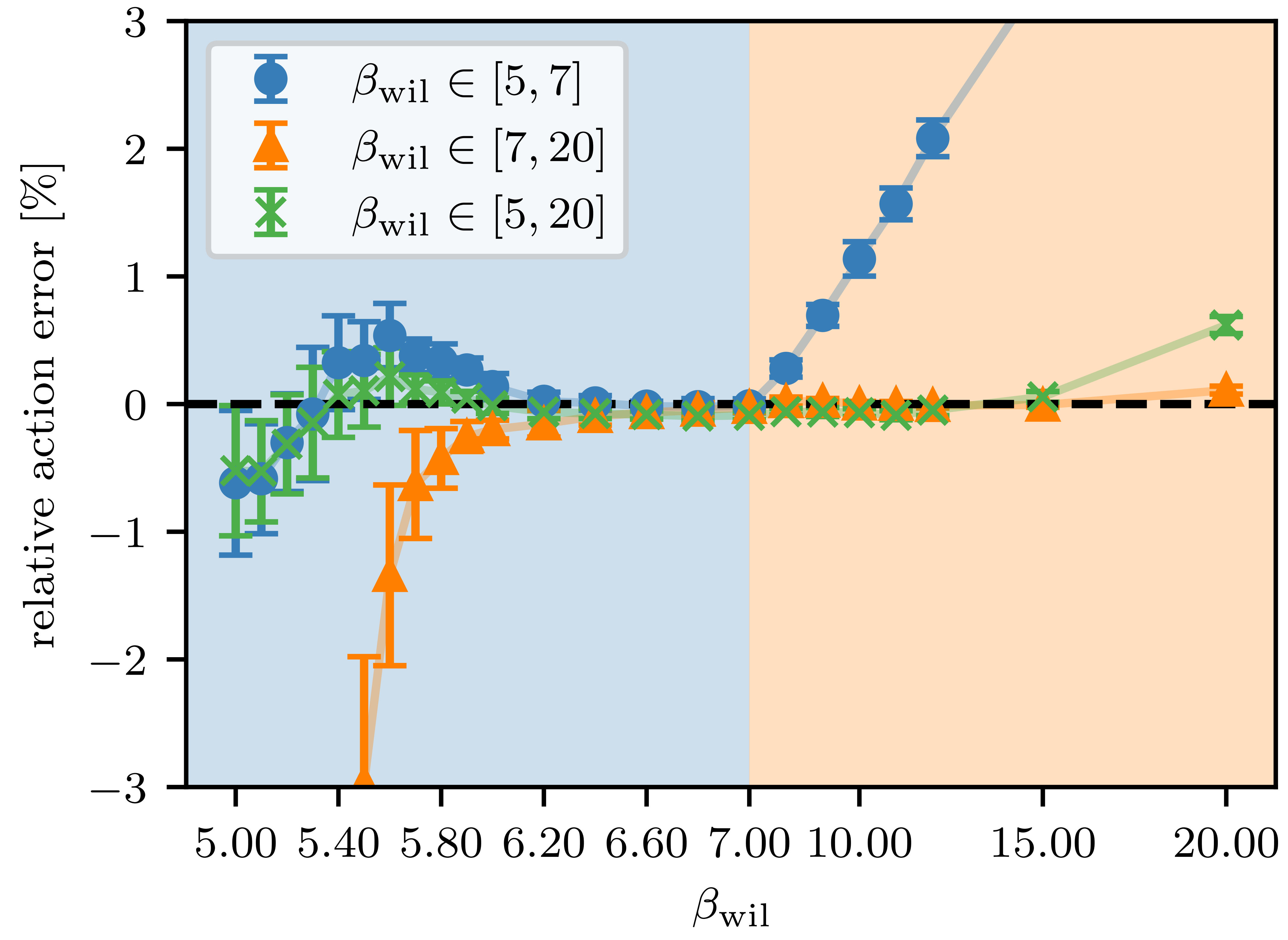
$$Q^S_\mu = \frac{1}{6}\sum_{\lambda\neq\mu} S^{(\lambda)}_\mu - U_\mu\,, \qquad Q^{(\nu)}_\mu = \frac{1}{4}\left(\sum_{\lambda\neq\mu,\nu} S^{(\lambda)}_\mu + \eta(x_\mu)\cdot S^{(\nu)}_\mu\right) - \left(1 + \frac{1}{2}\eta(x_\mu)\right) U_\mu\,,$$

