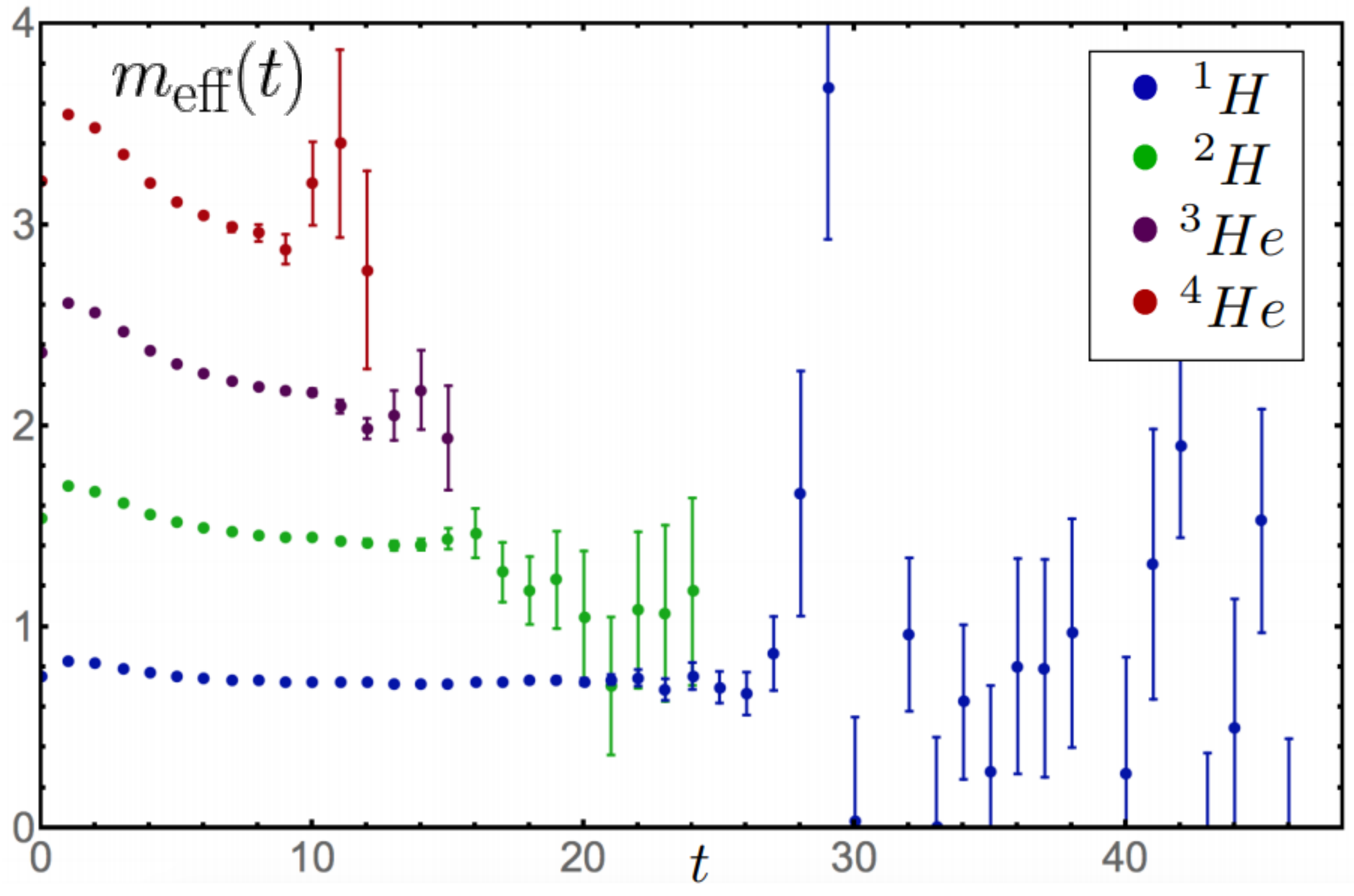
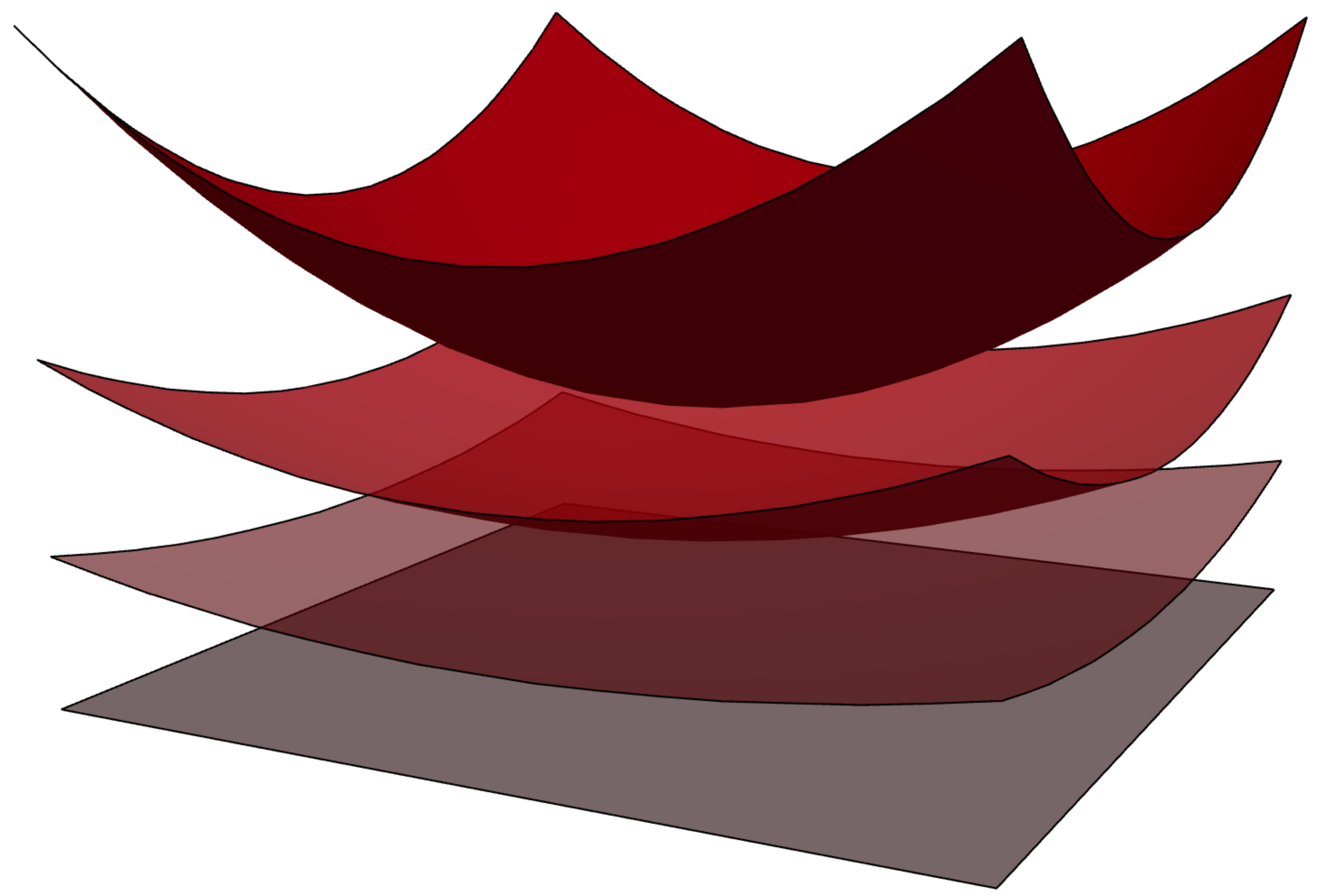
[Detmold, GK, Wagman, Warrington PRD102 (2020) 014514]
[Detmold, GK, Lamm, Wagman, Warrington PRD103 (2021) 094517]

Mitigating signal-to-noise problems using learned contour deformations





Gurtej Kanwar
University of Bern

+ W. Detmold / H. Lamm / Y. Lin / P. Shanahan
M. Wagman / N. Warrington

ML for LFT and Beyond (ECT*) | Jun 26, 2023

A “Signal-to-Noise Problem”

Can define the Signal-to-Noise (StN) ratio of a Monte Carlo estimator for $\langle \mathcal{O} \rangle$

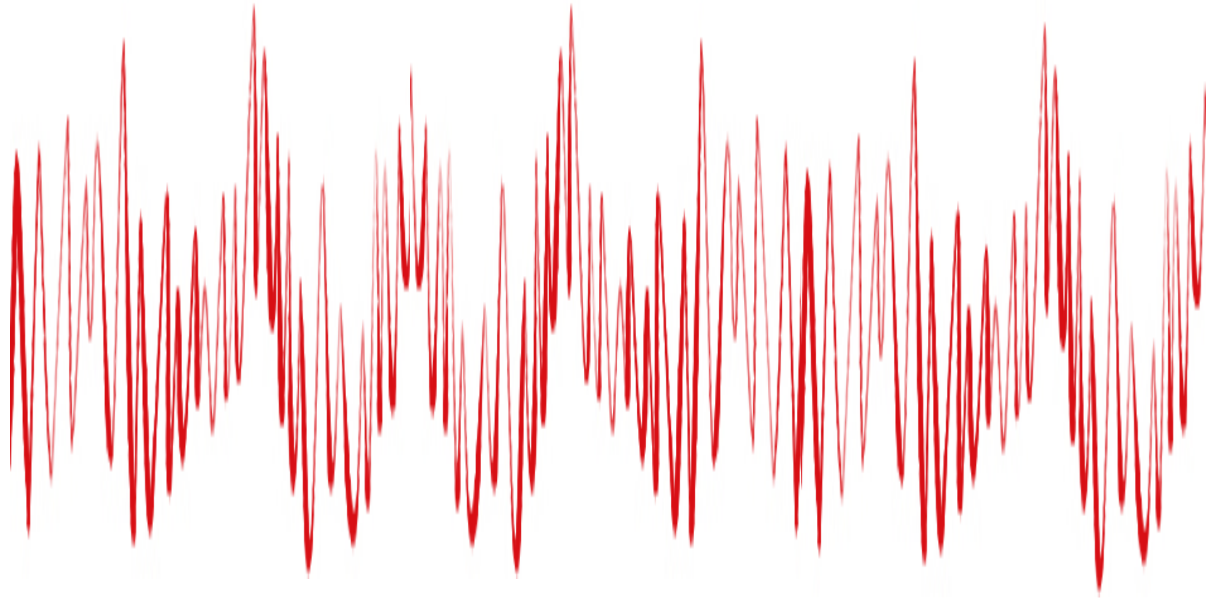
$$\text{StN} \equiv |\langle \mathcal{O} \rangle| / \mathrm{Err}[\mathcal{O}]$$

We have a **StN problem** if fluctuations dominate the mean …

$$\mathrm{Err}[\mathcal{O}] = \sqrt{\text{Var}[\mathcal{O}]/n} \quad \gg \quad |\langle \mathcal{O} \rangle|$$

NOISE SIGNAL

… leading to $\mathrm{StN} \ll 1$



SIGNAL?

NOISE?

Correlation functions

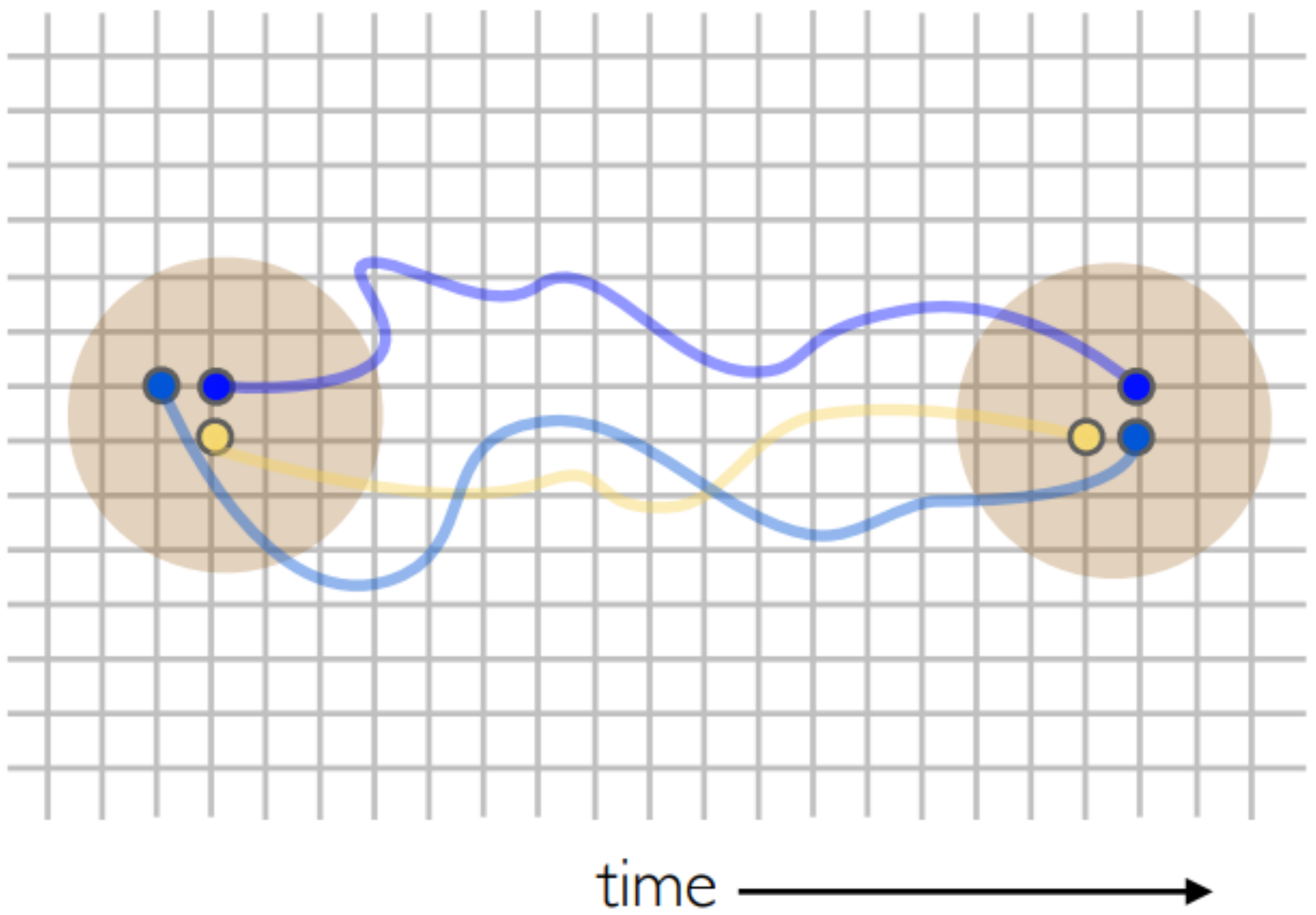
Imaginary-time correlation functions inform us of the spectrum of the theory

$$\langle \mathscr{A}(t)\mathscr{A}^\dagger(0)\rangle = \sum_n Z_n e^{-E_n t} \xrightarrow{t \gg (\Delta E)^{-1}} Z_0\, e^{-E_0 t}$$

Operators designed to create/ annihilate state(s) of interest

Ground state energy (e.g. particle mass)

[Detmold, INT-14-57W]

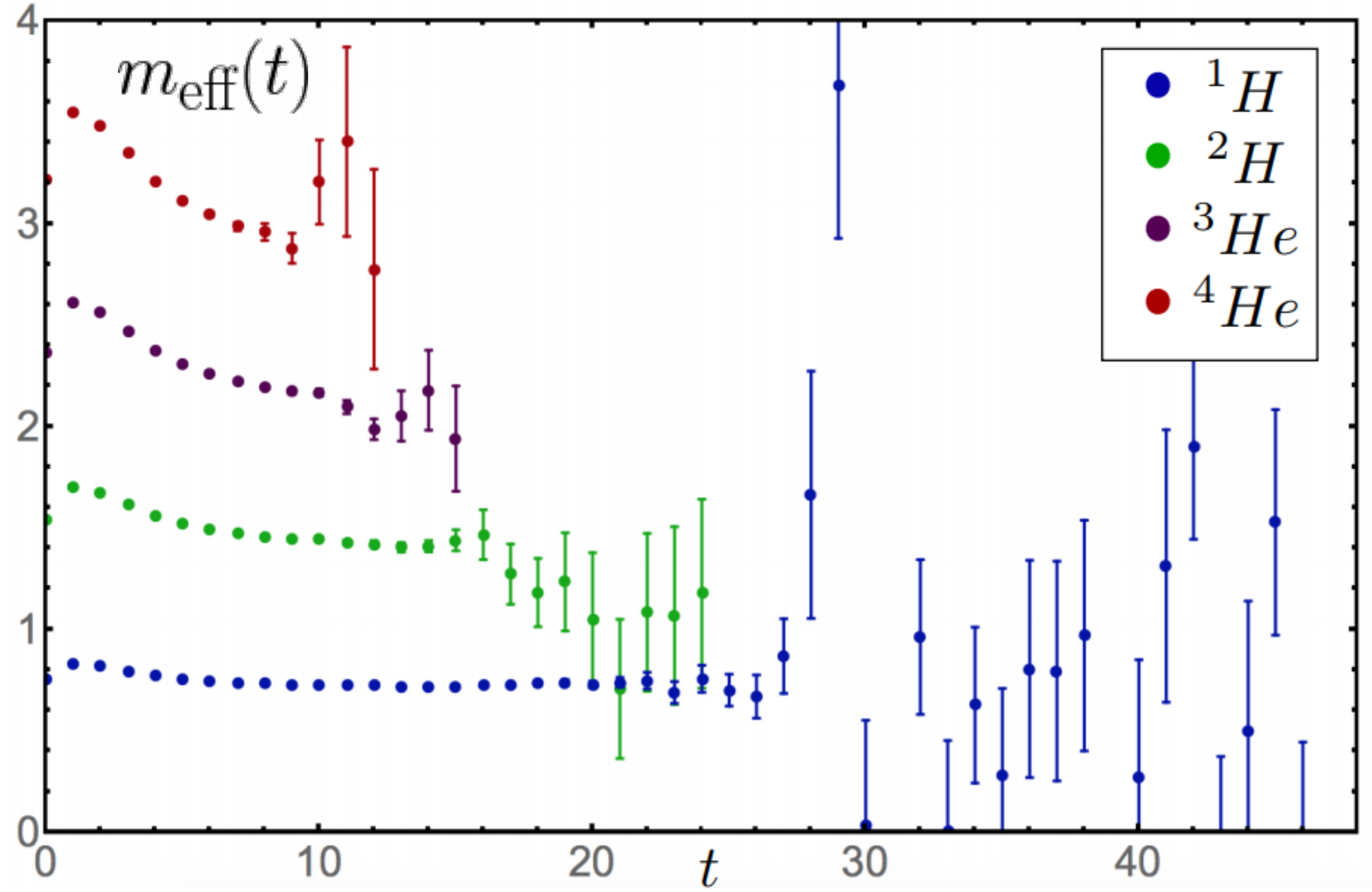


E.g. for the nucleon in lattice QCD

Matrix elements, form factors, etc. accessible via additional operator insertions.

