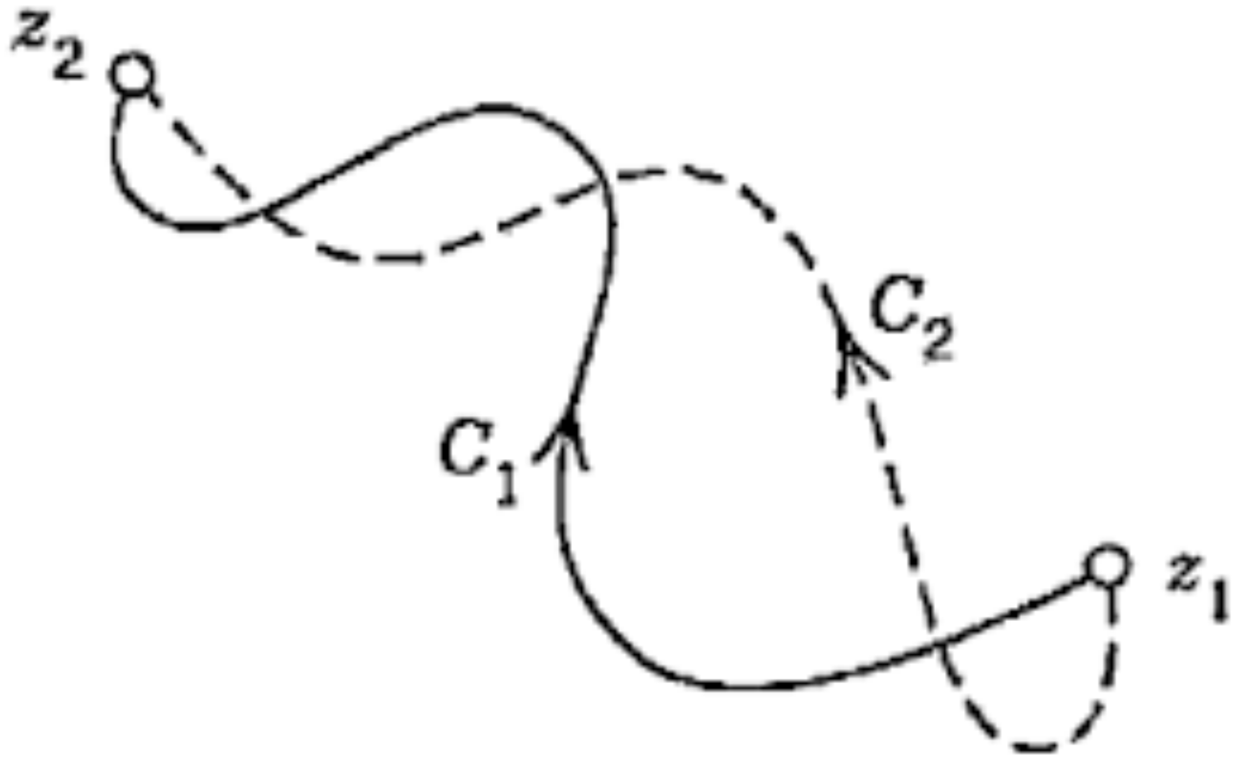
Neural-network contour deformations

for the

signal-to-noise problem



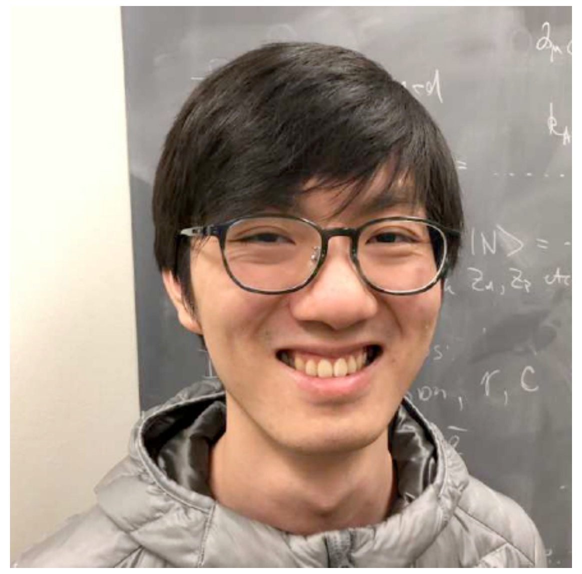
Detmold, GK, Wagman **PRD98 (2018) 074511, PoS LATTICE2018 176**
Detmold, GK, Wagman, Warrington **PRD102 (2020) 014514**
Detmold, GK, Lamm, Wagman, Warrington **PRD103 (2021) 094517**
Detmold, GK, Lin, Shanahan, Wagman **2309.00600, NeurIPS ML4PS (2023)**

Gurtej Kanwar
Institute for Theoretical Physics, AEC, U. Bern

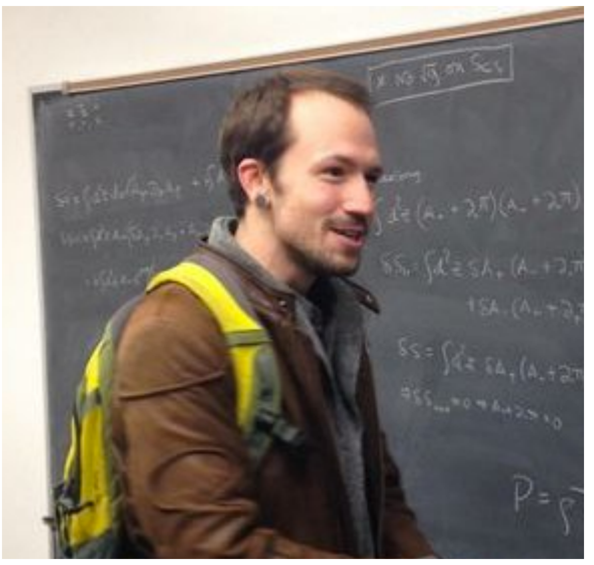
July 24-26, 2024
ML-LFT Workshop, Swansea

Massachusetts Institute of Technology

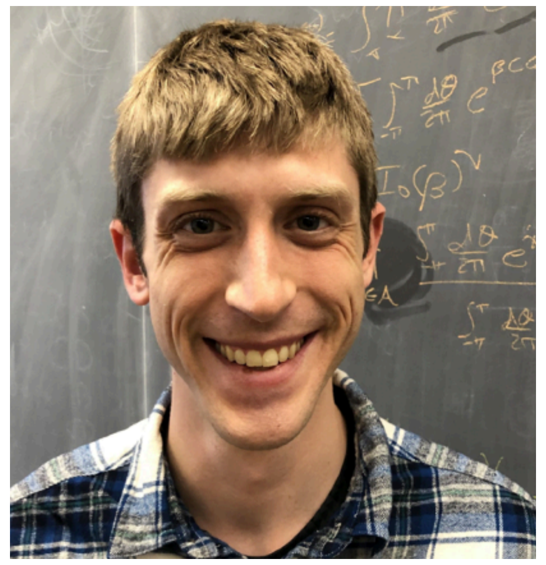

Will Detmold


Yin Lin


Phiala Shanahan


Neill Warrington

Fermilab


Mike Wagman


Hank Lamm

LQFT correlation functions

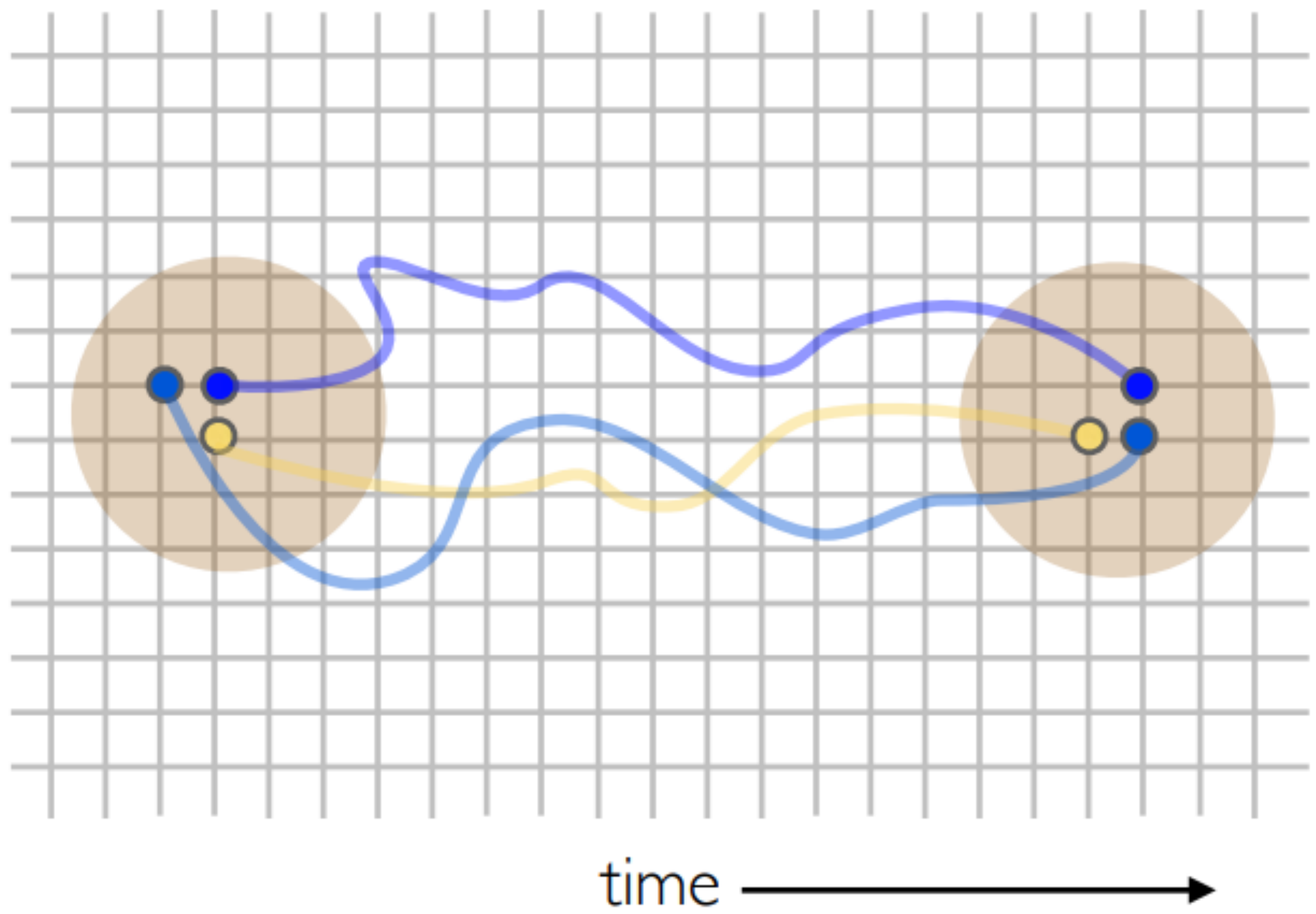
Imaginary-time correlation functions inform us of the spectrum of the theory

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

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.

Noise problems in correlation functions

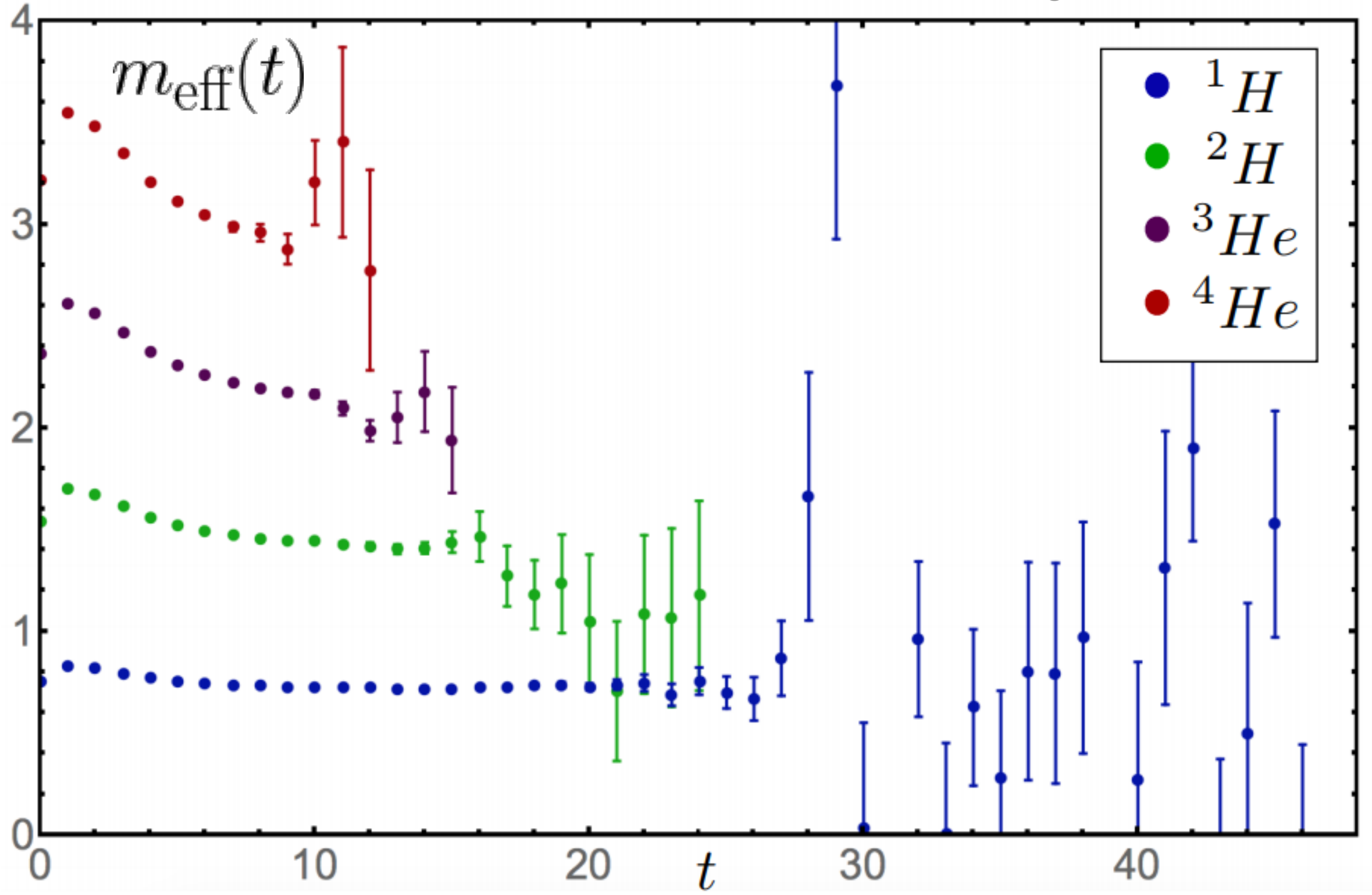
$$m_{\text{eff}}(t) = -\partial_t \log \left\langle \mathscr{A}(t)\mathscr{A}^\dagger(0) \right\rangle$$

$$= -\partial_t \log[\sum_n Z_n e^{-E_n t}] \xrightarrow{t\Delta E \gg 1} E_0$$

Must find **plateau***

- Small t: Excited-state contamination
- Large t: **Signal-to-noise ratio vanishes exponentially!**

Wagman, Lattice 2018



* Or regime dominated by only a few states

Noise problem = sign problem

$$C(t) \equiv \mathscr{A}(t)\mathscr{A}^\dagger(0)$$

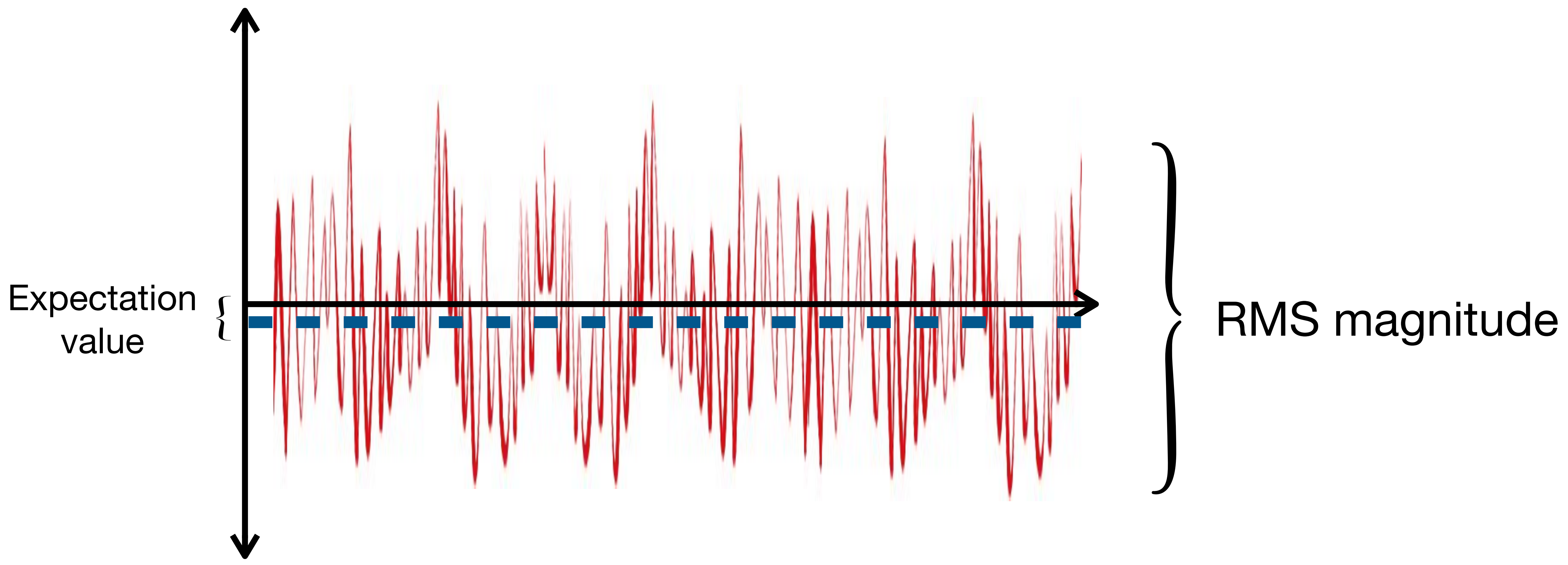
Intuitively a noise problem implies …

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

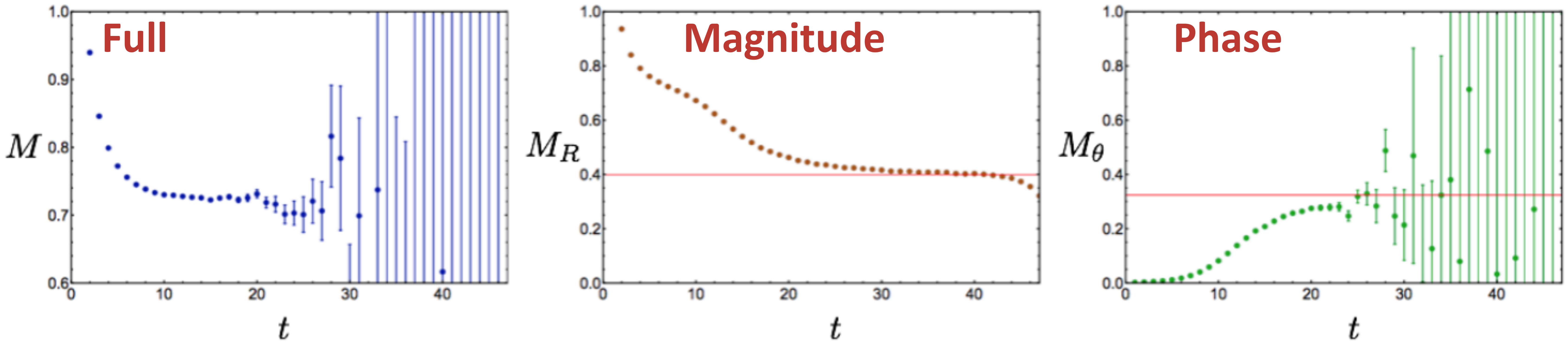
The RMS magnitude …

… is exponentially larger than the expectation value.

