

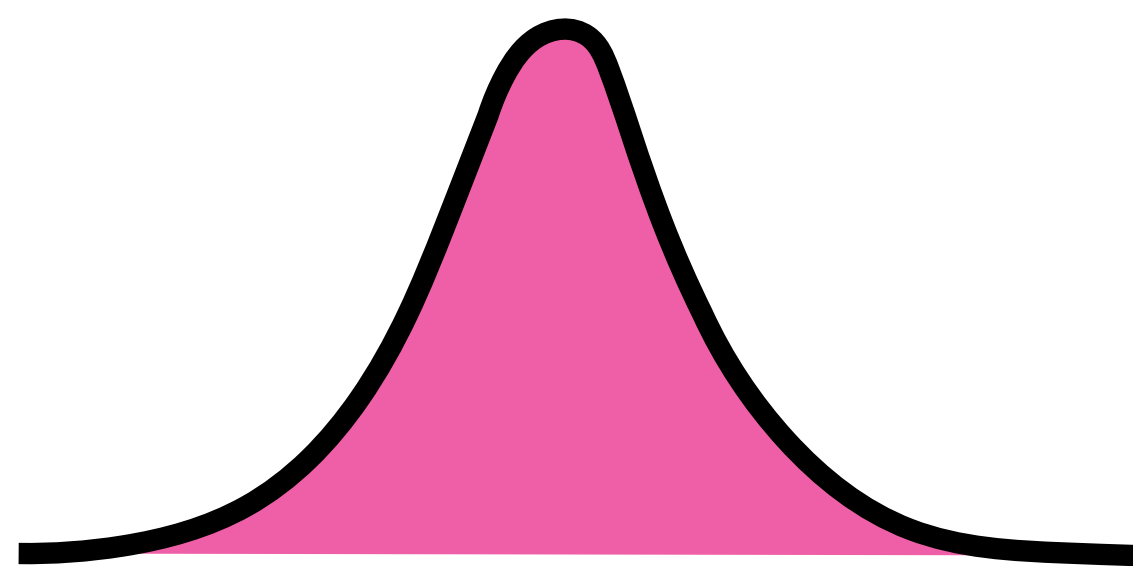
Learning the generating functional

Fernando Romero-López

fernando.romero-lopez@unibe.ch

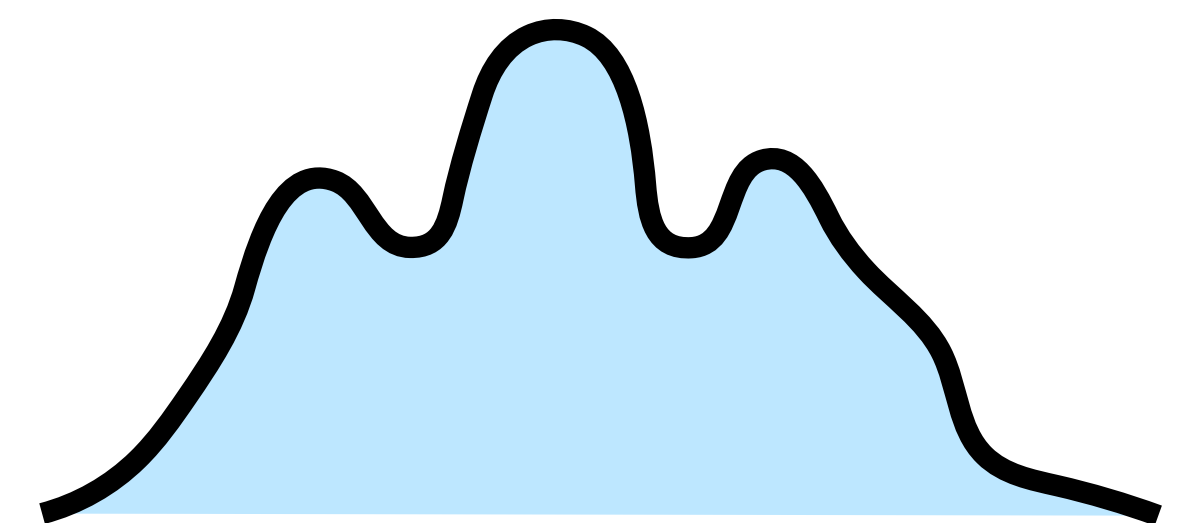
Lattice Conference @ UMD

July 14th



u^b

UNIVERSITÄT
BERN



Acknowledgements



• Phiala Shanahan



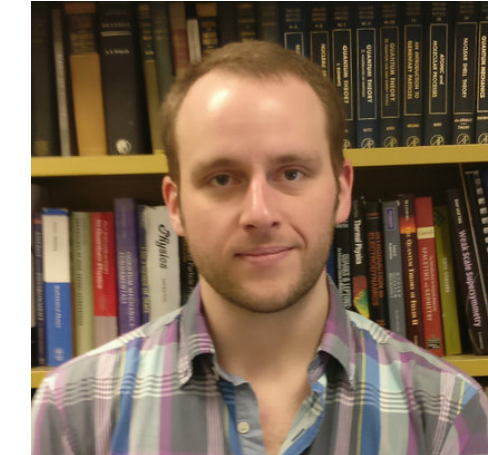
• Yang Fu



• Julian Urban

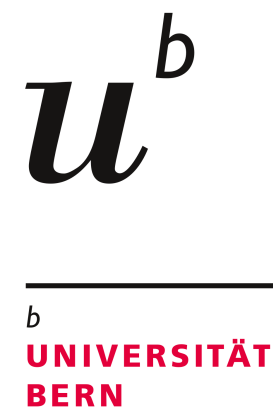


• Denis Boyda



• Dan Hackett

[See plenary by Dan Hackett!]



• Fernando Romero-Lopez



• Ryan Abbot



• Gurtej Kanwar

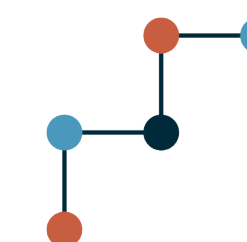
Learning the generating functional for variance reduction in lattice QCD

Ryan Abbott,¹ Yang Fu,^{2,3} Daniel C. Hackett,⁴ Gurtej Kanwar,⁵
Fernando Romero-López,⁶ and Phiala E. Shanahan^{2,3}

arXiv:2606.15986



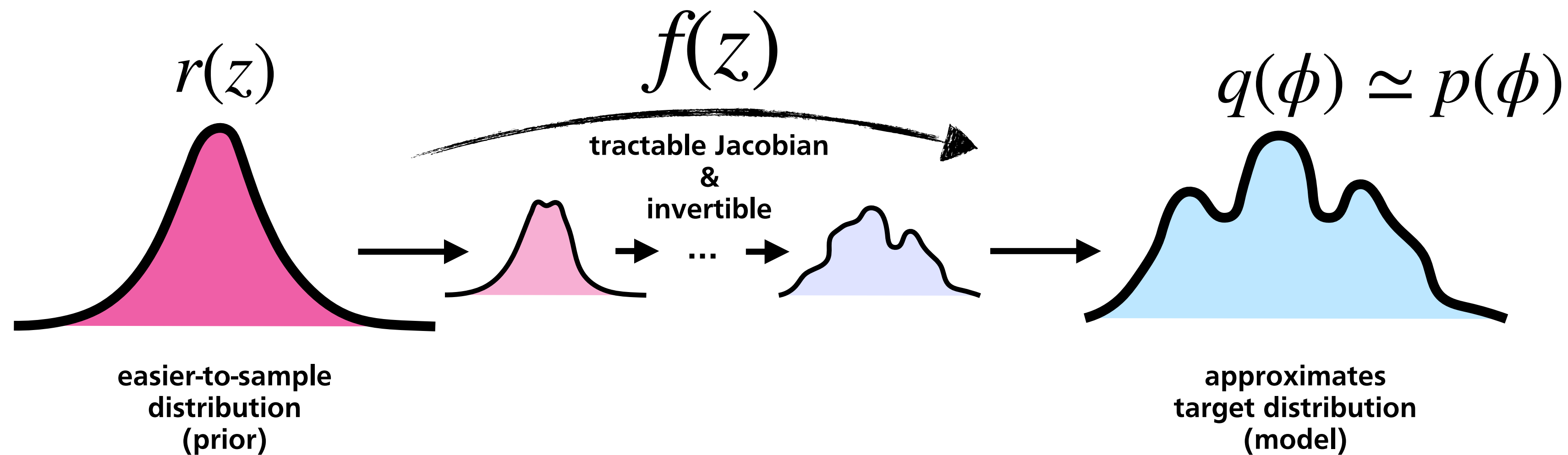
Platform for Advanced Scientific Computing



**Schweizerischer
Nationalfonds**

Flows in Lattice QCD

- **Normalizing flows** provide a promising tool to accelerate lattice QCD calculations
[Rezende, Mohamed, 1505.05770]



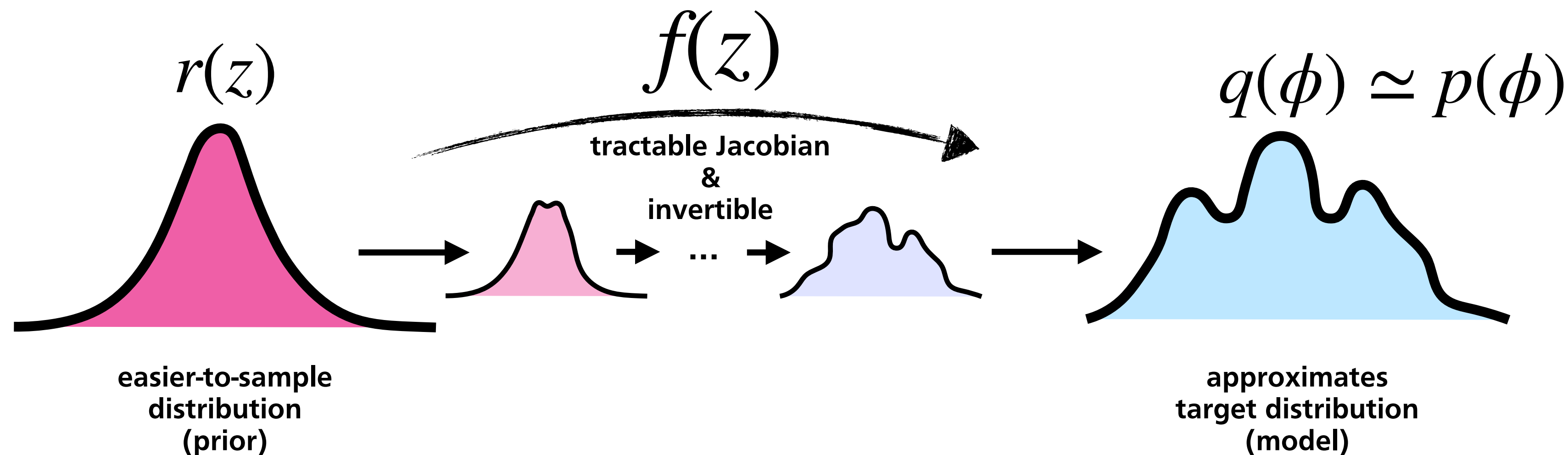
"Trainable change of variables"

Model probability

$$q(\phi) = r(z) \left| \det \frac{\partial f(z)}{\partial z} \right|^{-1}$$

Flows in Lattice QCD

- **Normalizing flows** provide a promising tool to accelerate lattice QCD calculations
[Rezende, Mohamed, 1505.05770]



"Trainable change of variables"

Model probability

$$q(\phi) = r(z) \left| \det \frac{\partial f(z)}{\partial z} \right|^{-1}$$

- Several approaches have focussed in sampling: high potential but difficult (volume scaling)
[Abbott et al, 2305.02402, 2404.11674]
[Related talks by Cellini, Dionisio, Freche, Granados-Pinto, Kreit, Parato, Schuh, Singha, and Vega] [See plenary by Hackett]
- Recently, new applications have targeted **variance reduction** of observables
[Bacchio, 2305.07932], [Catumba, Ramos 2502.15570], [Abbott et al, 2401.10874, 2603.02984, 2606.15986]
[See related talk by Butti]

Improved observables

- Variance reduction for observables that can be formulated with derivatives with respect to action parameters

$$\frac{d\langle\mathcal{O}\rangle}{d\alpha}$$

action parameter



- ▶ For instance, matrix element using **Feynman-Hellmann** techniques: sigma terms, hadron structure.