StN problems in correlation functions

[Wagman, Lattice 2018]



Must find **plateau***

- Small t: Excited-state contamination
- Large t: **Noise problem!**

THIS TALK

* Or regime dominated by only a few states

Variance analysis

[Parisi Phys. Rept. 103 (1984) 203]
[Lepage TASI Proceedings (1989) 97]

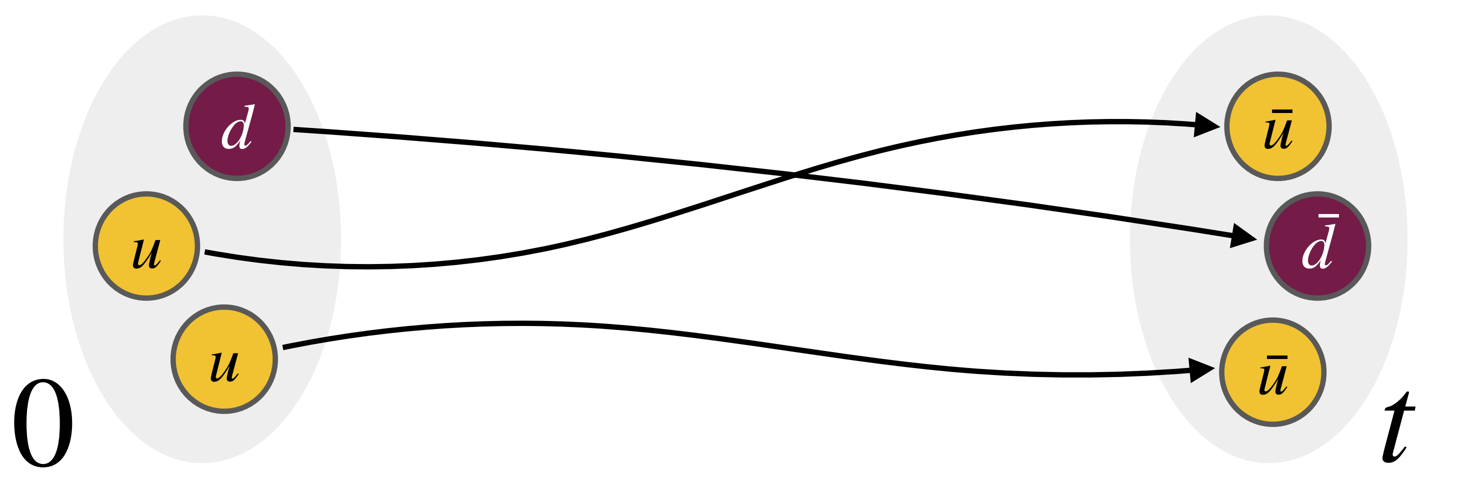
Variance scaling related to physical states

$$\mathrm{Var}[\mathscr{A}(t)\mathscr{A}^\dagger(0)] \sim \left\langle \underline{\mathscr{A}(t)\mathscr{A}^\dagger(t)}\,\underline{\mathscr{A}^\dagger(0)\mathscr{A}(0)} \right\rangle \sim e^{-M_{A^\dagger A}t}$$

Interpret as annihilating / creating a physical state

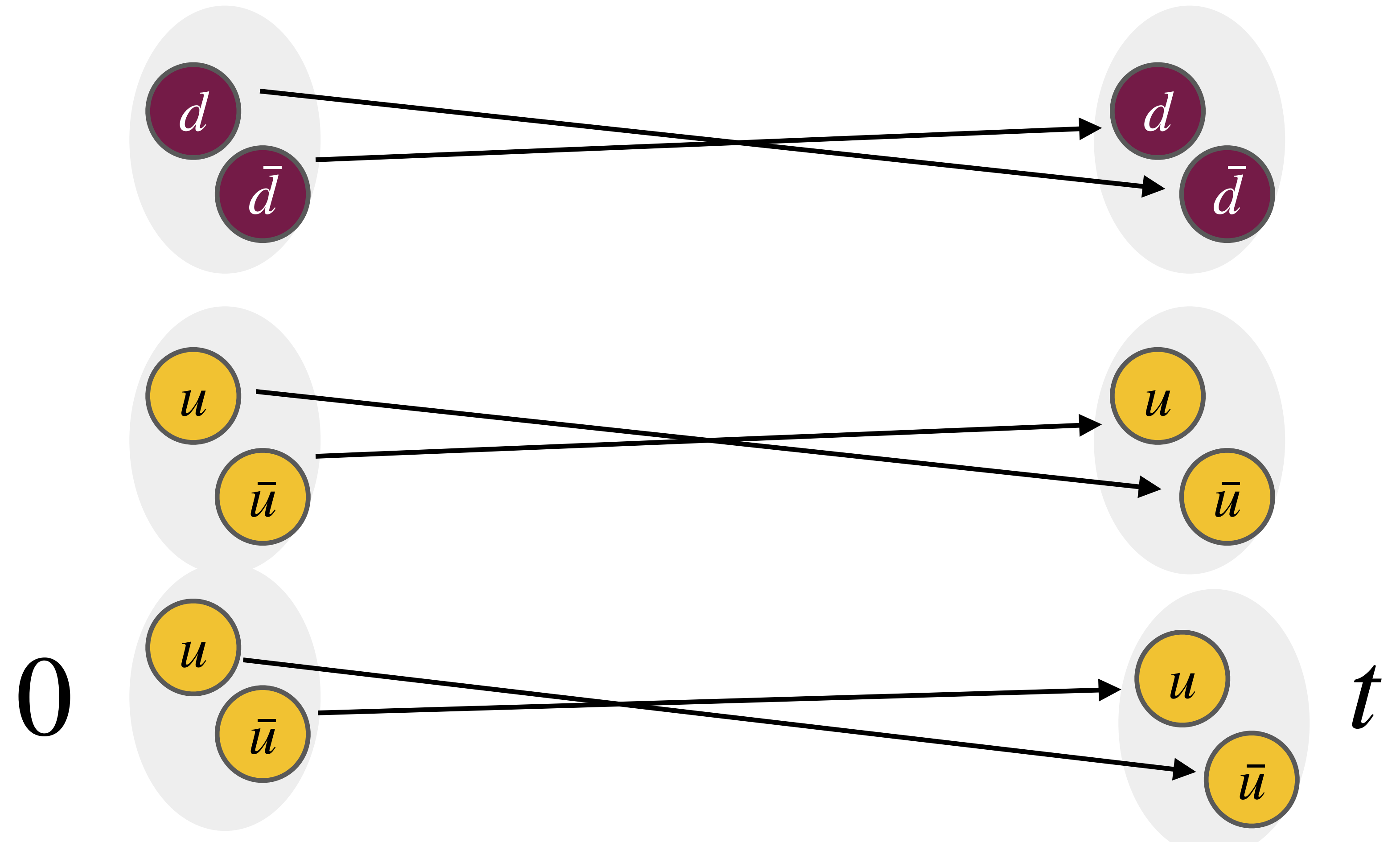
Example: StN of the nucleon given by

Signal:



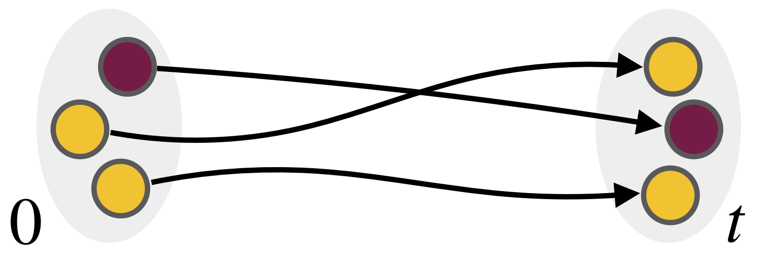
$$\left\langle \mathscr{N}(t)\mathscr{N}^\dagger(0) \right\rangle \sim e^{-M_N t}$$

Variance:

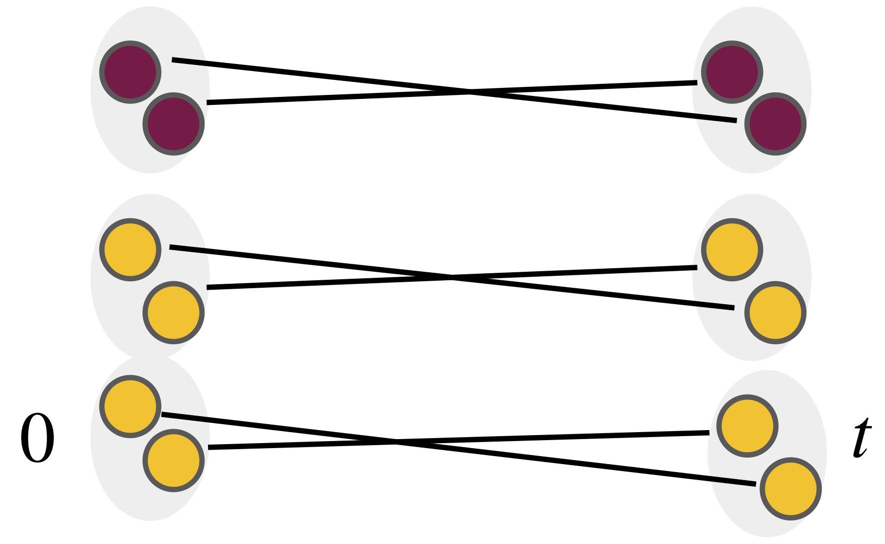


$$\mathrm{Var}[\mathscr{N}(t)\mathscr{N}^\dagger(0)] \sim e^{-3M_\pi t}$$

Variance analysis: nucleon



$$\left\langle \mathcal{N}(t)\mathcal{N}^\dagger(0) \right\rangle \sim e^{-M_N t}$$



$$\text{Var}[\mathcal{N}(t)\mathcal{N}^\dagger(0)] \sim e^{-3M_\pi t}$$

$$\text{StN}[\mathcal{N}(t)\mathcal{N}^\dagger(0)] = \frac{\left\langle \mathcal{N}(t)\mathcal{N}^\dagger(0) \right\rangle}{\sqrt{\text{Var}[\mathcal{N}(t)\mathcal{N}^\dagger(0)]}} \sim e^{-(M_N - 3M_\pi/2)t}$$

Exponentially bad!

Similar problems affect 3-point and higher correlation functions!

Noise problem = sign problem

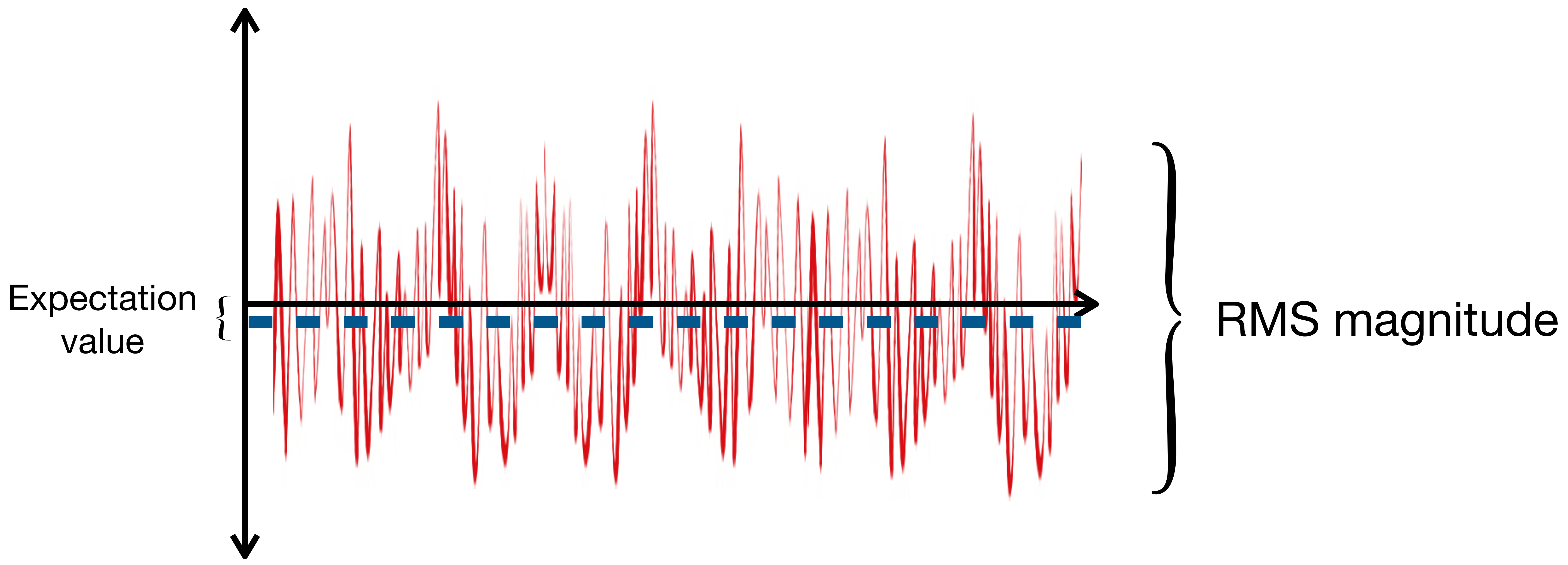
Intuitively a noise problem implies …

$$\mathrm{Err} \sim \sqrt{\langle |C(t)|^2 \rangle} \gg |\langle C(t) \rangle|$$

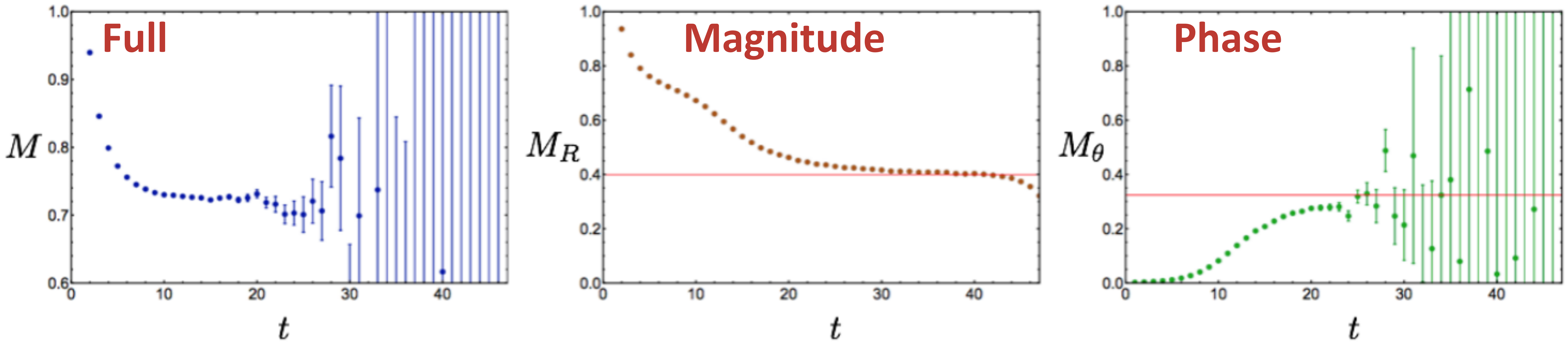
The RMS magnitude …

… is exponentially larger than the expectation value.

… which is a sign problem! (if the magnitude is well concentrated)



Noise problem = sign problem

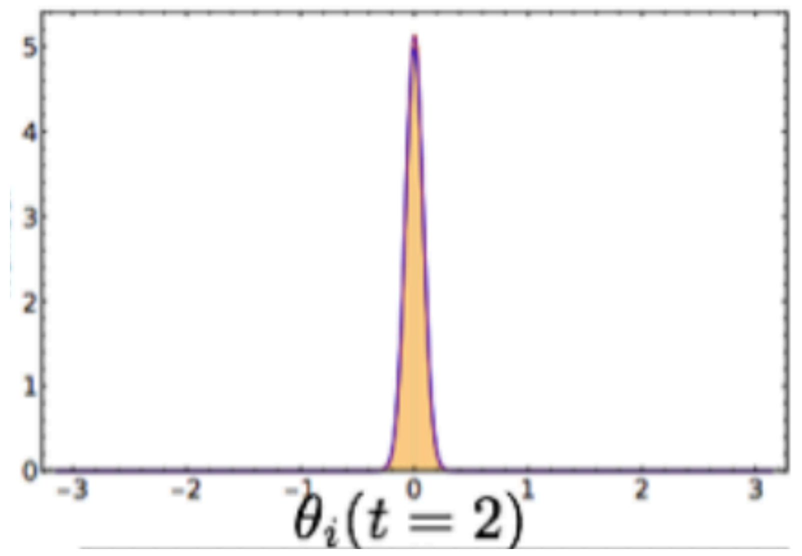


[Wagman & Savage, PRD96 (2017) 114508]

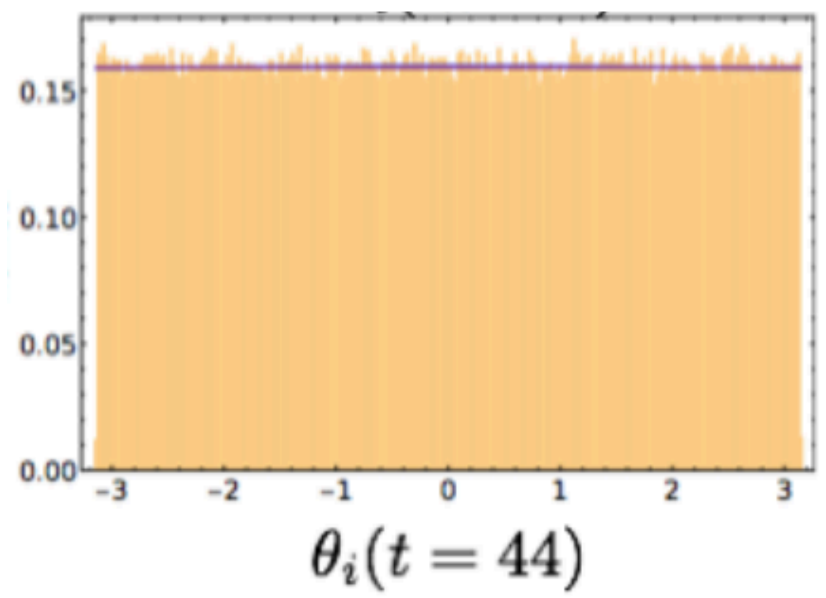
Noise in nucleon effective mass carried by complex phase

[Wagman 1711.00062]

Distribution of complex phases ~uniform at large baryon correlator separations.



vs.

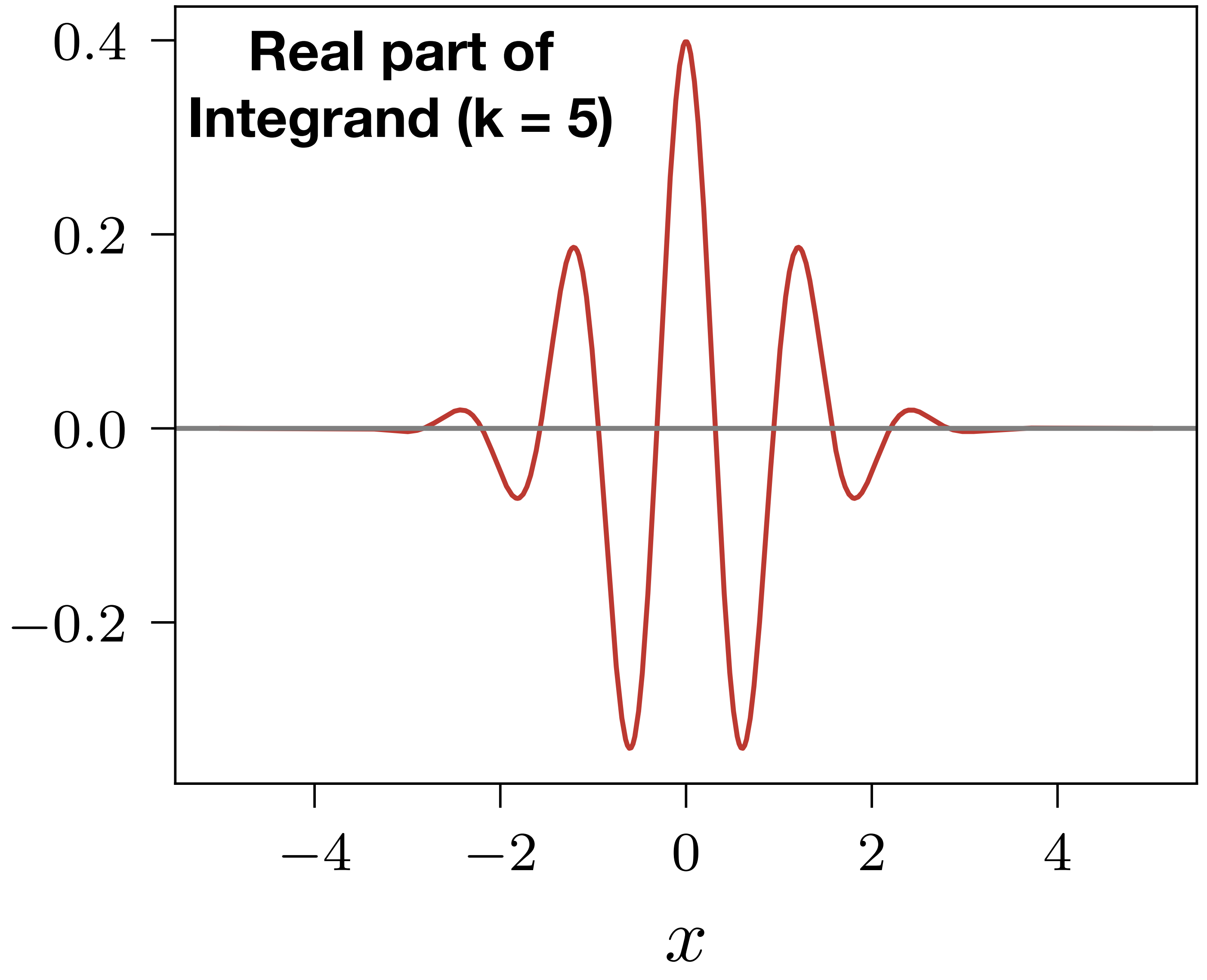


Case study: Gaussian StN problem

Toy example: Simple "observable" in a Gaussian "theory"

$$\left\langle e^{ikx} \right\rangle = \frac{1}{Z} \int_{-\infty}^{\infty} dx [e^{ikx}] e^{-x^2/2}$$

Monte Carlo approach: Sample from distribution $p(x) = e^{-x^2/2}/Z$



Signal: $\left\langle e^{ikx} \right\rangle \sim$ exp. small in k^2

Variance: $\mathrm{Var}[e^{ikx}] \sim \left\langle |e^{ikx}|^2 \right\rangle = 1$

StN and sign problem!

