Control Variates with Neural Networks

Hyunwoo Oh

in collaboration with Paulo Bedaque, Srijit Paul

based on *"Leveraging neural control variates for enhanced precision in lattice field theory"*
Phys. Rev. D **109**, 094519 (2024)

July 30 2024, Lattice 2024

UNIVERSITY OF
MARYLAND

Signal-to-Noise Problem

- It prevents first-principle calculations

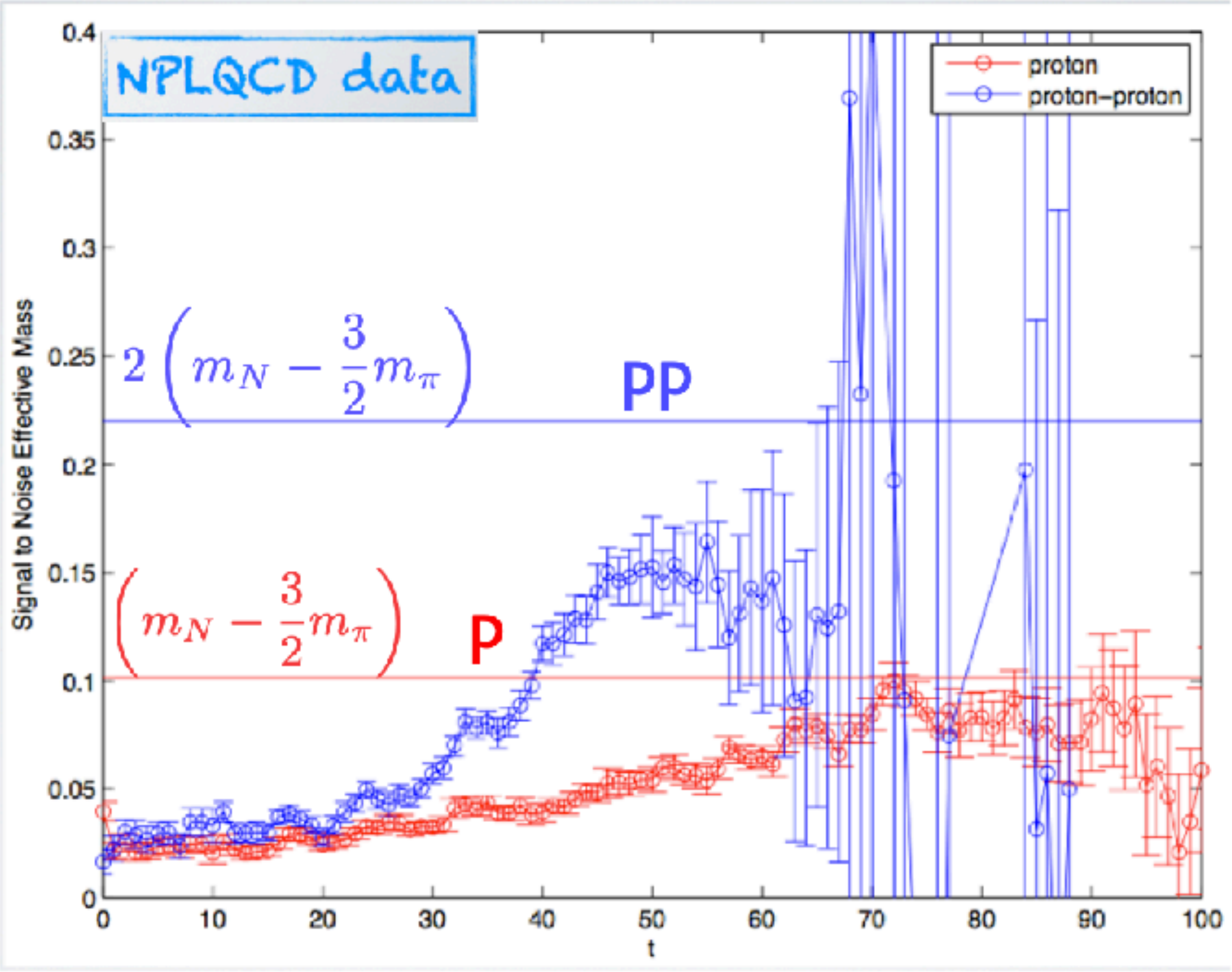
- Example: proton propagator

$$\overline{G_p(t)}/\epsilon_p(t) \sim \sqrt{N}\,\exp\left(-\left(m_p - 3m_\pi/2\right)t\right)$$

MC error

Number of configurations

- Monte Carlo only can improve errors by $1/\sqrt{N}$



Parisi (1984), Lepage (1989), M. Endres, D. Kaplan, J. Lee, A. Nicholson, arXiv:1112.4023

Control Variates (CV)

- If one define $\tilde{O} \equiv O - f$, where $\langle f \rangle = 0$,

$$\langle \tilde{O} \rangle = \langle O \rangle$$

- Its variance is

$$\mathrm{Var}(\tilde{O}) = \mathrm{Var}(O) + \langle f^2 \rangle - 2\langle Of \rangle$$

- $\mathrm{Var}(\tilde{O}) \leq \mathrm{Var}(O)$ if

$$\langle f^2 \rangle - 2\langle Of \rangle \leq 0$$

- Perfect control variates exist: $f_P = O - \langle O \rangle$

T. Bha, S. Lawrence, J. Yoo, Phys. Rev. D **109**, L031505 (2024)

Schwinger-Dyson Control Variates

- In general, it is hard to find observables with the expectation value zero.
- It was suggested to use lattice Schwinger-Dyson equation.

$$\int D\phi \frac{\delta}{\delta\phi}\left(g\,\mathrm{e}^{-S(\phi)}\right) = 0$$

- If $g : \mathbb{R}^n \to \mathbb{R}$,

$$f(\phi) = \sum_i \left(\frac{\partial g}{\partial \phi_i} - g \frac{\partial S}{\partial \phi_i} \right)$$

is a control variate with a proper boundary condition.