Improved observables

- Variance reduction for observables that can be formulated with derivatives with respect to action parameters

$$\frac{d\langle\mathcal{O}\rangle}{d\alpha}$$

action
parameter



- ▶ For instance, matrix element using **Feynman-Hellmann** techniques: sigma terms, hadron structure.

More generally: any N-point function via **generating functional**

Inspiration: [Catumba, Ramos 2502.15570]

Improved observables

- Variance reduction for observables that can be formulated with derivatives with respect to action parameters

$$\frac{d\langle \mathcal{O} \rangle}{d\alpha}$$

action
parameter



- ▶ For instance, matrix element using **Feynman-Hellmann** techniques: sigma terms, hadron structure.

More generally: any N-point function via **generating functional**

Inspiration: [Catumba, Ramos 2502.15570]

- Interpretation as **"control variate"**: same expectation value but difference variance

$$\langle \mathcal{Q} \rangle \rightarrow \langle \mathcal{Q} \rangle + \langle \nabla \cdot F - F \cdot \nabla S_0 \rangle \equiv \langle \mathcal{Q} \rangle + \langle \mathcal{L}(F) \rangle$$

$$\langle \mathcal{L}(F) \rangle = 0, \quad \text{Cov}(\mathcal{L}(F), \mathcal{Q}) \neq 0$$


Generating functional

- Variance reduction can be framed by means of the generating functional

$$Z_{\lambda_1, \dots, \lambda_n} = \int DU e^{-S_{\text{QCD}}(U) + \sum_i \lambda_i \mathcal{O}_i} \xrightarrow{\text{Ratio of partition functions}} \frac{Z_{\lambda_1, \dots, \lambda_n}}{Z_0} = \langle e^{\sum_i \lambda_i \mathcal{O}_i} \rangle$$

expectation value
in $\lambda=0$ theory


Generating functional

- Variance reduction can be framed by means of the generating functional

$$Z_{\lambda_1, \dots, \lambda_n} = \int DU e^{-S_{\text{QCD}}(U) + \sum_i \lambda_i \mathcal{O}_i} \xrightarrow{\text{Ratio of partition functions}} \frac{Z_{\lambda_1, \dots, \lambda_n}}{Z_0} = \langle e^{\sum_i \lambda_i \mathcal{O}_i} \rangle$$

expectation value
in $\lambda=0$ theory


- ✓ **N-point function** can be generated via **derivatives with respect to small parameters**

$$\langle \mathcal{O}_1 \dots \mathcal{O}_n \rangle = \frac{\partial^n}{\partial \lambda_1 \dots \partial \lambda_n} \frac{Z_{\lambda_1, \dots, \lambda_n}}{Z_0} \Big|_{\lambda_i=0}$$

Generating functional

- Variance reduction can be framed by means of the generating functional

$$Z_{\lambda_1, \dots, \lambda_n} = \int DU e^{-S_{\text{QCD}}(U) + \sum_i \lambda_i \mathcal{O}_i} \xrightarrow{\text{Ratio of partition functions}} \frac{Z_{\lambda_1, \dots, \lambda_n}}{Z_0} = \langle e^{\sum_i \lambda_i \mathcal{O}_i} \rangle$$

expectation value
in $\lambda=0$ theory


- ✓ **N-point function** can be generated via **derivatives with respect to small parameters**

$$\langle \mathcal{O}_1 \dots \mathcal{O}_n \rangle = \left. \frac{\partial^n}{\partial \lambda_1 \dots \partial \lambda_n} \frac{Z_{\lambda_1, \dots, \lambda_n}}{Z_0} \right|_{\lambda_i=0}$$

- Computing ratios of partition functions can be seen as a **reweighing problem**

$$\frac{Z_{\lambda_1, \dots, \lambda_n}}{Z_0} = \langle e^{\sum_i \lambda_i \mathcal{O}_i} \rangle \equiv \langle \hat{w}_{\lambda_1, \dots, \lambda_n} \rangle$$

unnormalized
reweighing factor


Example: one-point functions

- Example: one-point function

$$\langle \mathcal{O} \rangle = \partial_\lambda \langle \hat{w}_\lambda \rangle = \partial_\lambda \langle e^{\lambda \mathcal{O}} \rangle = \lim_{\lambda \rightarrow 0} \frac{\langle \hat{w}_\lambda \rangle - 1}{\lambda}$$

$\langle \hat{w}_\lambda \rangle = \frac{Z_\lambda}{Z_0}$

Example: one-point functions

- Example: one-point function $\langle \mathcal{O} \rangle = \partial_\lambda \langle \hat{w}_\lambda \rangle = \partial_\lambda \langle e^{\lambda \mathcal{O}} \rangle = \lim_{\lambda \rightarrow 0} \frac{\langle \hat{w}_\lambda \rangle - 1}{\lambda}$
 $\langle \hat{w}_\lambda \rangle = \frac{Z_\lambda}{Z_0}$
- Introduce “Effective Sample Size per configuration” metric: $\text{ESS} = \frac{\langle w_\lambda \rangle^2}{\langle w_\lambda^2 \rangle}$ $\leftarrow$ Quantifies fluctuations in reweighing factors
 $\text{ESS} \in (0, 1]$
- The variance of a one-point function is:

$$\text{Var}[\mathcal{O}] = \text{Var} [\partial_\lambda \hat{w}_\lambda] = \lim_{\lambda \rightarrow 0} \frac{1}{\lambda^2} \left(\frac{1}{\text{ESS}} - 1 \right) = \text{const} + O(\lambda)$$

Example: one-point functions

- Example: one-point function $\langle \mathcal{O} \rangle = \partial_\lambda \langle \hat{w}_\lambda \rangle = \partial_\lambda \langle e^{\lambda \mathcal{O}} \rangle = \lim_{\lambda \rightarrow 0} \frac{\langle \hat{w}_\lambda \rangle - 1}{\lambda}$
 $\langle \hat{w}_\lambda \rangle = \frac{Z_\lambda}{Z_0}$
- Introduce “Effective Sample Size per configuration” metric: $\text{ESS} = \frac{\langle w_\lambda \rangle^2}{\langle w_\lambda^2 \rangle}$ $\leftarrow$ Quantifies fluctuations in reweighing factors
 $\text{ESS} \in (0, 1]$
- The variance of a one-point function is:

$$\text{Var}[\mathcal{O}] = \text{Var} [\partial_\lambda \hat{w}_\lambda] = \lim_{\lambda \rightarrow 0} \frac{1}{\lambda^2} \left(\frac{1}{\text{ESS}} - 1 \right) = \text{const} + O(\lambda)$$

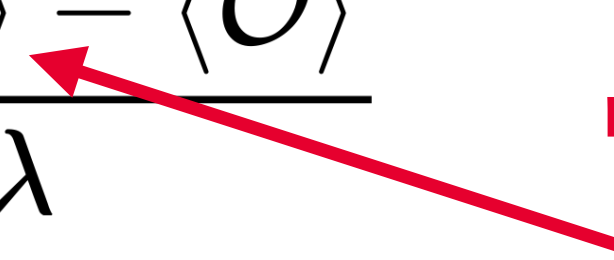
Natural transition into the flow approach: **increasing the ESS leads to lower variance!**

N-point functions

- Method 1: “Derivative trick” to obtain (N+1)-point function from N-point function
[Catumba, Ramos 2502.15570]

$$\langle \mathcal{O} \mathcal{J} \rangle - \langle \mathcal{O} \rangle \langle \mathcal{J} \rangle = \left. \frac{d\langle \mathcal{O} \rangle_\lambda}{d\lambda} \right|_{\lambda=0} = \langle \mathcal{O} \partial_\lambda w_\lambda \rangle|_{\lambda=0} \simeq \frac{\langle \mathcal{O} w_\lambda \rangle - \langle \mathcal{O} \rangle}{\lambda}$$

Normalized reweighing factor

$$w_\lambda = \frac{\hat{w}_\lambda}{\langle \hat{w}_\lambda \rangle}$$


N-point functions

- Method 1: “Derivative trick” to obtain (N+1)-point function from N-point function
[Catumba, Ramos 2502.15570]

$$\langle \mathcal{O} \mathcal{J} \rangle - \langle \mathcal{O} \rangle \langle \mathcal{J} \rangle = \left. \frac{d\langle \mathcal{O} \rangle_\lambda}{d\lambda} \right|_{\lambda=0} = \langle \mathcal{O} \partial_\lambda w_\lambda \rangle|_{\lambda=0} \simeq \frac{\langle \mathcal{O} w_\lambda \rangle - \langle \mathcal{O} \rangle}{\lambda}$$

Normalized
reweighing factor
 $w_\lambda = \frac{\hat{w}_\lambda}{\langle \hat{w}_\lambda \rangle}$

- Method 2: generating functional approach to generate N-point functions from zero-point functions

$$\langle \mathcal{O}_k \mathcal{O}_\ell \rangle = \frac{d^2 \langle \hat{w}_{\lambda_k, \lambda_\ell} \rangle}{d\lambda_k d\lambda_\ell} \quad \text{where} \quad \hat{w}_{\lambda_k, \lambda_\ell} = e^{\lambda_k \mathcal{O}_k + \lambda_\ell \mathcal{O}_\ell}$$

N-point functions

- Method 1: “Derivative trick” to obtain (N+1)-point function from N-point function
[Catumba, Ramos 2502.15570]

$$\langle \mathcal{O} \mathcal{J} \rangle - \langle \mathcal{O} \rangle \langle \mathcal{J} \rangle = \left. \frac{d\langle \mathcal{O} \rangle_\lambda}{d\lambda} \right|_{\lambda=0} = \langle \mathcal{O} \partial_\lambda w_\lambda \rangle|_{\lambda=0} \simeq \frac{\langle \mathcal{O} w_\lambda \rangle - \langle \mathcal{O} \rangle}{\lambda}$$

Normalized
reweighing factor

$$w_\lambda = \frac{\hat{w}_\lambda}{\langle \hat{w}_\lambda \rangle}$$

- Method 2: generating functional approach to generate N-point functions from zero-point functions

$$\langle \mathcal{O}_k \mathcal{O}_\ell \rangle = \frac{d^2 \langle \hat{w}_{\lambda_k, \lambda_\ell} \rangle}{d\lambda_k d\lambda_\ell} \quad \text{where} \quad \hat{w}_{\lambda_k, \lambda_\ell} = e^{\lambda_k \mathcal{O}_k + \lambda_\ell \mathcal{O}_\ell}$$

- Generating functional: Higher potential for variance reduction: $\text{Var}[\mathcal{O}_k \mathcal{O}_\ell] \simeq \text{Var}[\mathcal{O}_k] \text{Var}[\mathcal{O}_\ell] + \dots$
(For the independent variables, e.g. at large time separations)

Learned generating functional

- We train a flow to learn the insertion on an operator:

model prob
density 

$$q(V) = r(U) |\det J_f|^{-1} \approx \frac{1}{Z_\lambda} e^{-S(V) + \lambda \mathcal{O}(V)}.$$

Learned generating functional

- We train a flow to learn the insertion on an operator:

model prob density $\longrightarrow$ $q(V) = r(U) |\det J_f|^{-1} \approx \frac{1}{Z_\lambda} e^{-S(V) + \lambda \mathcal{O}(V)}.$

- Generating functional approach still holds

$$\langle O \rangle = \langle \partial_\lambda \hat{w}_\lambda \rangle \simeq \frac{\langle \hat{w}_\lambda \rangle - 1}{\lambda}$$

unnormalized reweighing factor on the flowed fields!