Case study: Gaussian StN problem

Deformation approach:

$$\left\langle e^{ikx} \right\rangle = \frac{1}{Z} \int_{\mathbb{R}} dx\, [e^{ikx}] e^{-\frac{1}{2}x^2}$$

1. Analytically continue & deform contour

$$= \frac{1}{Z} \int_{\mathbb{R}+ik'} d\tilde{x}\, [e^{ik\tilde{x}}] e^{-\frac{1}{2}\tilde{x}^2}$$

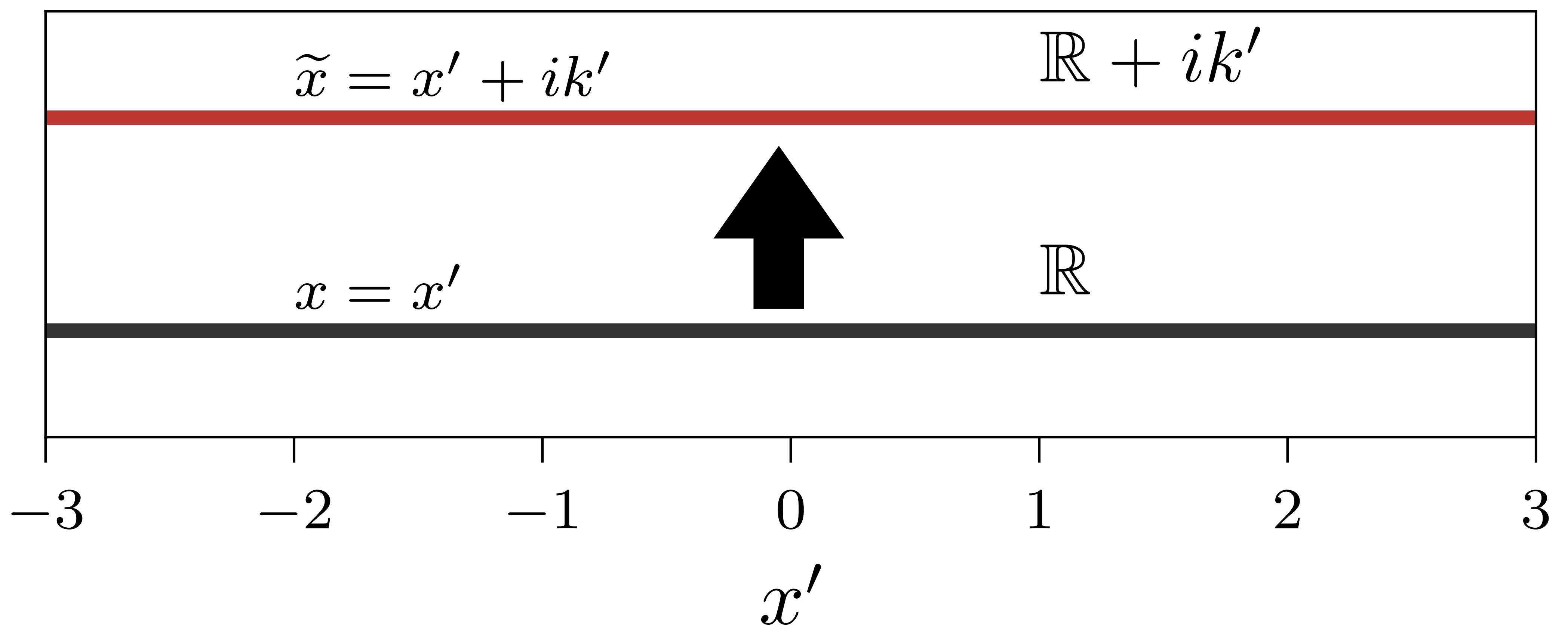
2. Give coordinates to new contour

$$= \frac{1}{Z} \int_{-\infty}^{\infty} dx'\, [e^{ikx'-kk'}] e^{-\frac{1}{2}x'^2 - ix'k' + \frac{1}{2}k'^2}$$

3. New observable w.r.t original MC weights

$$= \left\langle e^{ix'(k-k')} e^{\frac{1}{2}k'^2 - kk'} \right\rangle$$

Result: $\left\langle e^{ikx} \right\rangle = \left\langle e^{ix'(k-k')} e^{\frac{1}{2}k'^2 - kk'} \right\rangle$

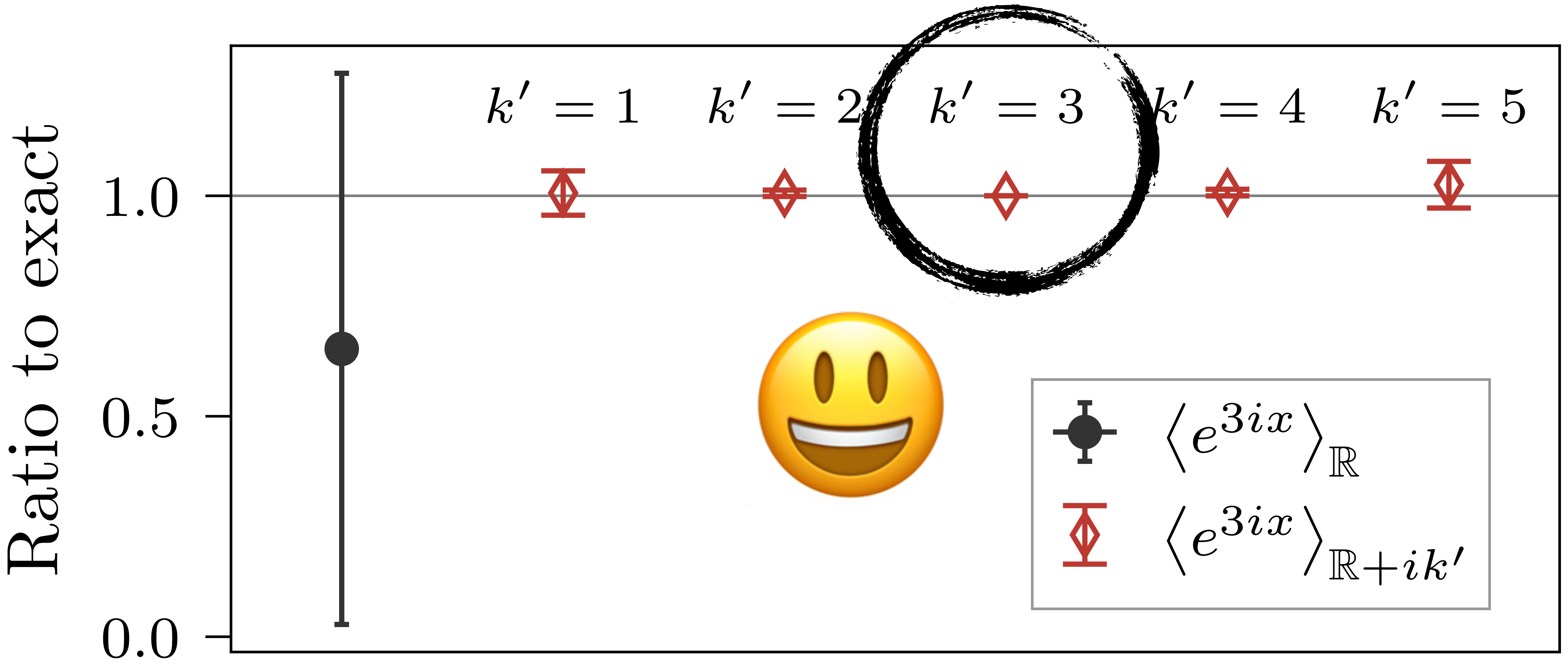


Case study: Gaussian StN problem

Result: $\left\langle e^{ikx} \right\rangle = \left\langle e^{ix'(k-k')} e^{\frac{1}{2}k'^2 - kk'} \right\rangle$

Less severe sign problem: Deformed observable $[e^{ikx'-kk'}]$ has smaller magnitude for $kk' > 0$.

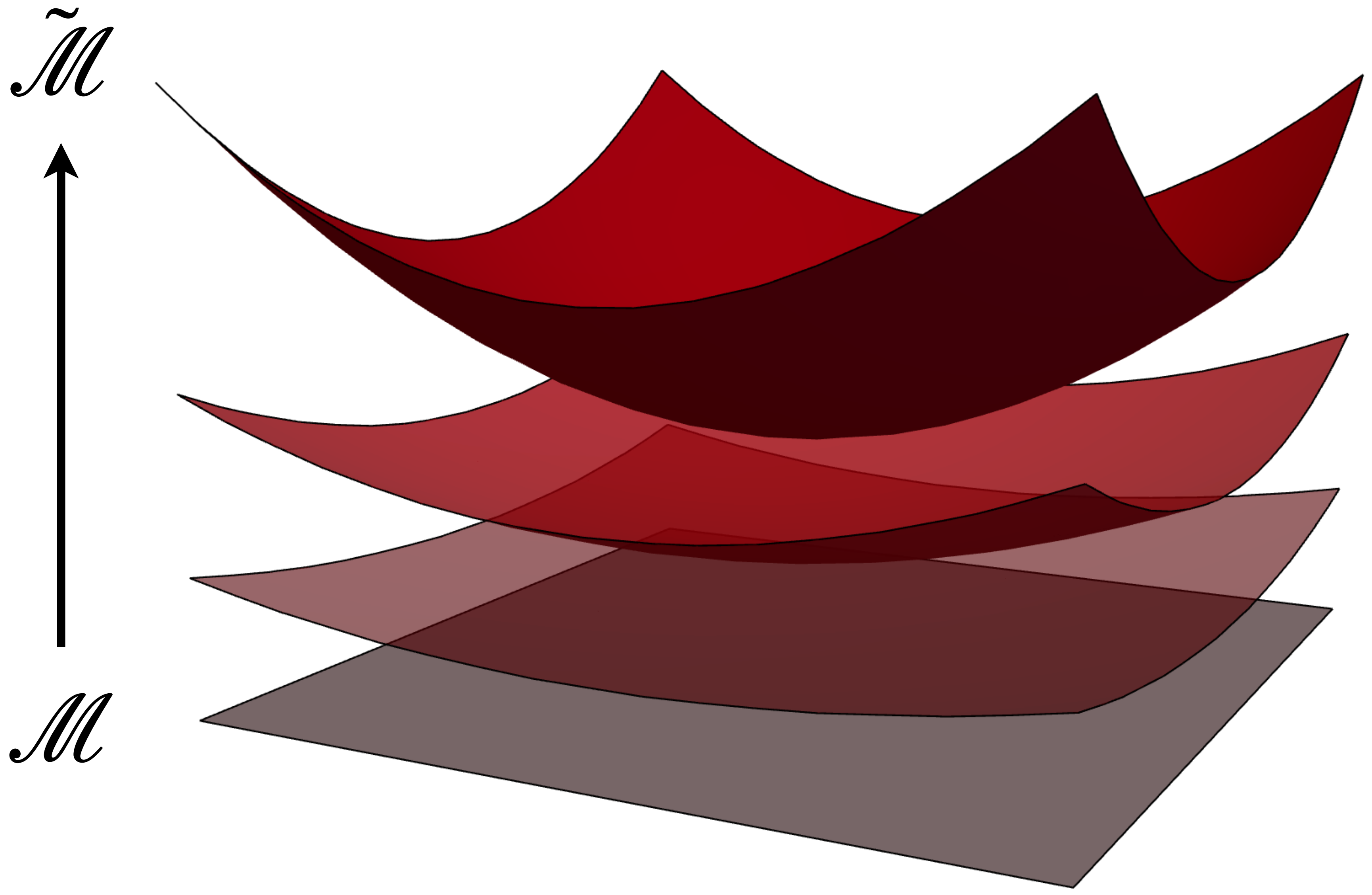
Exactness preserved: Anti-correlated phase fluctuations from the deformed action $e^{-\frac{1}{2}x'^2 - ix'k' + \frac{1}{2}k'^2}$.



Deforming the path integral

Just high-dimensional contour deformation...

$$\langle \mathcal{O} \rangle = \int_{\mathcal{M}} \mathcal{D}U e^{-S(U)} \mathcal{O}(U) = \int_{\widetilde{\mathcal{M}}} \mathcal{D}\widetilde{U} e^{-S(\widetilde{U})} \mathcal{O}(\widetilde{U})$$



Many works applying this to “extensive” sign problems (e.g. non-zero density, real time)

Cristoforetti+ PRD86(074506), PRD88(051501), PRD89(114505)
Aarts PRD88(094501)
Alexandru+ PRD93(014504), JHEP05(053), PRD96(094505), PRD98(054514), PRD98(034506), PRD97(094510), PRL121(191602)
Fujii+ JHEP12(125)
Tanizaki+ NJP18(033002)
Mori+ PTEP2018(023B04), PRD99(014033)
Alexandru+ PRL117(081602), PRD95(114501)
Mou+ JHEP11(135)
...

[Image credit: Neill Warrington]

Analytic continuation & holomorphy

Write Boltzmann weight e^{-S} and observable $\mathcal{O}$ in terms of **real field variables**...

- For $SU(N)$, we use angular parameters $\Omega \equiv (\phi_1, \ldots, \theta_1, \ldots)$, $\phi_i \in [0,2\pi]$ and $\theta_i \in [0,\pi/2]$

  $(N^2 - 1)$ angles

  [Bronzan PRD38 (1988) 1994]

... then **analytically continue**.

- Complexified angular params extends $SU(N) \to SL(N, \mathbb{C})$
- Adjoints should be rewritten $U^\dagger \to U^{-1}$

SU(3) parameterization

$$U(\Omega) = \begin{pmatrix} c_1 c_2 e^{i\phi_1} & s_1 e^{i\phi_3} & c_1 s_2 e^{i\phi_4} \\ s_2 s_3 e^{-i(\phi_4+\phi_5)} - s_1 c_2 c_3 e^{i(\phi_1+\phi_2-\phi_3)} & c_1 c_3 e^{i\phi_2} & -c_2 s_3 e^{-i(\phi_1+\phi_5)} - s_1 s_2 c_3 e^{i(\phi_2-\phi_3+\phi_4)} \\ -s_1 c_2 s_3 e^{i(\phi_1-\phi_3+\phi_5)} - s_2 c_3 e^{-i(\phi_2+\phi_4)} & c_1 s_3 e^{i\phi_5} & c_2 c_3 e^{-i(\phi_1+\phi_2)} - s_1 s_2 s_3 e^{-i(\phi_3-\phi_4-\phi_5)} \end{pmatrix}$$

where $s_i \equiv \sin(\theta_i)$, $c_i \equiv \cos(\theta_i)$.

Path integral deformations for observables

- Deform the path integral **numerator** only:

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int_{\tilde{\mathcal{M}}} e^{-S(\tilde{U})} \mathcal{O}(\tilde{U}) \xrightarrow{\text{manifold coordinates}} \frac{1}{Z} \int_{\mathcal{M}} J(U)\, e^{-S(\tilde{U}(U))} \mathcal{O}(\tilde{U}(U))$$

- Define a new observable $\mathcal{Q}(U)$

$$\mathcal{Q}(U) \equiv e^{-[S_{\text{eff}}(U) - S(U)]} \mathcal{O}(\tilde{U}(U))$$

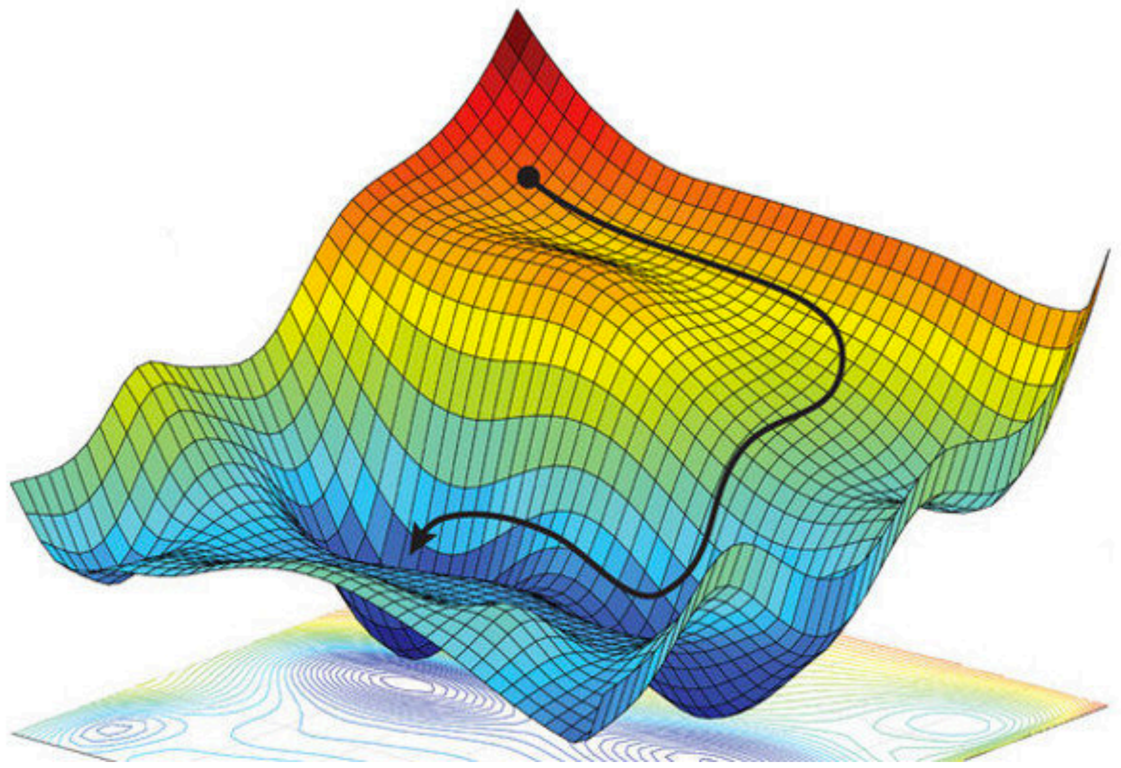
where $S_{\text{eff}}(U) \equiv S(\tilde{U}(U)) - \log J(U)$

- $\mathcal{Q}(U)$ has identical expectation value, but different variance:

$$\langle \mathcal{Q}(U) \rangle = \langle \mathcal{O}(U) \rangle$$
$$\text{Var}[\mathcal{Q}(U)] \neq \text{Var}[\mathcal{O}(U)]$$

Optimizing the variance

Choice of $\tilde{U}(U)$ defines $\mathcal{Q}(U)$ and determines the variance.



[Image credit: 1805.04829]

Parameterize $\tilde{U}(U)$ using ML techniques, then **numerically minimize variance**.