- Similarly, for $g : \mathbb{R}^n \to \mathbb{R}^n, f(\phi) = \nabla \cdot g - g \cdot \nabla S$

T. Bha, S. Lawrence, J. Yoo, Phys. Rev. D **109**, L031505 (2024)

Example 1: 2D Real Scalar Field

- Model:

$$S = \frac{1}{2} \sum_{x,\hat{\mu}} \left(\phi_{x+\hat{\mu}} - \phi_x \right)^2 + \frac{m^2}{2} \sum_x \phi_x^2 + \frac{\lambda}{4!} \sum_x \phi_x^4$$

- Observable: correlation function

$$O(t) = \sum_{t'} \left(\sum_x \phi(t' + t, x) \right) \left(\sum_x \phi(t', x) \right)$$

Parametrize Control Variates

- Ansatz for control variates from the knowledge of free theory was suggested

$$g(\phi) = \sum_x a_x \phi_x + \cdots \quad \text{and} \quad f = \sum_i \left(\frac{\partial g}{\partial \phi_i} - g \frac{\partial S}{\partial \phi_i} \right)$$

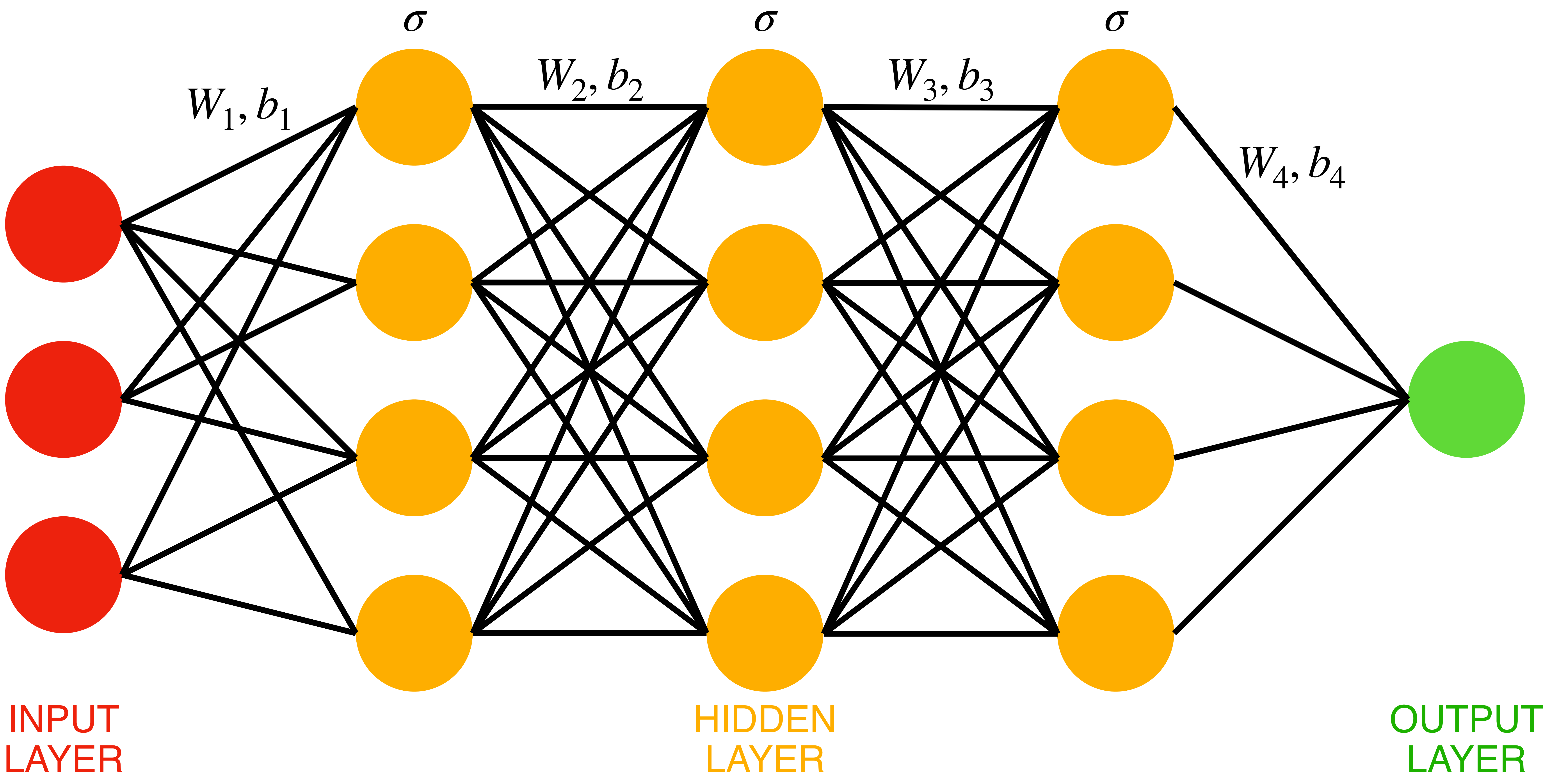
- Instead of using educated guess, **parametrize g as a neural network**

$$g(\phi) = NN(\phi)$$

T. Bha, S. Lawrence, J. Yoo, Phys. Rev. D **109**, L031505 (2024), P. Bedaque, HO, Phys. Rev. D **109**, 094519 (2024)

Neural Network

- $O_n = \sigma(W_n I_n + b_n) = I_{n+1}$



Imposing Translational Symmetry

- $f(T_x[\phi]) = f(\phi)$ should be imposed.

translation operator

$$\leftrightarrow g(T_y[\phi])_x = g[\phi]_{x+y} \text{ (covariance)}$$

- Define a function $g_0 : \mathbb{R}^n \to \mathbb{R}$

$$g(\phi)_x \equiv g_0(T_x[\phi]) \quad \text{and} \quad f(\phi) = \nabla \cdot g - g \cdot \nabla S$$

- It can be easily shown that $g(\phi)$ is translational covariant.

P. Bedaque, HO, Phys. Rev. D **109**, 094519 (2024)

Imposing Z_2 Symmetry

- $f(-\phi) = f(\phi)$ should be imposed:

$$\leftrightarrow g(-\phi) = -\, g(\phi)$$

- It requires zero biases for layers:

$$\phi_i \to W_{ij}\phi_j$$

- It requires an odd activation function:

$$\sigma(x) = \mathrm{arcsinh}(x)$$

- $\sigma(-x) = -\,\sigma(x)$ and $\sigma(\pm\infty) = \pm\,\infty$.

P. Bedaque, HO, Phys. Rev. D **109**, 094519 (2024)

Minimize Variance

- Natural choice of loss function is the variance:

$$L(w) = \langle (O - f)^2 \rangle - \langle O - f \rangle^2$$

neural network parameters

$= \langle O \rangle^2$

- However, if overfitting happens, $O(\phi_i) = f(\phi_i)$ for all training samples i,

very common in ML, dangerous

$$\bar{f} = \frac{1}{N} \sum_{i=1}^{N} f(\phi) = \overline{O} \neq 0$$

- Add tunable parameter μ to avoid overfitting:

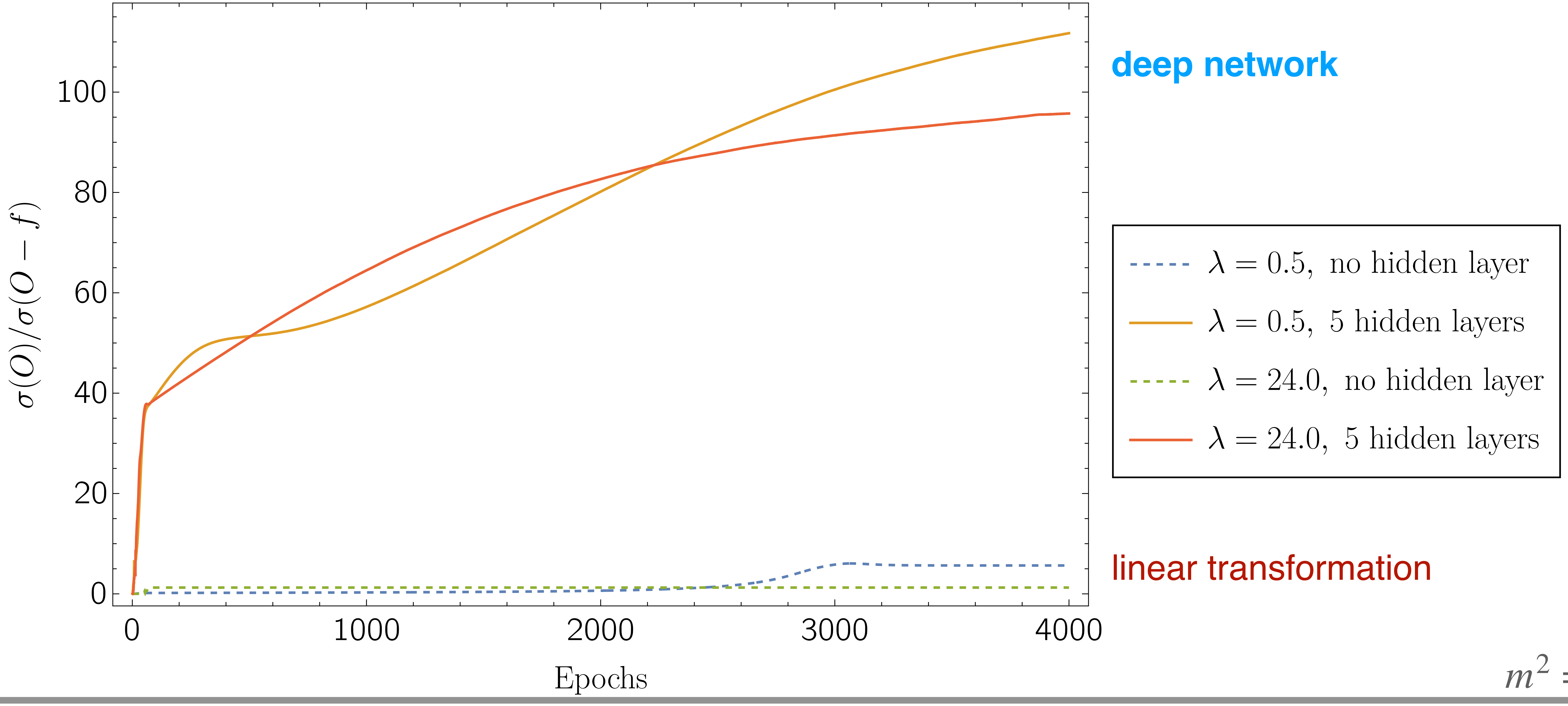
not used for estimation

$$L(w, \mu) = \langle (O - f - \mu)^2 \rangle$$

P. Bedaque, HO, Phys. Rev. D **109**, 094519 (2024)

Variance Reduction

- Observable: $O(L_0/2)$
- 20×20 lattice, 10^4 samples for training, 10^3 samples for estimation