- Importantly, reweighing factors includes now the Jacobian of the flow

$$\hat{w}_\lambda(U) \equiv e^{-S(V) + \lambda \mathcal{O}(V) + S(U)} |\det J_f|, \quad V = f(U)$$

Learned generating functional

- We train a flow to learn the insertion on an operator:

model prob density $\longrightarrow$ $q(V) = r(U) |\det J_f|^{-1} \approx \frac{1}{Z_\lambda} e^{-S(V) + \lambda \mathcal{O}(V)}.$

- Generating functional approach still holds

$$\langle O \rangle = \langle \partial_\lambda \hat{w}_\lambda \rangle \simeq \frac{\langle \hat{w}_\lambda \rangle - 1}{\lambda}$$

unnormalized reweighing factor on the flowed fields!

- Importantly, reweighing factors includes now the Jacobian of the flow

$$\hat{w}_\lambda(U) \equiv e^{-S(V) + \lambda \mathcal{O}(V) + S(U)} |\det J_f|, \quad V = f(U)$$

- **Guarantee of exactness:** flows do not introduce any bias

$$\langle \hat{w}_\lambda \rangle = \frac{Z_\lambda}{Z_0} \quad \text{for any invertible flow}$$

Perfect flow limit

- Recall the one-point function expression:

$$\langle O \rangle = \langle \partial_\lambda \hat{w}_\lambda \rangle \simeq \frac{\langle \hat{w}_\lambda \rangle - 1}{\lambda}$$

- A perfect flow, i.e. ESS=1, leads to noiseless estimators

$$\text{Var}[\mathcal{O}] = \text{Var}[\partial_\lambda \hat{w}_\lambda] = \lim_{\lambda \rightarrow 0} \frac{1}{\lambda^2} \left(\frac{1}{\text{ESS}} - 1 \right) \stackrel{\text{perfect flow}}{=} 0$$

- Key point:

$$\hat{w}_\lambda(U) \xrightarrow{\text{perfect flow}} \frac{Z_\lambda}{Z_0} \text{ configuration by configuration.}$$

Perfect flow limit

- Recall the one-point function expression:

$$\langle O \rangle = \langle \partial_\lambda \hat{w}_\lambda \rangle \simeq \frac{\langle \hat{w}_\lambda \rangle - 1}{\lambda}$$

- A perfect flow, i.e. $\text{ESS}=1$, leads to noiseless estimators

$$\text{Var}[\mathcal{O}] = \text{Var}[\partial_\lambda \hat{w}_\lambda] = \lim_{\lambda \rightarrow 0} \frac{1}{\lambda^2} \left(\frac{1}{\text{ESS}} - 1 \right) \stackrel{\text{perfect flow}}{=} 0$$

- Key point:

$$\hat{w}_\lambda(U) \xrightarrow{\text{perfect flow}} \frac{Z_\lambda}{Z_0} \text{ configuration by configuration.}$$

- Exact flow must exist as a connection between nearby distributions

► In practice, hard to achieve but high-quality flows are possible

$$\text{ESS}_{\text{flow}} > \text{ESS}_{\text{identity}}$$

Improved two-point functions

○ Two-point function as: $C(t) = \langle \mathcal{G}(t)\mathcal{G}(0) \rangle_c = \langle \mathcal{G}(t)\mathcal{G}(0) \rangle - \langle \mathcal{G}(t) \rangle \langle \mathcal{G}(0) \rangle$

finite-difference
second derivative $\longrightarrow$

$$C(t) \simeq \frac{1}{\lambda\lambda'} \left[\langle \hat{w}_{\lambda,\lambda'} \rangle - \langle \hat{w}_{\lambda,0} \rangle \langle \hat{w}_{0,\lambda'} \rangle \right]$$

$\langle \hat{w}_{\lambda,\lambda'} \rangle = \langle e^{\lambda\mathcal{G}(0) + \lambda'\mathcal{G}(t)} \rangle$

Improved two-point functions

○ Two-point function as: $C(t) = \langle \mathcal{G}(t)\mathcal{G}(0) \rangle_c = \langle \mathcal{G}(t)\mathcal{G}(0) \rangle - \langle \mathcal{G}(t) \rangle \langle \mathcal{G}(0) \rangle$

finite-difference
second derivative $\longrightarrow$

$$C(t) \simeq \frac{1}{\lambda\lambda'} \left[\langle \hat{w}_{\lambda,\lambda'} \rangle - \langle \hat{w}_{\lambda,0} \rangle \langle \hat{w}_{0,\lambda'} \rangle \right]$$

$\langle \hat{w}_{\lambda,\lambda'} \rangle = \langle e^{\lambda\mathcal{G}(0) + \lambda'\mathcal{G}(t)} \rangle$

- ✓ Perfect flow limit exists: can obtain noiseless N-points functions
- ✓ In practice, significant variance reduction can be achieved

Improved two-point functions

○ Two-point function as: $C(t) = \langle \mathcal{G}(t)\mathcal{G}(0) \rangle_c = \langle \mathcal{G}(t)\mathcal{G}(0) \rangle - \langle \mathcal{G}(t) \rangle \langle \mathcal{G}(0) \rangle$

finite-difference
second derivative $\longrightarrow$

$$C(t) \simeq \frac{1}{\lambda\lambda'} \left[\langle \hat{w}_{\lambda,\lambda'} \rangle - \langle \hat{w}_{\lambda,0} \rangle \langle \hat{w}_{0,\lambda'} \rangle \right]$$

$\langle \hat{w}_{\lambda,\lambda'} \rangle = \langle e^{\lambda\mathcal{G}(0) + \lambda'\mathcal{G}(t)} \rangle$

✓ Perfect flow limit exists: can obtain noiseless N-points functions

✓ In practice, significant variance reduction can be achieved

○ Practical implementation via “factorized flow” approximation: single flow used for several insertions

$$f_{\lambda,\lambda'}(U) \simeq T^{-1}(t, 0) \circ f_{\lambda'} \circ T(t, 0) \circ f_{\lambda}(U).$$

double flow $\simeq$ translated flow $\circ$ flow

► Interpretation: local fluctuations dominate the variance

Architecture: residual layers

- Use an equivariant flow architecture based on the original Trivializing Maps [Lüscher, arXiv:0907.5491]

[Abbott et al, 2305.02402]

See also: [Bacchio et al, 2212.08469]

[Gerdes et al, 2410.13161]

[Nagai, Tomiya, 2103.11965]

Untraced Wilson loops
that start and end at x

$$U'_\mu(x) = e^{F(U)} U_\mu(x)$$

$$F = \sum_i \delta_i P(W^i_{\mu\nu})$$

trainable

Traceless-antihermitian projection

Architecture: residual layers

- Use an equivariant flow architecture based on the original Trivializing Maps [Lüscher, arXiv:0907.5491]

[Abbott et al, 2305.02402]

See also: [Bacchio et al, 2212.08469]

[Gerdes et al, 2410.13161]

[Nagai, Tomiya, 2103.11965]

Untraced Wilson loops
that start and end at x

$$U'_\mu(x) = e^{F(U)} U_\mu(x)$$

$$F = \sum_i \delta_i \text{P}(W^i_{\mu\nu})$$

trainable

Traceless-antihermitian projection

- Split lattice in **active + frozen** variables, and update only active (upper triangular Jacobian)

Architecture: residual layers

- Use an equivariant flow architecture based on the original Trivializing Maps [Lüscher, arXiv:0907.5491]

[Abbott et al, 2305.02402]

See also: [Bacchio et al, 2212.08469]

[Gerdes et al, 2410.13161]

[Nagai, Tomiya, 2103.11965]

Untraced Wilson loops
that start and end at x

$$U'_\mu(x) = e^{F(U)} U_\mu(x)$$

$$F = \sum_i \delta_i P(W^i_{\mu\nu})$$

trainable

Traceless-antihermitian projection

- Split lattice in **active + frozen** variables, and update only active (upper triangular Jacobian)

- Build arbitrary loops "convoluting" the frozen links

$$F(V)$$

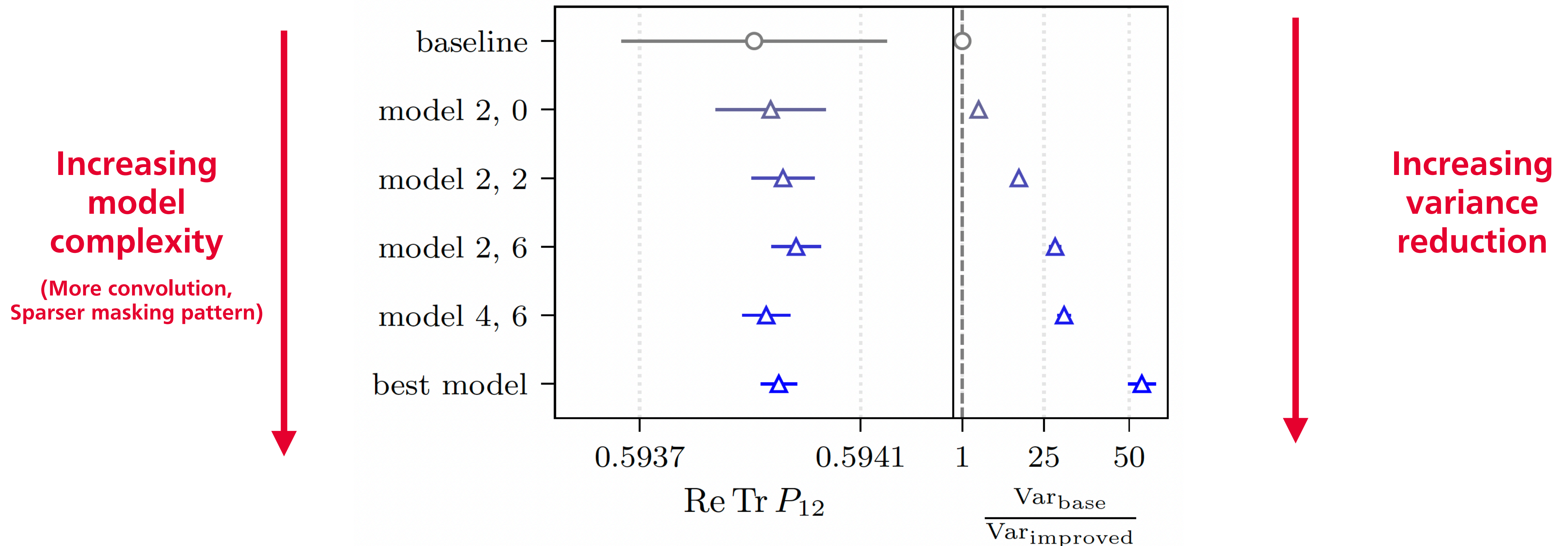
Force built from
convoluted links

$$V_\mu^{(1)} = U_\mu + \eta_i^1 \times (S_{x,\mu\nu}^R)^\dagger + \eta_i^2 \times W_{x,\mu\nu}^R U_\mu$$