Gradients of variance w.r.t. params defined using **original Monte Carlo ensemble**:

$$\nabla_{\vec{\omega}} \mathrm{Var}[\mathrm{Re}\,\mathcal{Q}] = \langle \nabla_{\vec{\omega}} (\mathrm{Re}\,\mathcal{Q})^2 \rangle = 2 \langle \mathrm{Re}\,\mathcal{Q}\,\mathrm{Re}\,\nabla_{\vec{\omega}} \mathcal{Q} \rangle$$

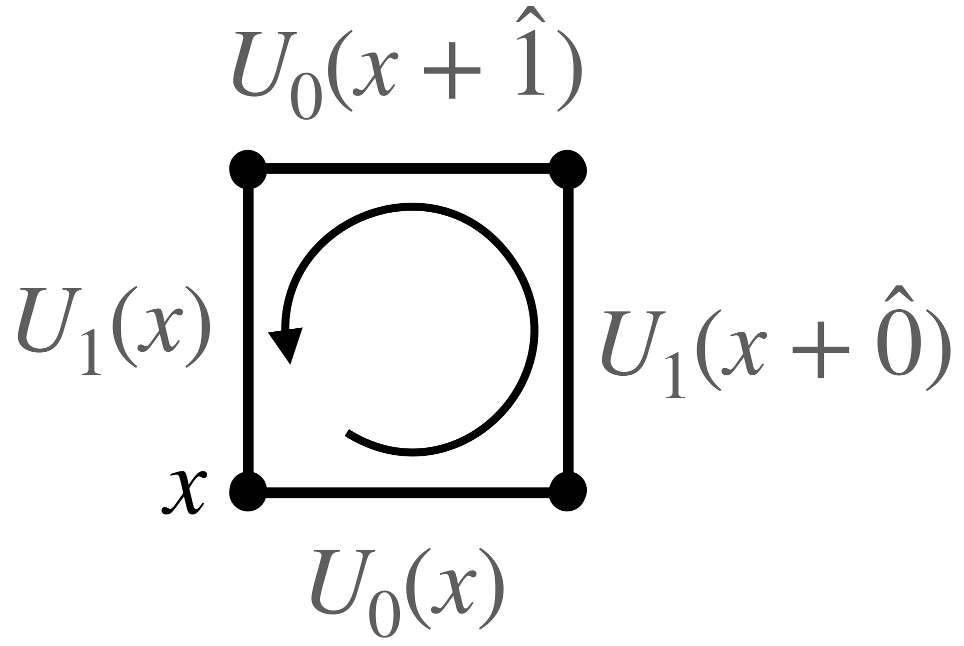
$$= 2 \left\langle (\mathrm{Re}\,\mathcal{Q}) \mathrm{Re} \left(\mathcal{Q} \left[-\nabla_{\vec{\omega}} S_{\mathrm{eff}} + \frac{\nabla_{\vec{\omega}} \mathcal{O}(\tilde{U})}{\mathcal{O}(\tilde{U})} \right] \right) \right\rangle$$

$SU(N)$ pure-gauge theory

Simple action for the theory given in terms of **plaquettes**

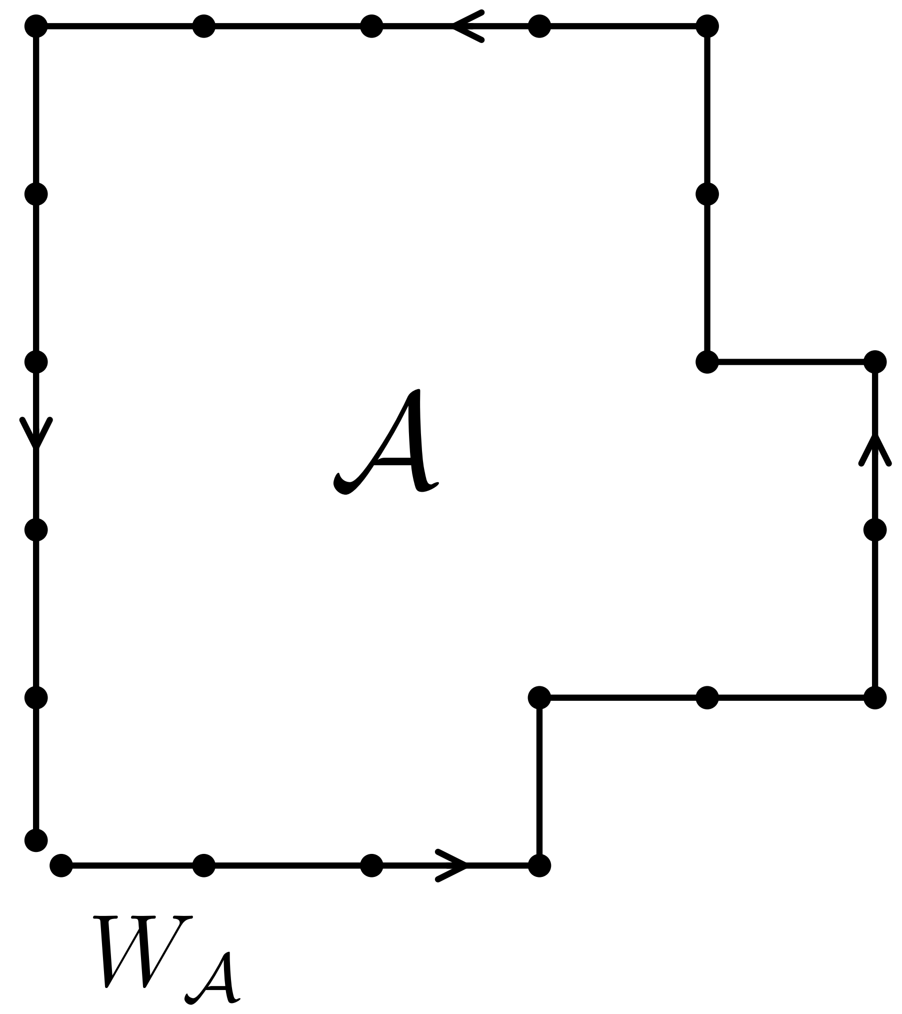
$$S[P] = -\frac{1}{g^2}\sum_x \mathrm{tr}(P_x + P_x^{-1}) \xrightarrow{\text{ctm.}} \frac{1}{2g^2}\mathrm{tr}(F^2)$$

$$P_x = U_0(x)U_1(x+\hat{0})U_0^\dagger(x+\hat{1})U_0^\dagger(x)$$

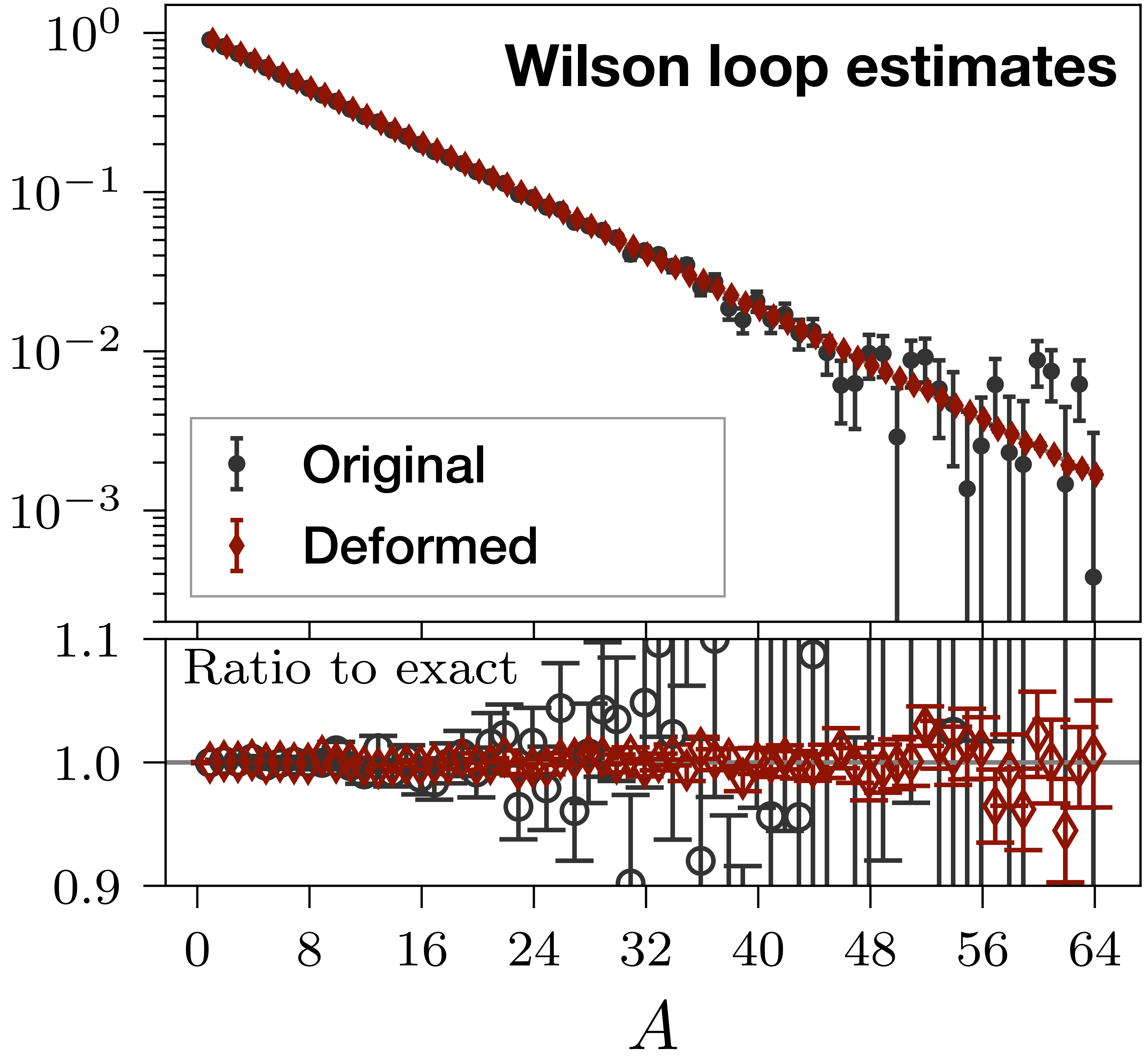


Wilson loop observables $W_{\mathscr{A}}$:

- Generalize plaquettes to arbitrary loops
- Probe of confinement (area vs perimeter law scaling)



1+1D SU(3) Wilson loops

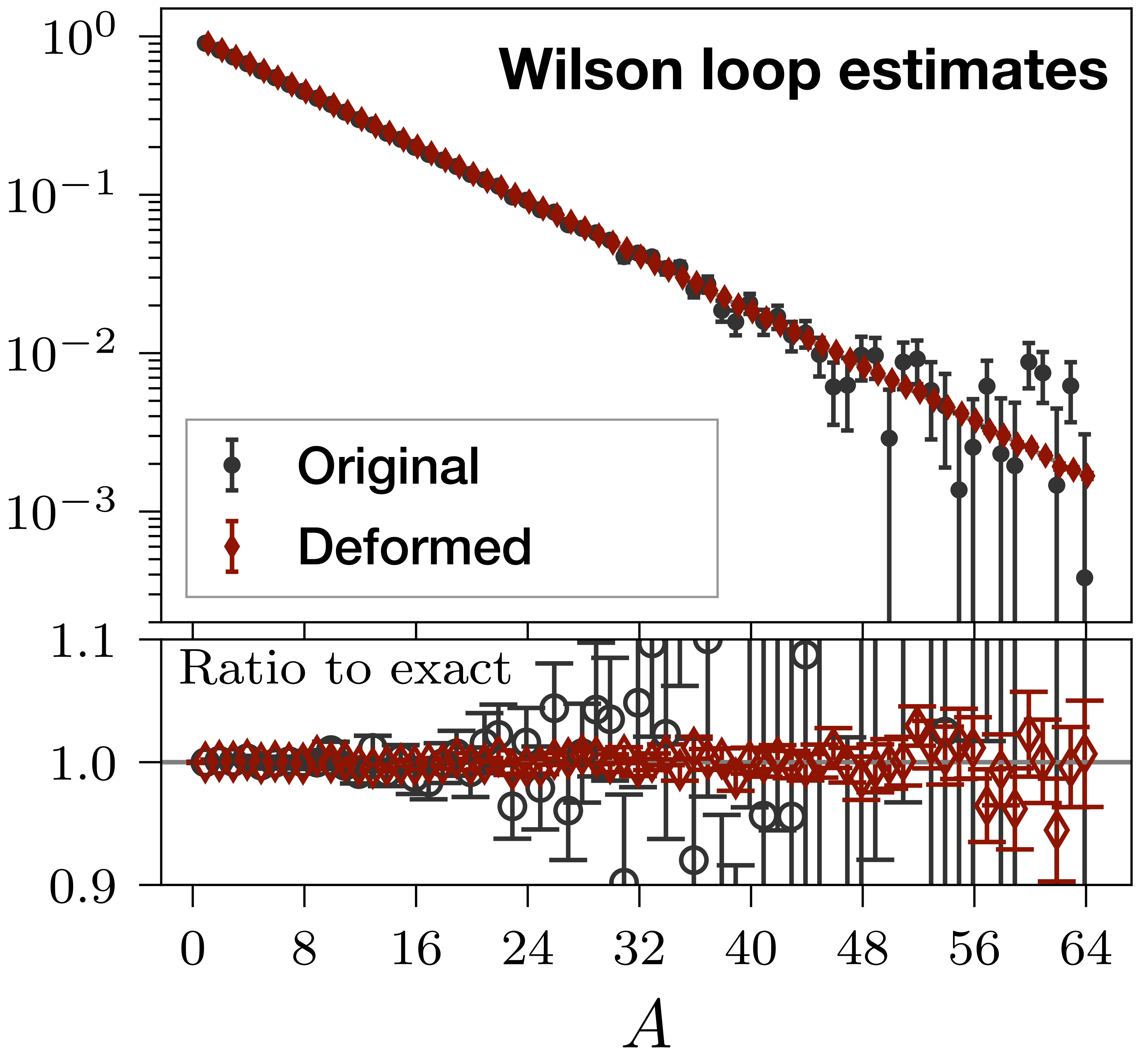


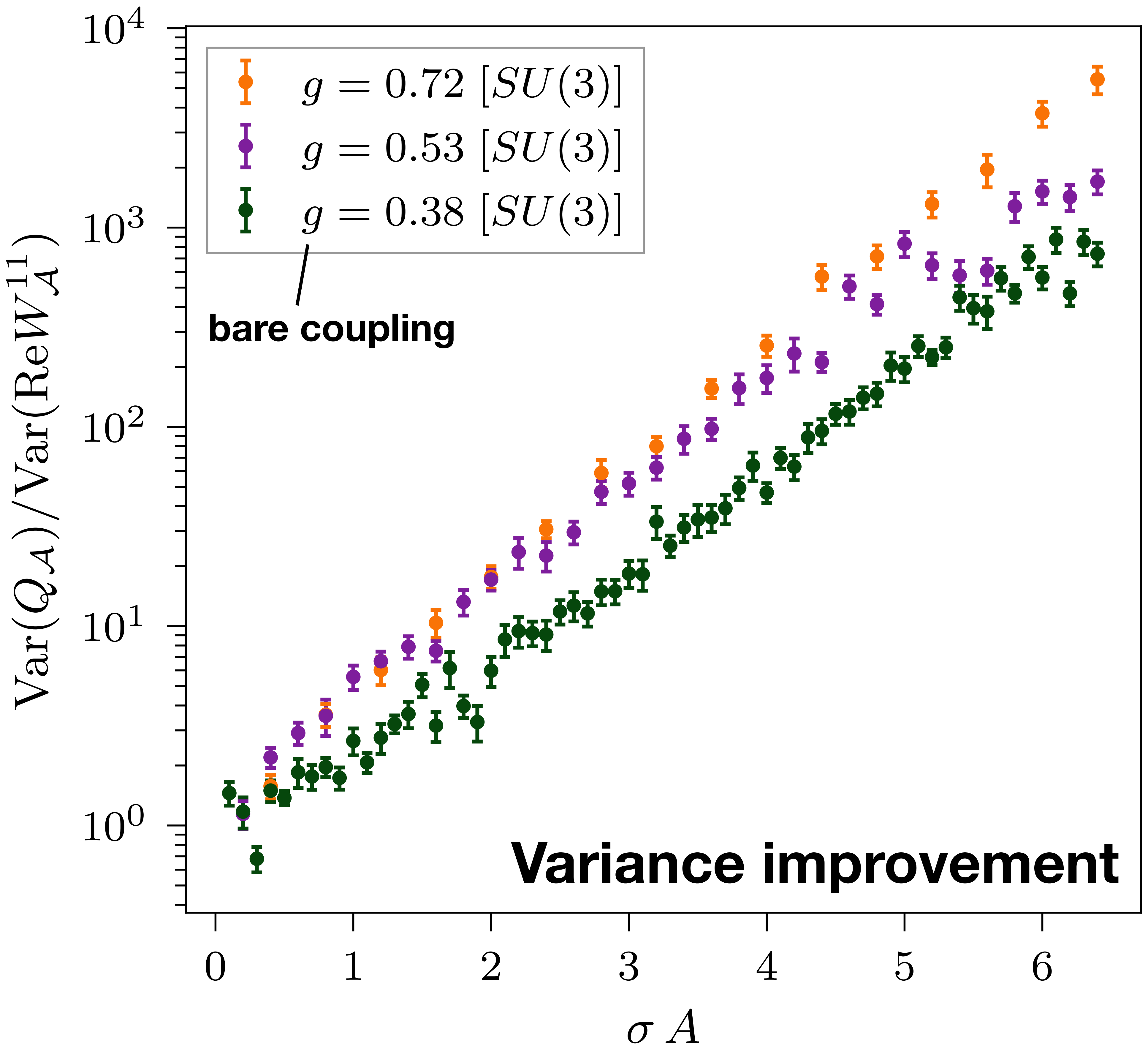
"Constant shift" deformations:

- Deform only periodic angular variables $\phi_i \in [0,2\pi]$
- $\tilde{\phi}_i = \phi_i + i\lambda$
- Field-independent, but spacetime dependent $\Longrightarrow O(V)$ learnable parameters

1+1D SU(3) Wilson loops

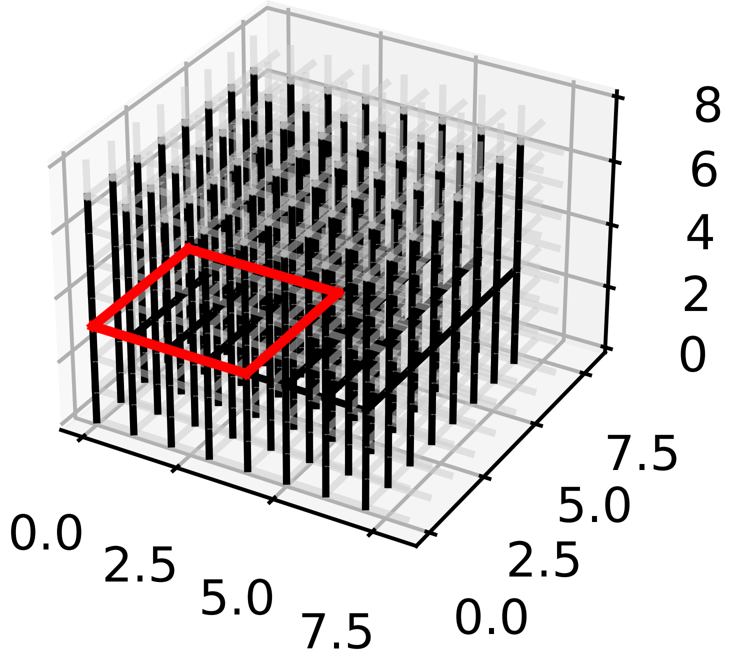
Significant variance reduction at large A for all couplings, no bias.