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



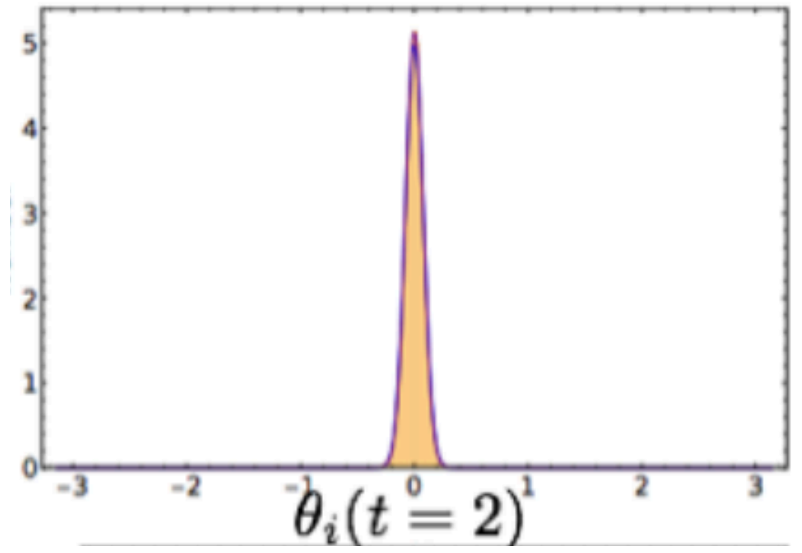
Noise problem = sign problem



[Wagman & Savage, PRD96 (2017) 114508]

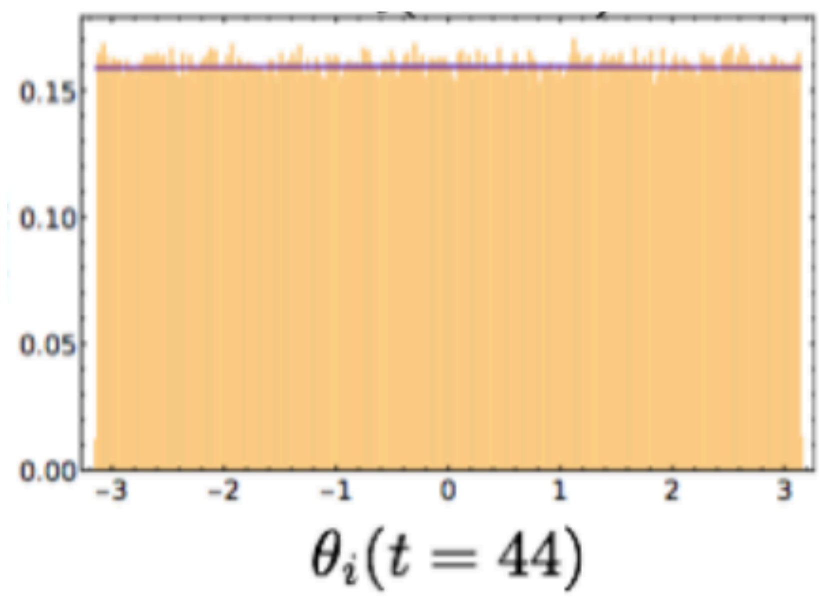
Noise in nucleon effective mass carried by complex phase

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



vs.

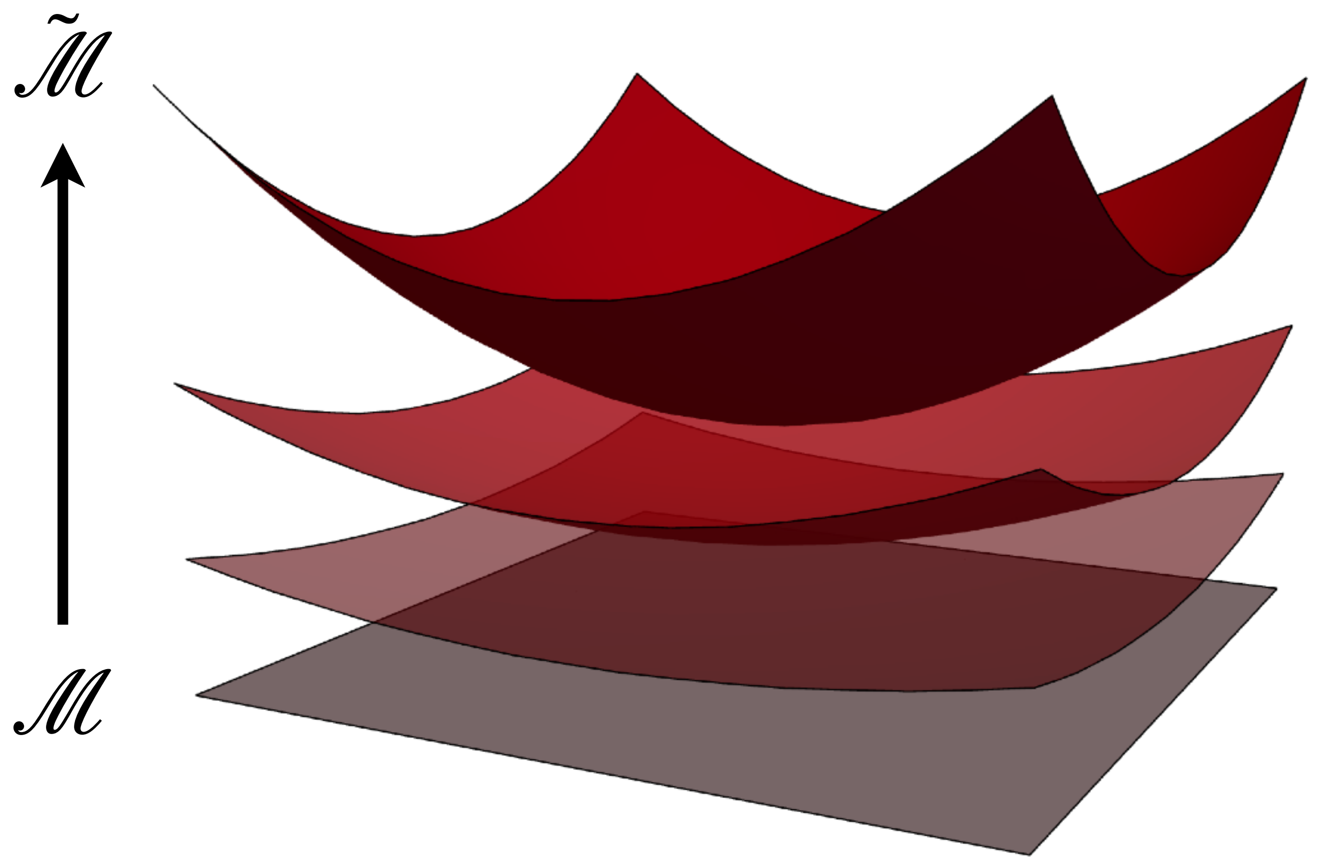
[Wagman 1711.00062]



Path integral deformations

Lattice integrands are often holomorphic, allowing the integration contour to be deformed without bias.

$$\langle \mathcal{O} \rangle = \frac{1}{Z}\int_{\mathcal{M}} e^{-S(\phi)}\mathcal{O}(\phi) = \frac{1}{Z}\int_{\tilde{\mathcal{M}}} e^{-S(\tilde{\phi})}\mathcal{O}(\tilde{\phi})$$



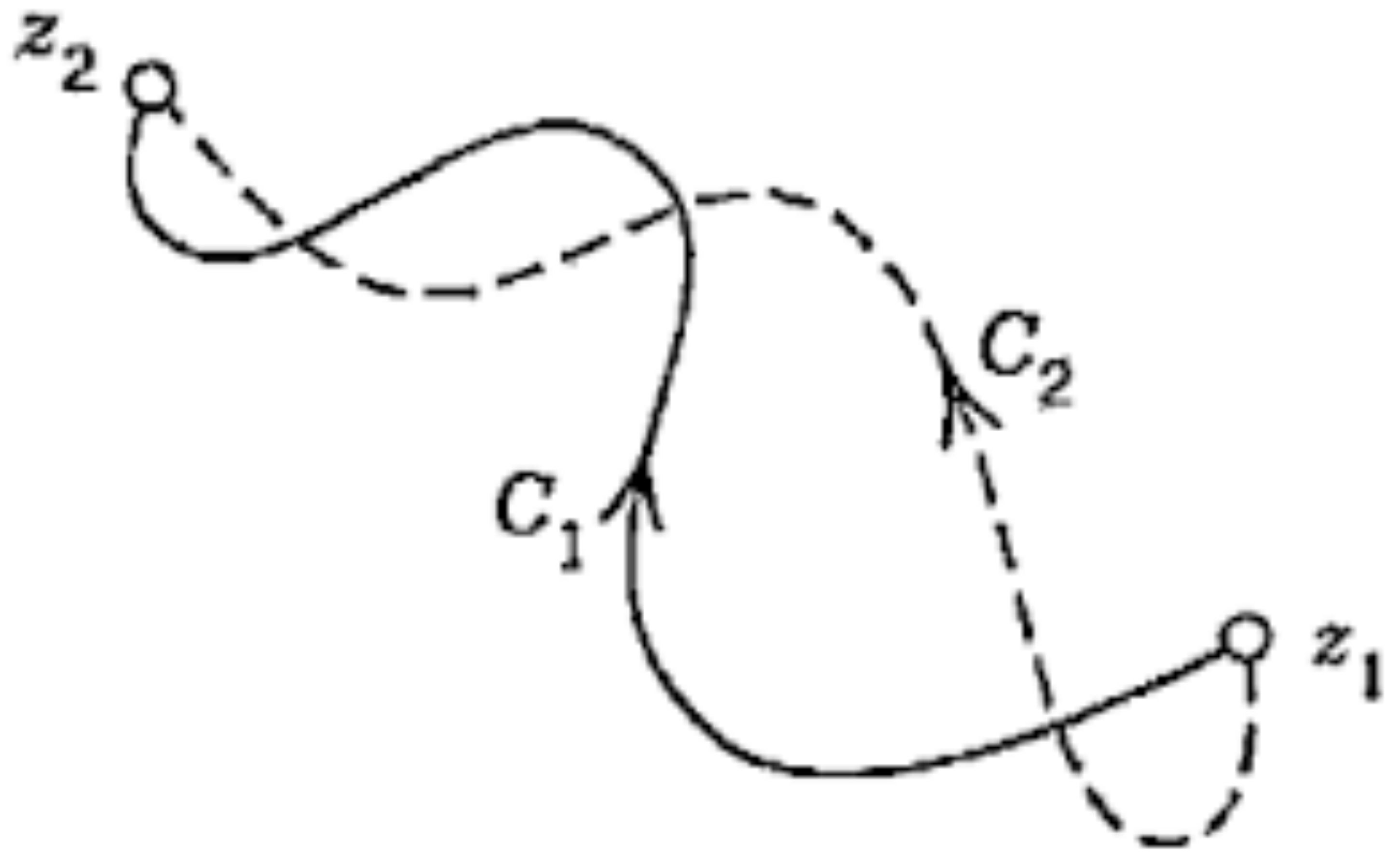
Connection to Lefschetz thimbles, Complex Langevin

Many works applying this to oscillatory path integrals originating from complex $S(\phi)$ (e.g. non-zero density, real time)

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

[Image credit: Neill Warrington]

Can we leverage contour deformations to shrink $\mathrm{Var}[\mathcal{O}]$ while preserving $\langle \mathcal{O} \rangle$?



Case study: Gaussian StN problem

Toy example: Simple observable in a Gaussian "theory"

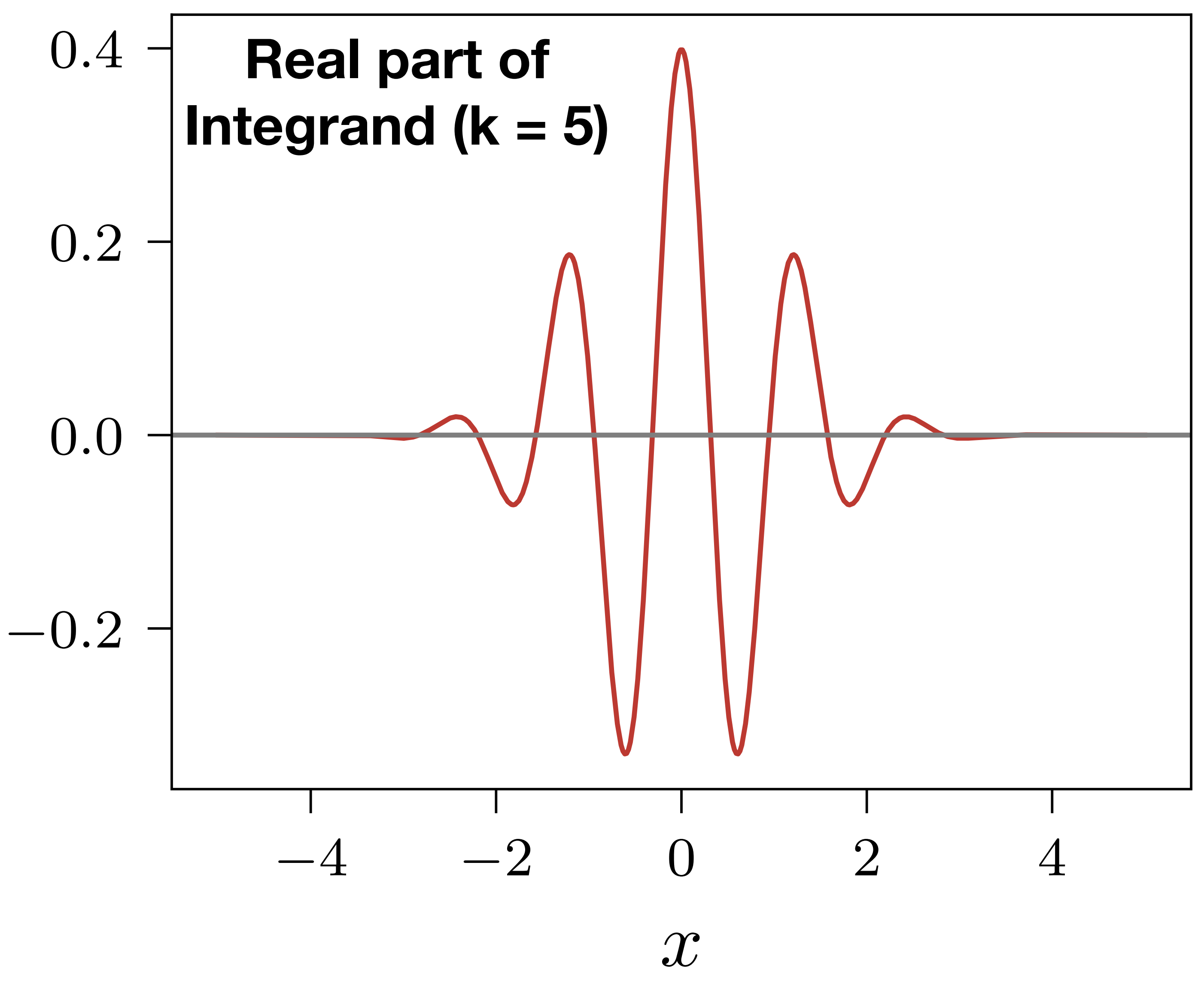
$$\left\langle e^{ikx} \right\rangle = \frac{1}{Z} \int_{-\infty}^{\infty} dx [e^{ikx}] e^{-x^2/2}$$

Monte Carlo approach: Sample $p(x) = e^{-x^2/2}/Z$, measure e^{ikx}

Signal: $\left\langle e^{ikx} \right\rangle = e^{-k^2/2}$

Variance: $\left\langle |e^{ikx}|^2 \right\rangle = 1$

StN and sign problem!