[Similar to L-CNN, Favoni et al, 2012.12901]

Variance-reduced plaquette

- Consider a fixed-timeslice plaquette operator $\mathcal{G}(t) = \sum_{\vec{x}} P_{12}(\vec{x}, t)$ $S_{\text{YM}} \rightarrow S_{\text{YM}} + \lambda \mathcal{O}$
 - Significant variance reduction in one-point function at fixed number of configurations

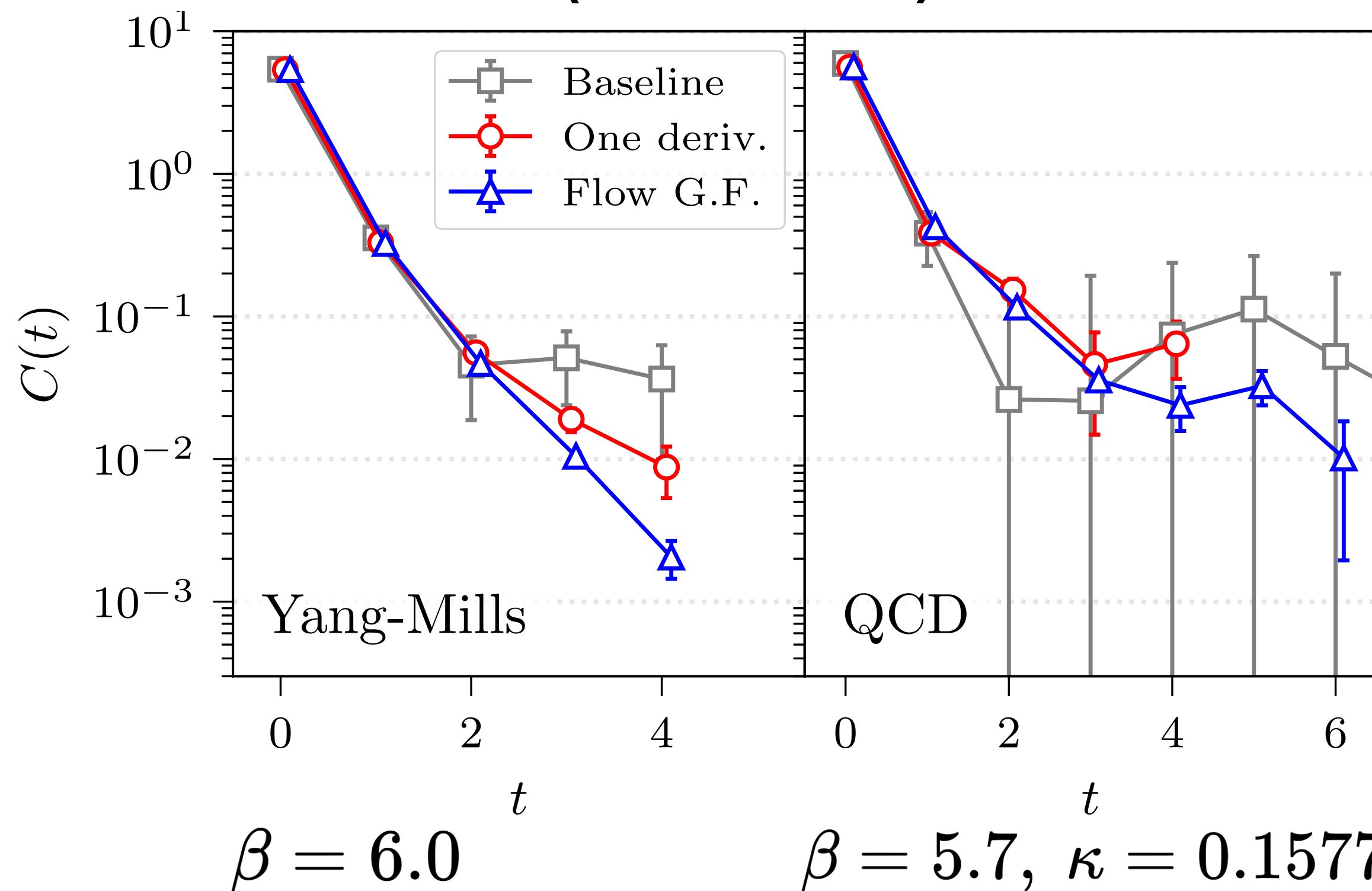


Glueballs in Yang-Mills & QCD

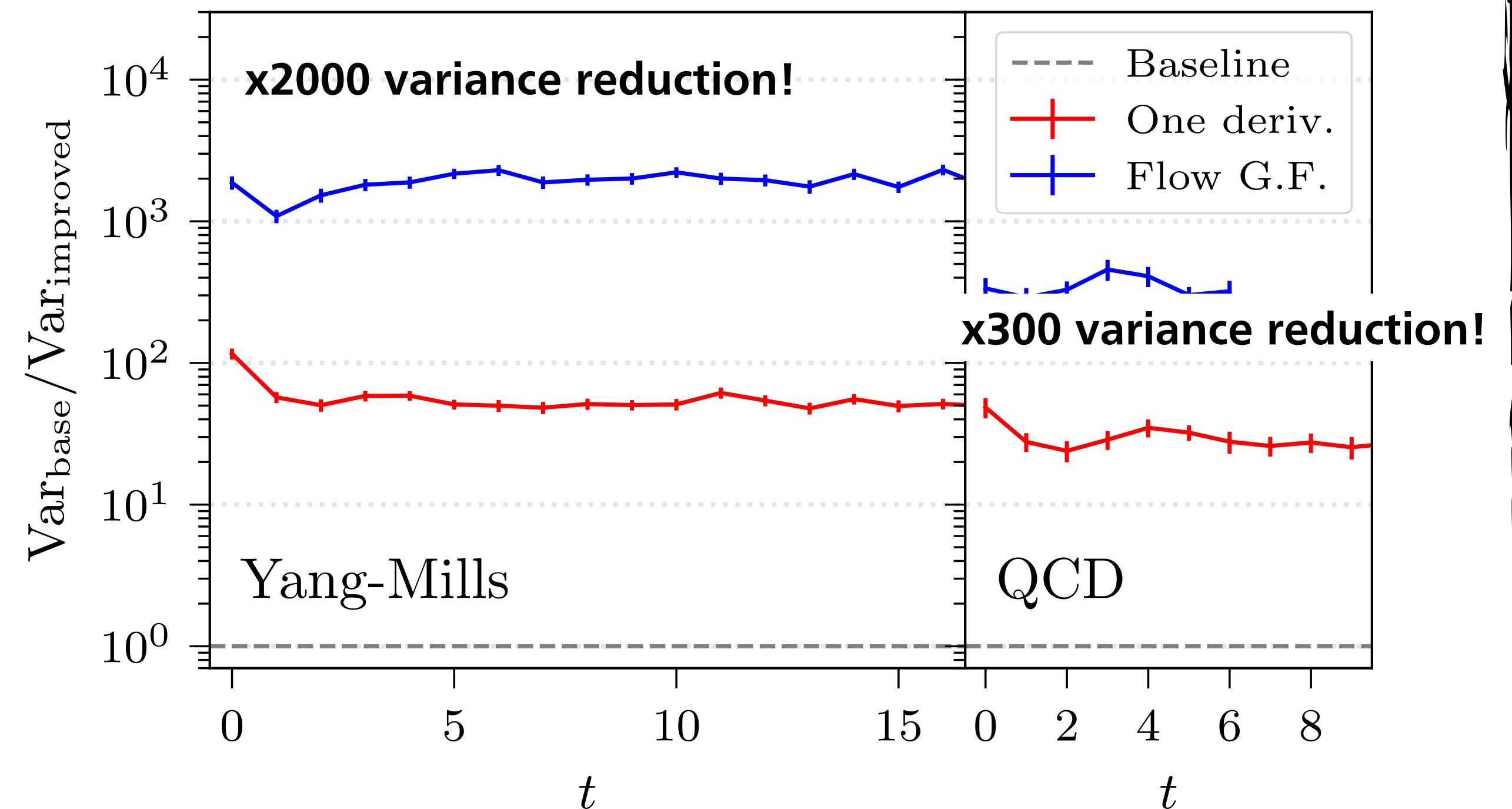
$$C(t) = \langle \mathcal{G}(t)\mathcal{G}(0) \rangle_c = \langle \mathcal{G}(t)\mathcal{G}(0) \rangle - \langle \mathcal{G}(t) \rangle \langle \mathcal{G}(0) \rangle$$

$$\mathcal{G}(t) = \sum_{\vec{x}} P_{12}(\vec{x}, t)$$

► Correlator ($V = 16^3 \times 32$)



► Variance reduction



The σ meson

- Understanding the sigma meson is a difficult target for hadron spectroscopy

[2304.03762 (HadSpec) , 1908.01847 (GWQCD)]