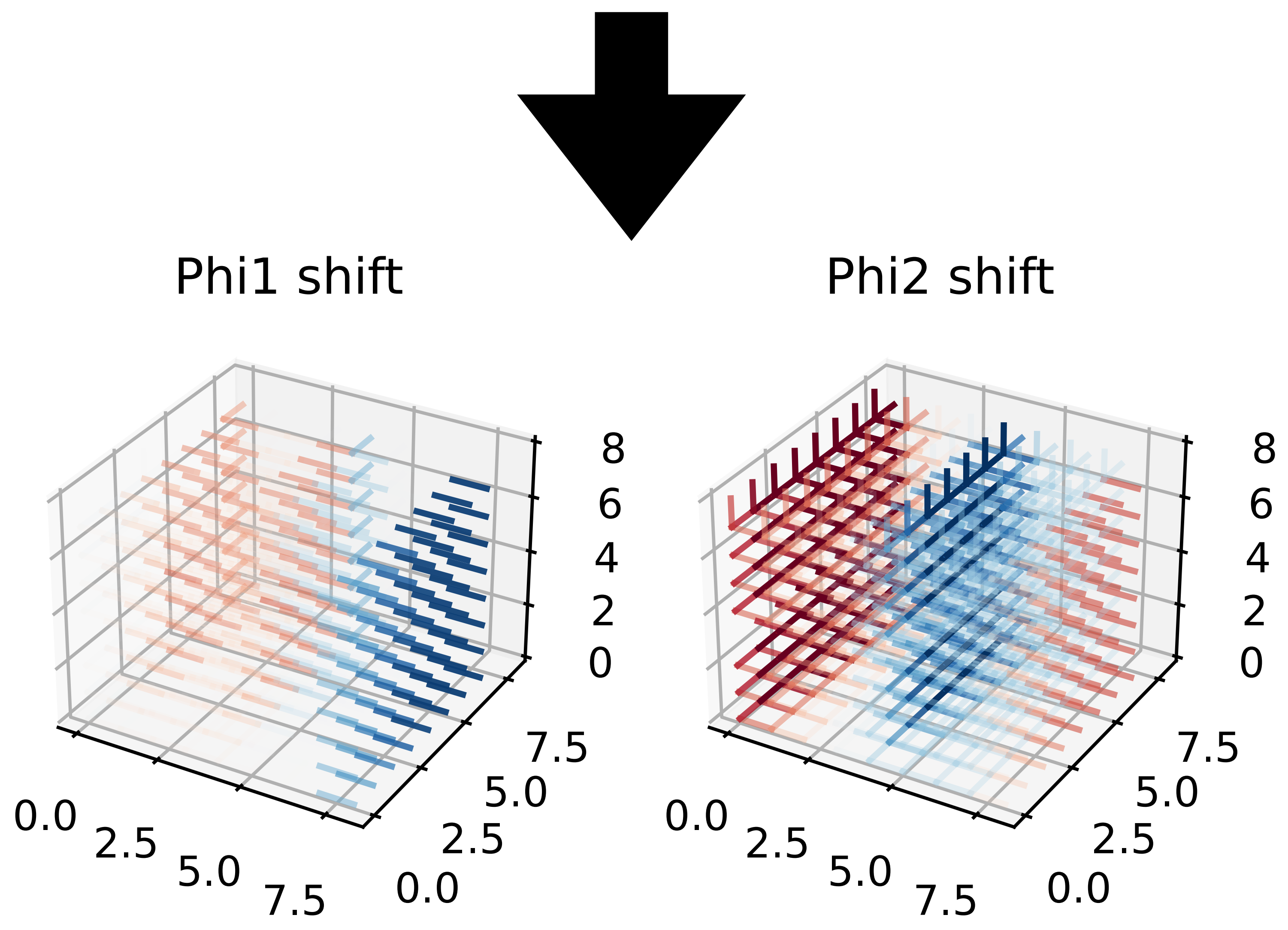


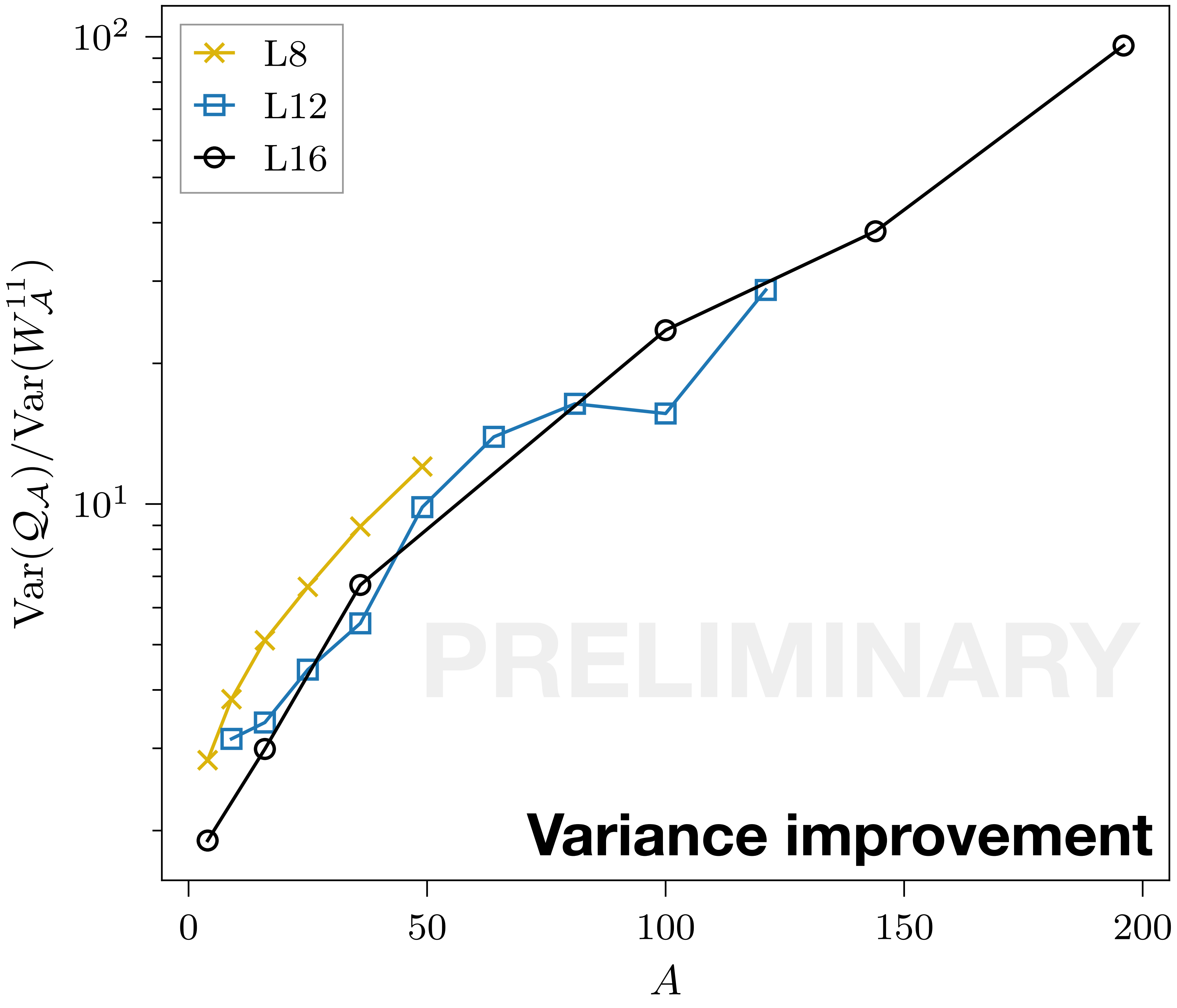
2+1D SU(2) Wilson loops

Wilson loop and gauge fixing scheme



Learned shifts

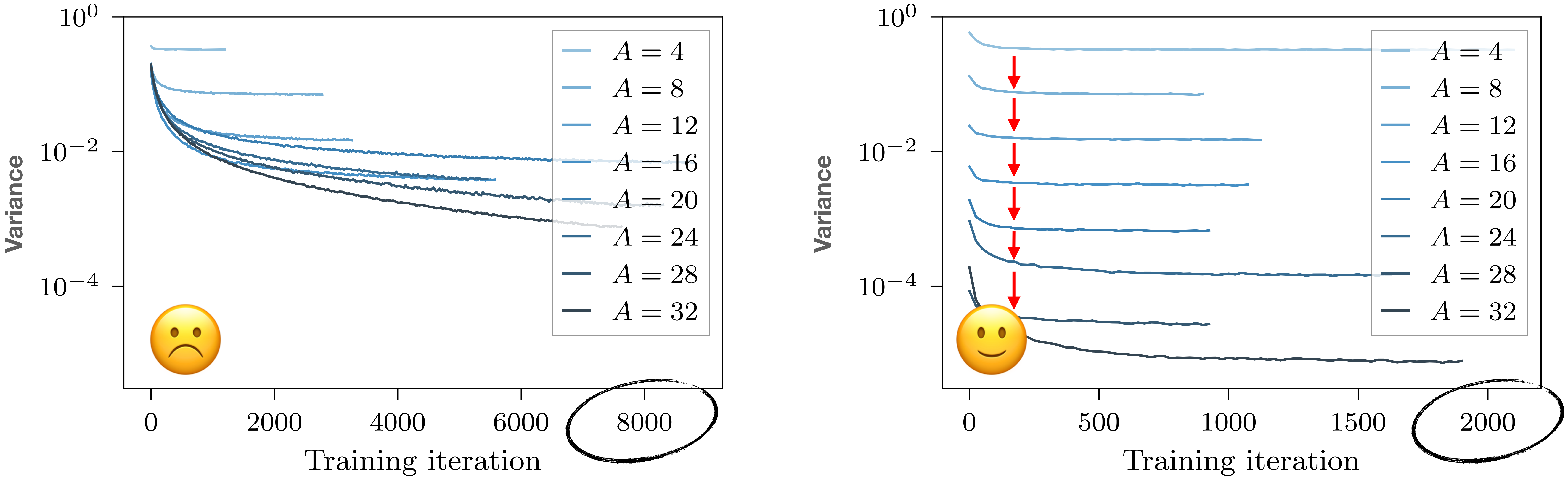




Takeaway: Transfer learning works well

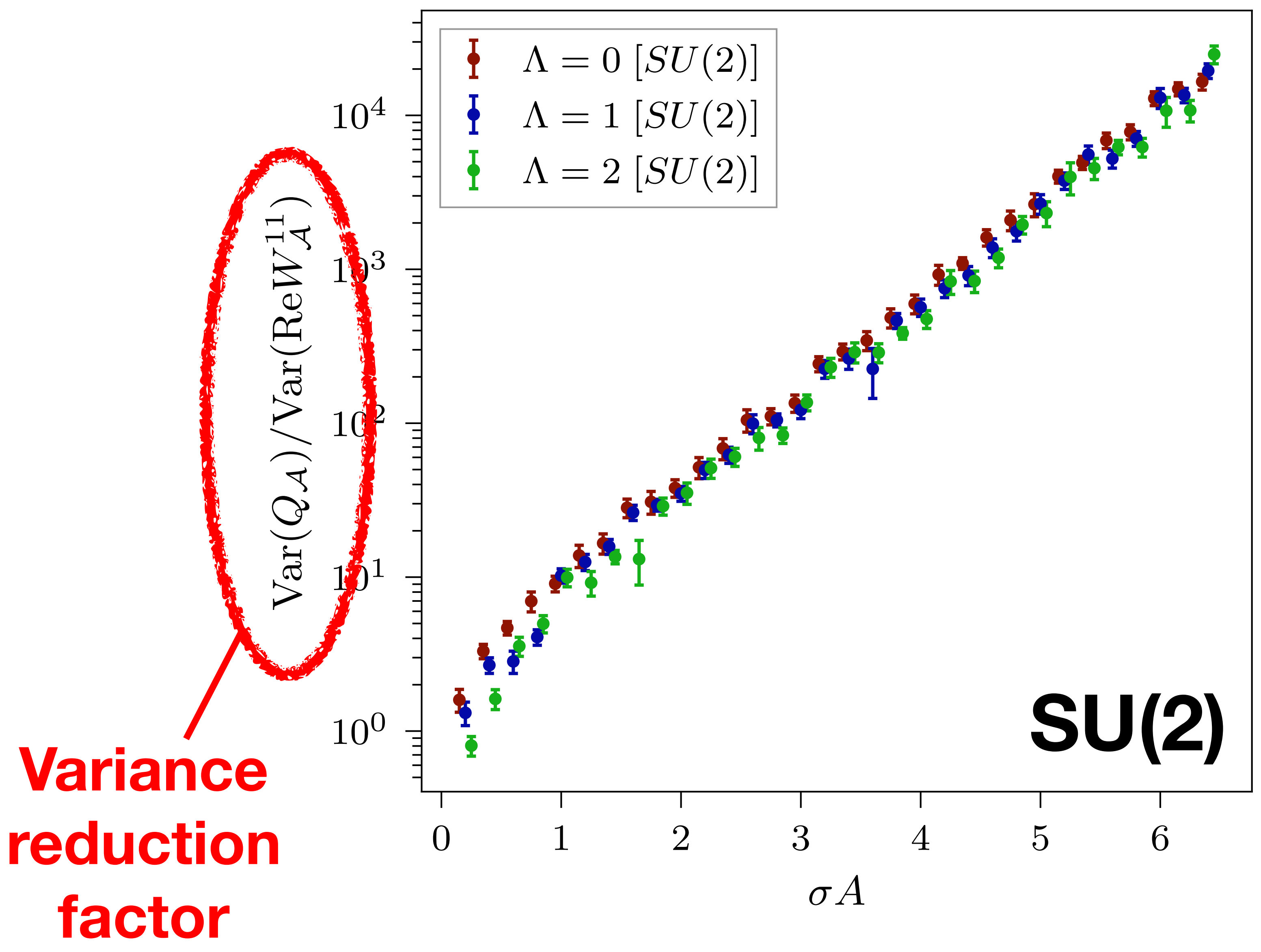
Challenge: Different optimal manifold per observable (e.g. per Wilson loop A)

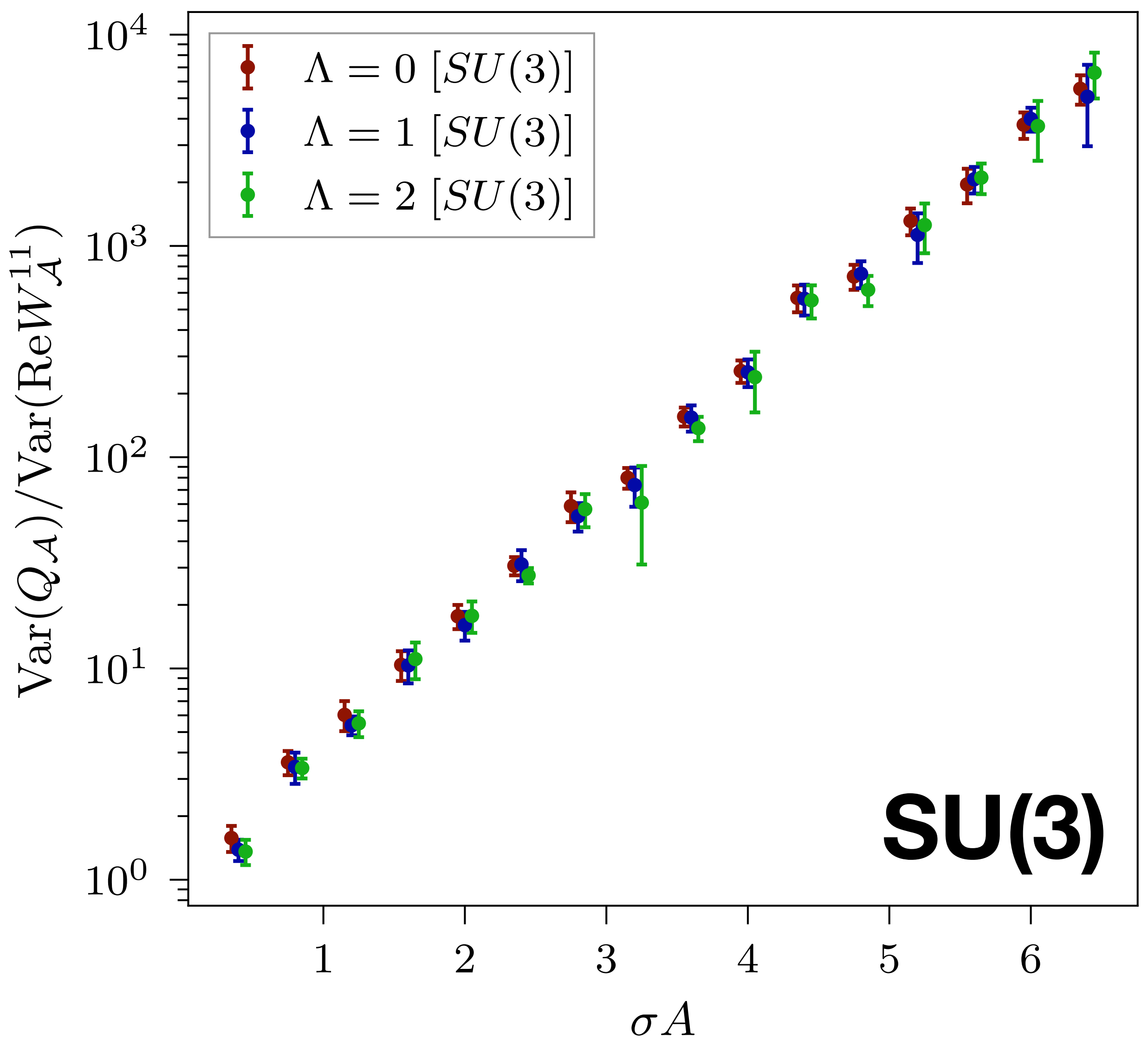
Solution: “Transfer” optimized manifold to initialize next search



Takeaway: (so far) simple manifolds work

- Contour deformations based on field-independent transformations only
  - Using a Fourier expansion to introduce field-dependence not useful





- Deformations are related to normalizing flows…
  can we use NNs & other methods from this domain?

See e.g. [Rodekamp+ **PRB106 (2022) 125139**]

Summary

Thanks!
Questions?

Using complex analysis we can …

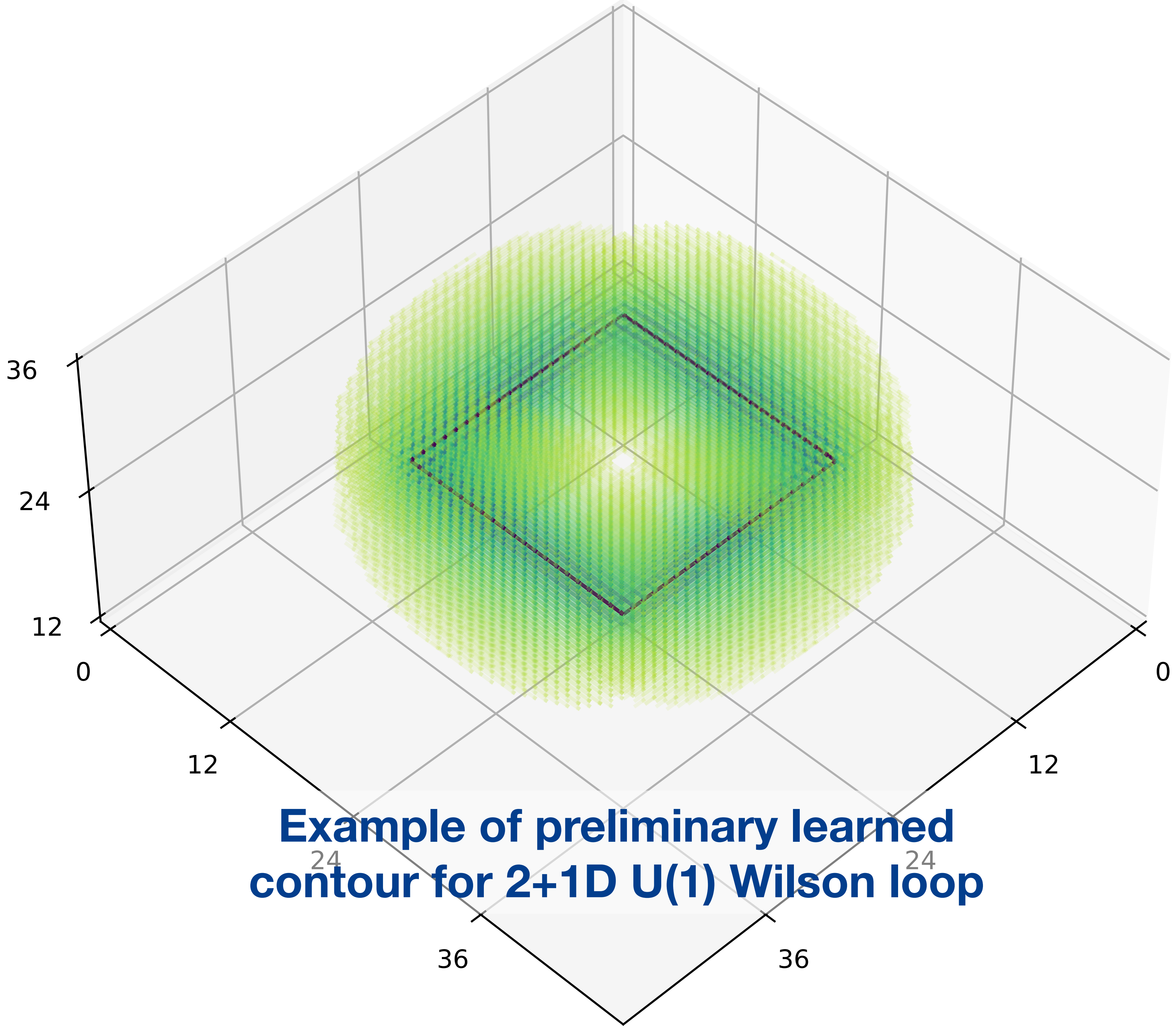
- **Deform observables** $\mathcal{O} \to \mathcal{Q}$, where $\langle \mathcal{O} \rangle = \langle \mathcal{Q} \rangle$ but $\mathrm{Var}[\mathcal{O}] \neq \mathrm{Var}[\mathcal{Q}]$.