deep network

linear transformation

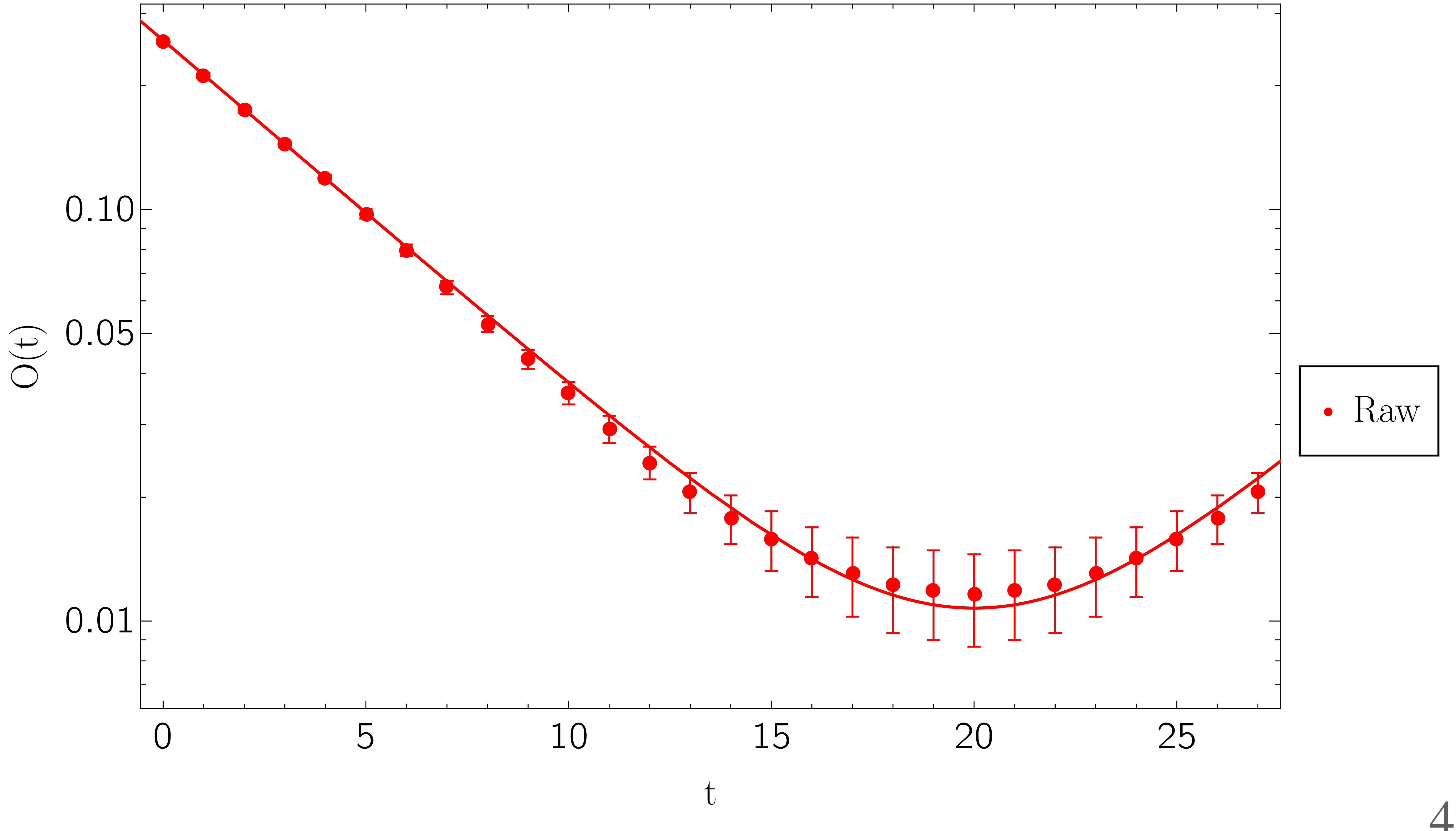
$m^2 = 0.1$

P. Bedaque, HO, Phys. Rev. D **109**, 094519 (2024)

Mass Fit

- Find a mass fit using

$$O(t) = A(e^{-mt} + e^{-m(L_0 - t)})$$



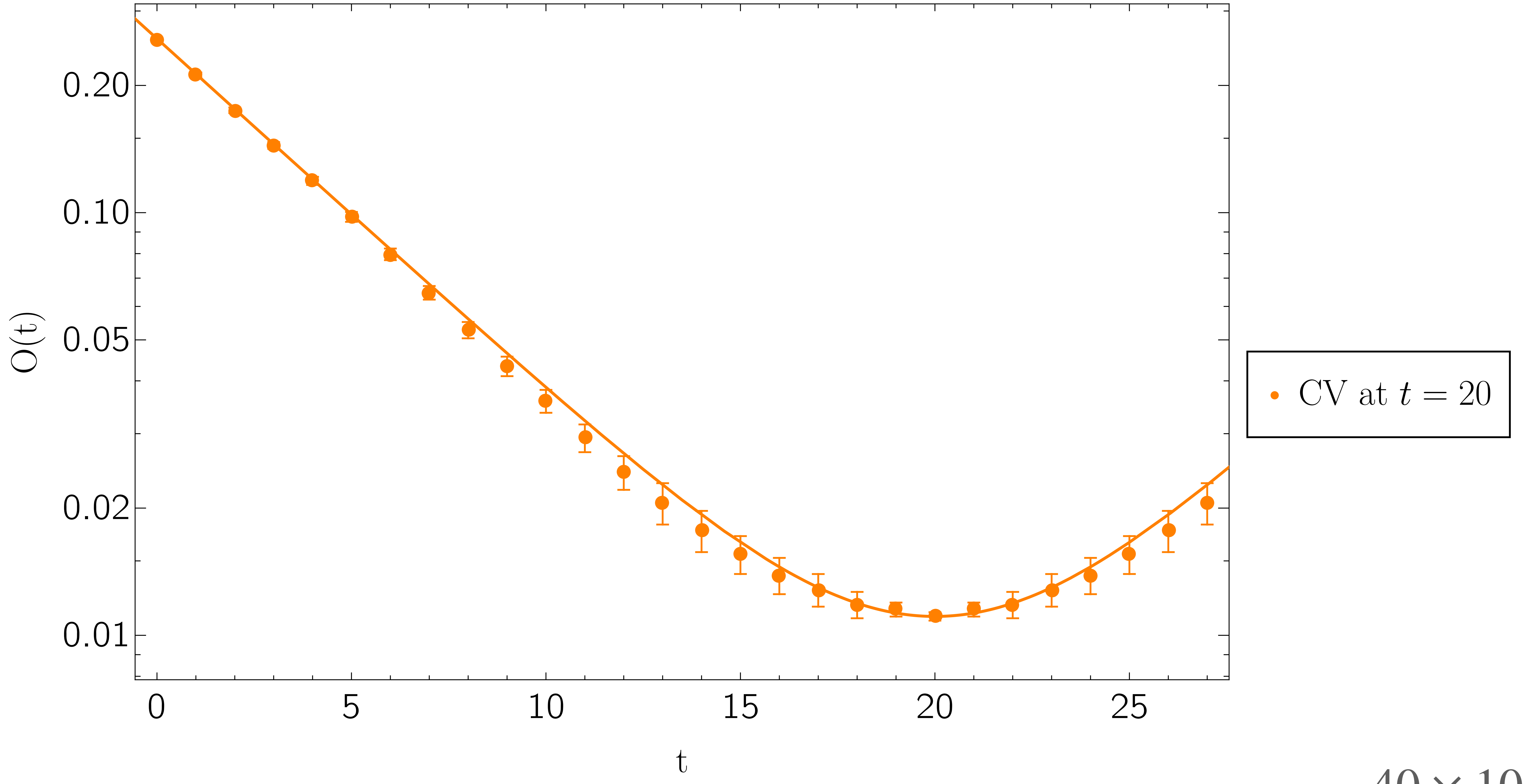
$40 \times 10,\ m^2 = 0.01,\ \lambda = 0.1$