- Standard GEVP studies include correlators of the form

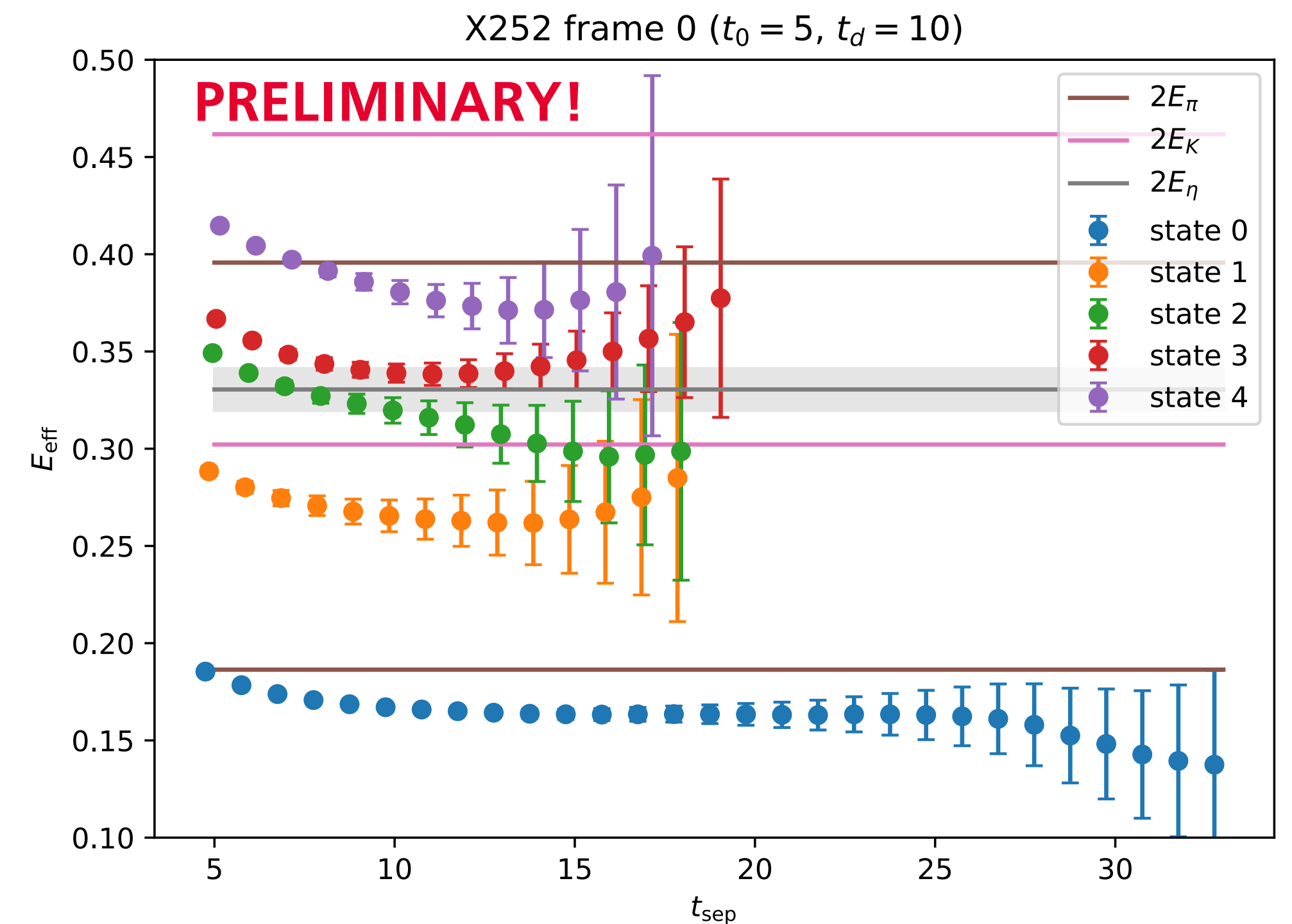
$$C(t) = \langle \sigma^\dagger(t) \sigma(0) \rangle = \text{[diagram: bubble diagram]} - \text{[diagram: two tadpoles]} - \text{[diagram: two tadpoles]}$$

$\sigma \equiv \bar{u}u + \bar{d}d$

- Noise is constant, while signal decays $\sim e^{-m_\sigma t}$

- Generating functional approach directly applicable

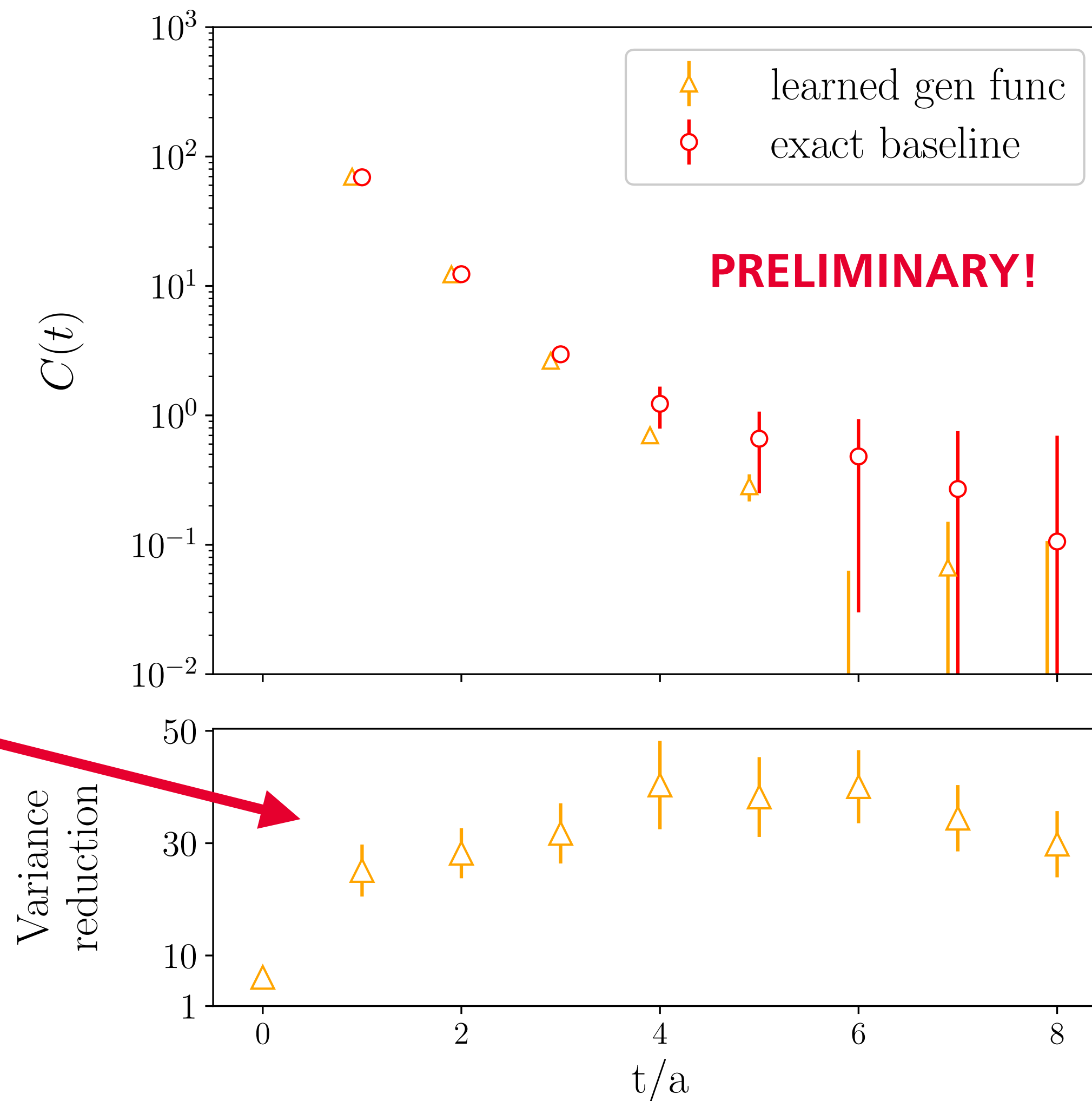
$$S_f \rightarrow S_f + \lambda_0 \sum_{\vec{x}} \bar{q}(\vec{x}, 0) q(\vec{x}, 0)$$



[see poster by Miguel Salg (Bern)]



Improved $\langle \sigma \sigma \rangle$ correlator



QCD with $N_f = 2$ Wilson fermions

$\beta = 5.7, \kappa = 0.1577$

$aM_\pi \sim 0.25, M_\pi \sim 600 \text{ MeV}$

$4^3 \times 16$

(Exact evaluation of propagators
and fermion determinants)

Higher variance reduction
at longer time separation

Factor ~30x variance reduction!

Summary and outlook

✓ The **generating functional** provides a natural framework for variance reduction

✓ Flows learn the insertion of operators: **zero-variance limit is possible!**

✓ Examples with $> O(100)$ variance reduction: glueballs

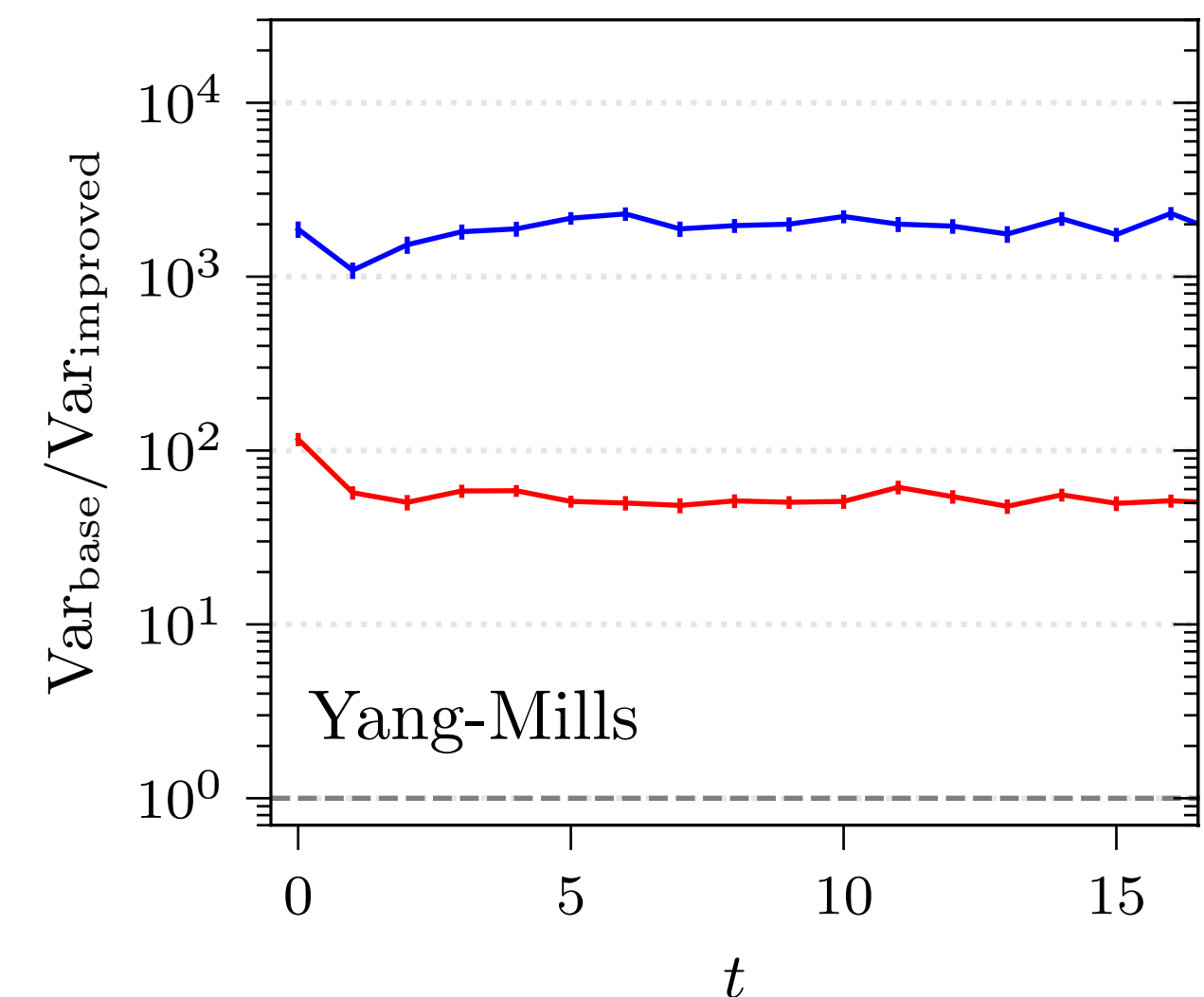
✓ Promising preliminary results in meson correlators: σ meson

○ On-going training for glueball correlators in QCD with $M_\pi \sim 170$ MeV

► Already achieved variance reduction: >100

○ Next steps: more observables and state-of-the-art QCD scales!