Case study: Gaussian StN problem

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

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

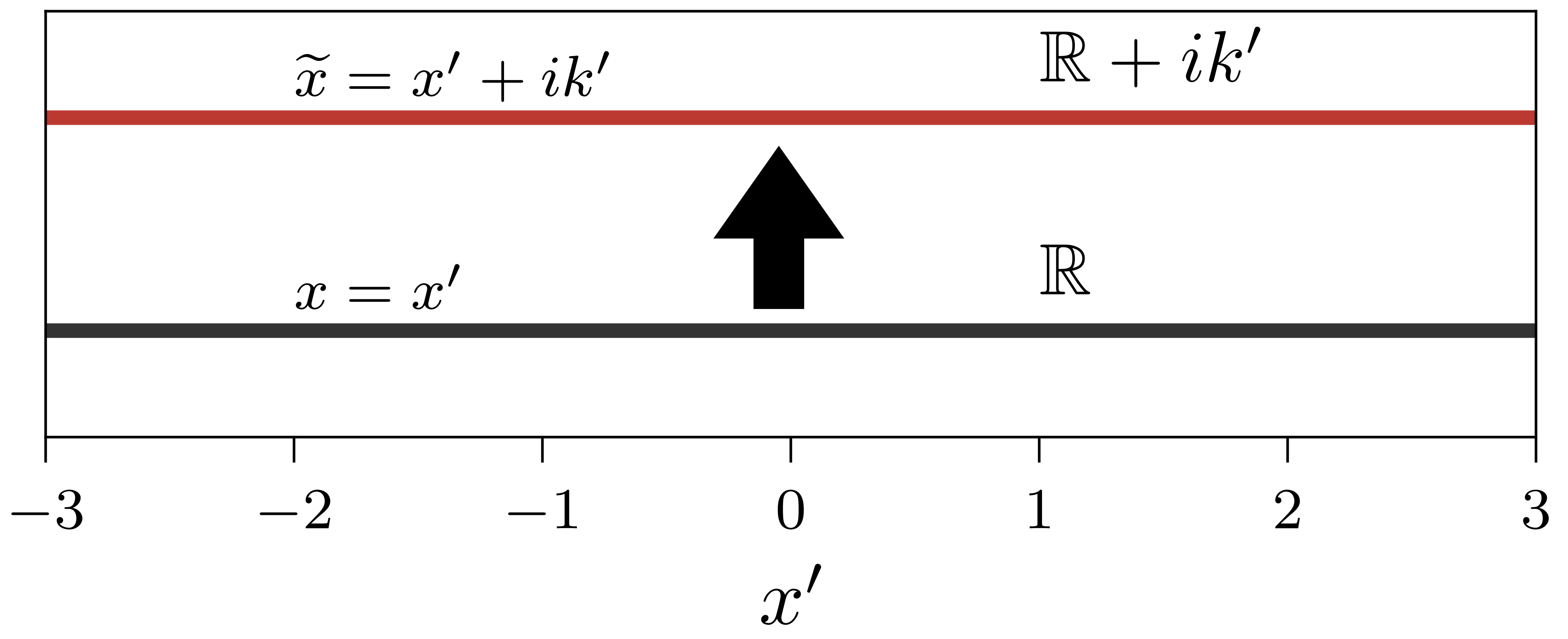
$$= \frac{1}{Z} \int_{\mathbb{R}+ik'} d\tilde{x} \, [e^{ik\tilde{x}}] e^{-\frac{1}{2}\tilde{x}^2}$$

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

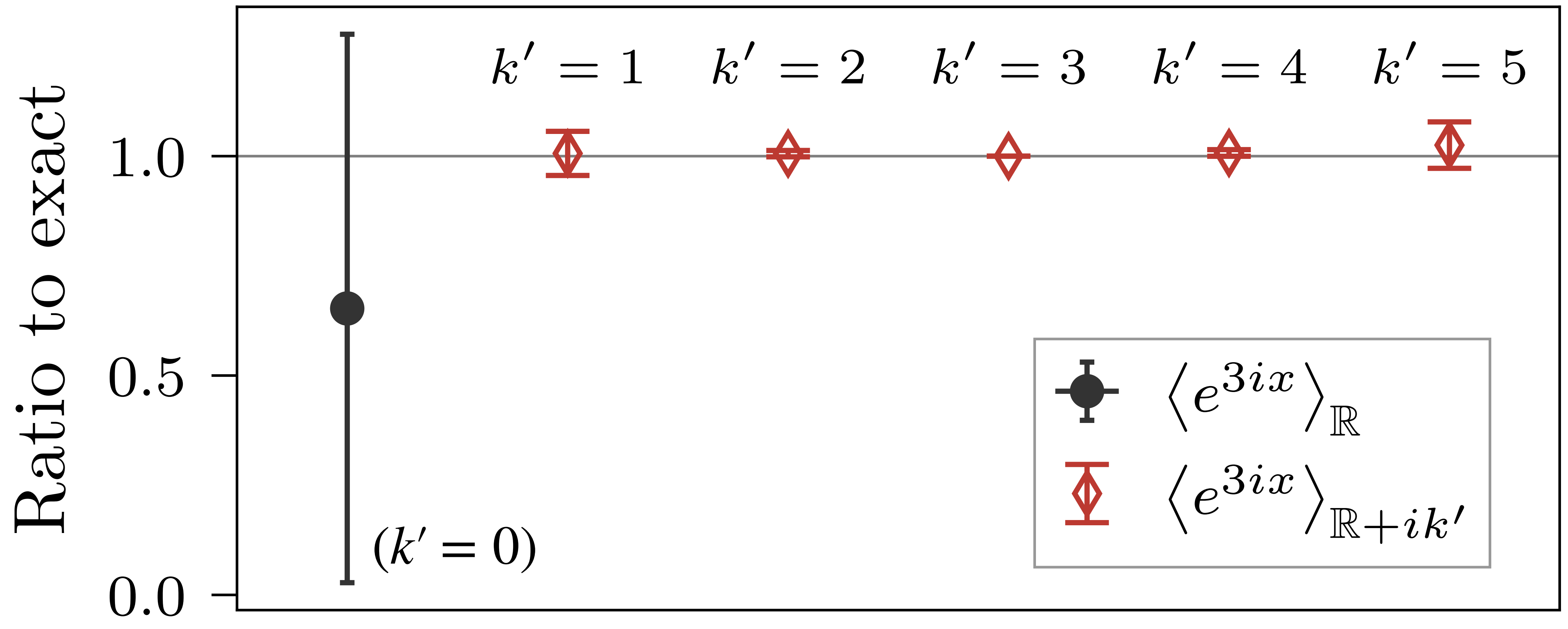
$$= \left\langle e^{ix'(k-k')} e^{\frac{1}{2}k'^2 - kk'} \right\rangle$$



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 $kk' > 0$.



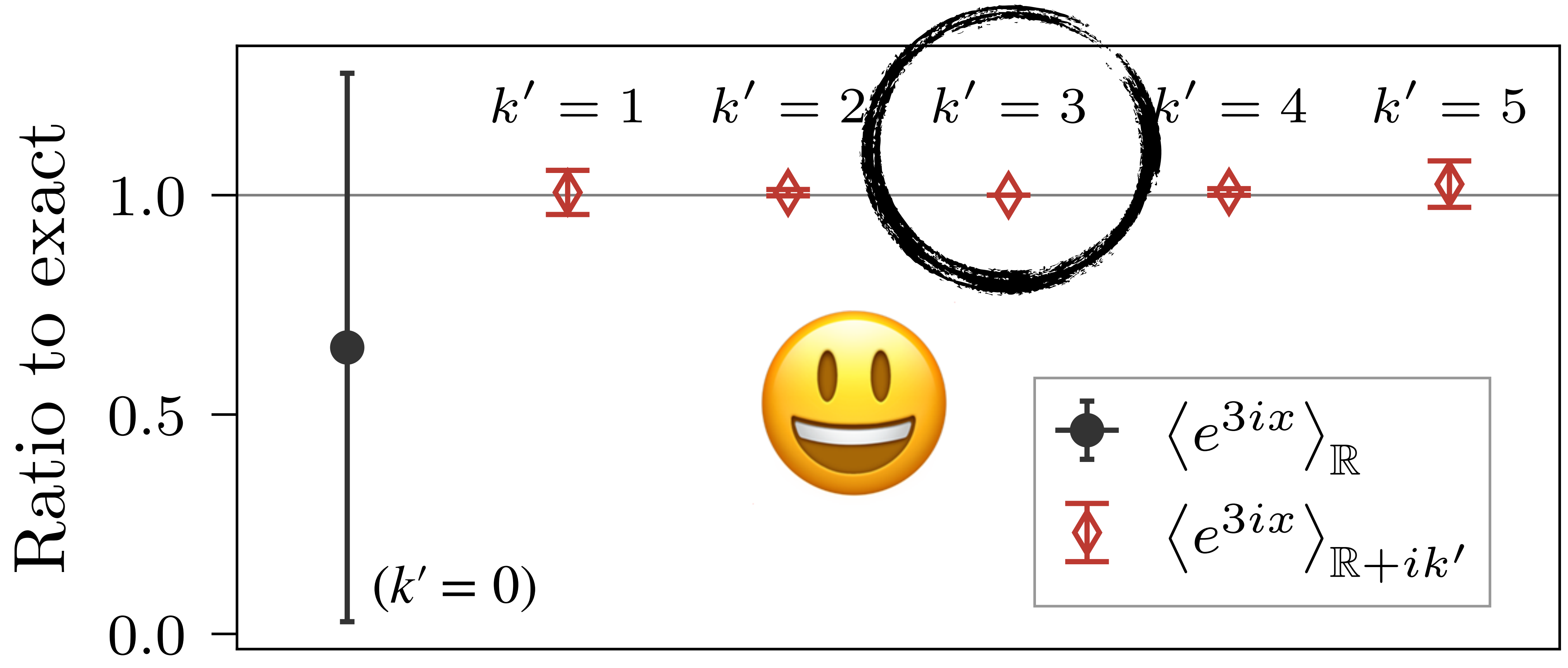
Exactness preserved: Anti-correlated phase fluctuations from the deformed action $e^{-\frac{1}{2}x'^2 - ix'k' + \frac{1}{2}k'^2}$.

Case study: Gaussian StN problem

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 $e^{-\frac{1}{2}x'^2 - ix'k' + \frac{1}{2}k'^2}$.

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



Integral deformations for noisy observables

Detmold, GK, Wagman, Warrington PRD102 (2020) 014514

- Deformed path integral defines a **modified observable**:

$$\mathcal{Q}(\phi) \equiv \det J(\phi) e^{-[S(\tilde{\phi}(\phi)) - S(\phi)]} \mathcal{O}(\tilde{\phi}(\phi))$$

- Identical Monte Carlo expectation values $\langle \cdot \rangle$, different variance:

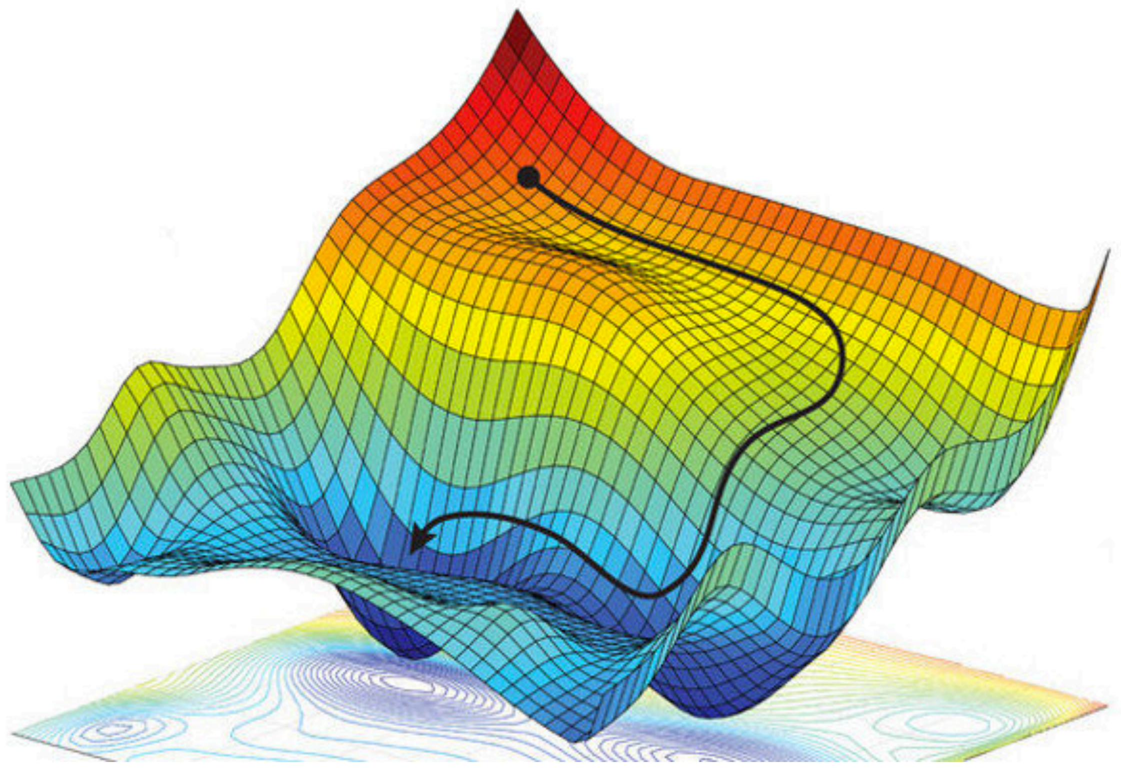
$$\langle \mathcal{Q}(\phi) \rangle = \langle \mathcal{O}(\phi) \rangle$$

$$\mathrm{Var}[\mathcal{Q}(\phi)] \neq \mathrm{Var}[\mathcal{O}(\phi)]$$

Variance may be reduced by a good choice of deformation

Learning the integration contour

The choice of $f : \phi \mapsto \tilde{\phi}$ defines $\tilde{\mathscr{M}}$, $\mathscr{Q}(\phi)$, and the variance.



[Image credit: 1805.04829]

Parameterize $f(\phi; \omega)$ then **minimize variance**.

- Caveat: Complex analyticity
- Caveat: $SU(N)$ variables

Insight: gradients of variance w.r.t. ω can be defined using **original Monte Carlo ensemble**.

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

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

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

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 $\Omega \equiv \underbrace{(\phi_1, \ldots, \theta_1, \ldots)}_{(N^2-1)\text{ angles}}$, $\phi_i \in [0,2\pi]$ and $\theta_i \in [0,\pi/2]$

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

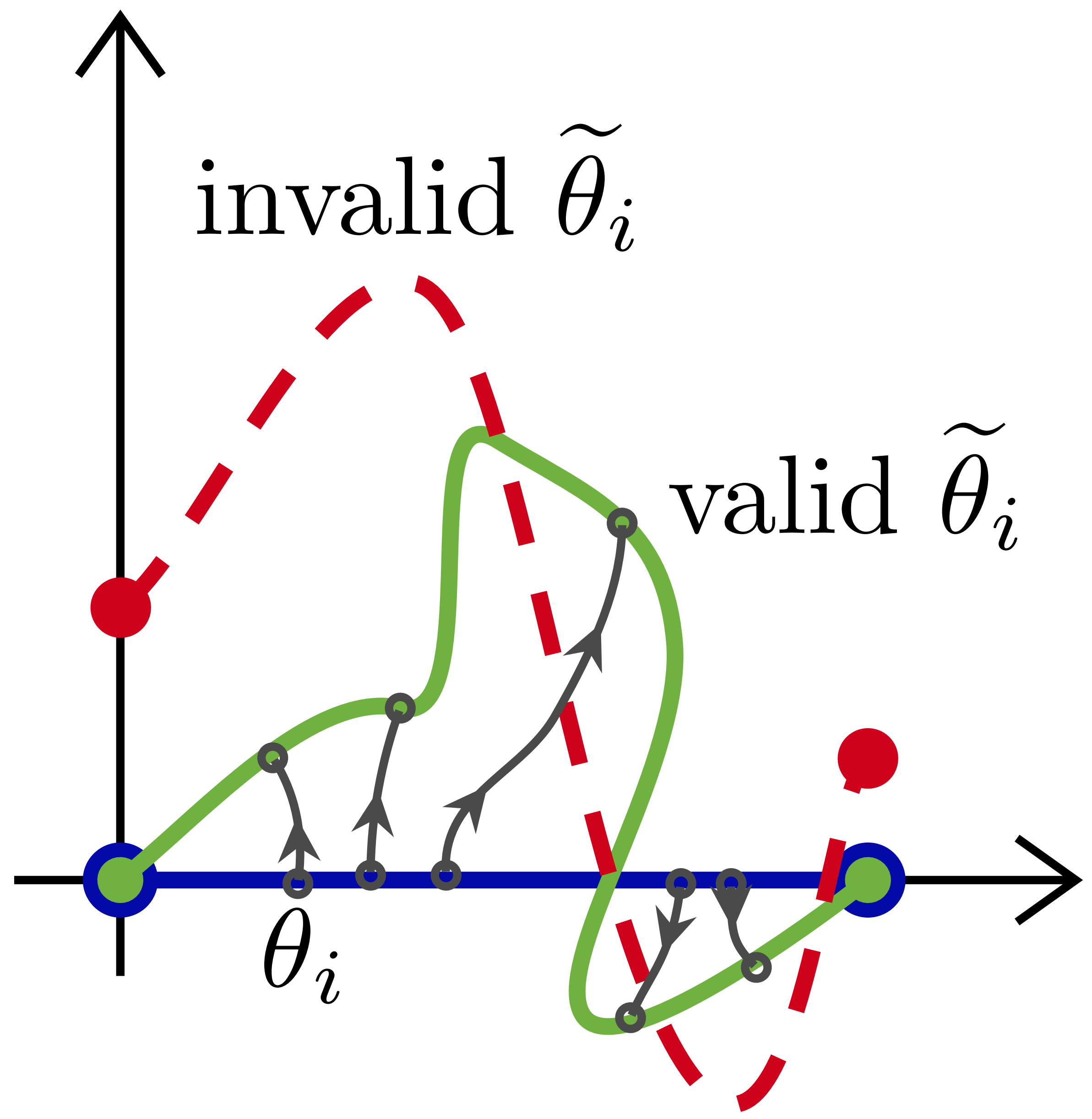
Deforming angular variables

Angular parameterization of $SU(N)$ has two types of angles:

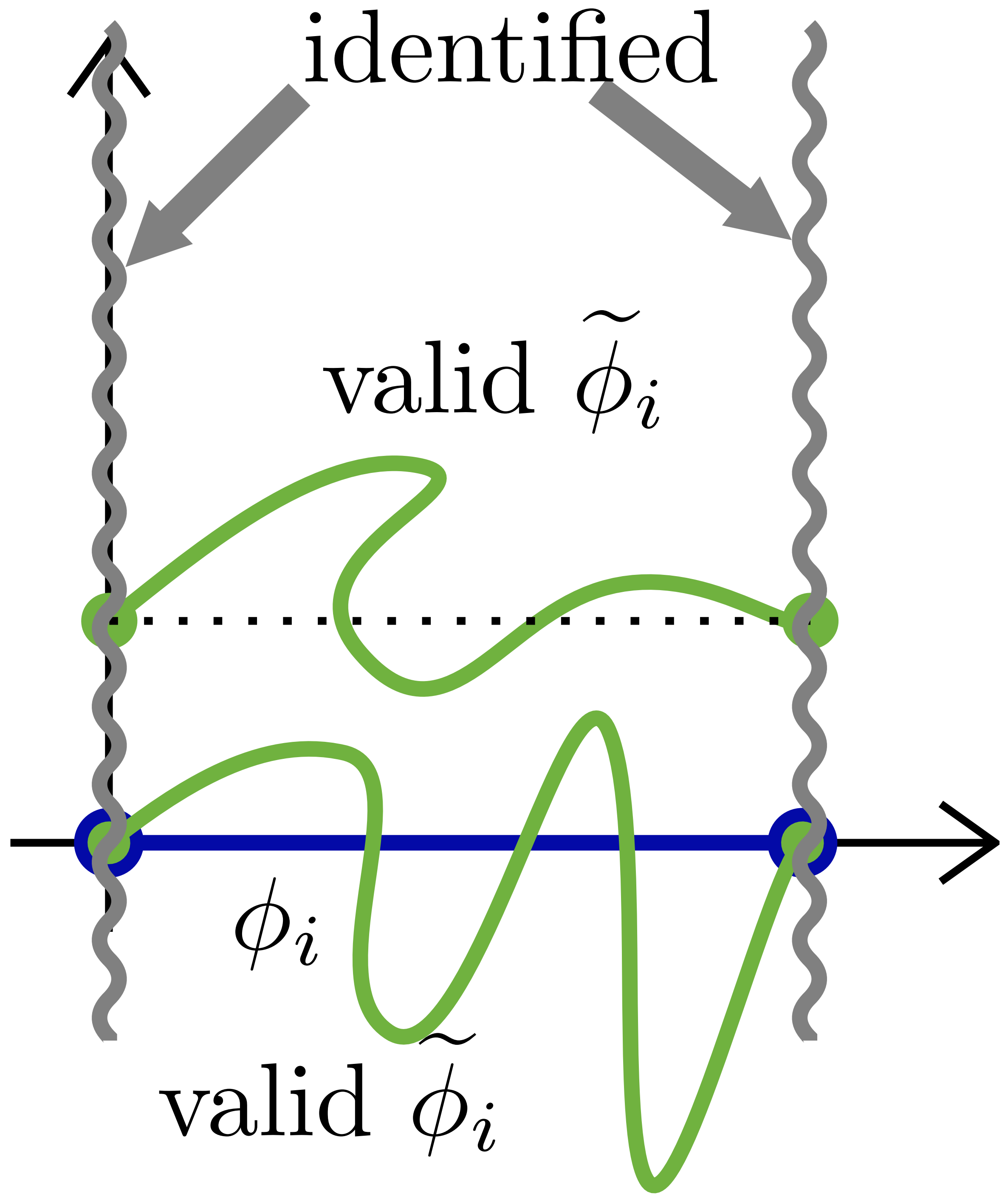
- **Azimuthal** angles $\phi_j \in [0,2\pi]$
- **Zenith** angles $\theta_i \in [0,\pi/2]$

In each case, must deal appropriately with **endpoints.**

Zenith angles

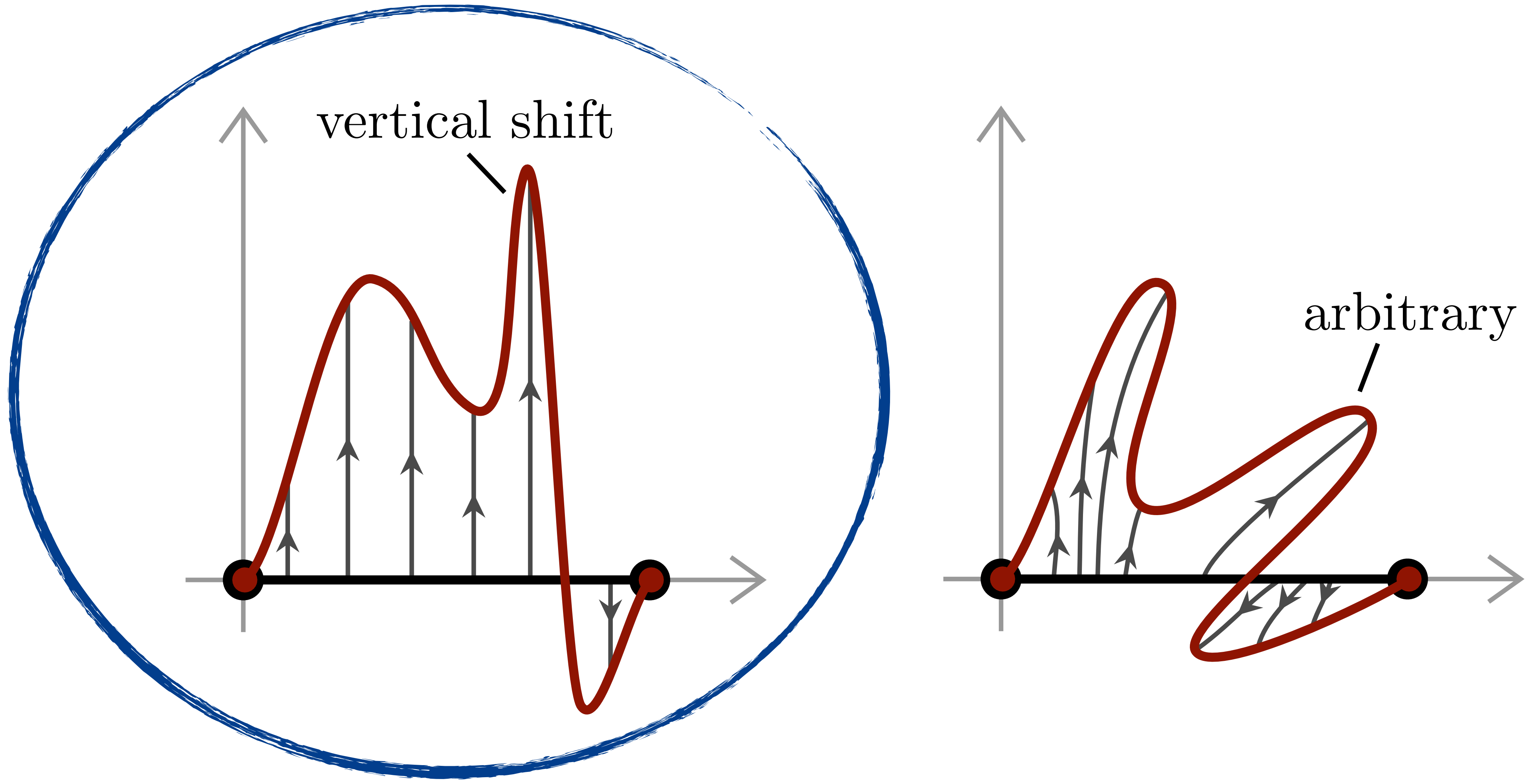


Azimuthal angles



Deformation

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



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

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

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

Original real part

Parameterized imaginary shift

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

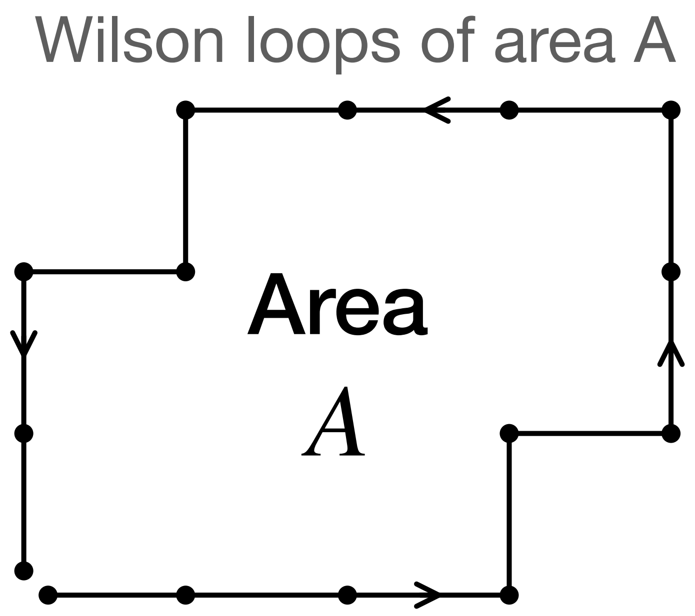
$$f_{\phi^a} = \sum_{m=1}^{\Lambda} \kappa_m^{xy;a}\sin(m\phi^a_y + \zeta_m^{xy;a})\left\{ 1 + \sum_{n=1}^{\Lambda}\left[\sum_{r=1}^{3} \lambda_{mn}^{xy;ar}\sin(2n\theta^r_y) + \sum_{\substack{s \neq a \\ s=1}}^{5}\eta_{mn}^{xy;as}\sin(n\phi^s_y + \chi_{mn}^{xy;as})\right]\right\}.$$

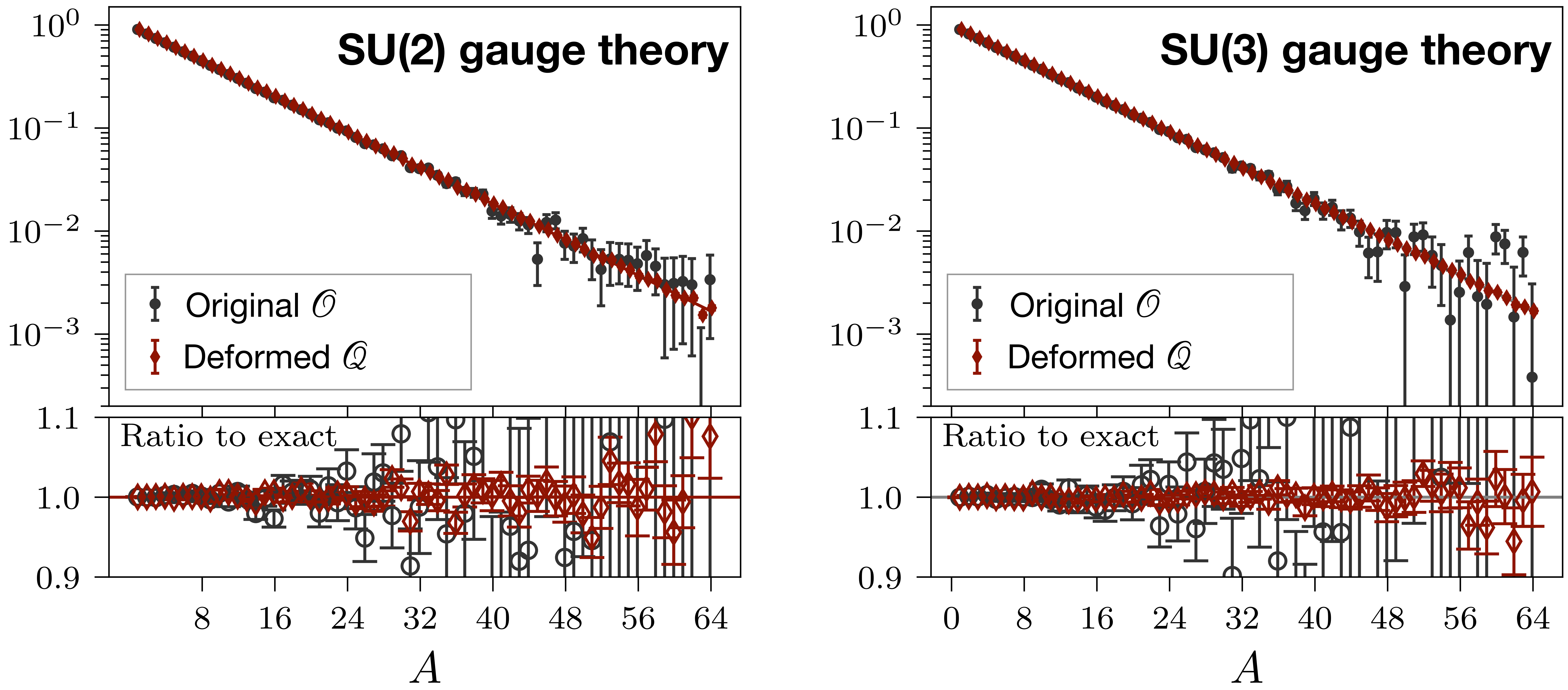
Triangular Jacobian: $f(\Omega)$ only allowed to depend on $y \le x$. Jacobian determinant calculable in $O(V)$.

This is key for scalability!

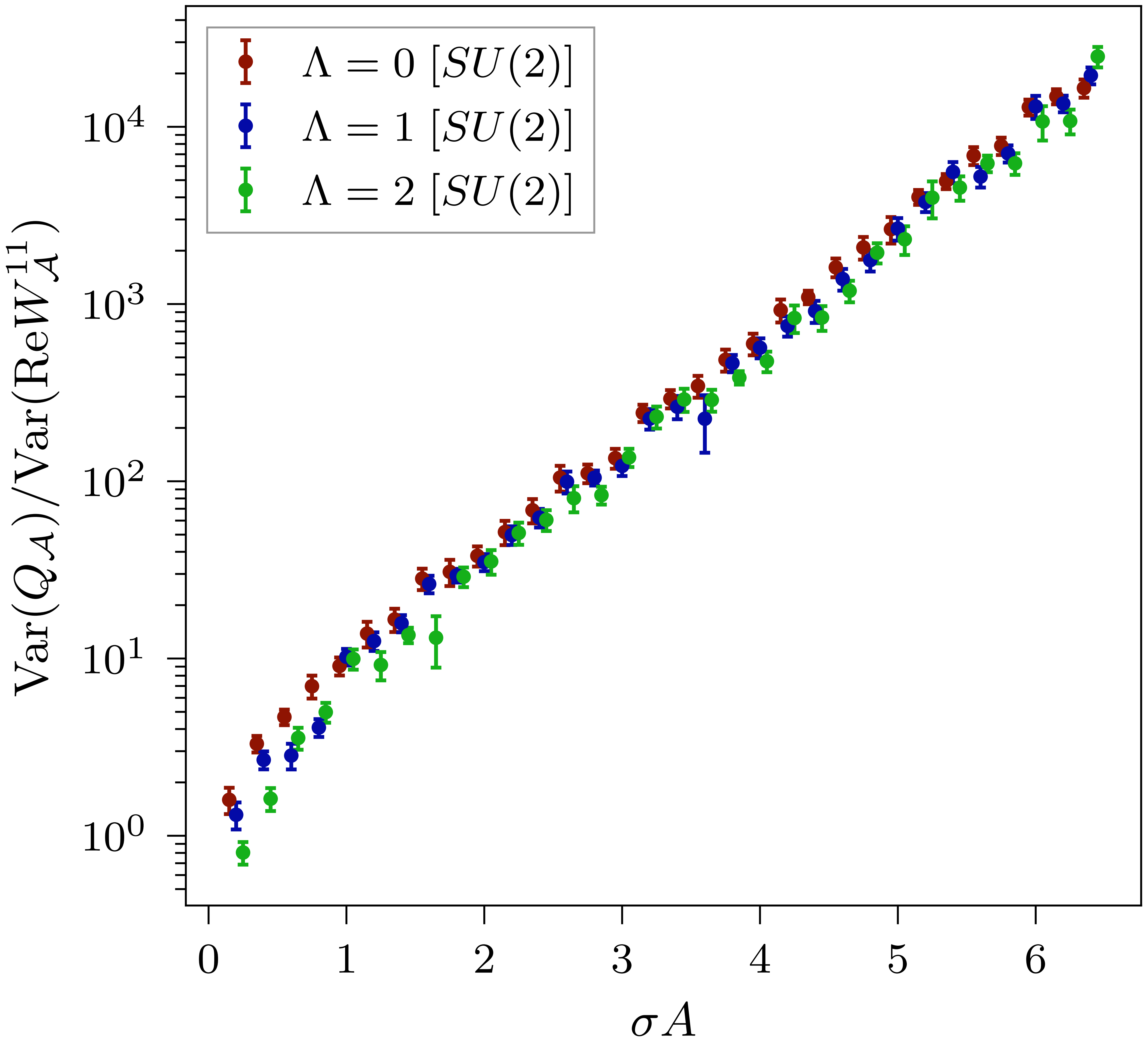
Deformations crush 1+1D noise problems

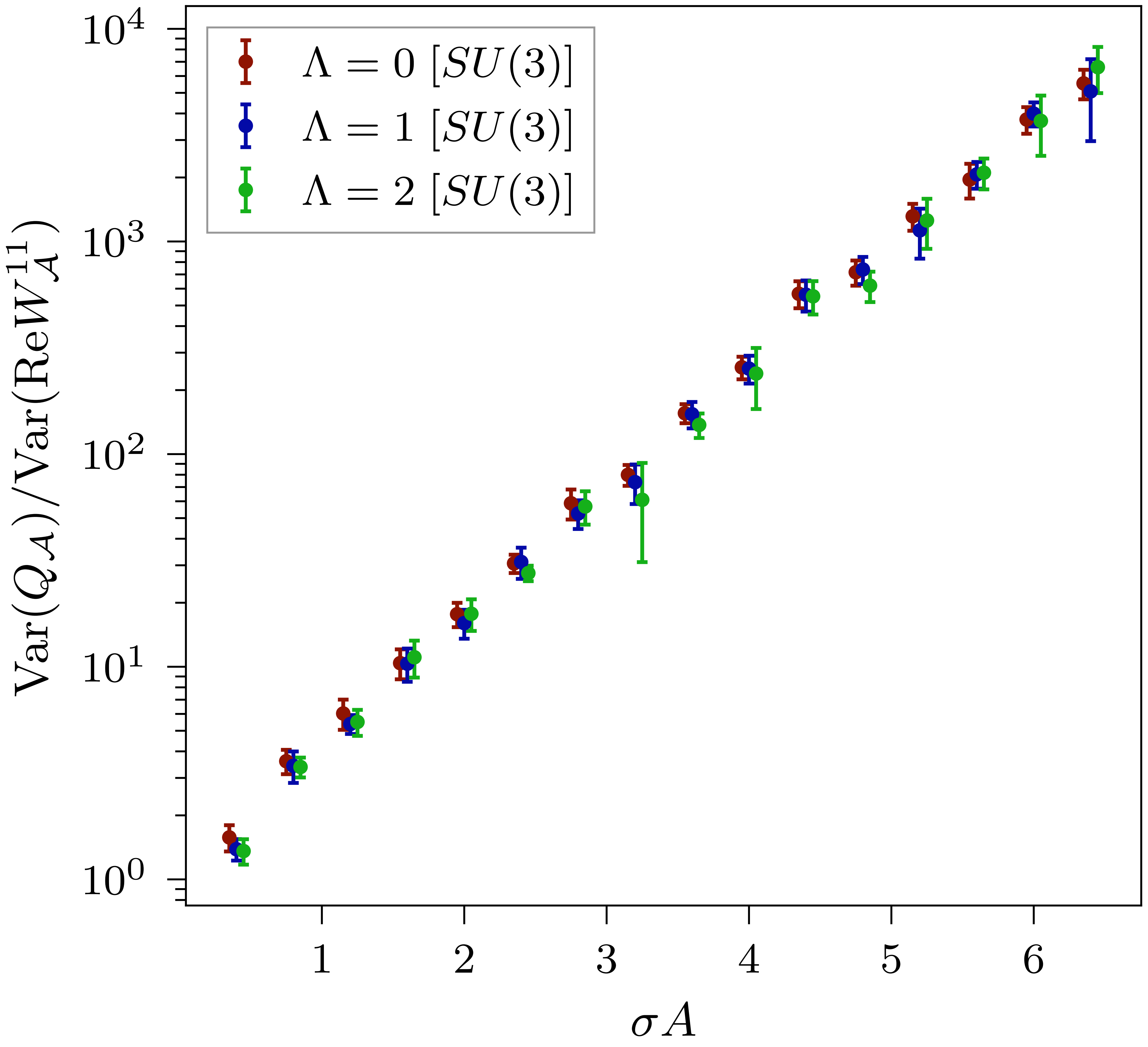
Detmold, GK, Lamm, Wagman, Warrington PRD103 (2021) 094517