P. Bedaque, HO, Phys. Rev. D **109**, 094519 (2024)

Mass Fit

- Find a mass fit using

$$O(t) = A(e^{-mt} + e^{-m(L_0 - t)})$$



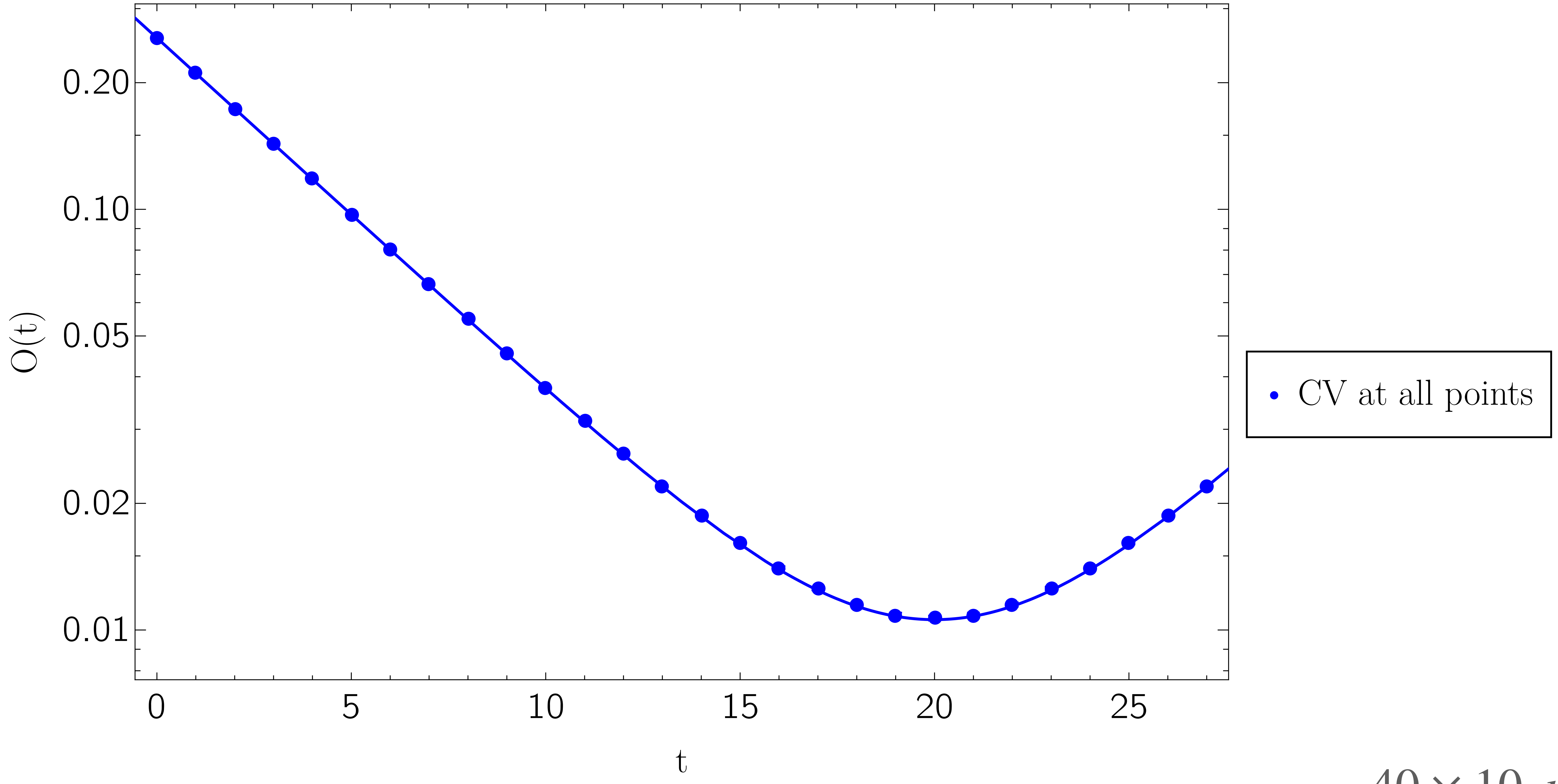
$40 \times 10,\, m^2 = 0.01,\, \lambda = 0.1$

P. Bedaque, HO, Phys. Rev. D **109**, 094519 (2024)

Mass Fit

- Transfer learning was implemented:

use CV at $t = 10$ as an initial network for training CV at $t = 9{,}11$



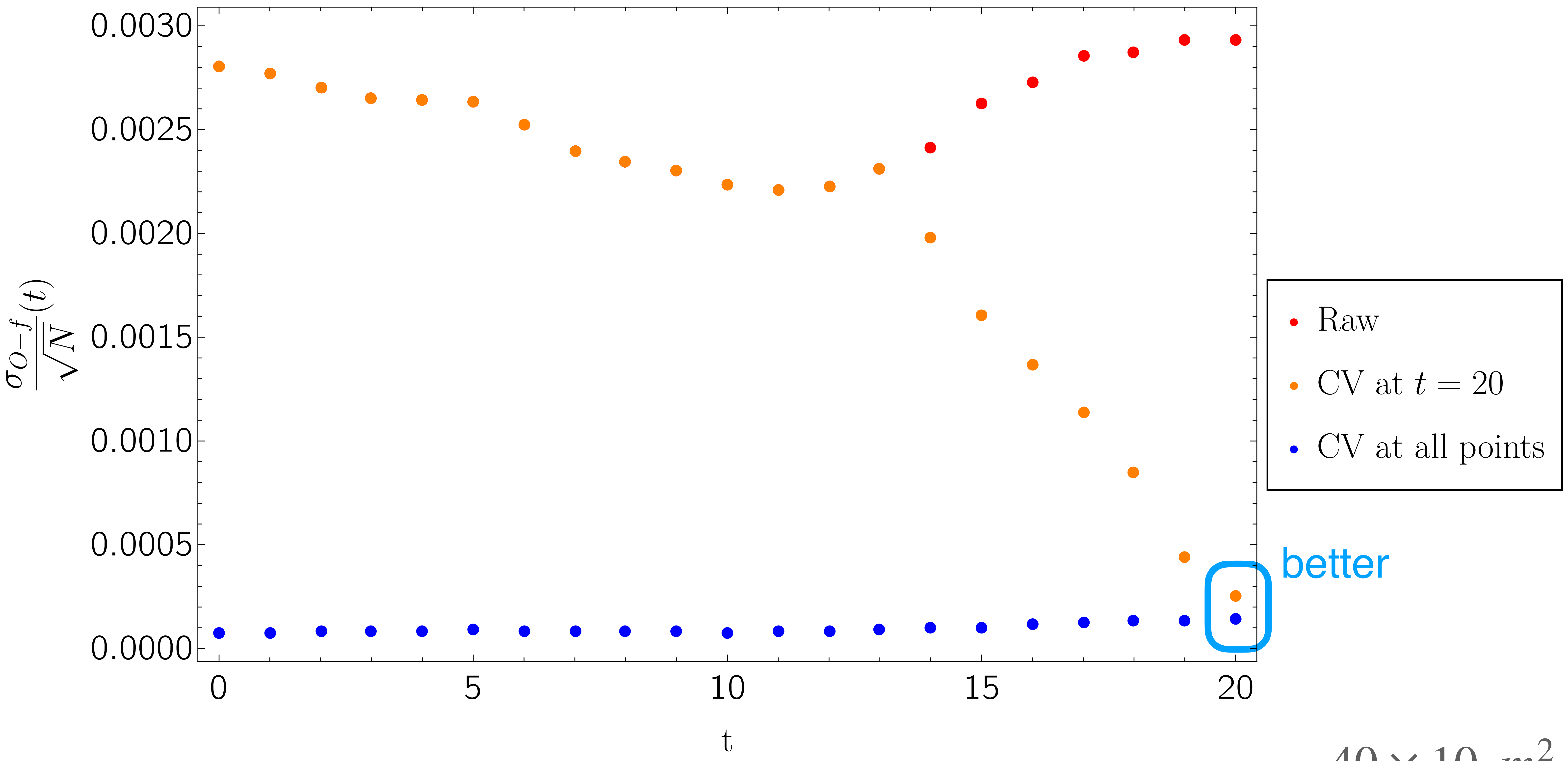
$40 \times 10,\, m^2 = 0.01,\, \lambda = 0.1$