► More complicated observables: smeared Wilson loops, other quark bilinears



Summary and outlook

✓ The **generating functional** provides a natural framework for variance reduction

✓ Flows learn the insertion of operators: **zero-variance limit is possible!**

✓ Examples with $> O(100)$ variance reduction: glueballs

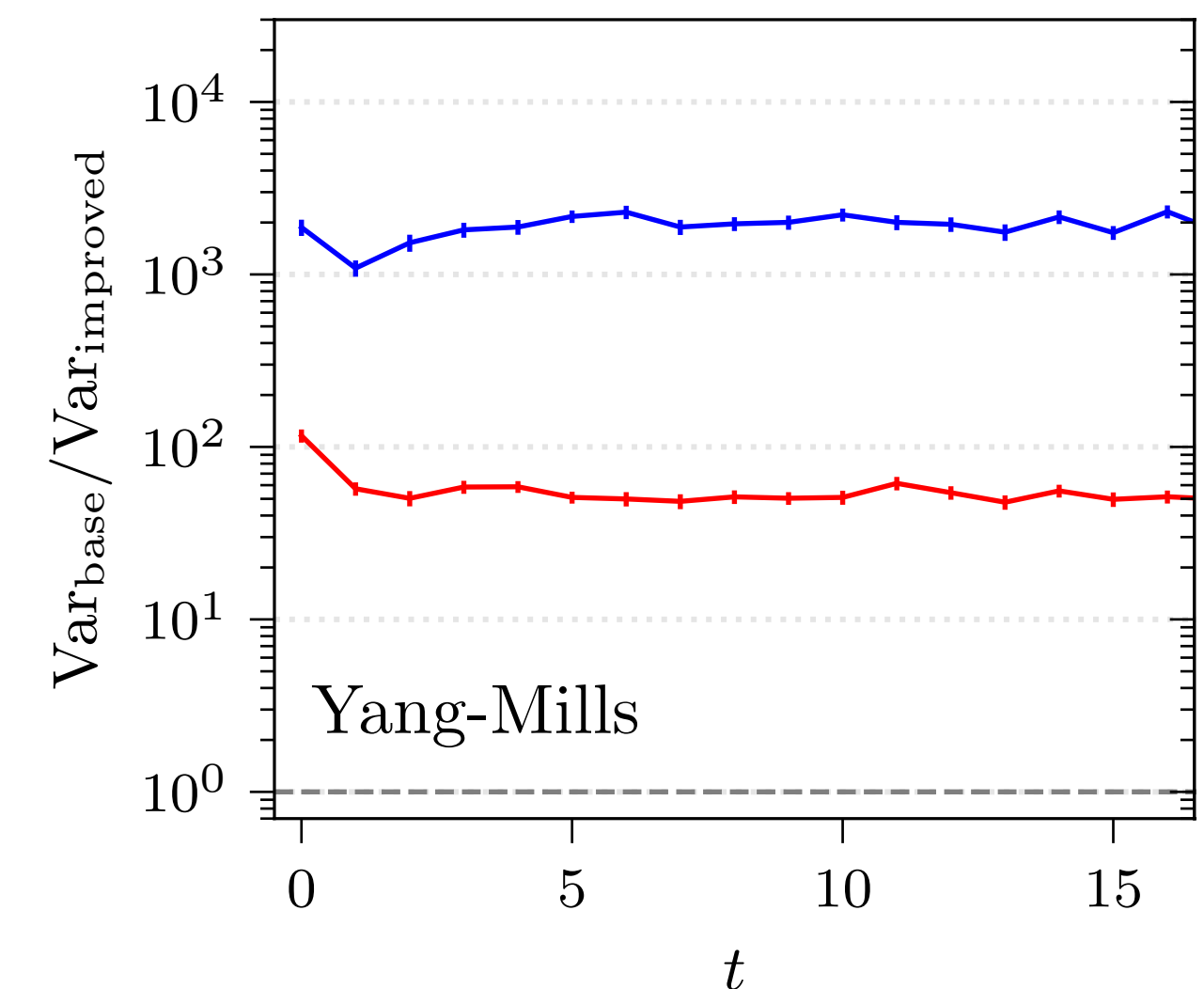
✓ Promising preliminary results in meson correlators: σ meson

○ On-going training for glueball correlators in QCD with $M_\pi \sim 170$ MeV

► Already achieved variance reduction: >100

○ Next steps: more observables and state-of-the-art QCD scales!

► More complicated observables: smeared Wilson loops, other quark bilinears

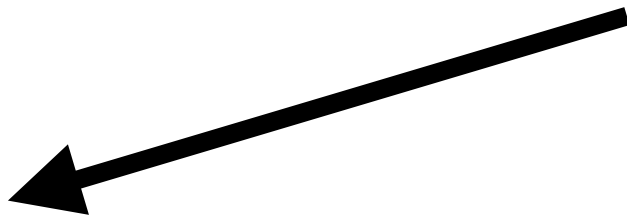


Thanks!

Back-up slides

Timeslice operators

- Consider an interpolating operator on a fixed time slice $\mathcal{O}(t) = \sum_{\vec{x}} \mathcal{O}(\vec{x}, t)$
- Action includes operator at a fixed timeslice:

$$S_{\lambda} = S_{\text{QCD}} + \lambda \mathcal{O}(t = 0)$$


insertion
at a fixed t

Timeslice operators

- Consider an interpolating operator on a fixed time slice $\mathcal{O}(t) = \sum_{\vec{x}} \mathcal{O}(\vec{x}, t)$

- Action includes operator at a fixed timeslice:

$$S_{\lambda} = S_{\text{QCD}} + \lambda \mathcal{O}(t = 0)$$

insertion
at a fixed t

- Residual layers need to be modified to add time-dependent coefficients

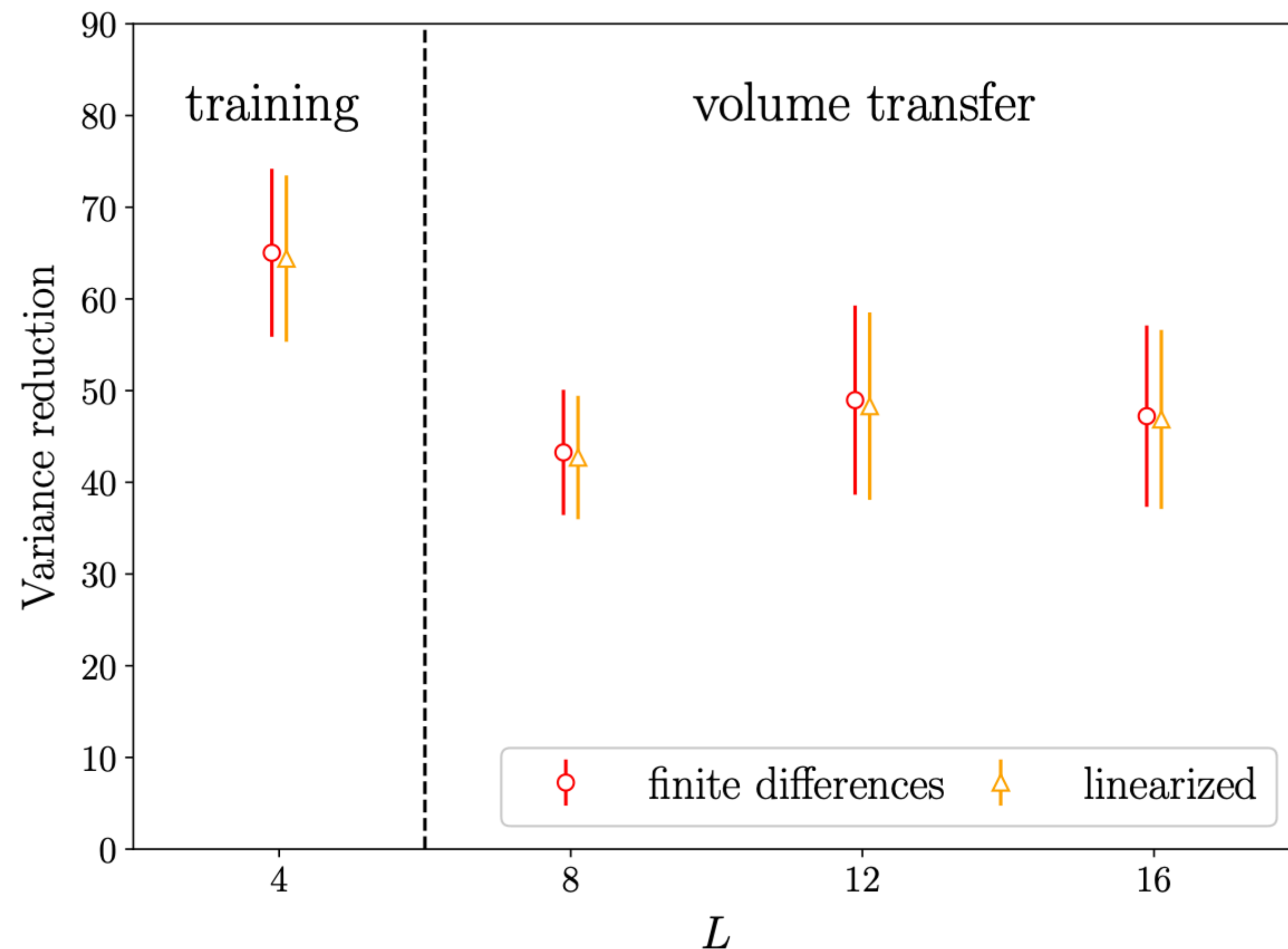
$$U'_{\mu}(\vec{x}, t) = e^{F(U|t, \mu)} U_{\mu}(\vec{x}, t)$$

"Force" in layers depends
on Euclidean time

- Volume transfer of models needs to account for broken symmetry

- ▶ Train at target temporal extent, and transfer only in spatial directions, e.g. $4^3 \times 32$
- ▶ No noticeable degradation with volume transfer [Abbott et al, 2603.02984]

Volume transfer & linearize



$$V = \{1 + \lambda F(U)\}U + O(\lambda^2) = e^{\lambda F(U)}U + O(\lambda^2)$$
$$\partial_\lambda w_\lambda = \nabla \cdot F - F \cdot \nabla S_0 + \mathcal{O} - \langle \mathcal{O} \rangle$$