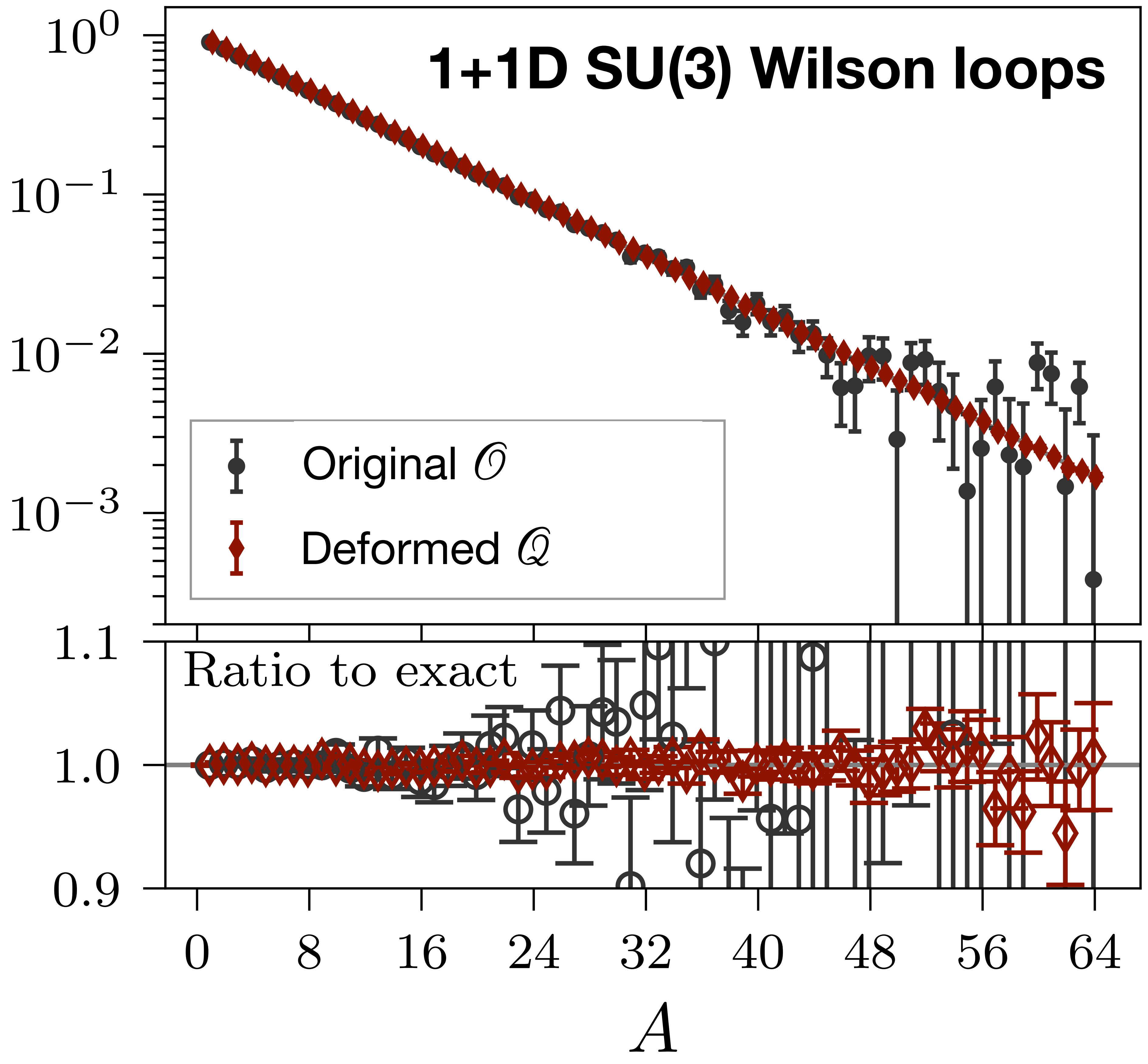
No scaling with Fourier cutoff





Restricted "constant shift" deformations

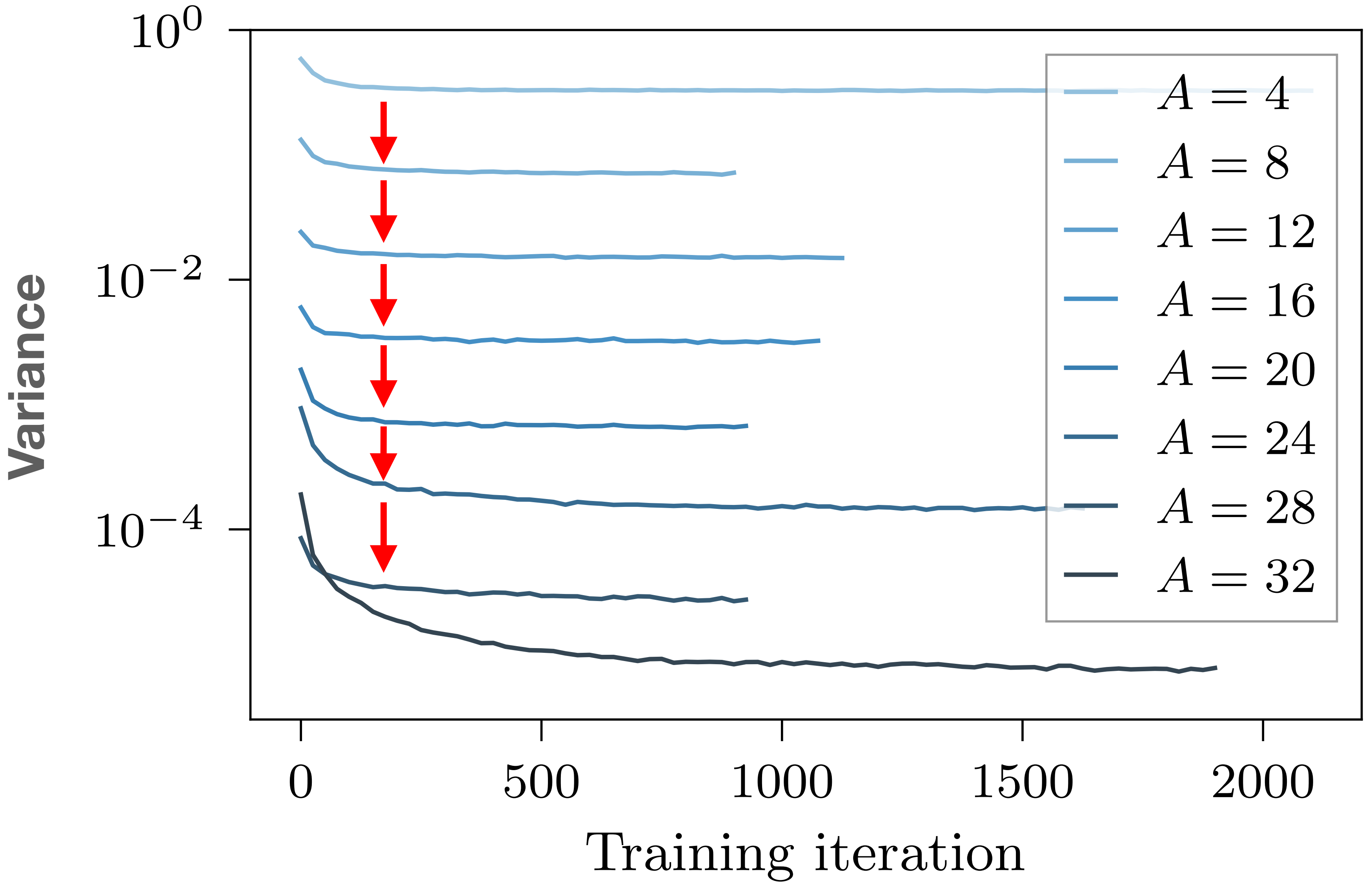
- Deform only periodic angular variables $\phi_i \in [0,2\pi]$
- $\tilde{\phi}_i = \phi_i + i\lambda$
- Field-independent, but spacetime dependent $\Longrightarrow O(V)$ learnable parameters



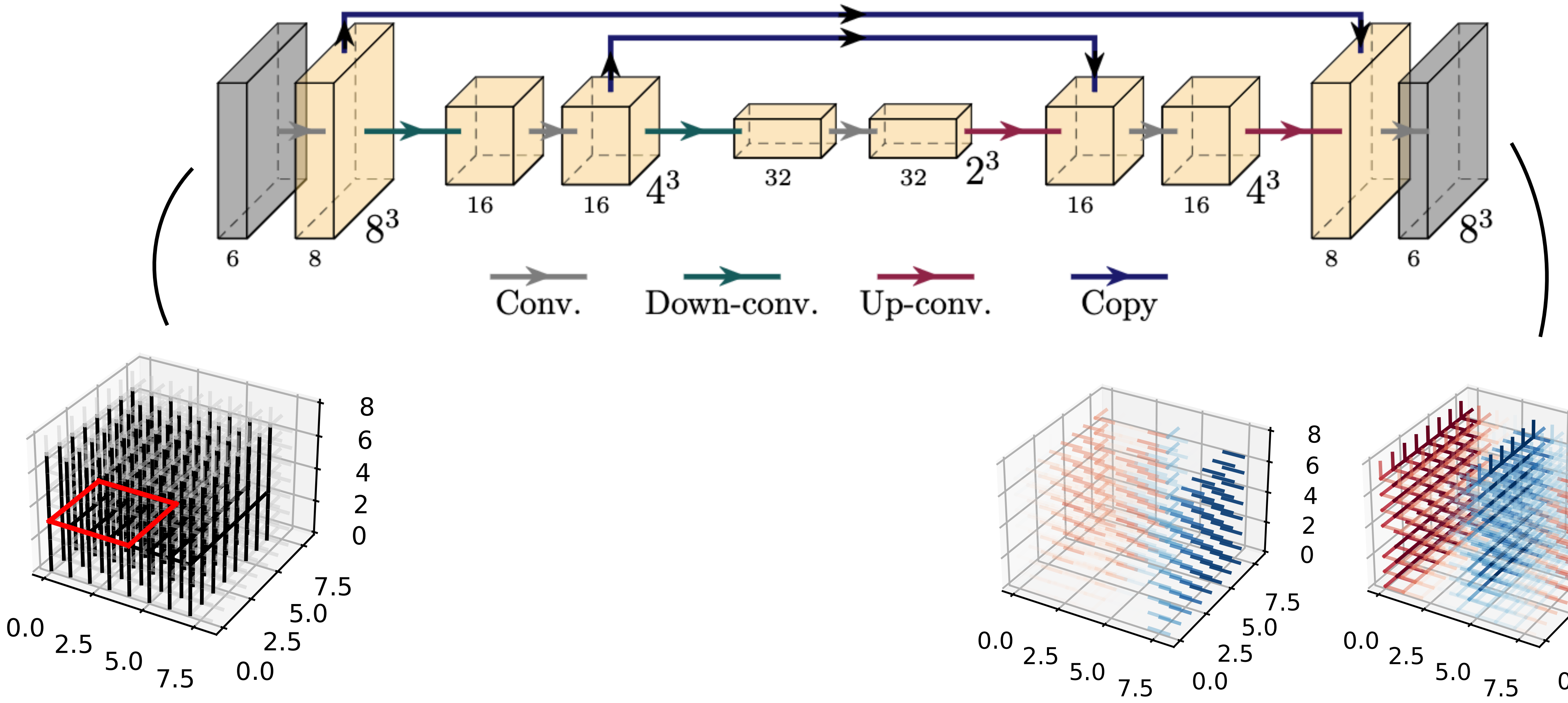
Direct vs. U-net parameterization

Direct parameterization with $O(V)$ learnable parameters leads to **overtraining** in typical problems!

1. Transfer learning



2. Indirect U-net parameterization



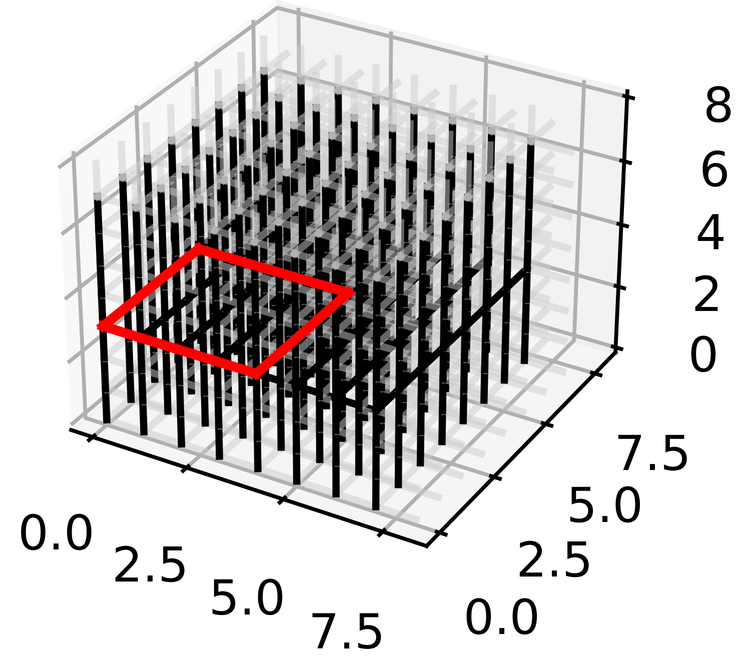
Wilson loop and gauge fixing scheme

Learned shifts

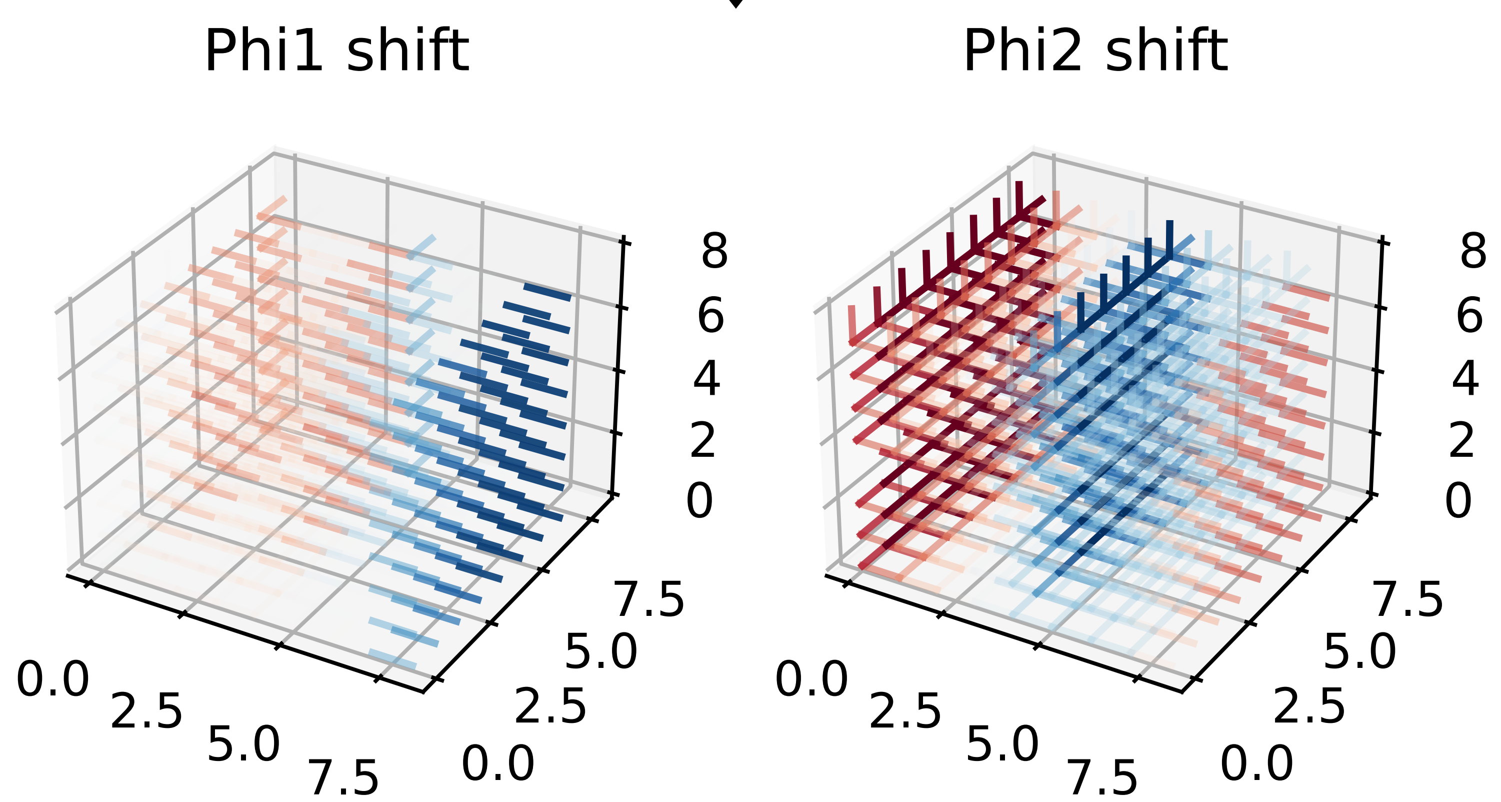
Detmold, GK, Lin, Shanahan, Wagman (2023) 2309.00600, NeurIPS ML4PS (2023)