P. Bedaque, HO, Phys. Rev. D **109**, 094519 (2024)

Mass Fit

- Transfer learning was implemented:

use CV at $t = 10$ as an initial network for training CV at $t = 9{,}11$

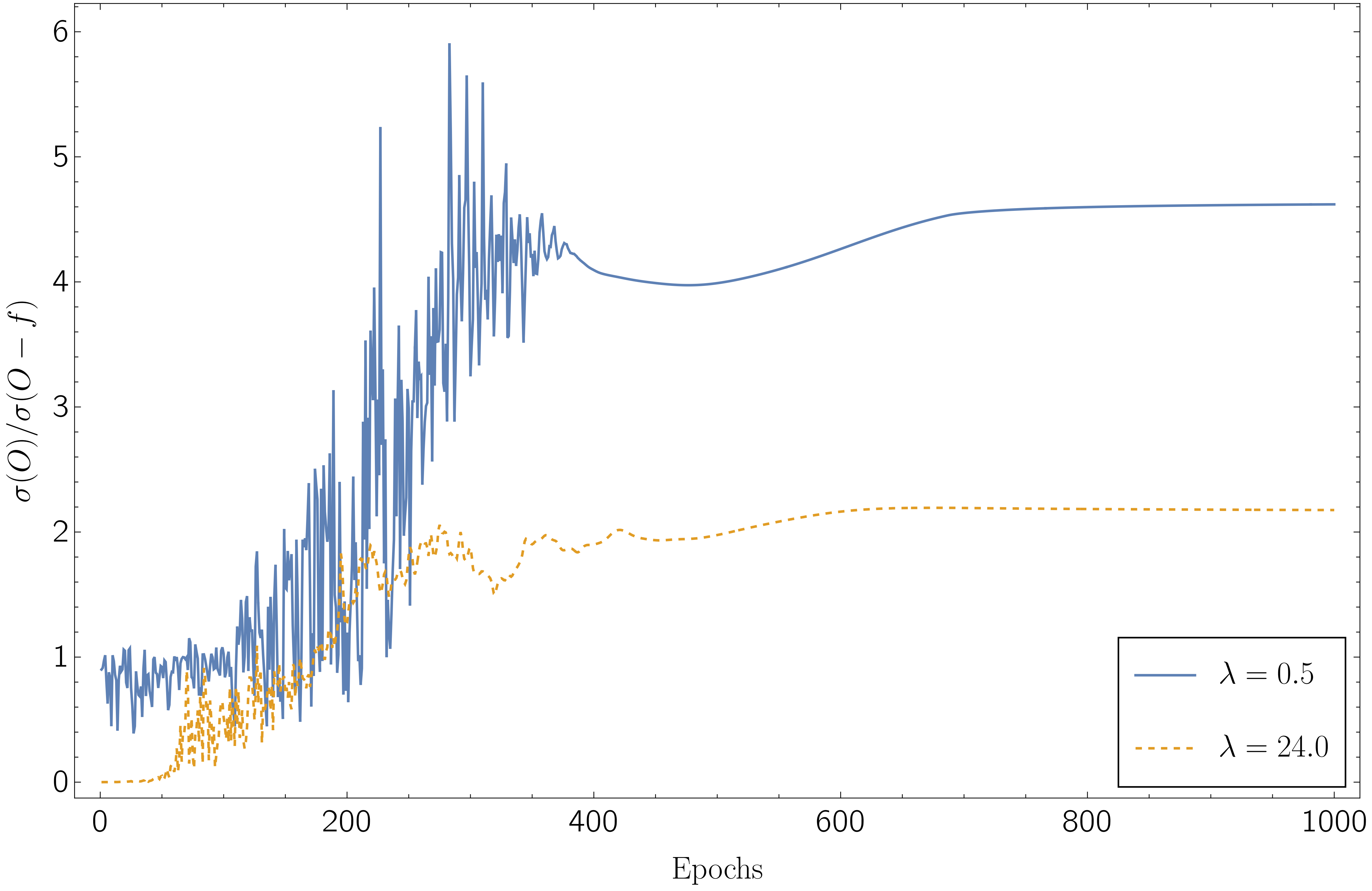


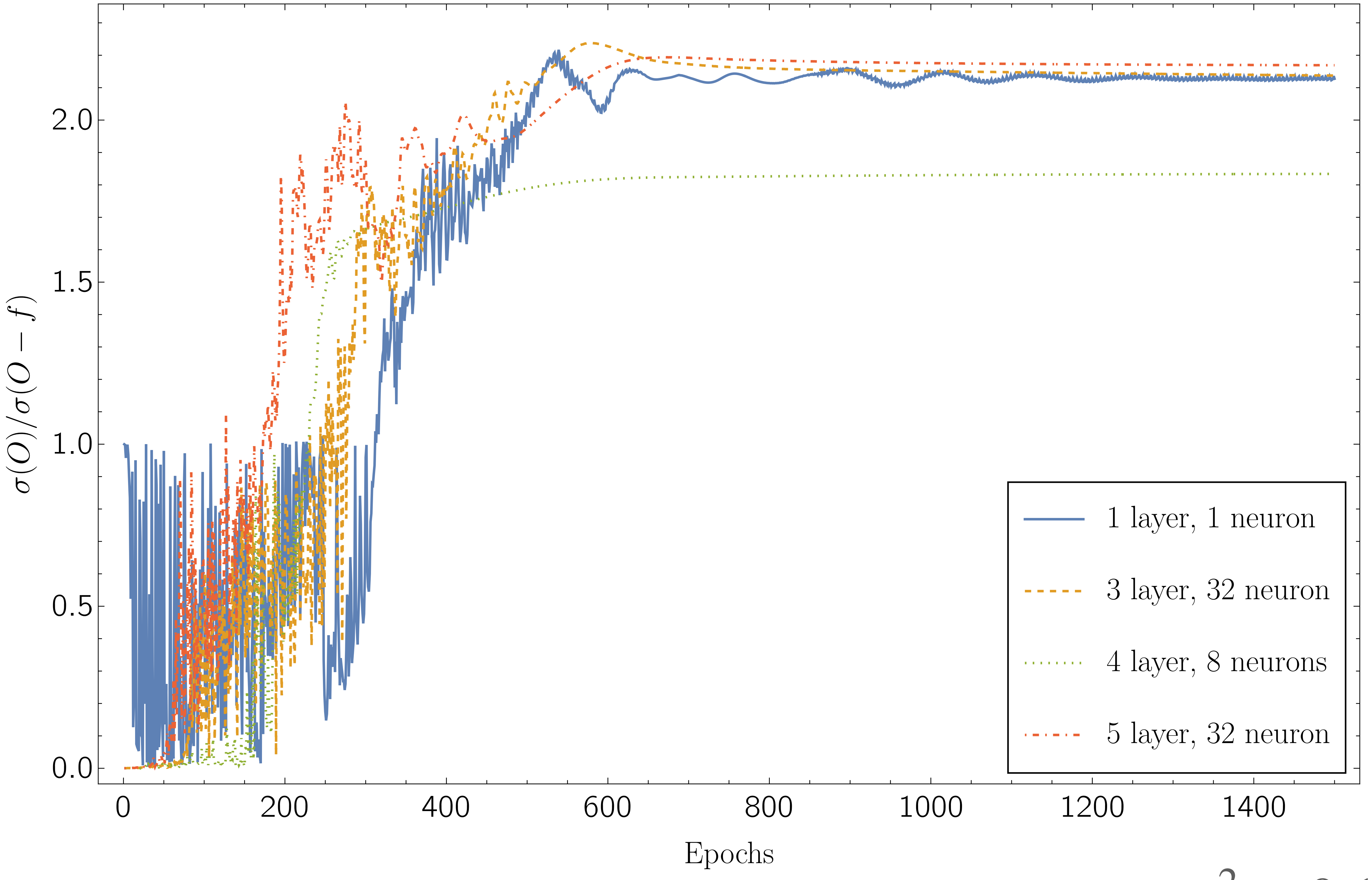
$40 \times 10,\, m^2 = 0.01,\, \lambda = 0.1$