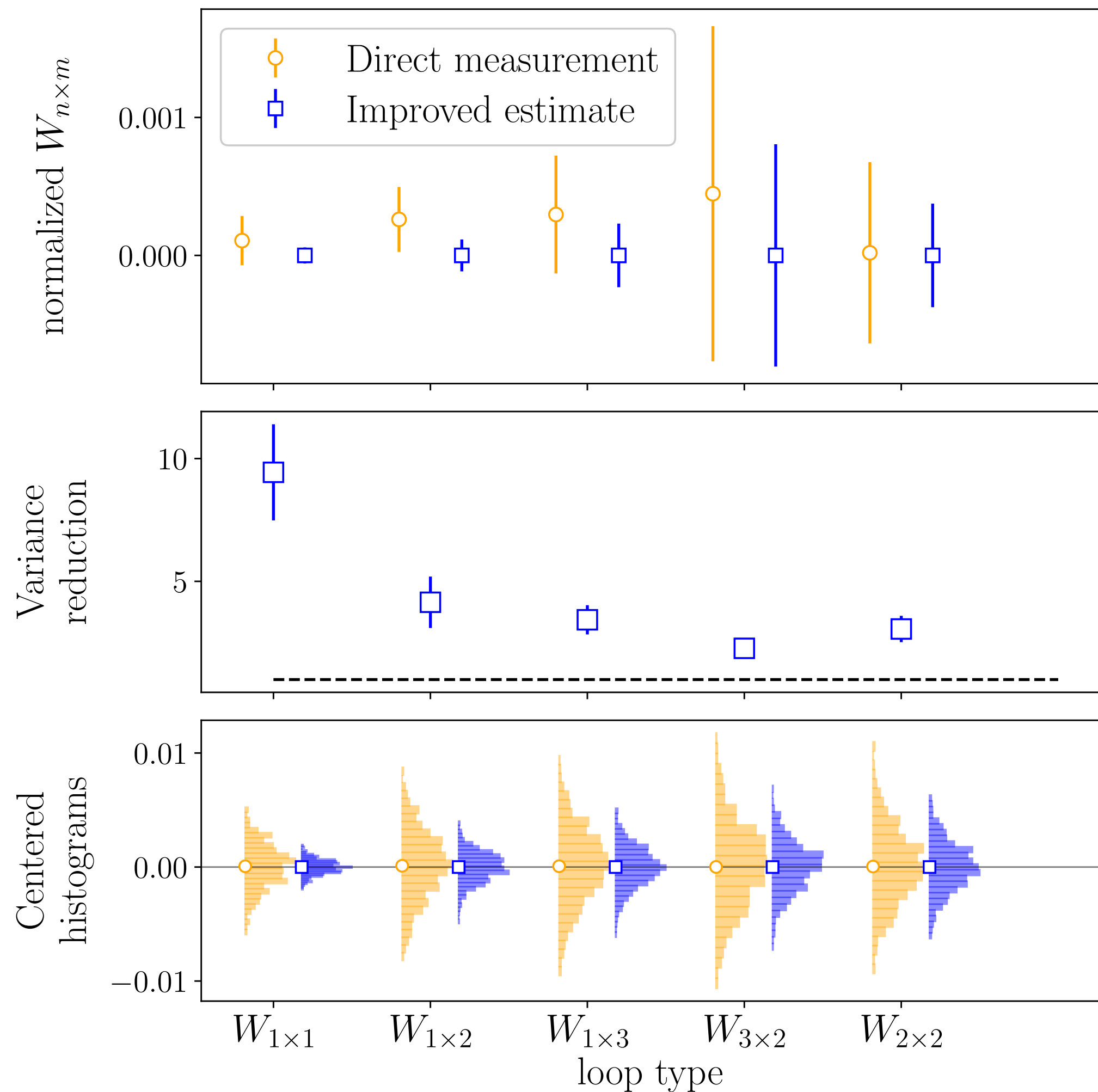
Other Wilson Loops

✓ Variance reduction works for larger loops

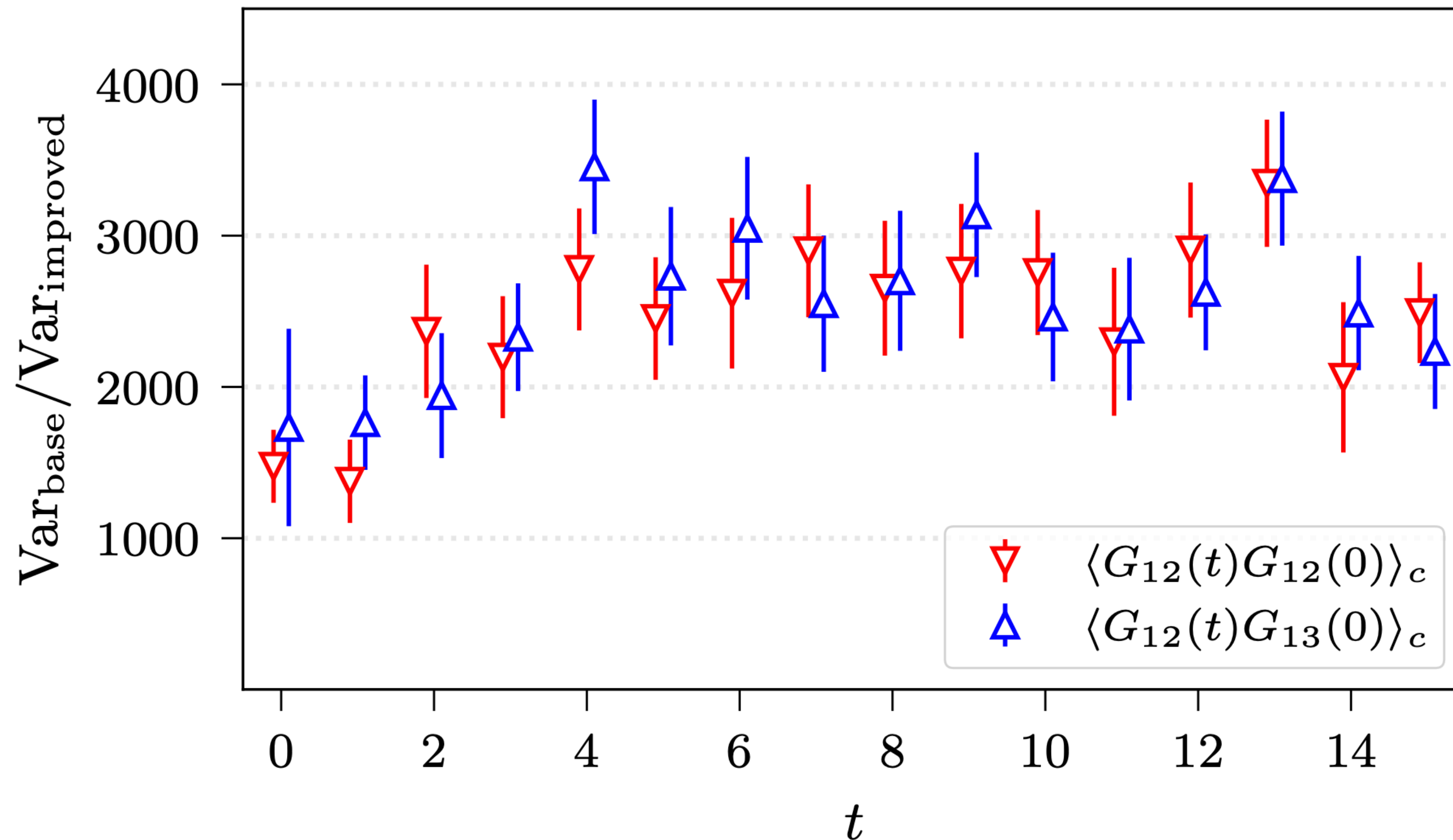
○ However, the gain is reduced

► Larger loops are more nonlocal

► Need different architectures?



Using symmetry



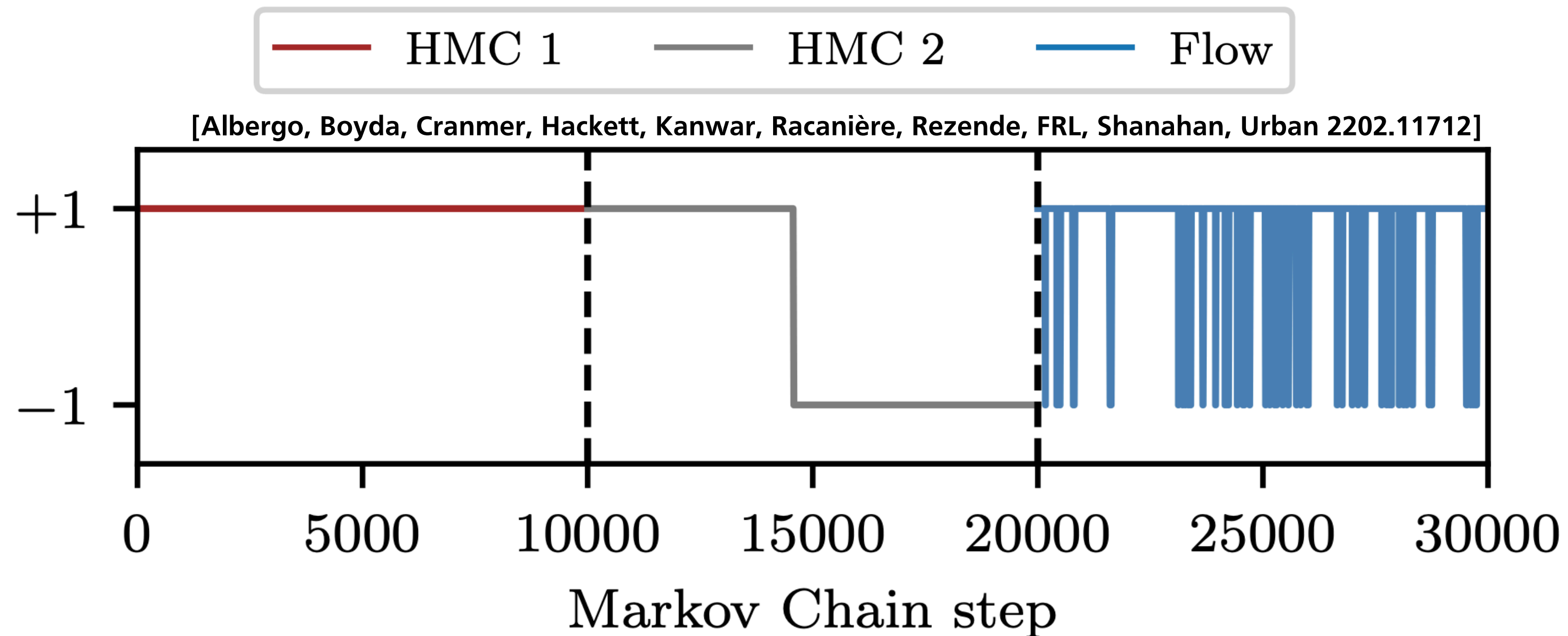
Flows for sampling

- Original LQCD applications of flows aimed to tackle Critical Slowing Down/Topology Freezing

[Albergo, Kanwar, Shanahan 1904.12072]

- Promising results in low-dimensional theories (Schwinger model)

Sign of the
fermion determinant
→ b



Our recent works

(1+1)d real scalar field theory

[\[Albergo, Kanwar, Shanahan 1904.12072\]](#)

[\[Hackett, Hsieh, Albergo, Boyda, JW Chen, KF Chen, Cranmer, Kanwar, Shanahan 2107.00734\]](#)

(1+1)d Abelian gauge theory

[\[Kanwar, Albergo, Boyda, Cranmer, Hackett, Racanière, Rezende, Shanahan 2003.06413\]](#)

(1+1)d non-Abelian gauge theory

[\[Kanwar, Albergo, Boyda, Cranmer, Hackett, Racanière, Rezende, Shanahan 2003.06413\]](#)

(1+1)d Yukawa model

i.e. real scalar field theory + fermions

[\[Albergo, Kanwar, Racanière, Rezende, Urban, Boyda, Cranmer, Hackett, Shanahan 2106.05934\]](#)

Schwinger model i.e. (1+1)d QED

[\[Albergo, Boyda, Cranmer, Hackett, Kanwar, Racanière, Rezende, FRL, Shanahan, Urban 2202.11712\]](#)

2D fermionic gauge theories with pseudofermions

[\[Abbott, Albergo, Boyda, Cranmer, Hackett, Kanwar, Racanière, Rezende, FRL, Shanahan, Tian, Urban 2207.08945\]](#)

QCD/SU(3) in the strong-coupling region

[\[Abbott et al, 2208.03832\]](#) [\[Abbott et al, 2305.02402\]](#)

...