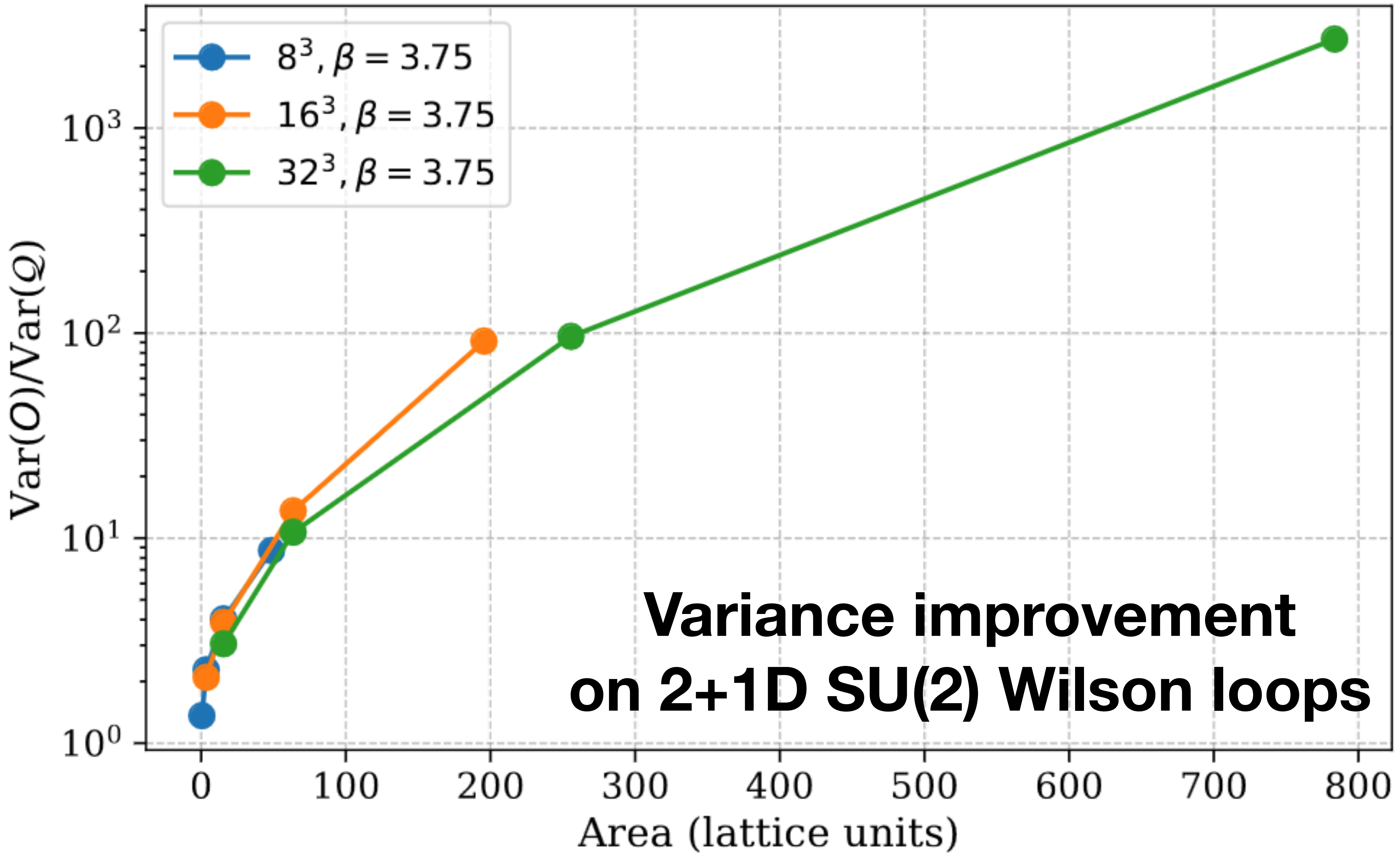
U-nets allow training for 2+1D problems

Wilson loop and
gauge fixing scheme



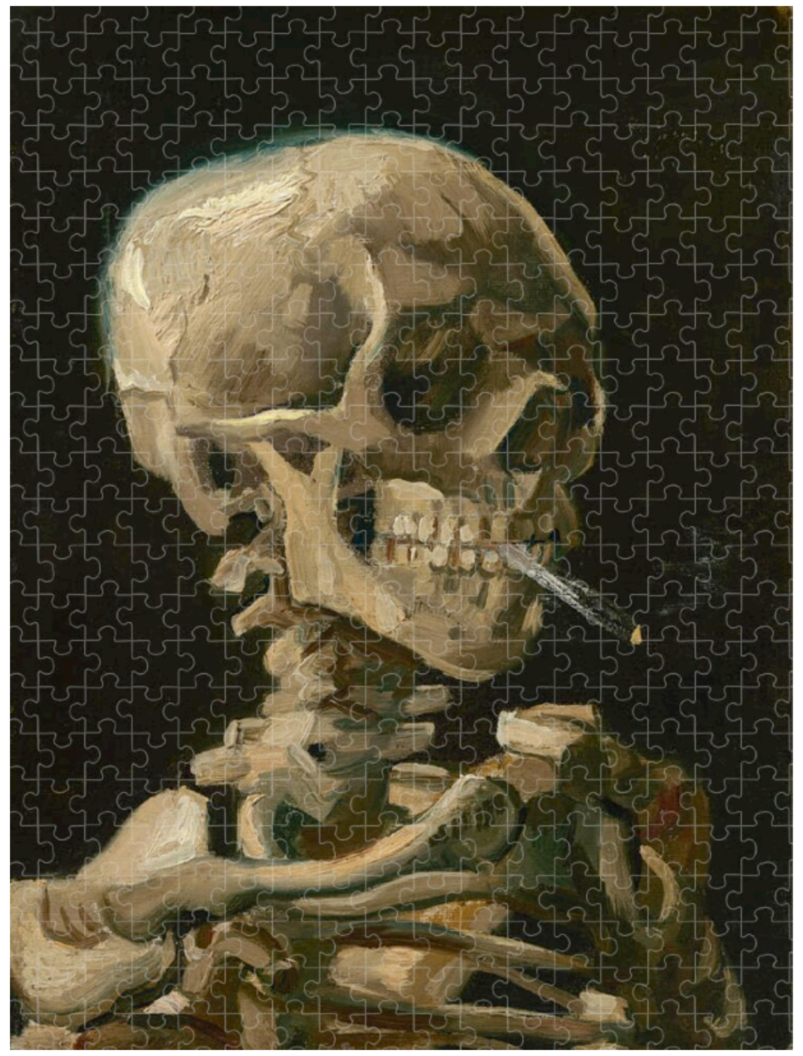
Learned
shifts





Variance improvement on 2+1D SU(2) Wilson loops

Gauge fixing skeletons in the closet



In 1+1D, could change basis to plaquettes plus gauge dofs:

$$\{U_\mu(x)\} \leftrightarrow \{U_p(x), \Omega(x)\}$$

This is **not uniquely possible** in 2+1D and higher!

- Either work with subset of plaquettes with a complicated change-of-basis
- Or, simpler, perform maximal-tree gauge fixing and use remaining links as dofs

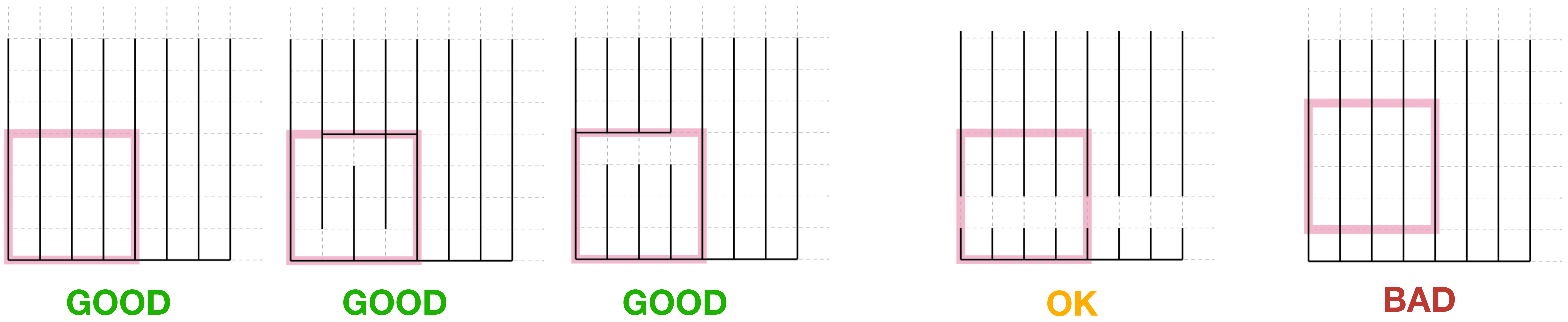
Have explored both options — No significant benefit to former, more flexibility provided by the latter. **Which maximal tree** is a hyperparameter to optimize.

Detmold, GK, Lin, Shanahan, Wagman (2024) In progress

Maximal tree gauges

Defined by a subset of lattice links containing **no closed loops** to be fixed to $\mathbb{I}$

Interplay between Wilson loop geometry and maximal tree choice

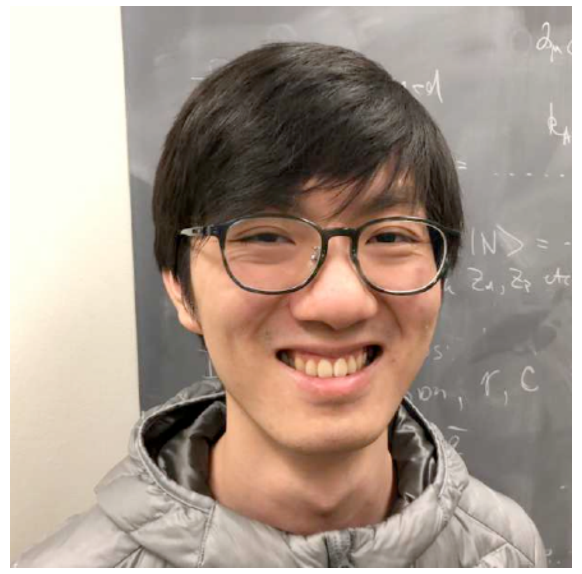


Exponential variance improvement with Wilson loop area

Worse performance as opening is moved lower

No variance improvements

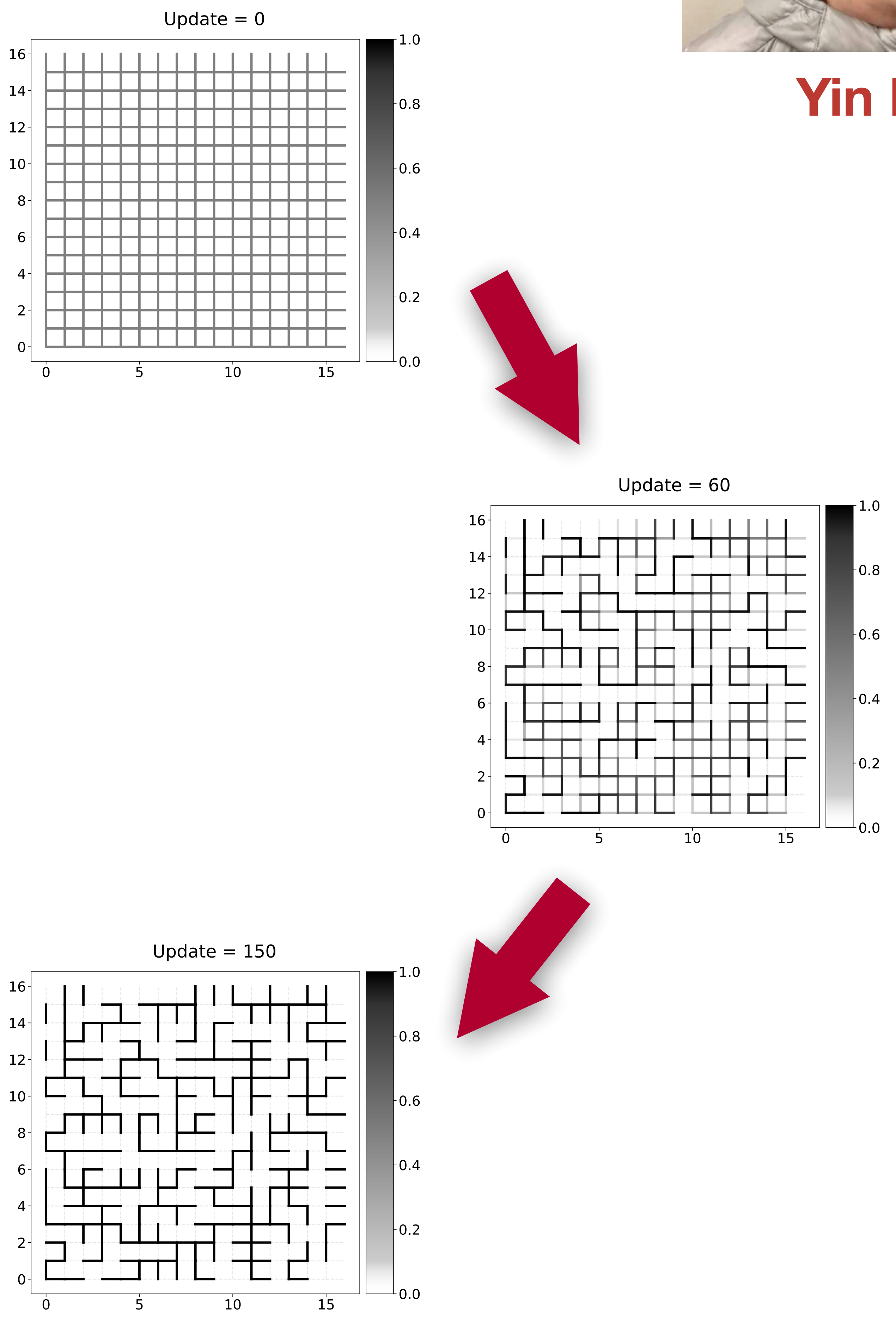
Could we “learn” the maximal tree?



Yin Lin

We define a parameterized gauge fixing functional

- Fix gauge by minimizing over gauge orbit
- Includes Coulomb, Landau, and all maximal tree gauges
- Can be optimized using adjoint state method



General gauge-fixing functional

$$E \propto -\sum_{x,\mu} \mathrm{Tr}\big(p_\mu(x) U_\mu^g(x)\big) \begin{cases} p_i(x) = 1 & \textbf{(Coulomb gauge)} \\ p_\mu(x) = 1 & \textbf{(Landau gauge)} \\ p_\mu(x) = k_\mu(x) \in \{0,1\} & \textbf{(Max. trees)} \end{cases}$$

See poster at Lattice ’24!