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

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

Machine learning the FP action: Results

Restricted training ranges:

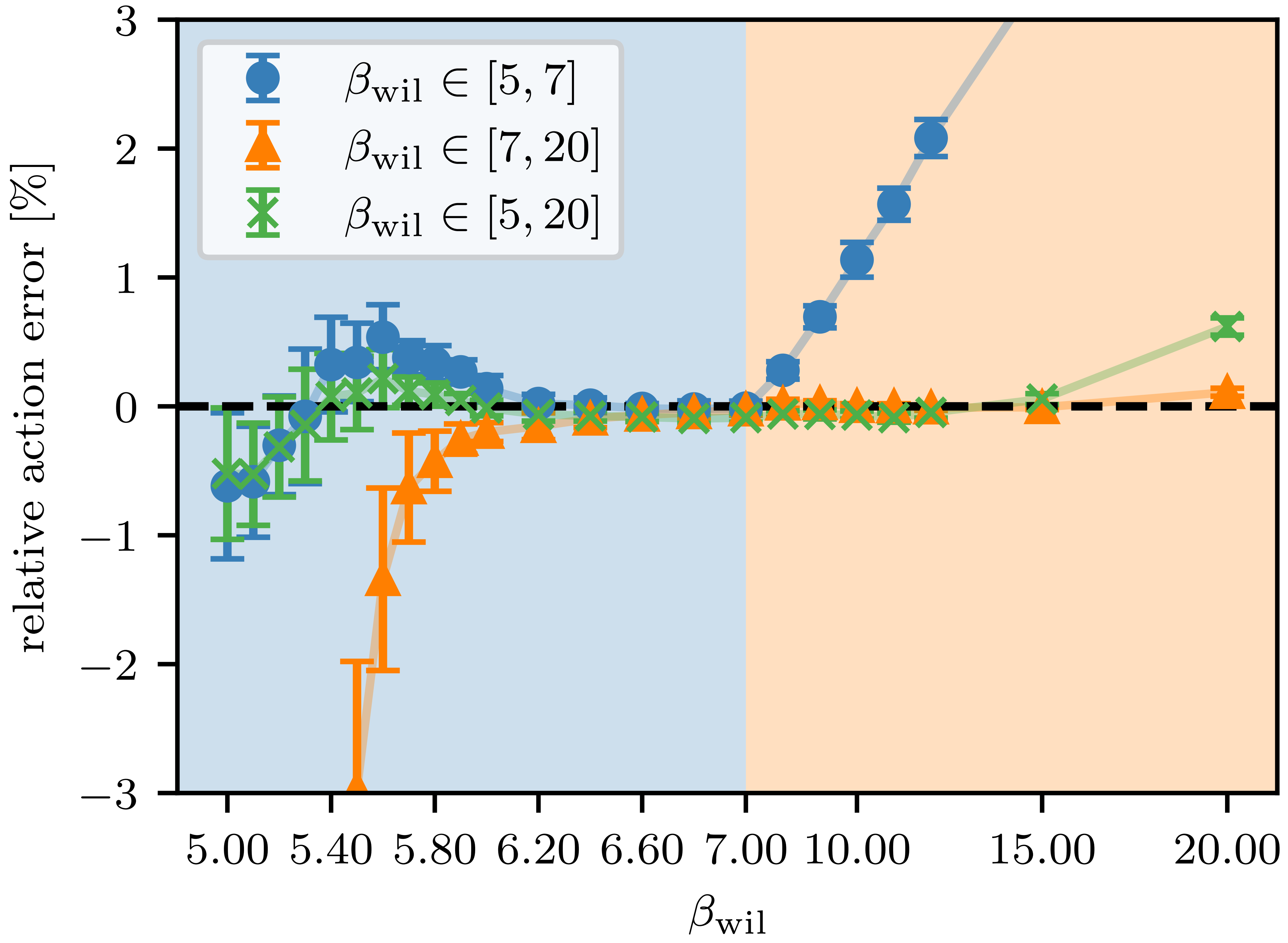


Transfer learning:

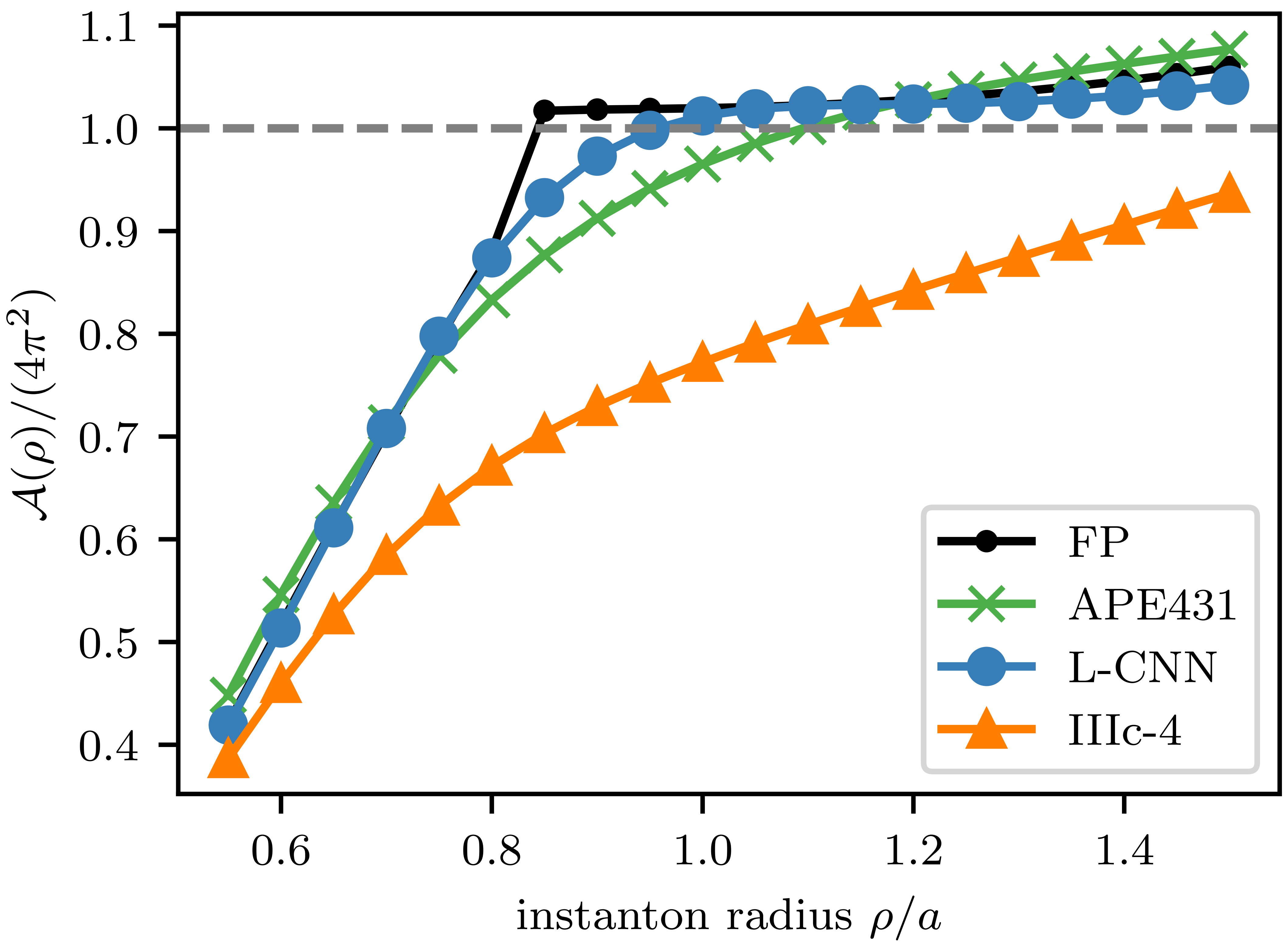
finetuned model	relative error (test data) 4^4	6^4	8^4
4^4	**0.178 %**	0.201 %	0.181 %
6^4	0.185 %	**0.196 %**	0.177 %
8^4	0.191 %	0.202 %	**0.176 %**
finetuned model	**derivative error (test data) 4^4**	**6^4**	**8^4**
4^4	$\mathbf{7.63 \times 10^{-2}}$	8.19×10^{-2}	8.22×10^{-2}
6^4	$\mathbf{7.39 \times 10^{-2}}$	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:



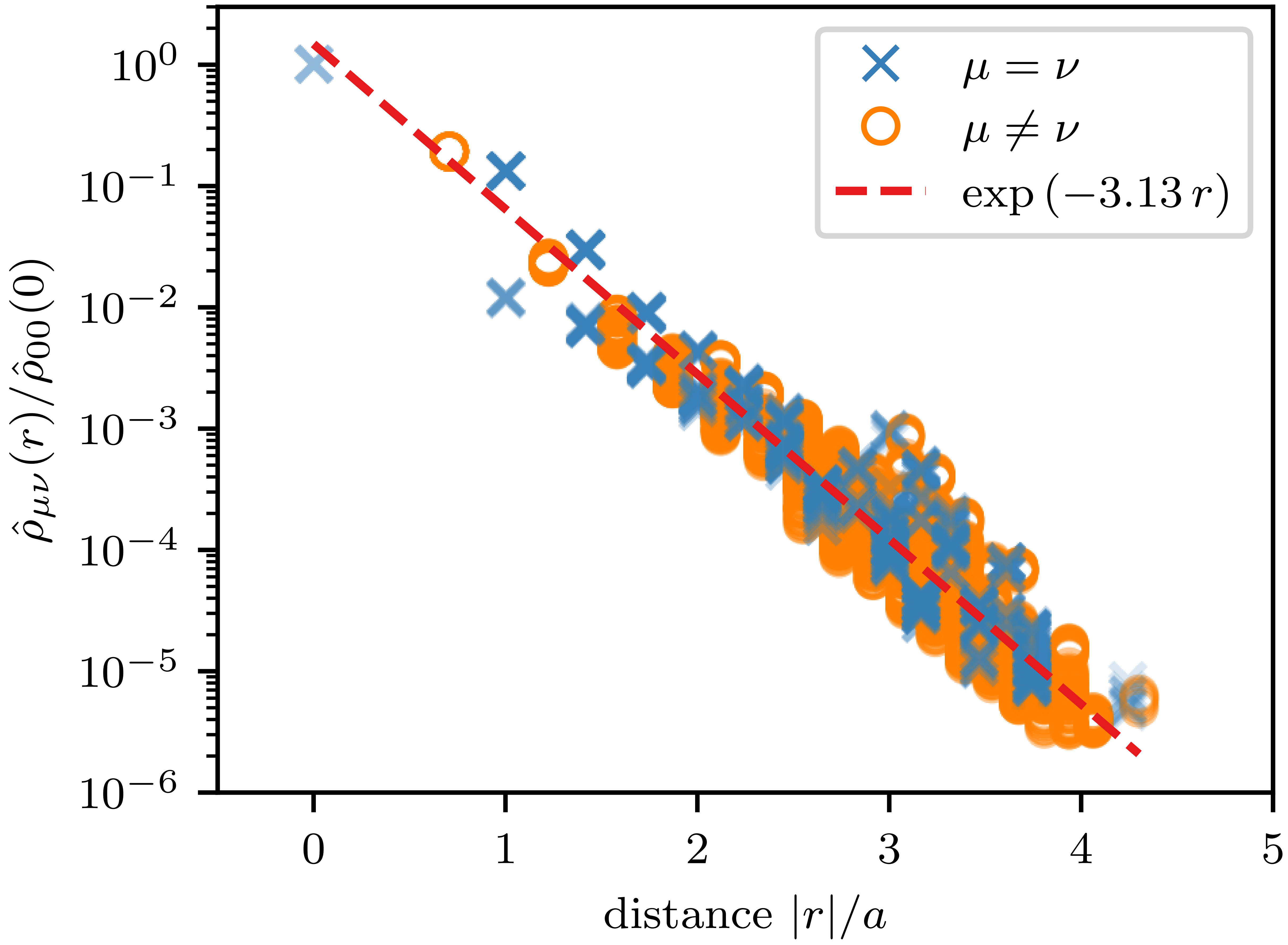
Finetuning on instantons:



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

Machine learning the FP action: Locality

Locality of L-CNN trained FP action:



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

where $D^{ab}_{\mu\nu}(x,y) = \dfrac{\delta^2 A}{\delta V^a_{x,\mu} \delta V^b_{y,\nu}}$

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

Machine learning the FP action: Symmetries

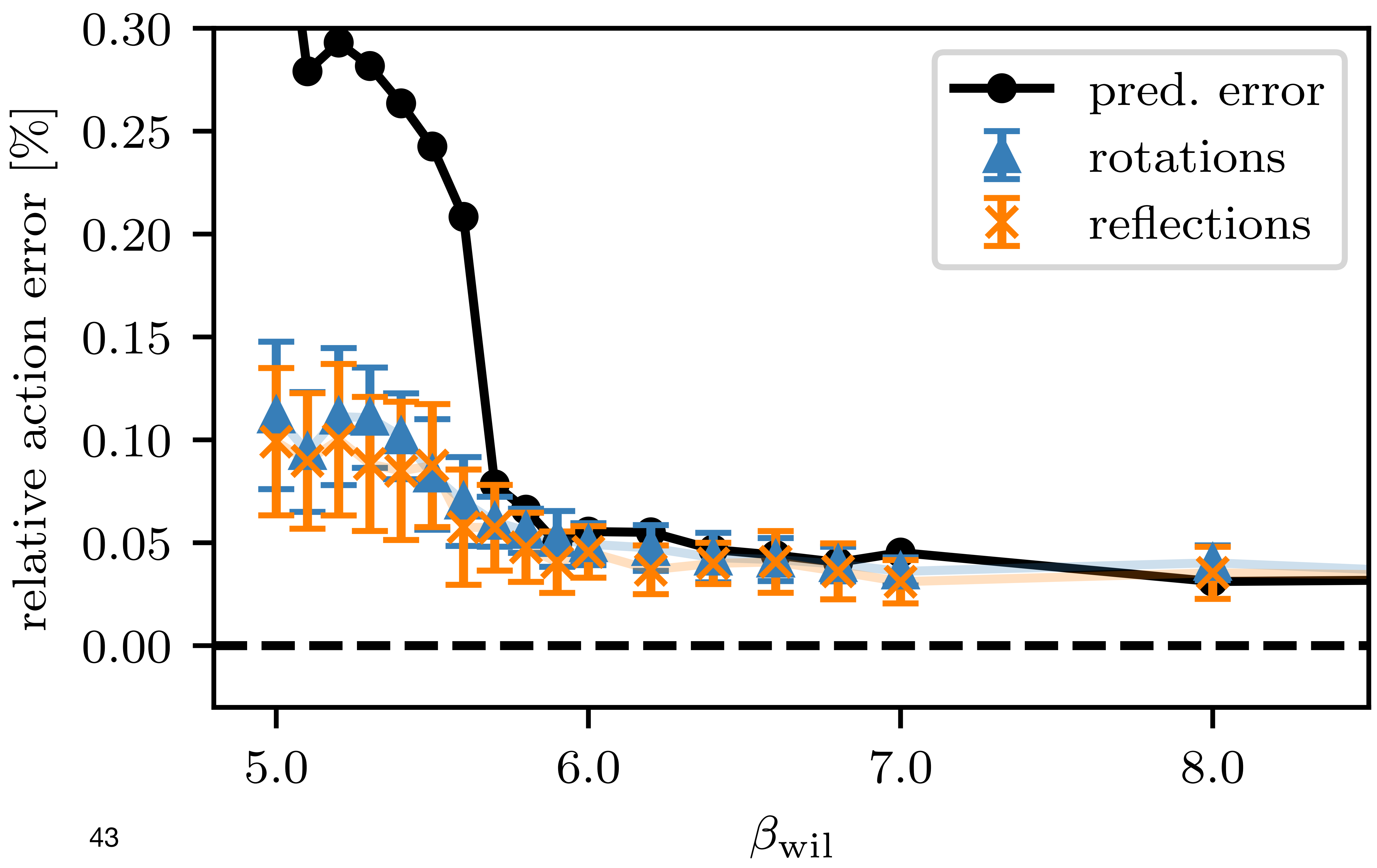
Test of lattice symmetries:

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!



HMC performance