Developments are still needed for large-scale QCD

Flows in 4D theories

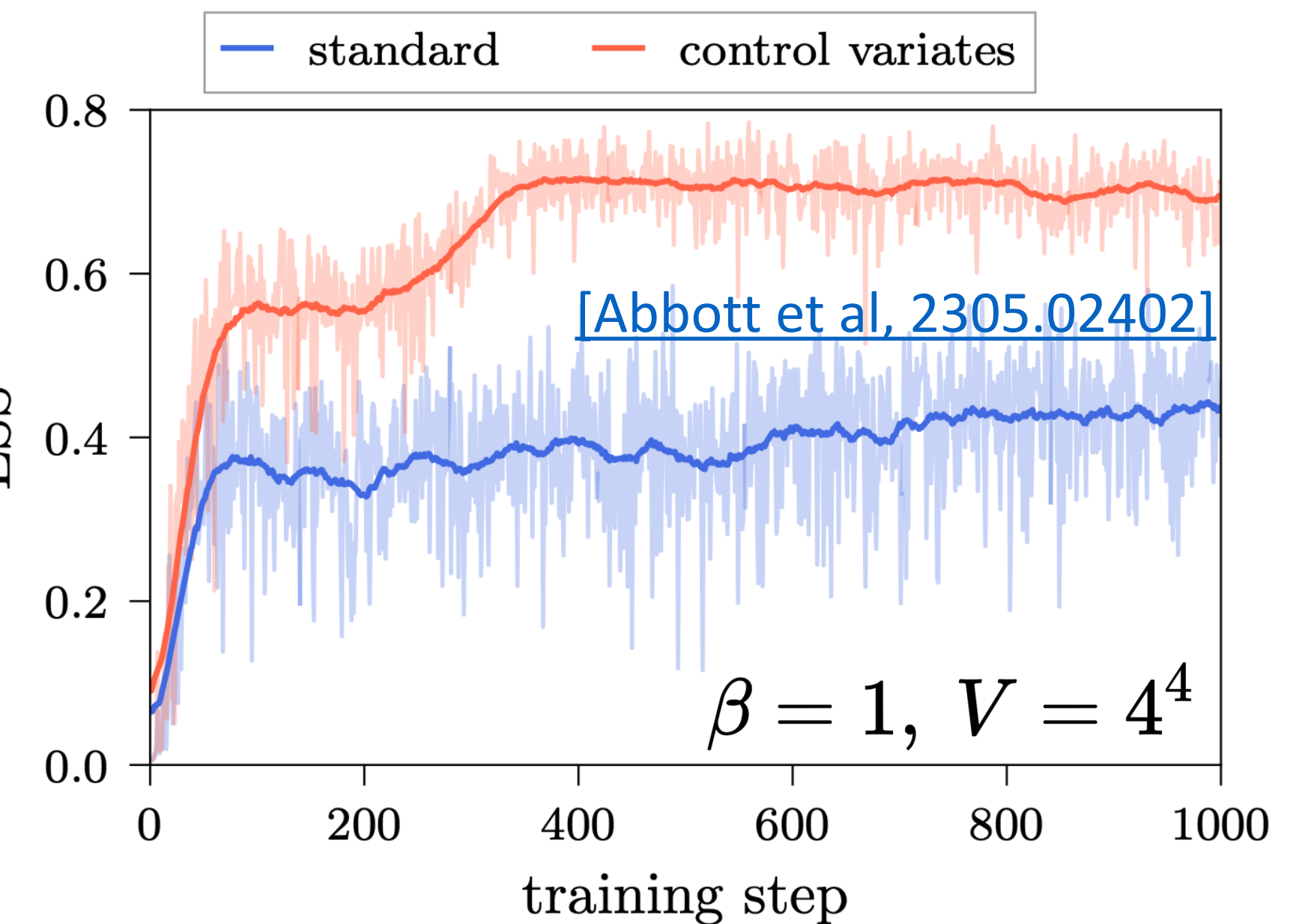
○ Suitable architectures for 4D gauge theories.

- ▶ However, direct sampling remains hard
- ▶ Need very high-quality models to reach large volumes
(Naive volume scaling is exponential)

model
quality
metric

$$\text{ESS} = \frac{\langle w \rangle^2}{\langle w^2 \rangle}$$

ESS



Flows in 4D theories

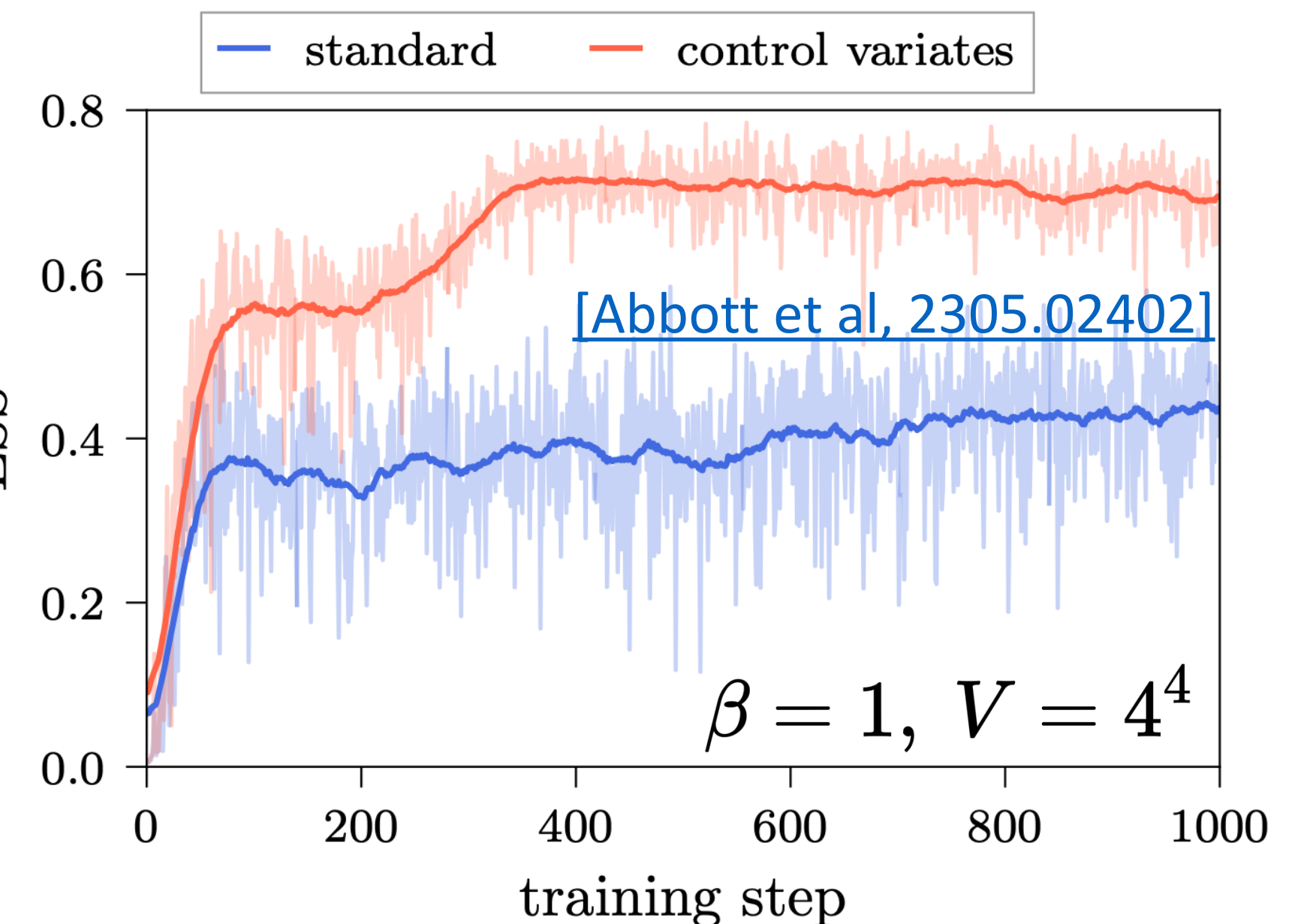
○ Suitable architectures for 4D gauge theories.

- ▶ However, direct sampling remains hard
- ▶ Need very high-quality models to reach large volumes
(Naive volume scaling is exponential)

model
quality
metric

$$\text{ESS} = \frac{\langle w \rangle^2}{\langle w^2 \rangle}$$

ESS



○ Instead, explore applications with smaller gap between theories: $\text{QCD} \rightarrow \text{QCD} + \lambda \mathcal{O}$

Connection to gradient flow

- Gradient flow minimizes the action on a gauge configuration by solving the differential equation
[M. Lüscher, arXiv:1006.4518]

$$\dot{B}_\mu = -\frac{\delta S(U)}{dB_\mu} \longrightarrow U_\mu^{(i+1)} = e^{\epsilon F(U^{(i)})} U_\mu^{(i)}$$

solve numerically with infinitesimal steps

Connection to gradient flow

- Gradient flow minimizes the action on a gauge configuration by solving the differential equation
[M. Lüscher, arXiv:1006.4518]

$$\dot{B}_\mu = -\frac{\delta S(U)}{dB_\mu} \longrightarrow U_\mu^{(i+1)} = e^{\epsilon F(U^{(i)})} U_\mu^{(i)}$$

solve numerically with infinitesimal steps

- The residual layers act as a single and finite step of “generalized” gradient flow

$$U'_\mu(x) = e^{F(U)} U_\mu(x)$$
$$F = \epsilon \times \sum_{\mu \neq \nu} P(W_{\mu\nu}^{1 \times 1})$$


Connection to gradient flow

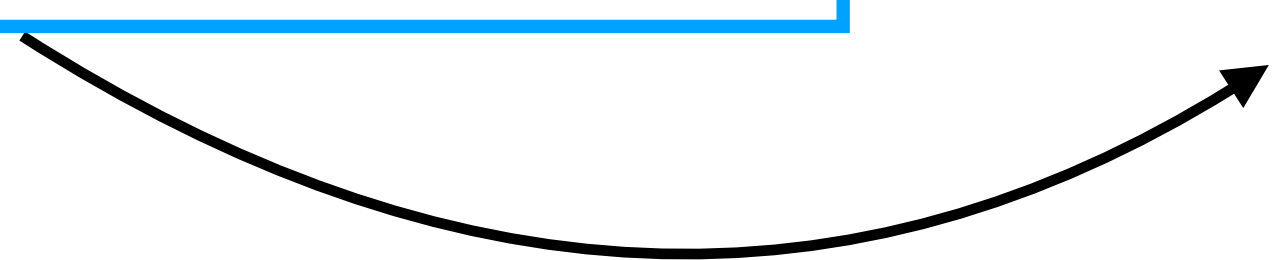
- Gradient flow minimizes the action on a gauge configuration by solving the differential equation
[M. Lüscher, arXiv:1006.4518]

$$\dot{B}_\mu = -\frac{\delta S(U)}{dB_\mu} \longrightarrow U_\mu^{(i+1)} = e^{\epsilon F(U^{(i)})} U_\mu^{(i)}$$

solve numerically with infinitesimal steps

- The residual layers act as a single and finite step of “generalized” gradient flow

$$U'_\mu(x) = e^{F(U)} U_\mu(x)$$

$$F = \epsilon \times \sum_{\mu \neq \nu} P(W_{\mu\nu}^{1 \times 1})$$


- Close to the original trivializing map proposal [M. Lüscher, arXiv:0907.5491]

$$\log J = -\frac{4}{3}\epsilon \sum_{\mu \neq \nu} \text{tr } W_{\mu\nu}^{1 \times 1} + O(\epsilon^2) \longrightarrow \text{Qualitatively induces a change in the gauge coupling} \quad \beta \longrightarrow \beta + \Delta\beta$$