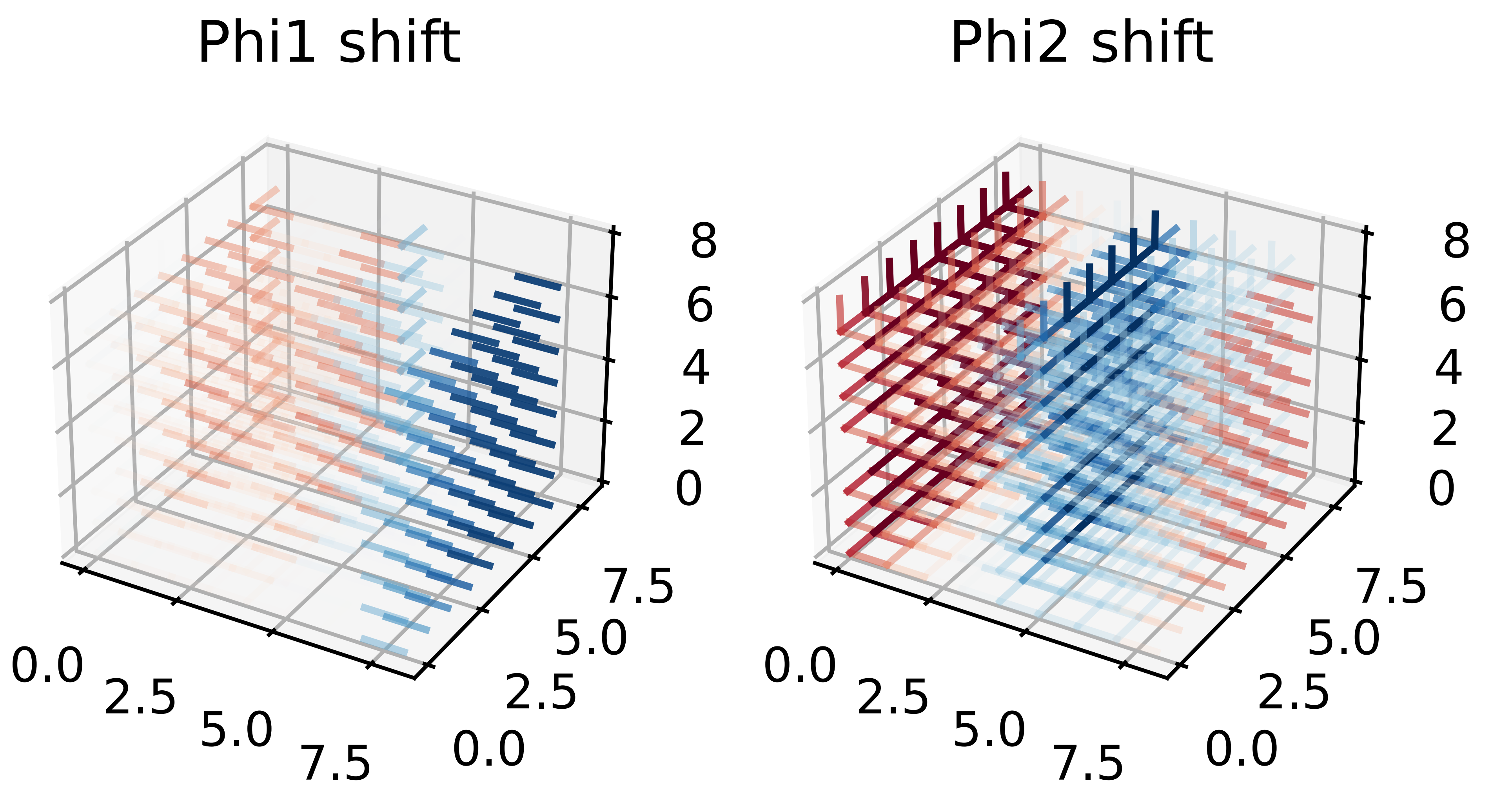
Summary

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.



Look out for a paper on 3d and 4d SU(N) results soon! Come chat at the poster session!

Summary

Using complex analysis we can …

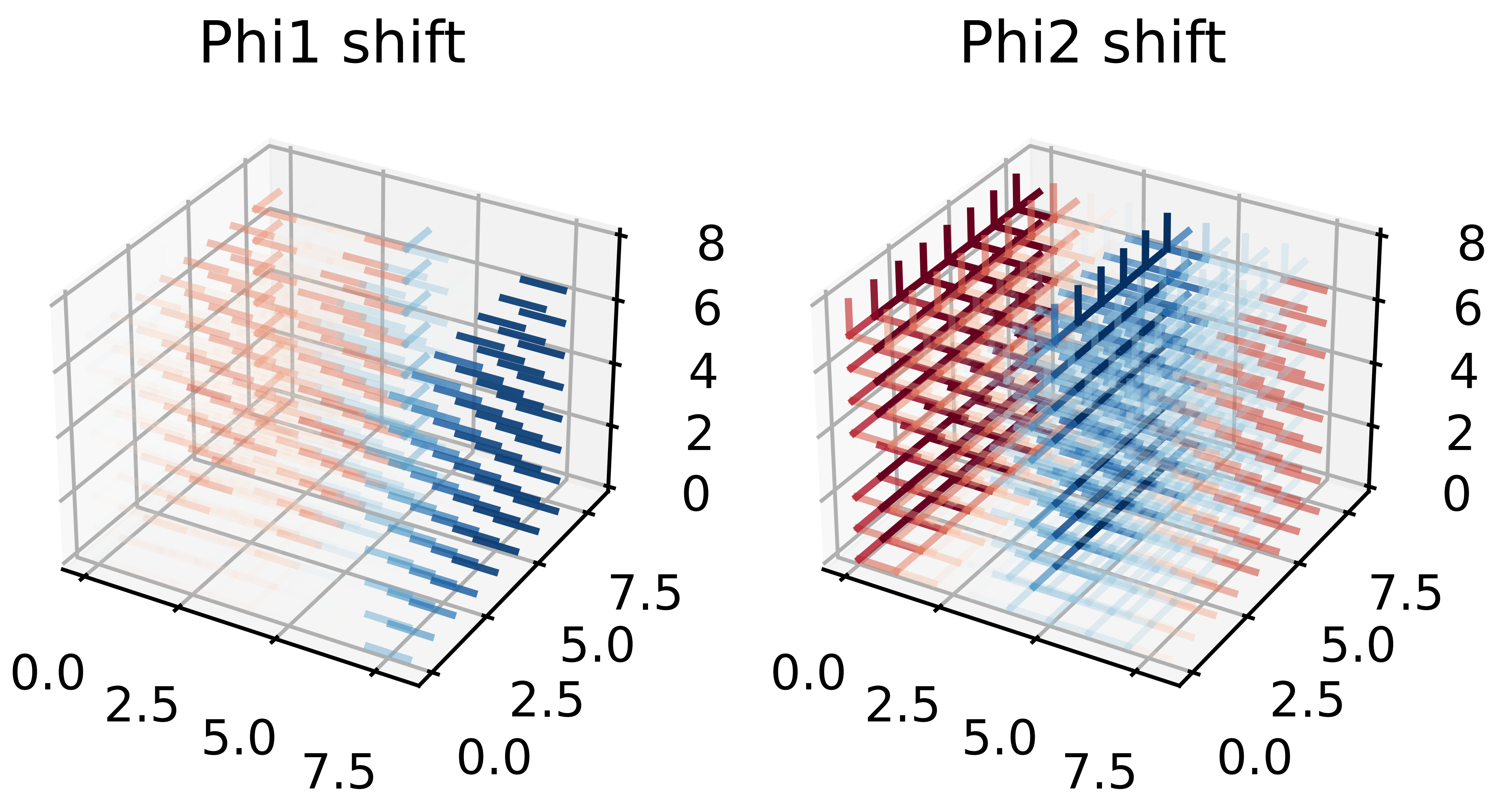
- **Deform observables $\mathcal{O} \to \mathcal{Q}$**, where $\langle \mathcal{O} \rangle = \langle \mathcal{Q} \rangle$ but $\mathrm{Var}[\mathcal{O}] \neq \mathrm{Var}[\mathcal{Q}]$.

  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.

Thanks! Questions?



Look out for a paper on 3d and 4d SU(N) results soon! Come chat at the poster session!

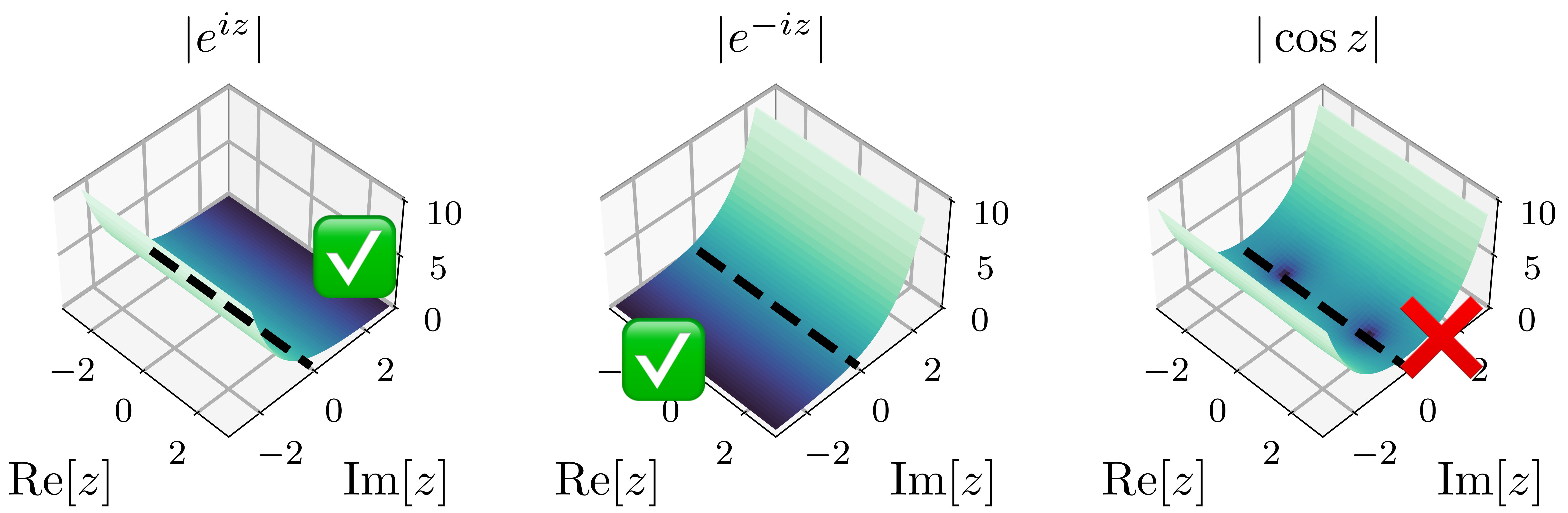
Backup slides

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

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

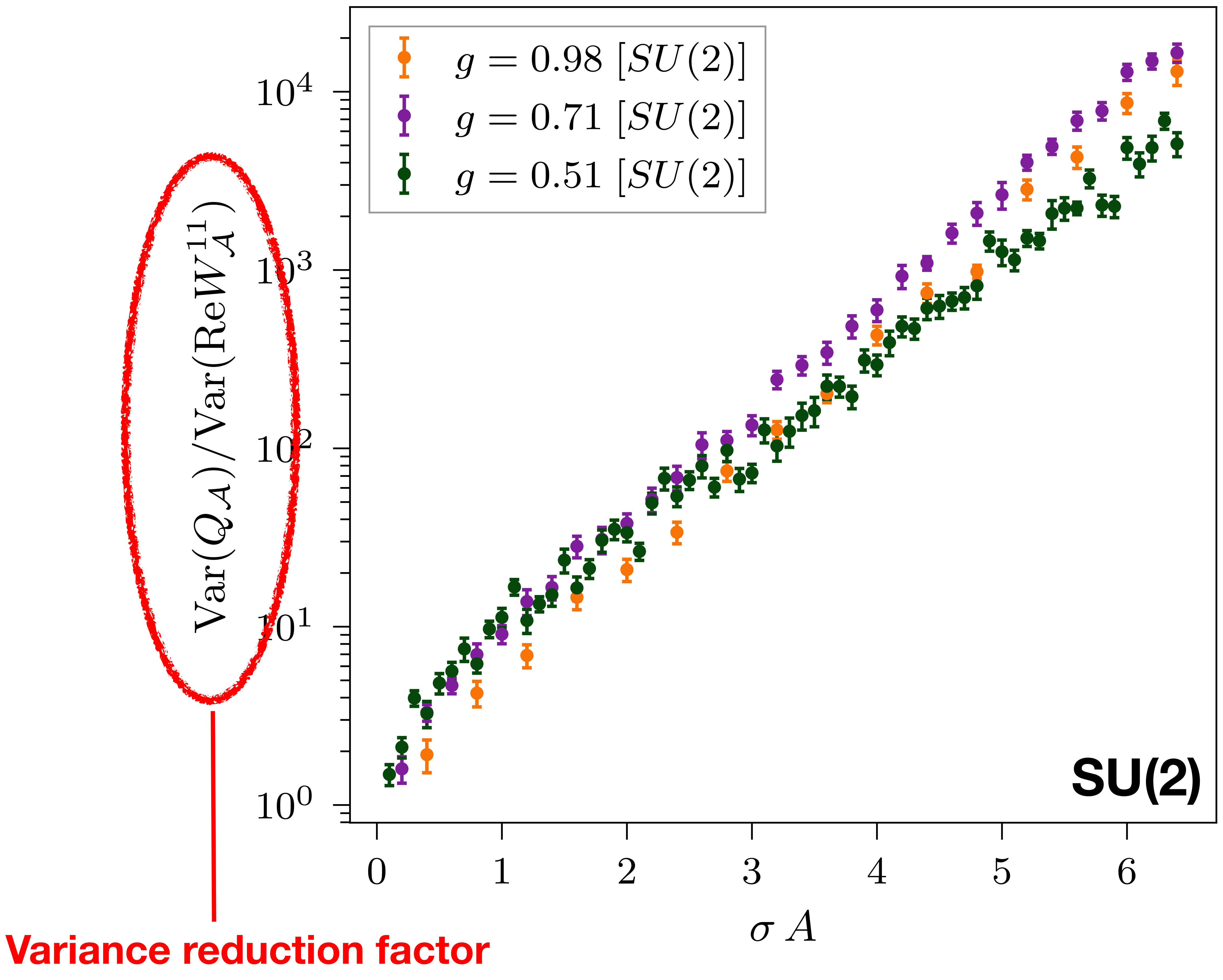
1D example: $$\int_0^{2\pi} \frac{dz}{2\pi} e^{iz}\, e^{\beta \cos z} = \int_0^{2\pi} \frac{dz}{2\pi} e^{-iz}\, e^{\beta \cos z} = \int_0^{2\pi} \frac{dz}{2\pi} \cos z\, e^{\beta \cos z} = I_1(\beta)$$

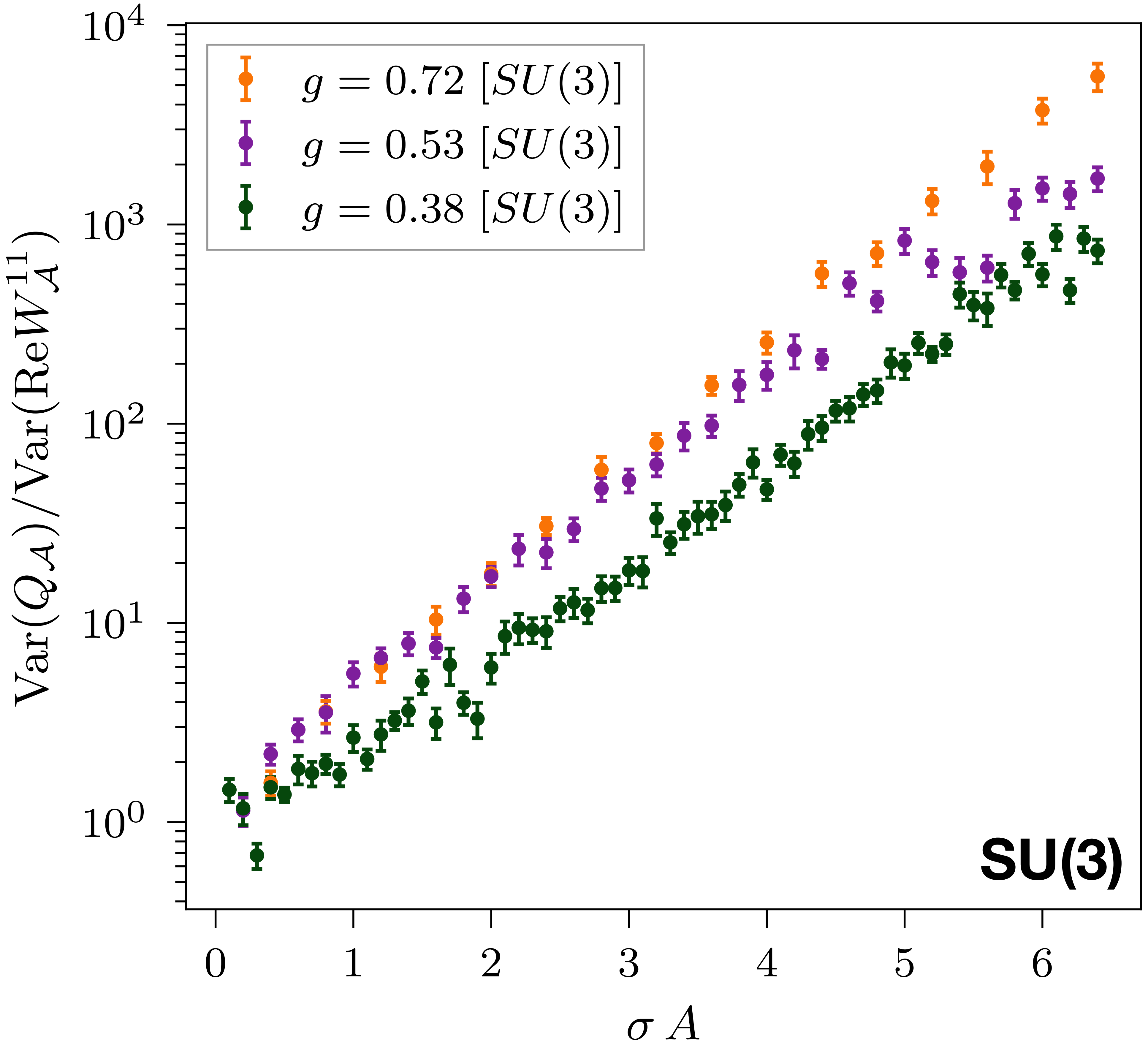
… however, some choices are better than others!