P. Bedaque, HO, Phys. Rev. D **109**, 094519 (2024)

Curse of Dimensionality

- 40×40 lattice, 10^3 samples for training, 10^3 samples for estimation
- L2 regularization was implemented: $\delta \sum_w w^2$ where w are network parameters





$m^2 = 0.1$

P. Bedaque, HO, Phys. Rev. D **109**, 094519 (2024)

Model 2: 2D U(1) Pure Gauge Theory

- Model:

$$Z = \int \prod_i dU_i \exp\left(-\beta \sum_i \big(1 - \cos(P_i)\big)\right)$$

where $\beta = \dfrac{2}{g^2}, P_i = U_1 U_2 U_3^\dagger U_4^\dagger$

- Observable: Wilson loop

$$O(A) = \prod_j P_j = \exp(i \sum_j \theta_j^P)$$

Open Boundary Condition with Gauge Fixing

- Integration measure can be written in terms of plaquette variables, and action and observable are separable:

$$Z = \int \prod_i dP_i \exp\left(\beta \sum_i \cos(P_i)\right)$$

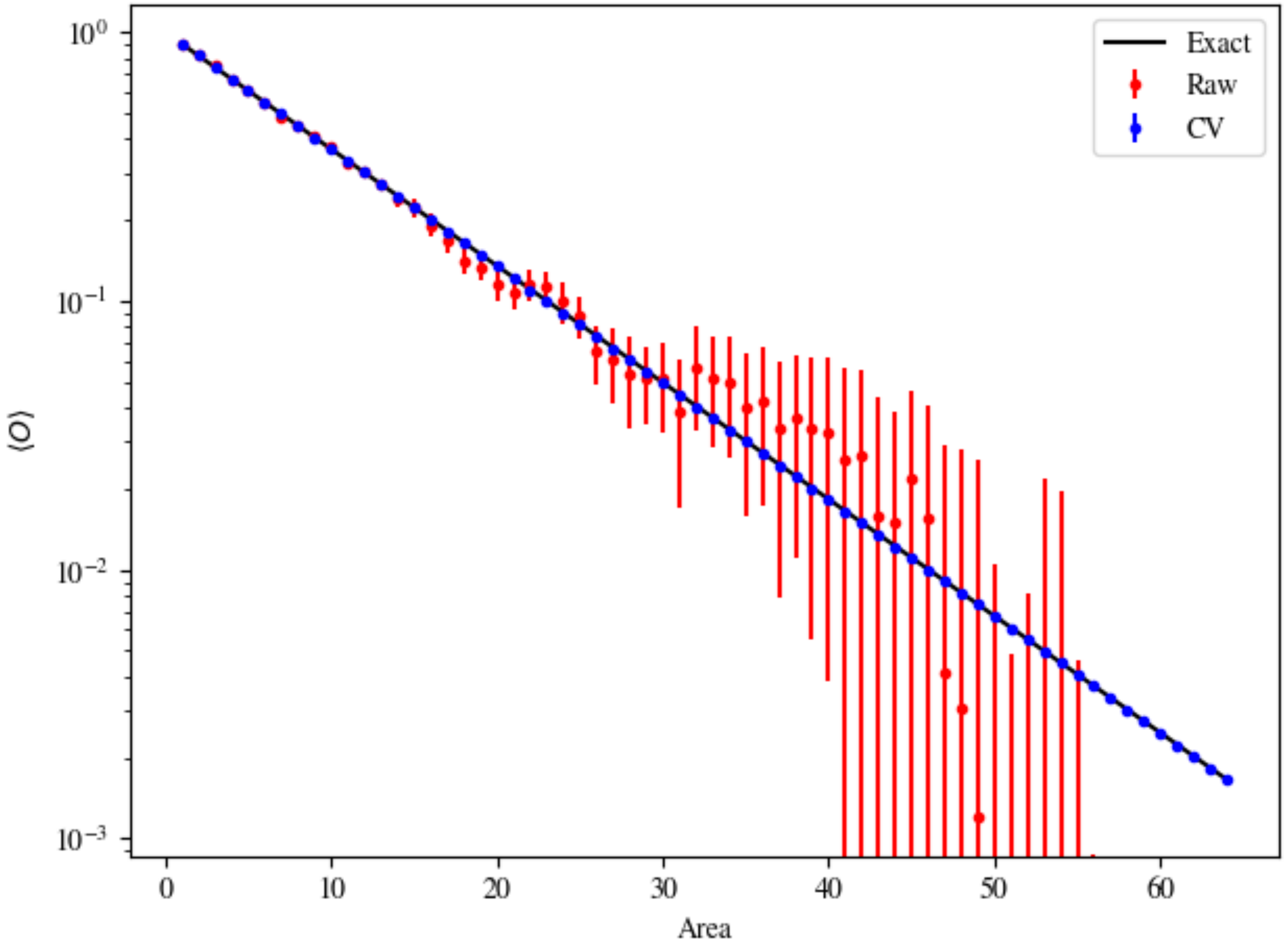
- Ansatz:

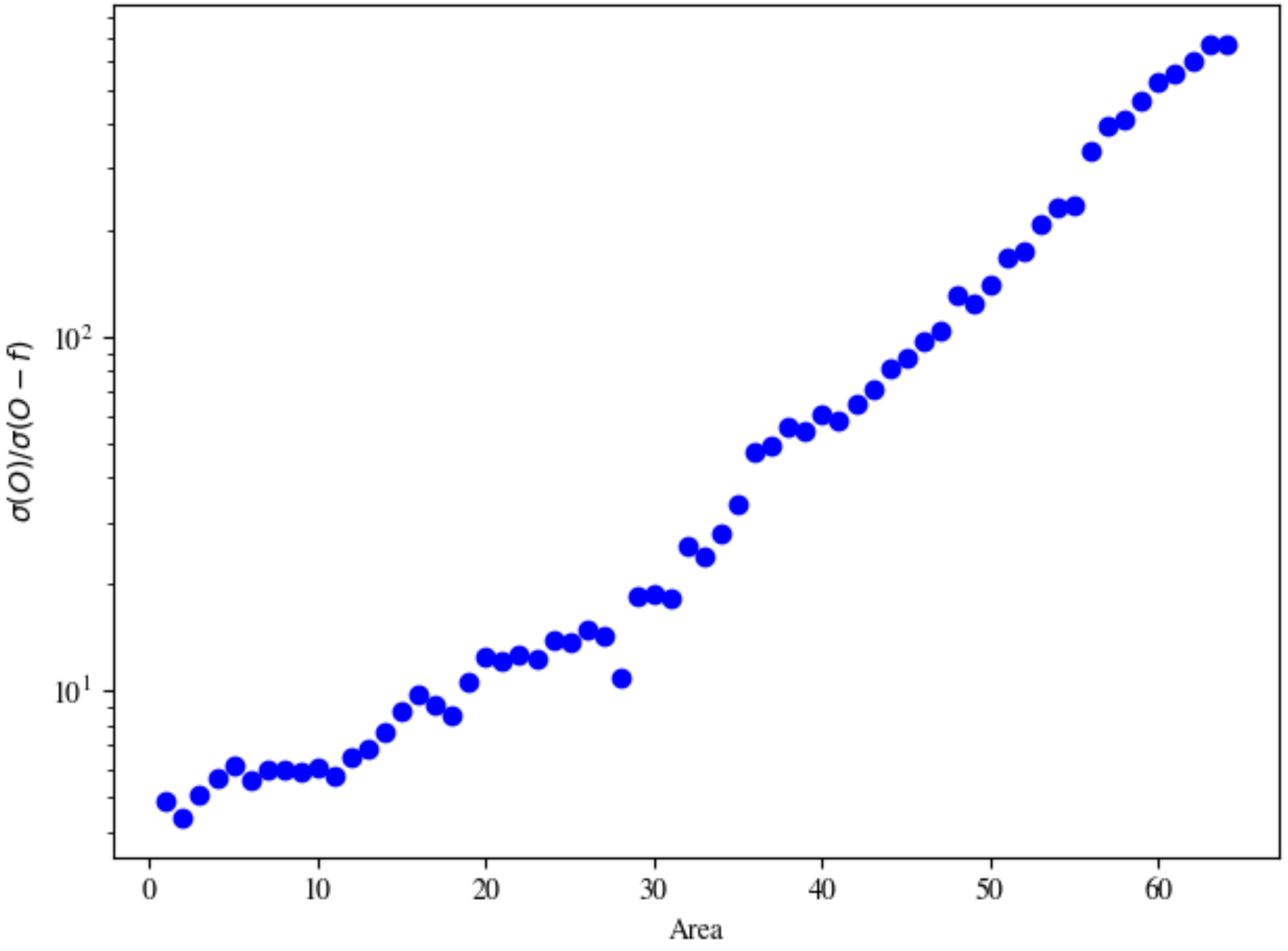
$$f(\phi) = \exp(i \sum_j \phi_j) - \prod_j \left(\exp(i\phi_j) - f_1(\phi_j)\right)$$

with $f_1(\phi) = \dfrac{dg}{d\phi} - g\dfrac{dS}{d\phi} \rightarrow \langle f_1 \rangle = 0 \Rightarrow \langle f \rangle = 0$

Open Boundary Condition with Gauge Fixing

- $\langle O(A) \rangle = \mu^A$ where $\mu \equiv \dfrac{I_1(\beta)}{I_0(\beta)}$



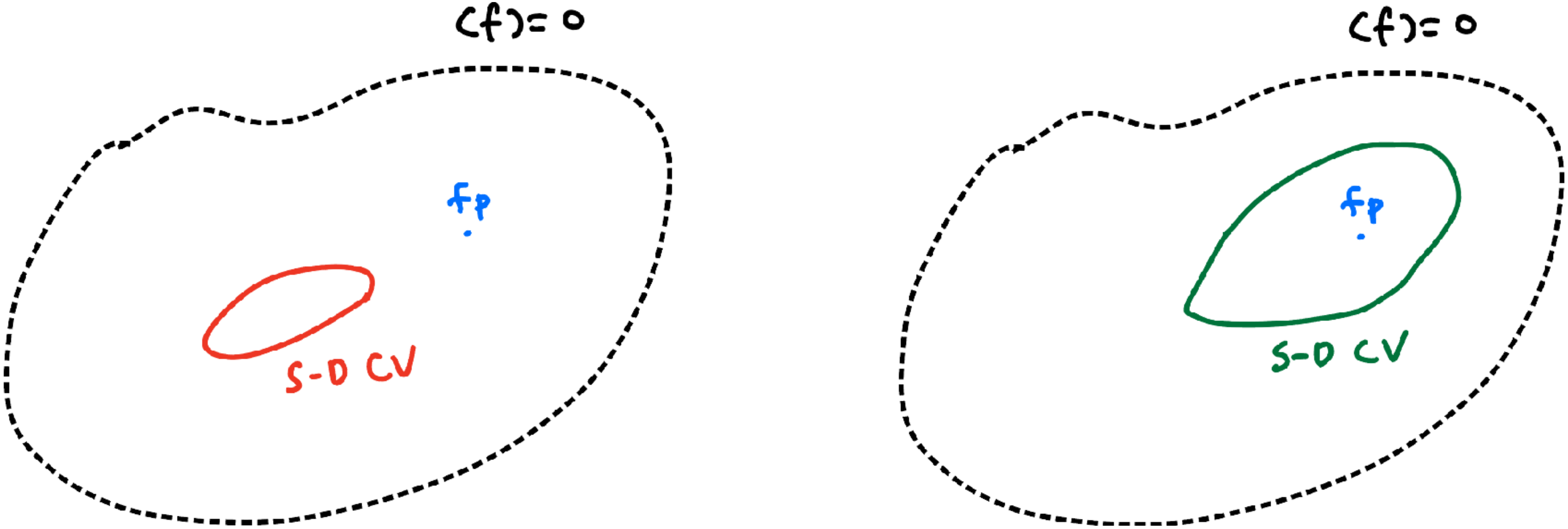


$\beta = 5.555$

Schwinger-Dyson CV Is Not the Most General

- Explore a part of the whole control variates space

$$\sum_i \left(\frac{\partial g}{\partial \phi_i} - g\frac{\partial S}{\partial \phi_i}\right) \overset{?}{=} f_p$$



Conclusion

- Summary

1. Control variates method is a promising way to reduce the variance of observables.
2. It can be formulated via neural networks
3. Showed a possibility on two toy models
4. Discussed issues when applying Schwinger-Dyson control variates

- Future study

1. Find control variates for plaquette correlators on 3D U(1) gauge theory
2. Find control variates using link variables
3. Find different constructions of control variates