  I.e., no systematic error!

- **Minimize variance numerically** (using existing MC samples).

- **Achieve far more precise** measurements in proof-of-principle applications to lattice field theories.

Future work: 3D & 4D, fermions, more advanced ML techniques?



Example of preliminary learned contour for 2+1D U(1) Wilson loop

Backup Slides

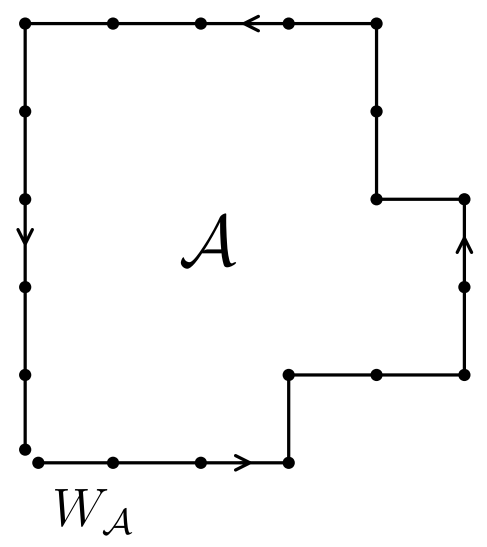
Application to 1+1D $U(1)$ gauge theory

Rewrite path integral in terms of **plaquette angles**

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \prod_x dP_x \, \mathcal{O}[P] \, e^{-S[\arg(P)]}$$

$$S[\phi] = -\beta \sum_x \cos(\phi_x)$$

Wilson loops are a sign-afflicted observable



$$W_{\mathscr{A}} = [\prod_{x \in \mathscr{A}} P_x] = e^{i \sum_{x \in \mathscr{A}} \theta_x}$$

Four lattice spacings

σ	V	g	β
0.4	64	0.52	1.843
0.3	64	0.47	2.296
0.2	64	0.40	3.124
0.1	64	0.30	5.555

U(1) deformation

Shift plaquette integral in complex direction

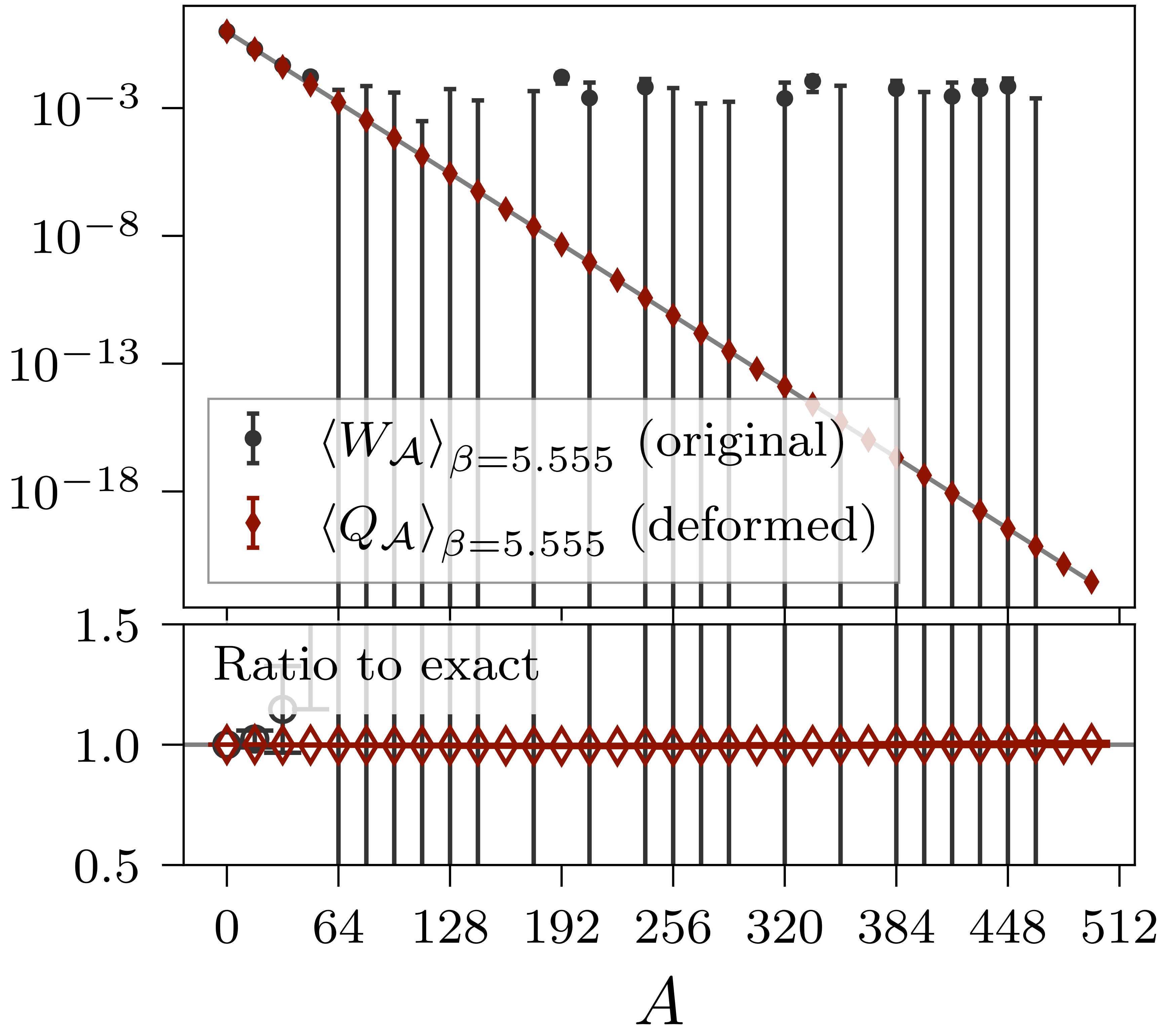
$$\tilde{\phi}_x = \phi_x + i\delta_x$$

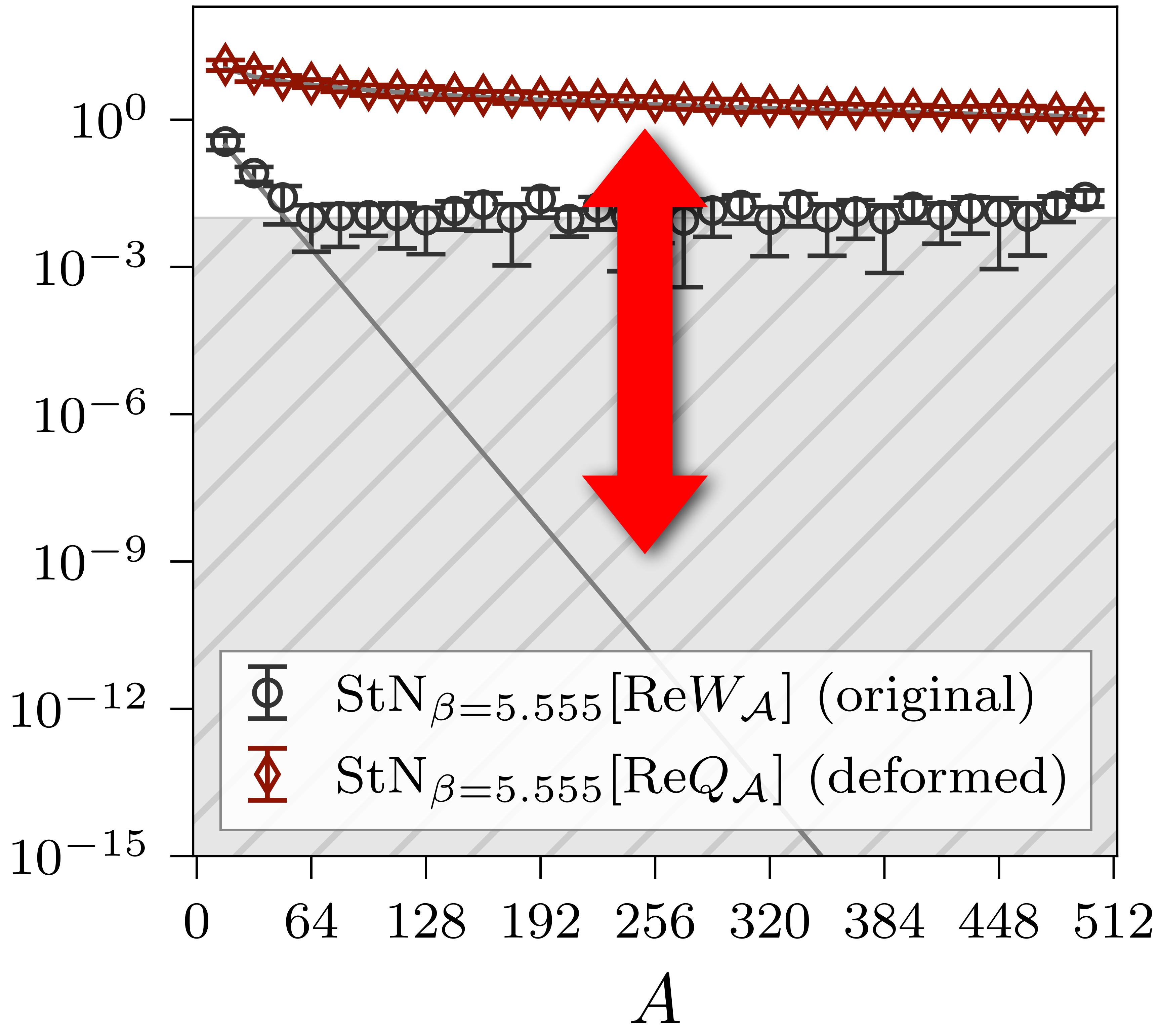
Trivial Jacobian!

[Image credit: Neill Warrington]

$U(1)$ results

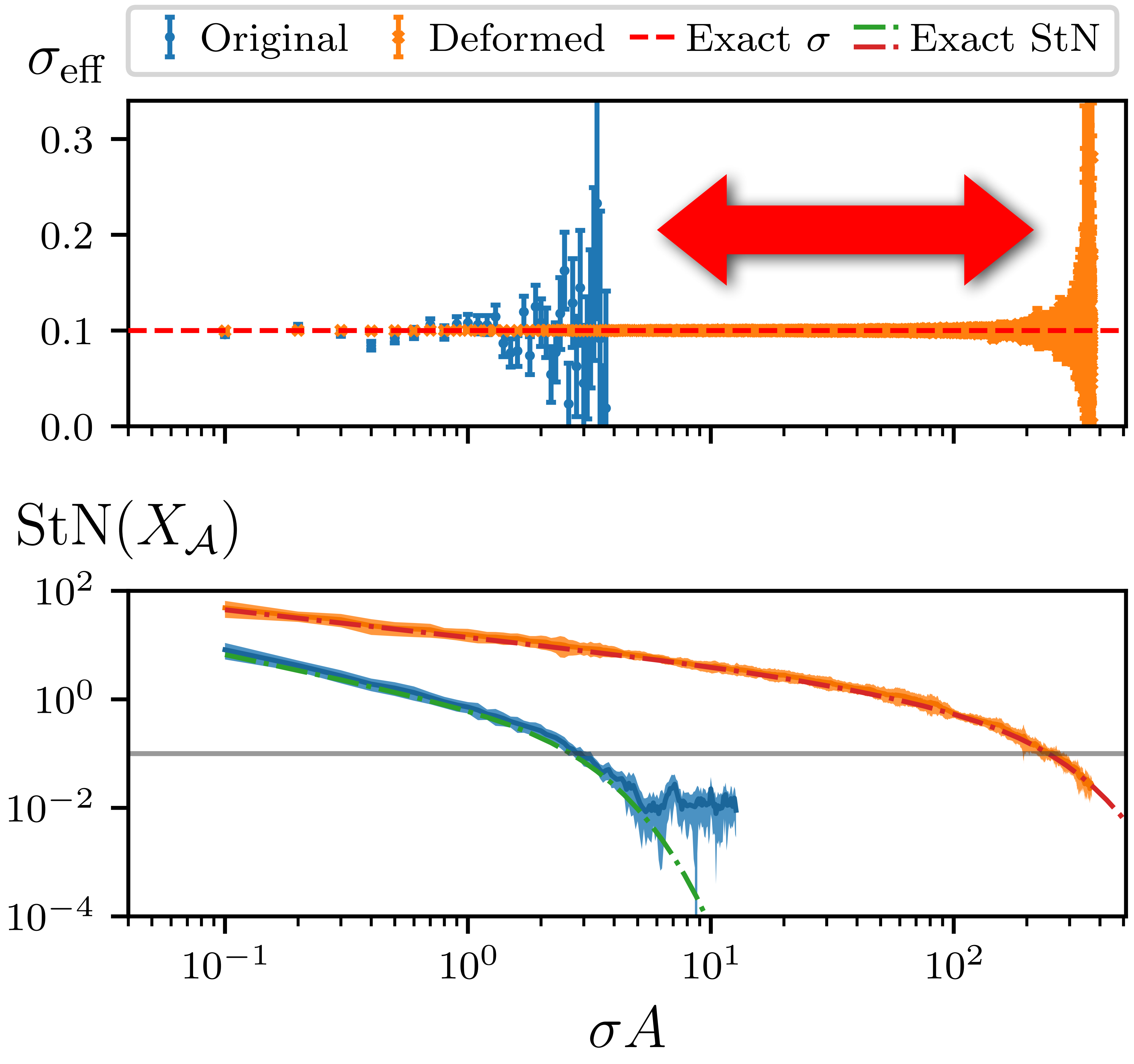
Exponential improvement in variance:





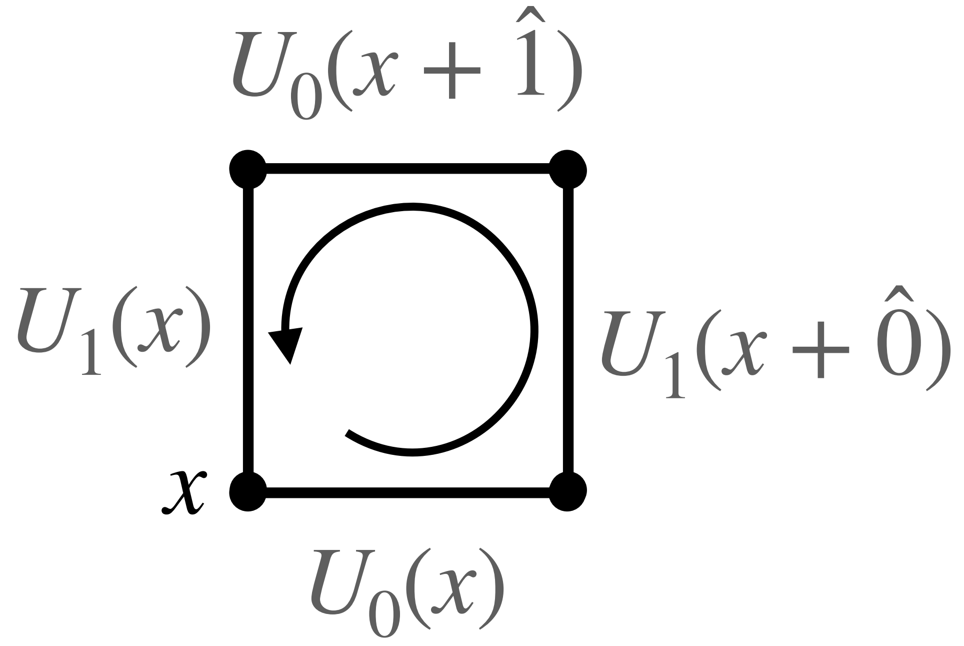
$U(1)$ results

Much longer plateaus in effective string tension σ_{eff}



1+1D $SU(N)$ gauge theory

Simple action for the theory given in terms of **plaquettes**

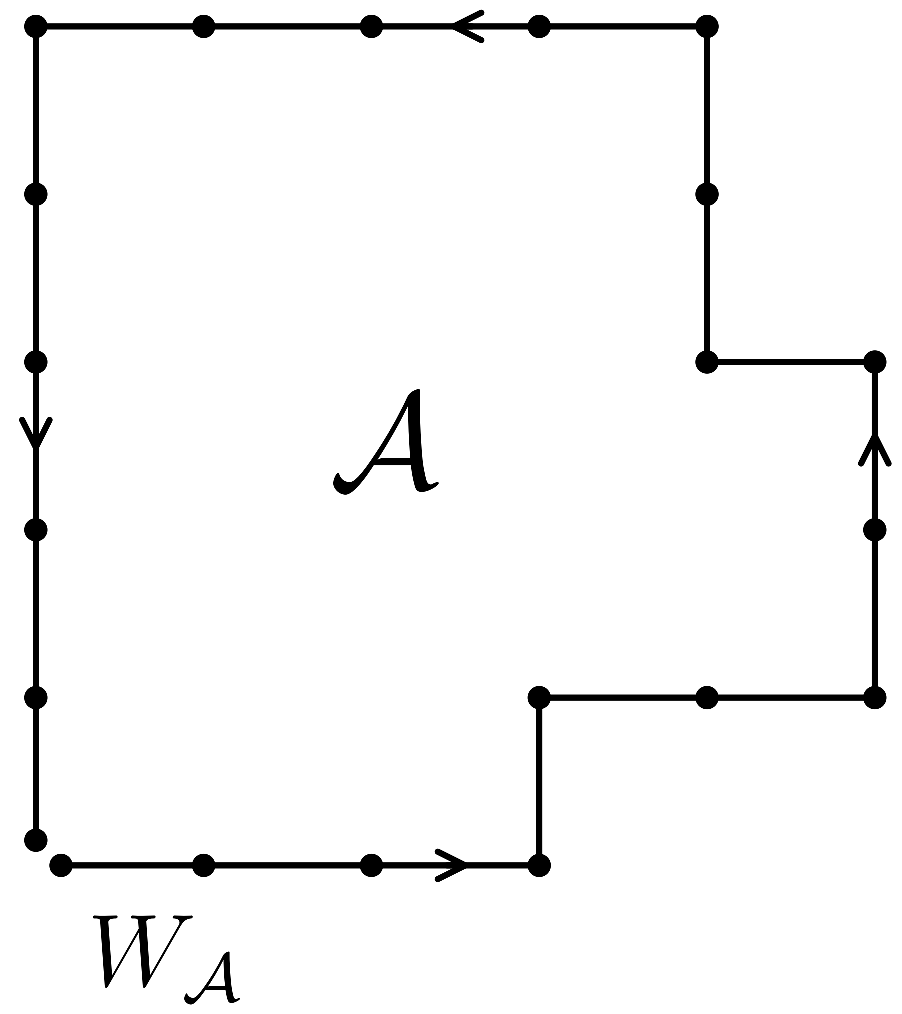


$$S[P] = -\frac{1}{g^2}\sum_x \mathrm{tr}(P_x + P_x^{-1}) \xrightarrow{\text{ctm.}} \frac{1}{2g^2}\mathrm{tr}(F^2)$$

$$P_x = U_0(x)U_1(x+\hat{0})U_0^\dagger(x+\hat{1})U_0^\dagger(x)$$

Wilson loop observables:

- Generalize plaquettes to arbitrary loops
- Probe of confinement (area vs perimeter law scaling)

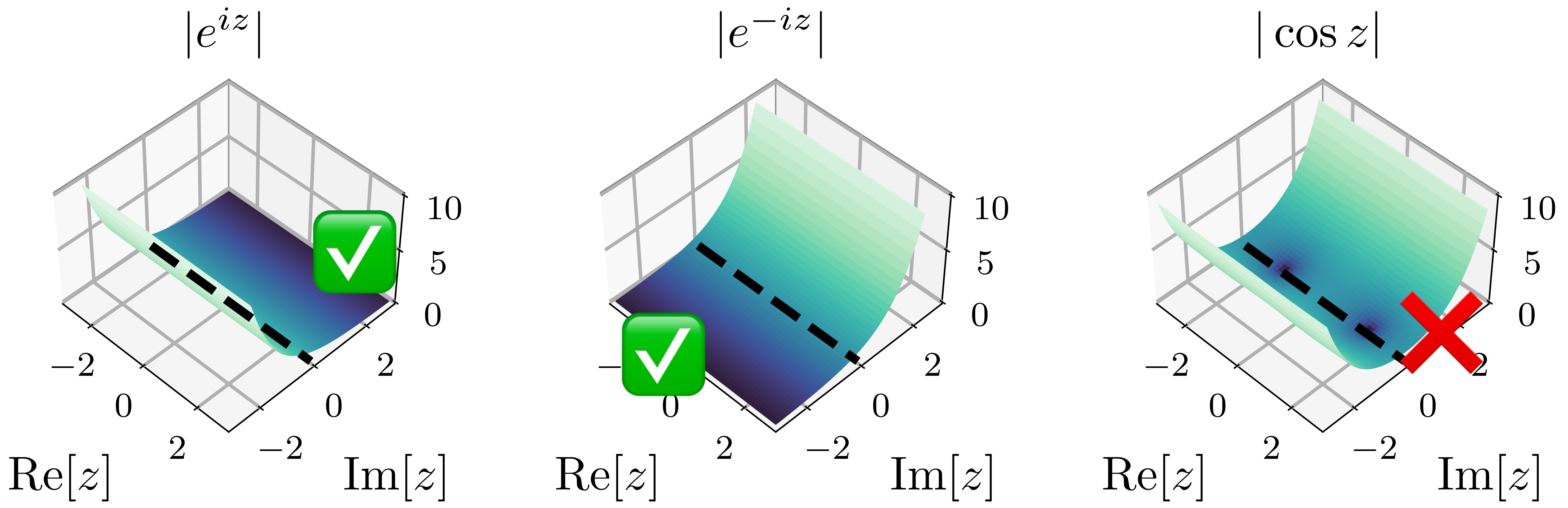


Why select $W^{11}_{\mathscr{A}}$?