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

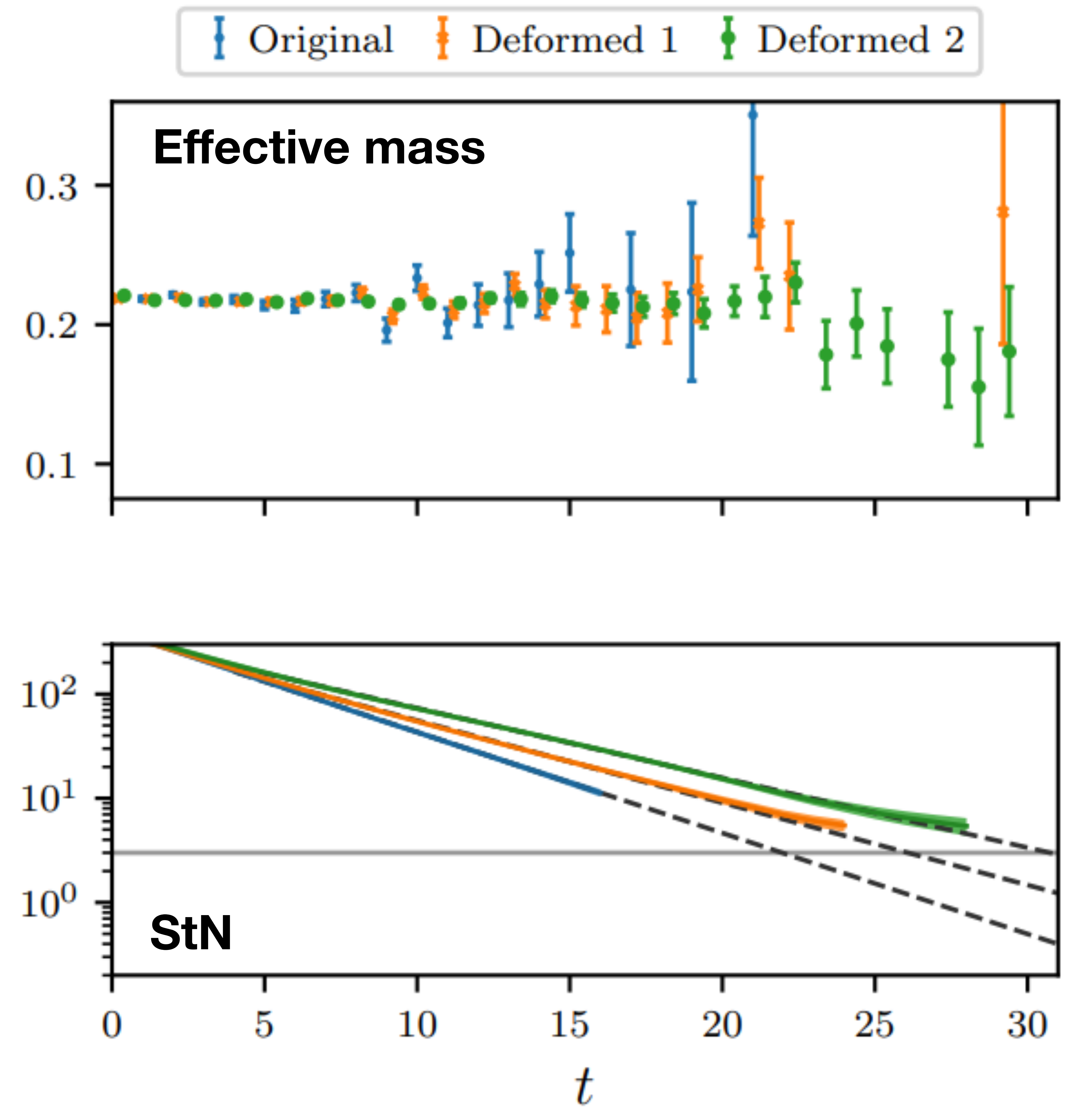
Using polar dofs in the path integral

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

and the holomorphic action

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

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



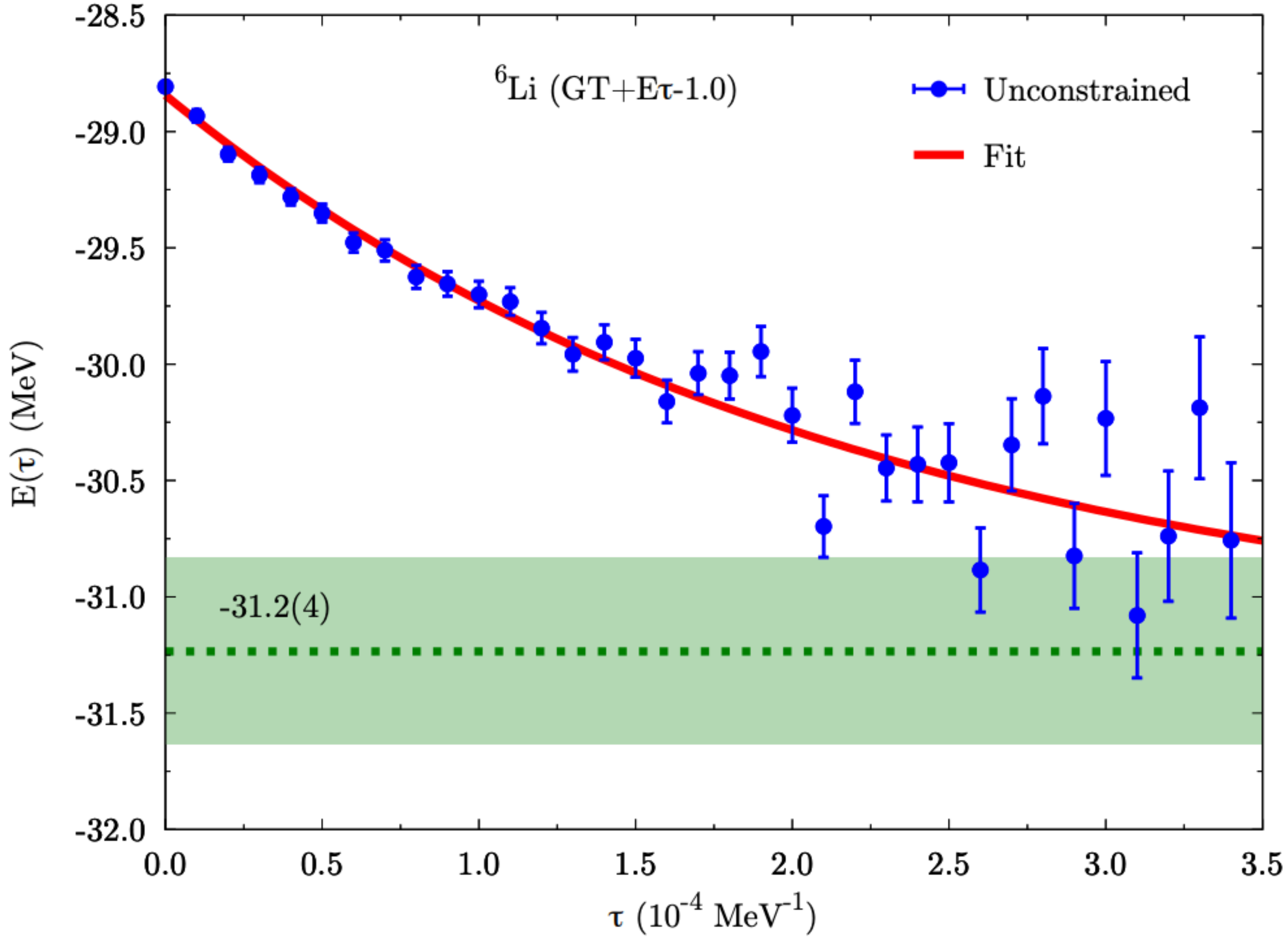
Non-lattice applications: GFMC/AFDMC

Large t: StN decays exponentially

Small t: Excited state effects

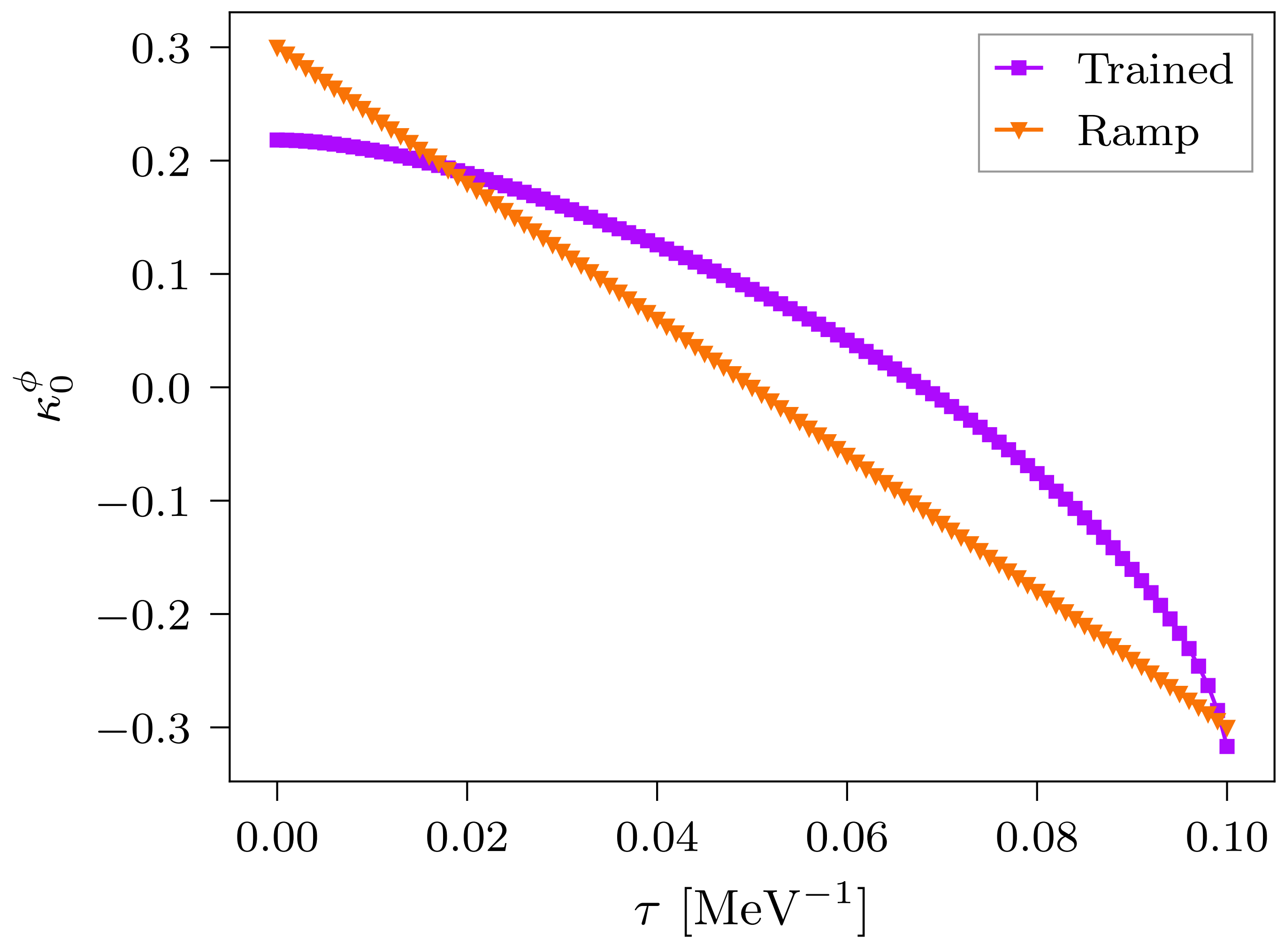
To extract physical information,
fit excited state model

[Gandolfi, et al. 2001.01374]

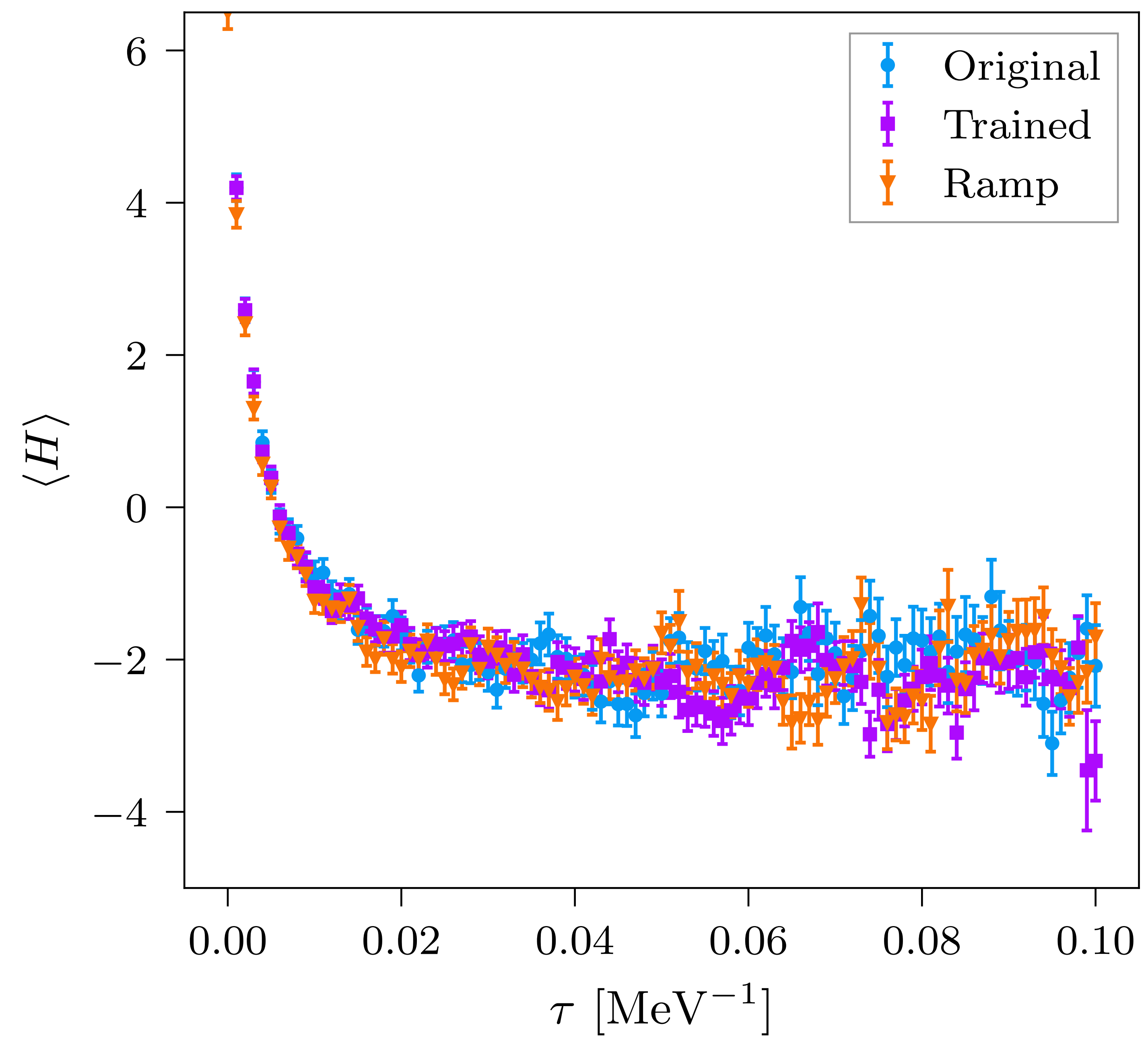


GFMC results

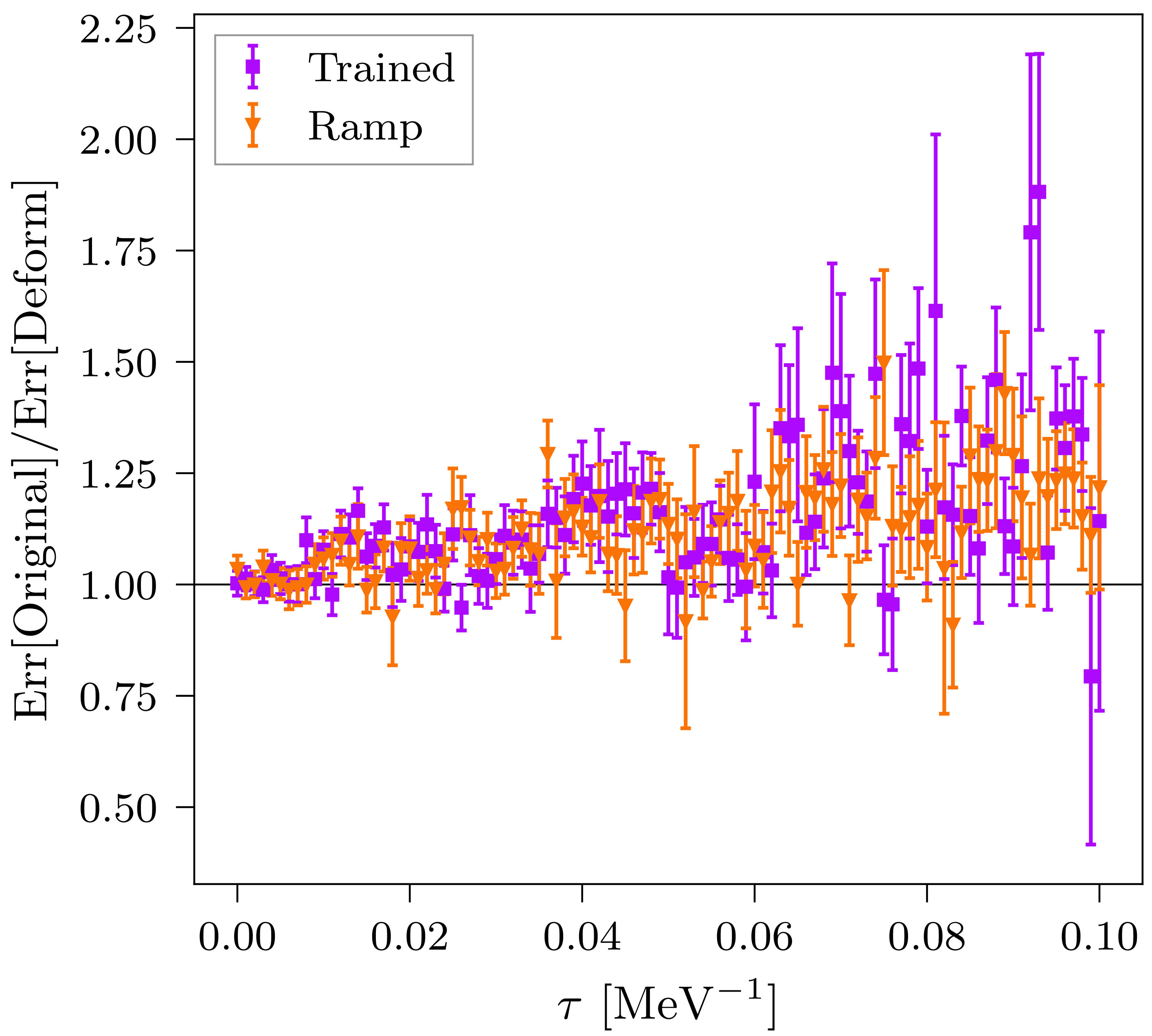
Learned imaginary shift vs Euclidean time



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 $\rho(\vec{q})$ significantly improved.