Any observable with equivalent expectation value can be taken as the **base observable** for deformation …

1D example: $$\int_0^{2\pi} \frac{dz}{2\pi} e^{iz} \, e^{\beta \cos z} = \int_0^{2\pi} \frac{dz}{2\pi} e^{-iz} \, e^{\beta \cos z} = \int_0^{2\pi} \frac{dz}{2\pi} \cos z \, e^{\beta \cos z} = I_1(\beta)$$

… however, some choices are better than others!



1+1D $SU(N)$ gauge theory

Path integral can be written over plaquettes directly

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \prod_x dP_x \, \mathcal{O}[P] \, e^{-S[P]} \qquad S[P] = -\frac{1}{g^2} \sum_x \mathrm{tr}(P_x + P_x^{-1})$$

Analytic continuation straightforward with the complexification $SU(N) \to SL(N, \mathbb{C})$.

Singularities of P_x^{-1} easily avoided by contour parameterization.

We study three lattice spacings and both SU(2) and SU(3):

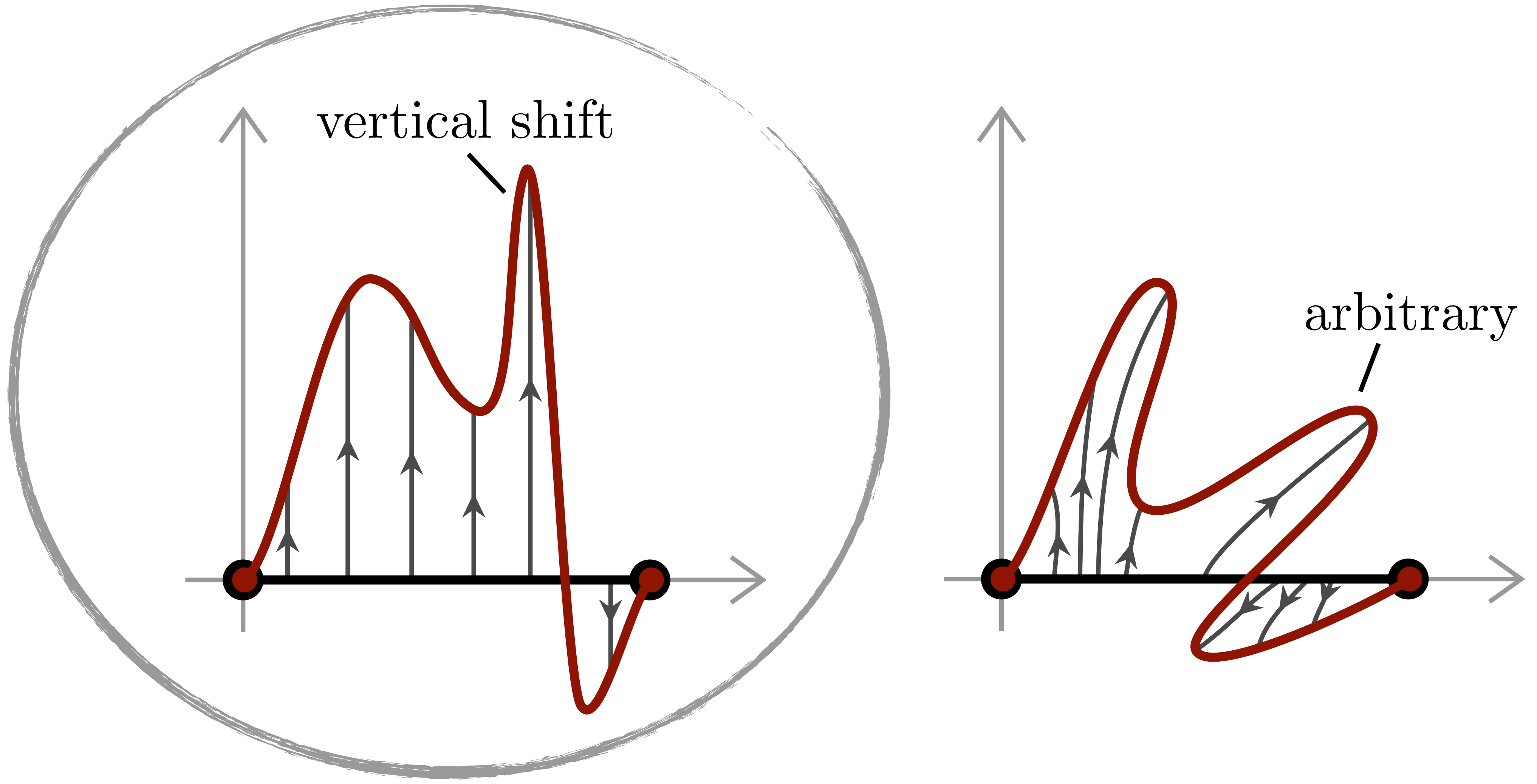
		$SU(2)$		$SU(3)$	
σ	V	g	β	g	β
0.4	16	0.98	4.2	0.72	11.7
0.2	32	0.71	8.0	0.53	21.7
0.1	64	0.51	15.5	0.38	41.8

(Parameters chosen to fix σV)

Deformation

Vertical deformations $\tilde{\Omega}_x = \Omega_x + if(\Omega)$

[Alexandru et al. PRD97 (2018) 094510]



Fourier series definition of $f(\Omega)$, using a subset of all possible terms

$$\tilde{\phi}^a_x = \phi^a_x + i\kappa_0^{x;\phi^a} + i\sum_{y\leq x} f_{\phi^a}(\Omega_y;\kappa^{xy},\lambda^{xy},\chi^{xy},\zeta^{xy}),$$

$$\tilde{\theta}^a_x = \theta^a_x + i\sum_{y\leq x} f_{\theta^a}(\Omega_y;\kappa^{xy},\lambda^{xy},\chi^{xy}).$$

Original real part — Parameterized imaginary shift

$$f_{\theta^a} = \sum_{m=1}^{\Lambda}\kappa_m^{xy;a}\sin(2m\theta_y^a)\Big\{1+\sum_{n=1}^{\Lambda}\Big[\sum_{\substack{r\neq a\\ r=1}}^{3}\lambda_{mn}^{xy;ar}\sin(2n\theta_y^r)+\sum_{s=1}^{5}\eta_{mn}^{xy;as}\sin(n\phi_y^s+\chi_{mn}^{xy;as})\Big]\Big\},$$

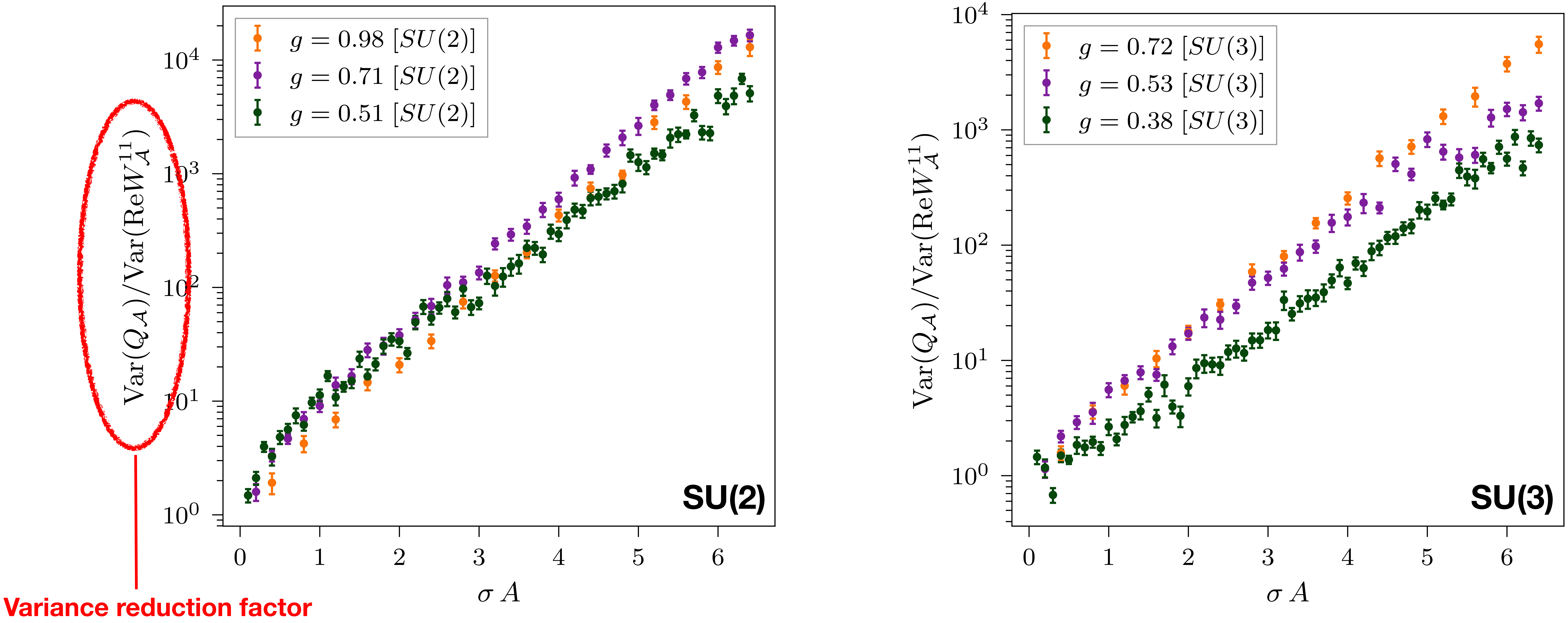
$$f_{\phi^a} = \sum_{m=1}^{\Lambda}\kappa_m^{xy;a}\sin(m\phi_y^a+\zeta_m^{xy;a})\Big\{1+\sum_{n=1}^{\Lambda}\Big[\sum_{r=1}^{3}\lambda_{mn}^{xy;ar}\sin(2n\theta_y^r)+\sum_{\substack{s\neq a\\ s=1}}^{5}\eta_{mn}^{xy;as}\sin(n\phi_y^s+\chi_{mn}^{xy;as})\Big]\Big\}.$$

Triangular Jacobian: $f(\Omega)$ only allowed to depend on $y \leq x$. Jacobian determinant calculable from diagonal elements.

This is key for scalability

SU(N) lattice spacing effects

Similar variance reduction effects across all 3 lattice spacings:



Complex scalar theory

Use phase-magnitude decomposition for variables $\phi_t = R_t e^{i\theta_t}$

Holomorphic: $$S = -2 \sum_{t=0}^{L-1} R_t R_{t+1} \cos(\theta_{t+1} - \theta_t) + V(R)$$

$$V(R) \;=\; \sum_t (2 + m^2) R_t^2 + \lambda R_t^4$$

Interested in correlation functions

$$G_t = \left\langle R_t R_0 e^{i\theta_t - i\theta_0} \right\rangle \equiv \langle C_t(R, \theta) \rangle$$

Deformation for scalar theory

Intuition: phase differences appear in action similarly to phases of Schwinger, use shifts into imaginary direction

$$\tilde{\theta}_t = \theta_t + i\delta_t^{(1)} + i\delta_t^{(2)} f_c(R_t R_{t+1}) + i\delta_t^{(3)} f_c(R_{t-1} R_t)$$

Extra terms inspired by small phase fluctuation expansion.

Results: 0+1D ϕ^4 correlators

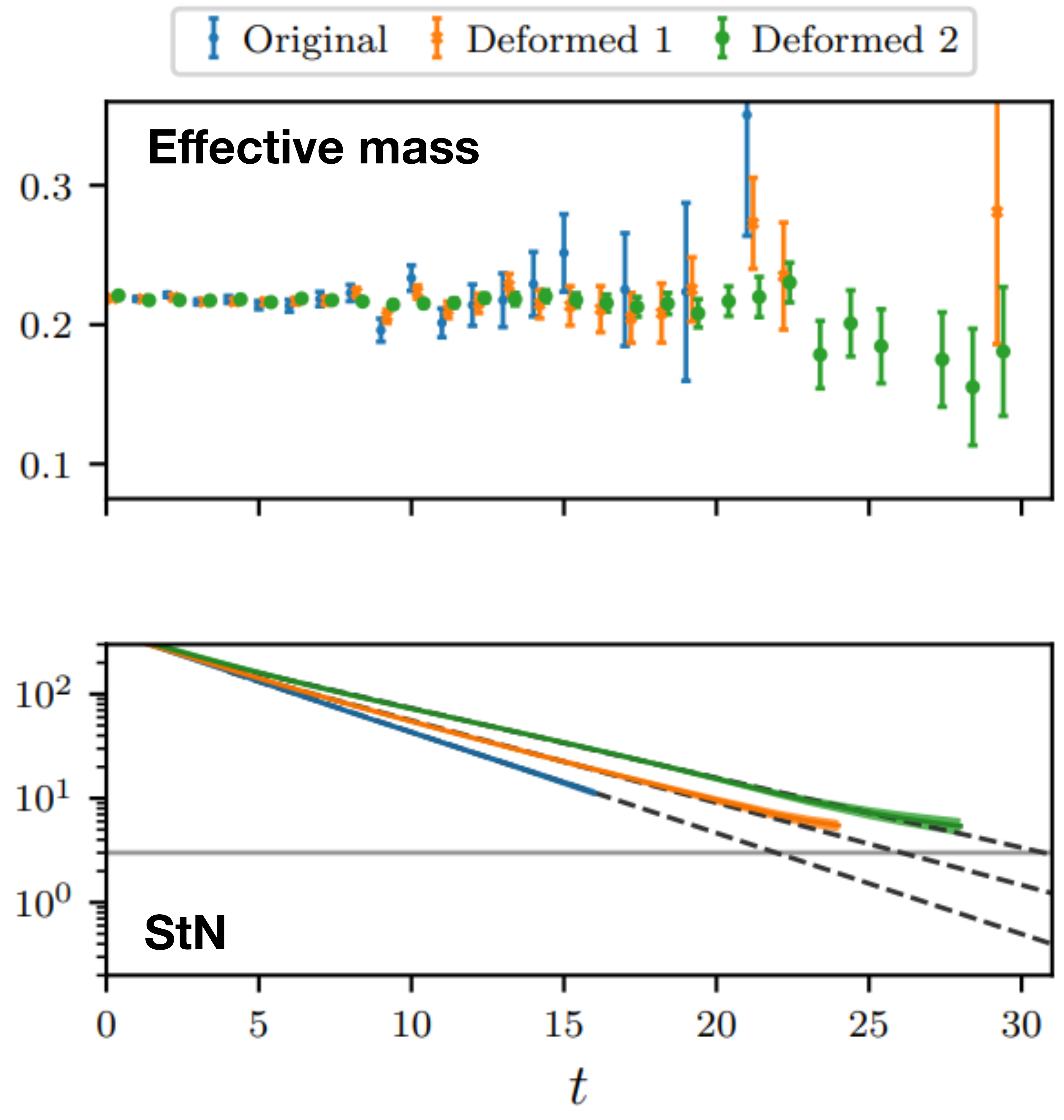
Using polar dofs in the path integral

$$\phi(t) = R(t)e^{i\theta(t)}$$

and the holomorphic action

$$S[R,\theta] = -2\sum_t R(t)\,R(t+1)\,\cos[\theta(t+1)-\theta(t)] + \sum_t V(R(t))$$

$$V(R) \equiv (2+m^2)\,R^2 + \lambda R^4$$

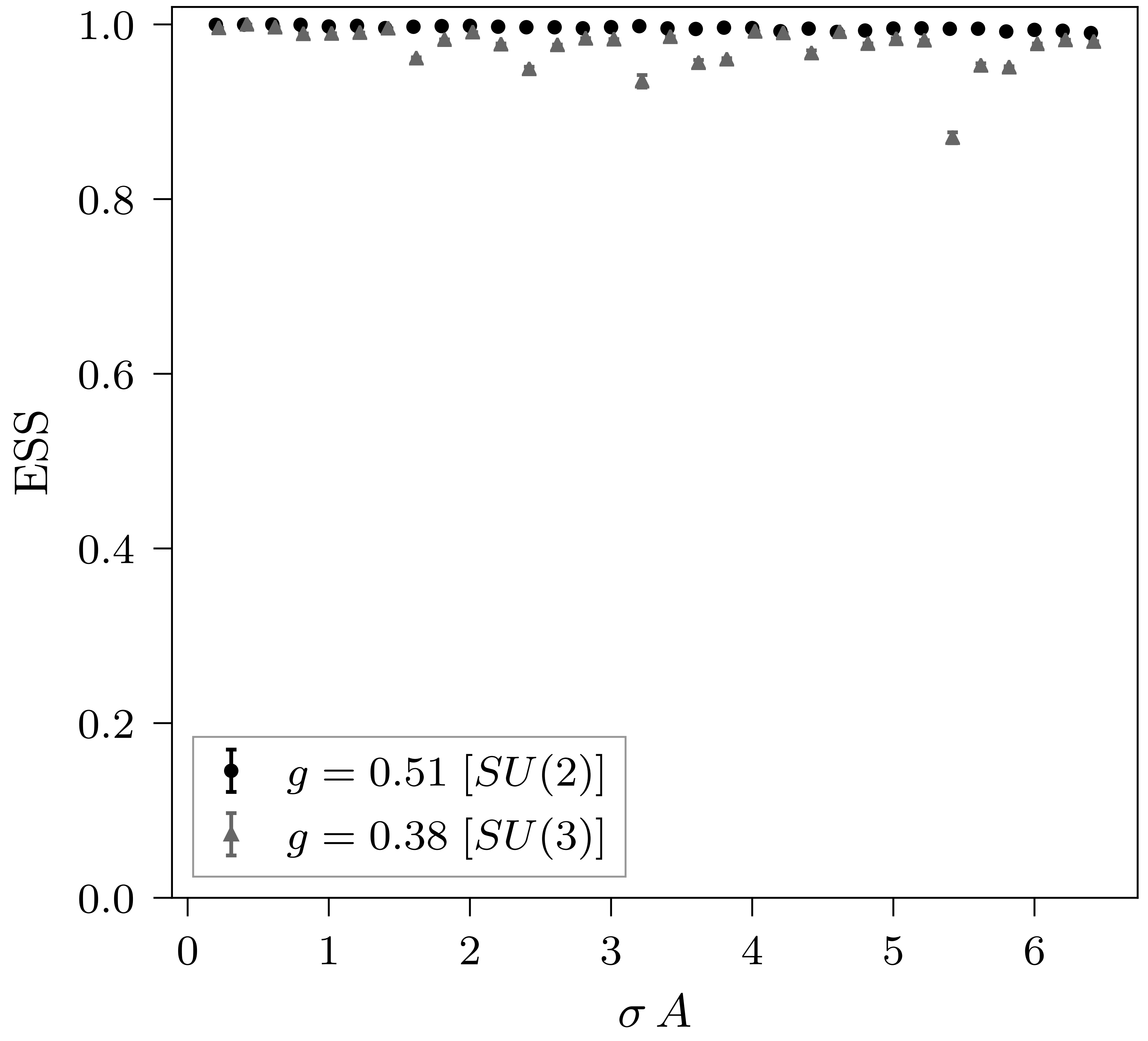


Overlap

“Effective sample size” (ESS) measures degree of overlap between MC distribution and observable path integral

- ESS = 1 implies observable magnitude is always 1, sample-by-sample, i.e. best possible importance sampling

We observe ESS consistently $\gtrsim 80\,\%$



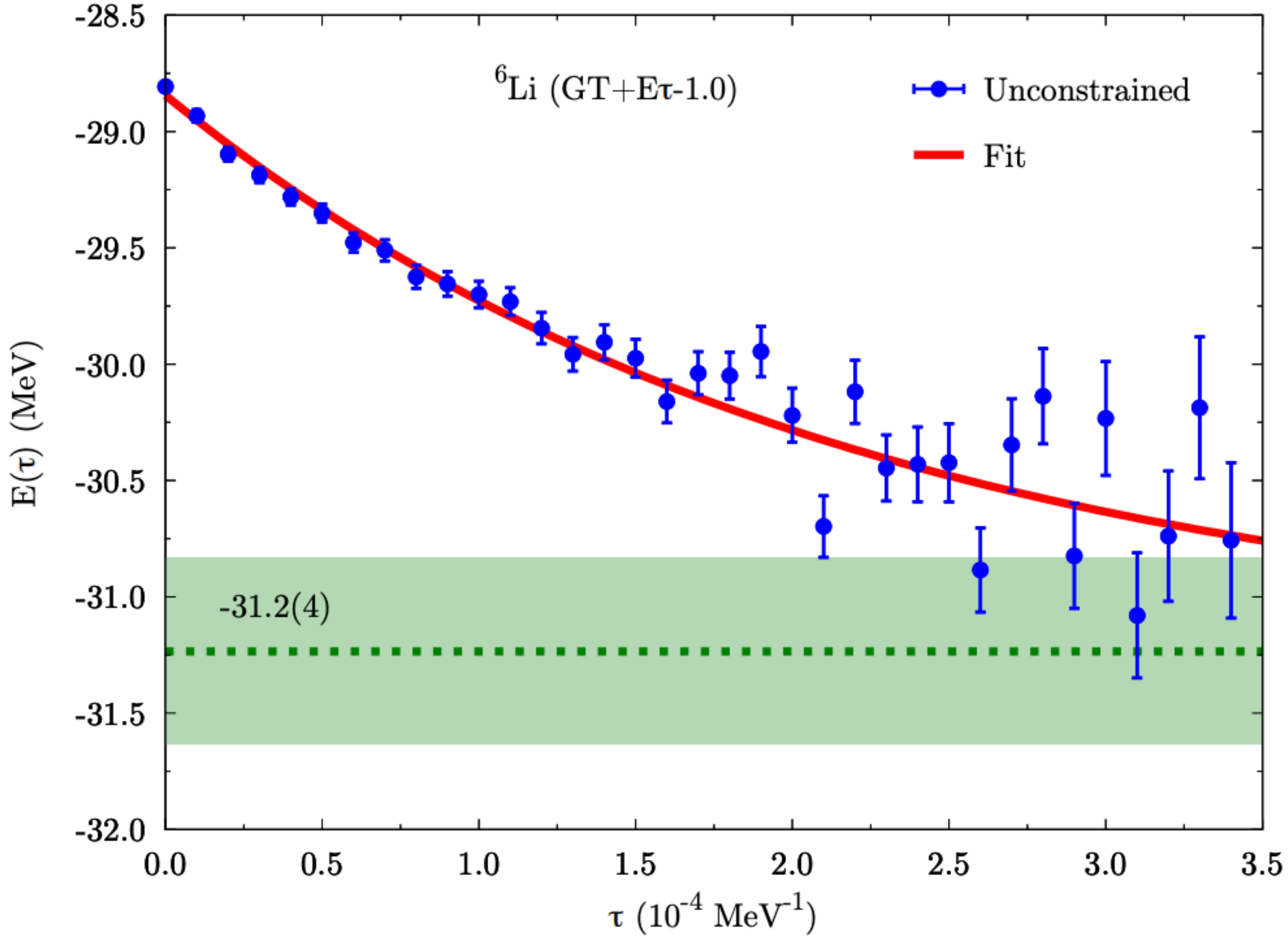
Non-lattice applications: GFMC/AFDMC

Large t: StN decays exponentially

Small t: Excited state effects

To extract physical information,
fit excited state model

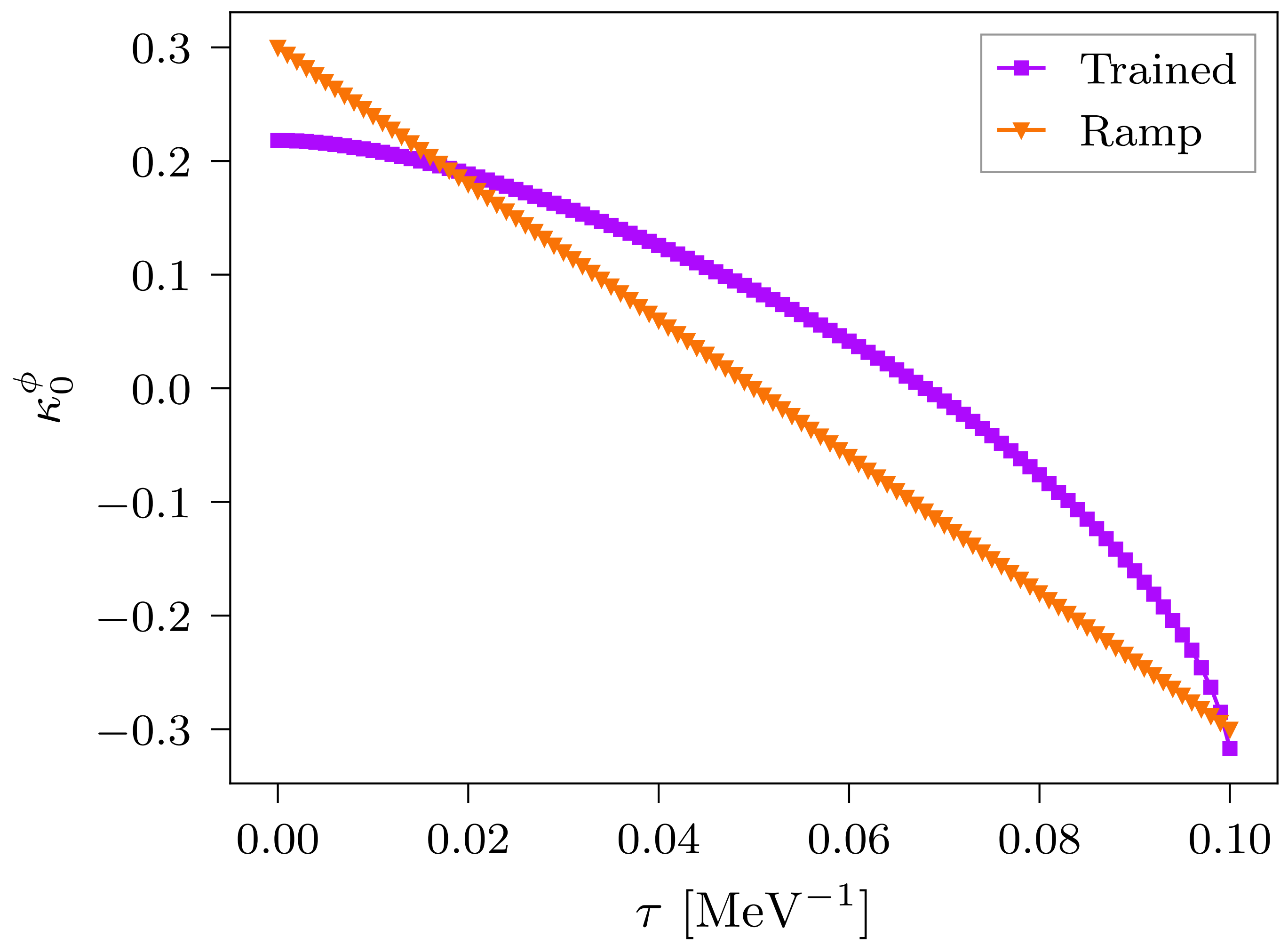
[Gandolfi, et al. 2001.01374]



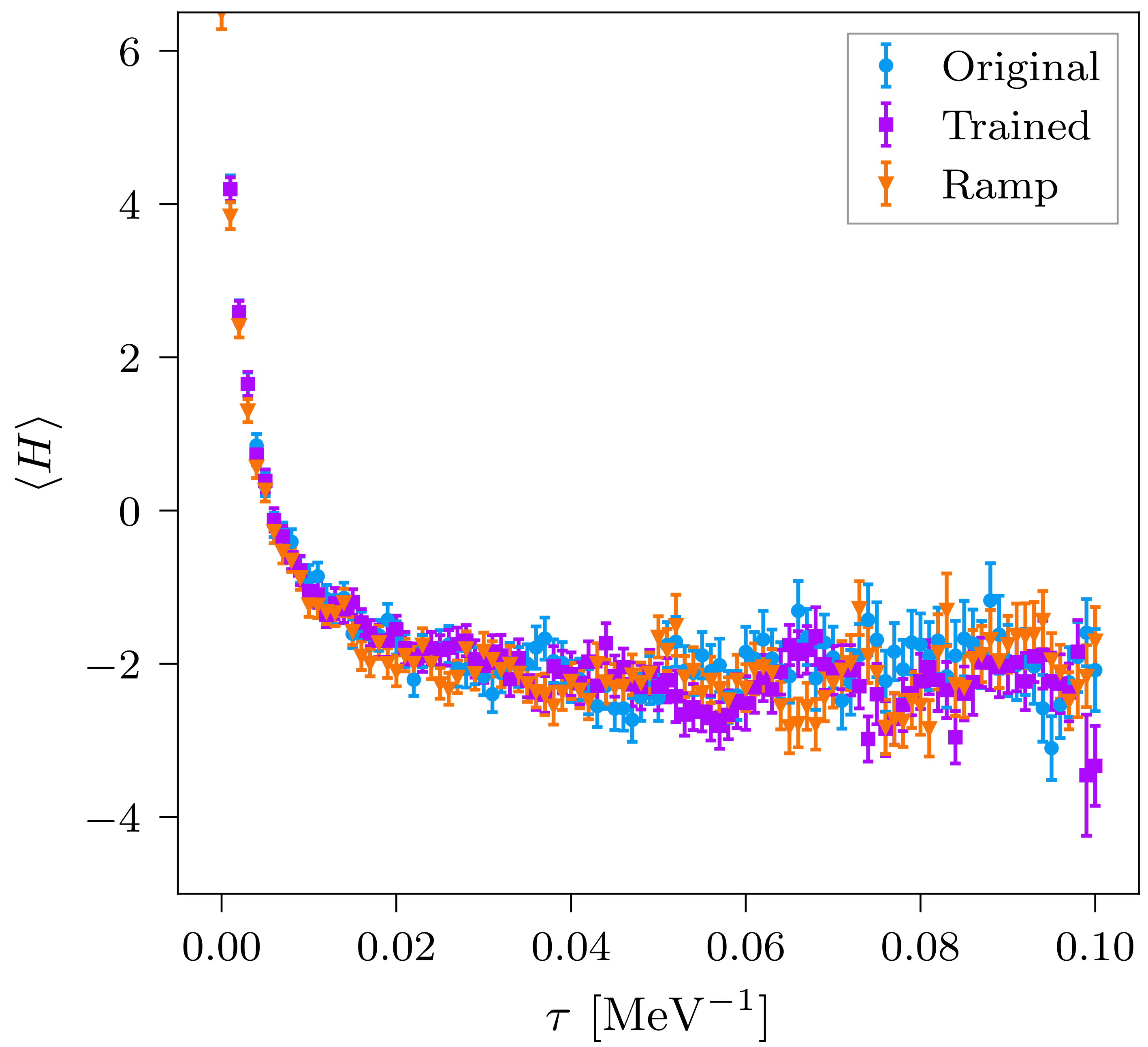
[GK, Lovato, Rocco, Wagman 2304.03229]

GFMC results

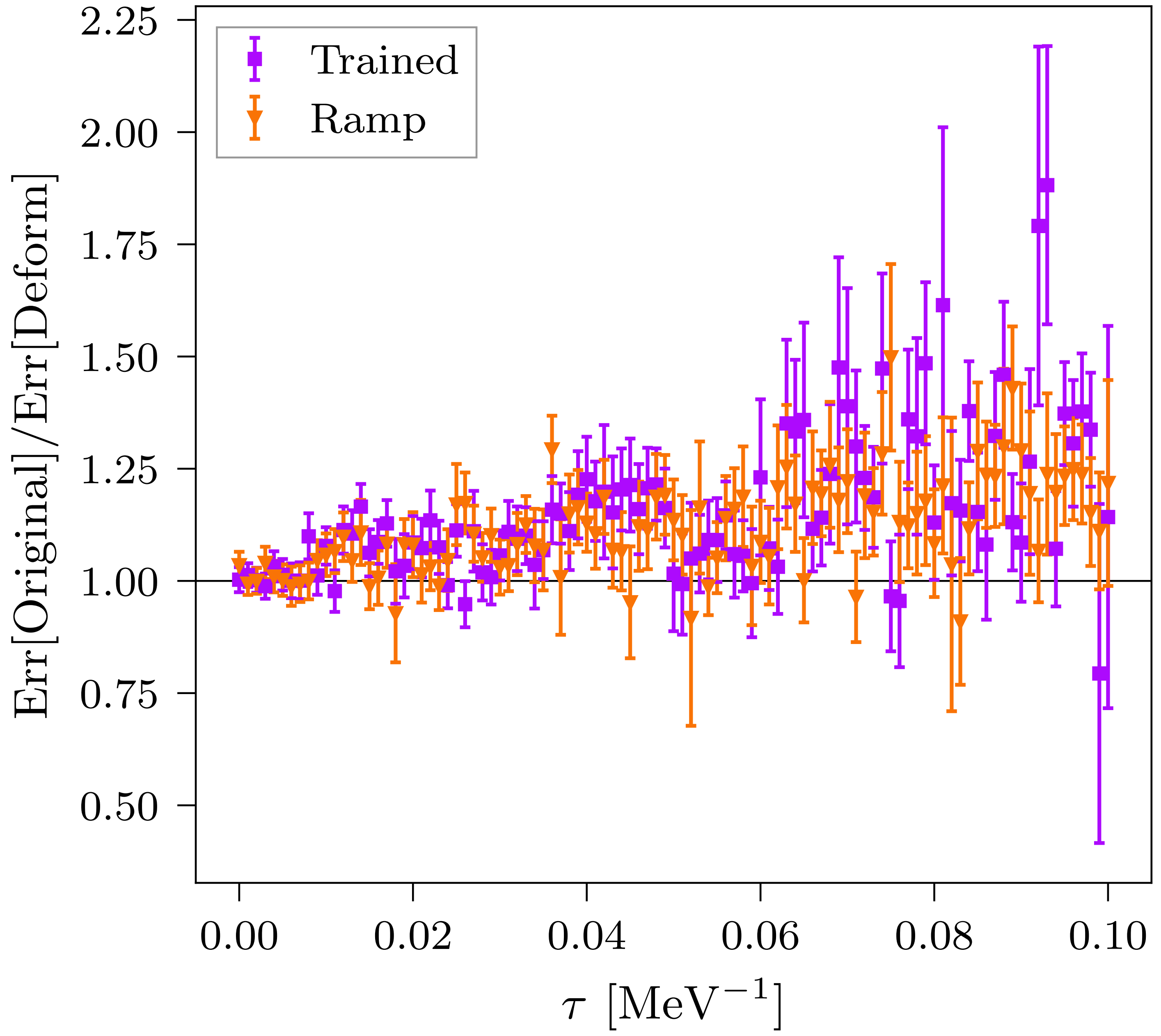
Learned imaginary shift vs Euclidean time



Measured deuteron binding energy



Improvement ratio



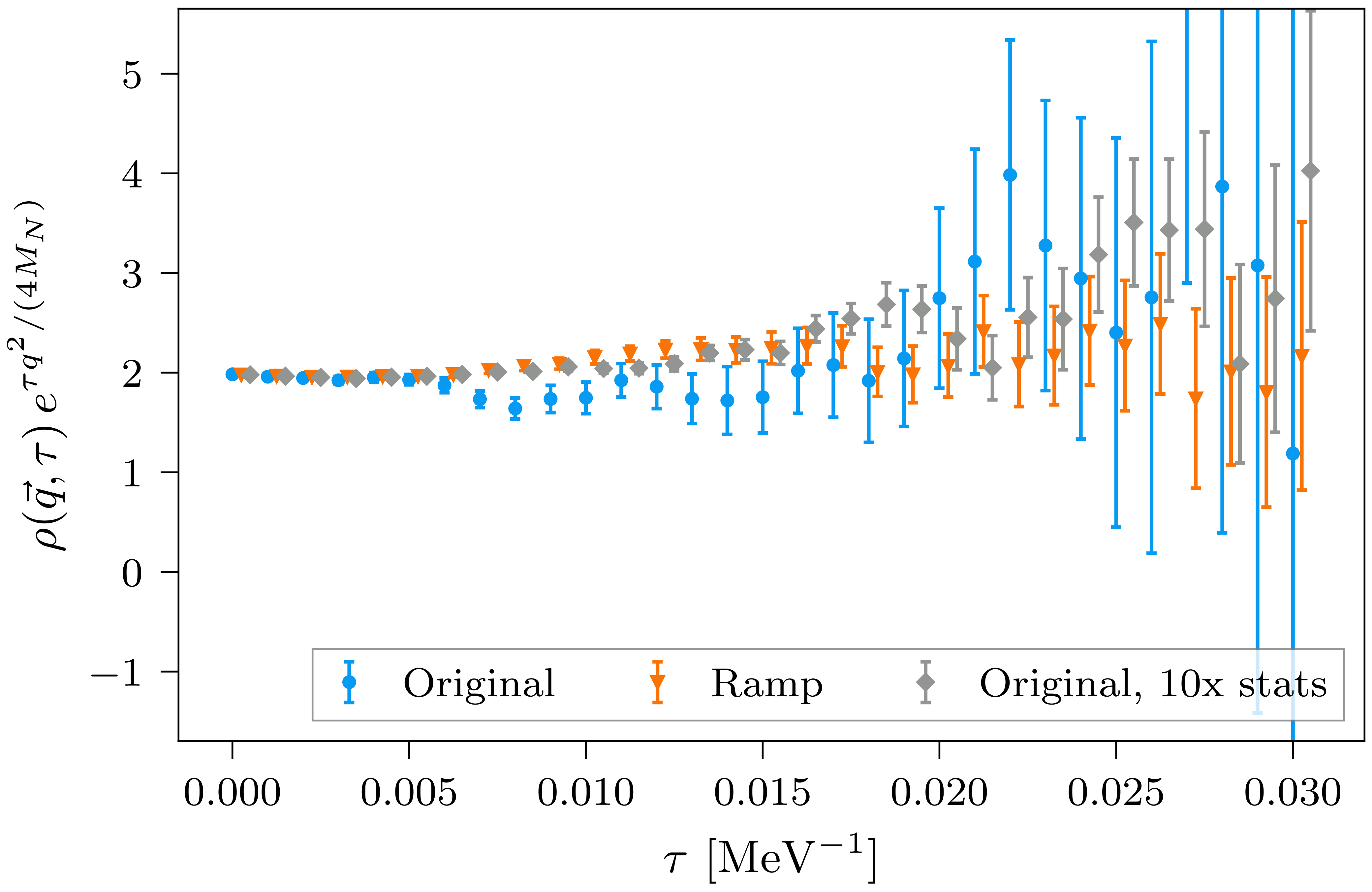
No spectacular results for $\langle H \rangle$, but …

[GK, Lovato, Rocco, Wagman 2304.03229]

GFMC results

… deuteron Euclidean density response $\rho(\overrightarrow{q})$ significantly improved.

$\vec{q} = (0, 0, 800)$ MeV



$\vec{q} = (600, 600, 600)$ MeV