STOCHASTIC NORMALIZING FLOWS FOR NEW THEORIES AND OBSERVABLES

ELIA CELLINI

24/07/24

ML meets LFT

Swansea

Based on:

M. Caselle, E.C., A. Nada, M. Panero
- JHEP 07 (2022) 015, arxiv:2201.08862

M. Caselle, E.C., A. Nada
- JHEP 02 (2024) 048, arxiv:2307.01107
- arxiv:2408.XXXX

A. Bulgarelli, E.C., K. Jansen, S. Kühn, A. Nada, S. Nakajima, K.A. Nicoli, M. Panero
- Arxiv:24XX.XXXX

UNIVERSITÀ DI TORINO

INFN

OUTLINE

1. **NON-EQUILIBRIUM MCMC**
2. **NORMALIZING FLOWS**
3. **STOCHASTIC NORMALIZING FLOWS**
4. **EFFECTIVE STRING THEORY**
5. **ENTANGLEMENT ENTROPY**

NON-EQUILIBRIUM MCMC

JARZYNSKI'S EQUALITY

General equality that relates **non-equilibrium experiments** and **equilibrium quantities**:

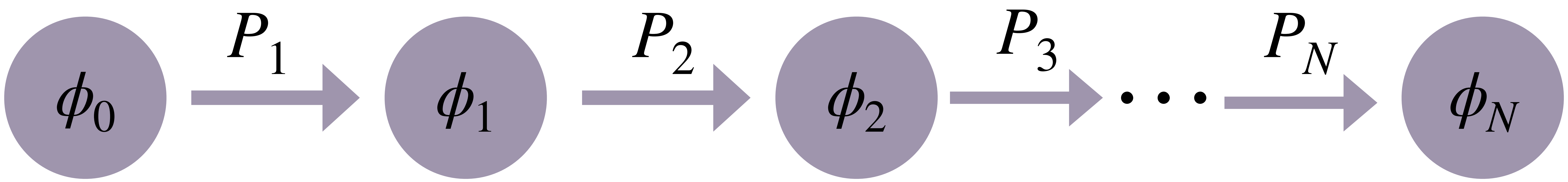
[Jarzynski; cond-mat/9610209]

$$\langle e^{-W} \rangle_f = \frac{Z_{fin}}{Z_{in}} = e^{-\Delta F}$$

We can prove (and exploit) this equality using as "physical" system a Markov Chain Monte Carlo (MCMC) algorithm!

NON-EQUILIBRIUM MCMC (NE-MCMC)

$$q_0 \simeq e^{-S_0} \xrightarrow{P_1} e^{-S_1} \xrightarrow{P_2} \dots \xrightarrow{P_N} e^{-S_N} \simeq p$$



1. **Thermalized** q_0 "**prior**"
2. $P_i \propto \exp(-S_i)$ c**hange along the processes** and satisfy detailed balance.
3. $p = \exp(-S_N)/Z_N \rightarrow$ "**target**" distribution

Remark: no thermalization during the processes.

NON-EQUILIBRIUM MCMC

Forward probability density:

$$q_0(\phi_0)\prod_{n=0}^{N-1} P[\phi_i \to \phi_{i+1}] = q_0(\phi_0)P_f[\phi_0, \ldots, \phi_N]$$

Reverse probability density:

$$p(\phi_N)\prod_{n=0}^{N-1} P[\phi_{i+1} \to \phi_i] = p(\phi_N)P_r[\phi_N, \ldots, \phi_0]$$

DISSIPATED WORK

Observe that:

$$\ln \frac{q_0(\phi_0)P_f[\phi_0, \ldots, \phi_N]}{p(\phi_N)P_r[\phi_N, \ldots, \phi_0]} = \underbrace{S_N(\phi_N) - S_0(\phi_0) - Q}_{\text{(dimensionless) Work } W} - \Delta F = W(\phi_0, \ldots, \phi_N) - \Delta F = W_d$$

(dimensionless) **Work** W

Where:

$$\underbrace{Q = \ln \frac{P_r[\phi_N, \ldots, \phi_0]}{P_f[\phi_0, \ldots, \phi_N]} = \sum_{n=0}^{N-1} \ln \frac{q_{n+1}(\phi_{n+1})}{q_{n+1}(\phi_n)}}_{\text{Detailed Balance}} = \sum_{n=0}^{N-1} \left(S_{n+1}(\phi_{n+1}) - S_{n+1}(\phi_n) \right)$$

Detailed Balance

CROOKS FLUCTUATION THEOREM

Thus:

$$\frac{q_0(\phi_0)P_f[\phi_0,\ldots,\phi_N]}{p(\phi_N)P_r[\phi_N,\ldots,\phi_0]} = \frac{\mathscr{P}_f(W_d)}{\mathscr{P}_r(-W_d)} = e^{W_d}$$

→ **Crooks Theorem**

[Crooks; cond-mat/9901352]

Observe also:

$$1 = \int \prod_{i=0}^{N} d\phi_i q_0(\phi_0)P_f[\phi_0,\ldots,\phi_N]\left(\frac{p(\phi_N)P_r[\phi_N,\ldots,\phi_0]}{q_0(\phi_0)P_f[\phi_0,\ldots,\phi_N]}\right) = \langle e^{-W_d}\rangle_f$$

JARZYNSKI'S EQUALITY

$$1 = \langle e^{-W_d} \rangle_f$$

Jarzynski's equality

[Jarzynski; cond-mat/9610209]

$$\langle e^{-W} \rangle_f = e^{-\Delta F}$$

$$\langle \mathcal{O} \rangle_{\phi \sim p} = \langle \mathcal{O} e^{-W_d} \rangle_f$$

Non-Equilibrium Ensemble

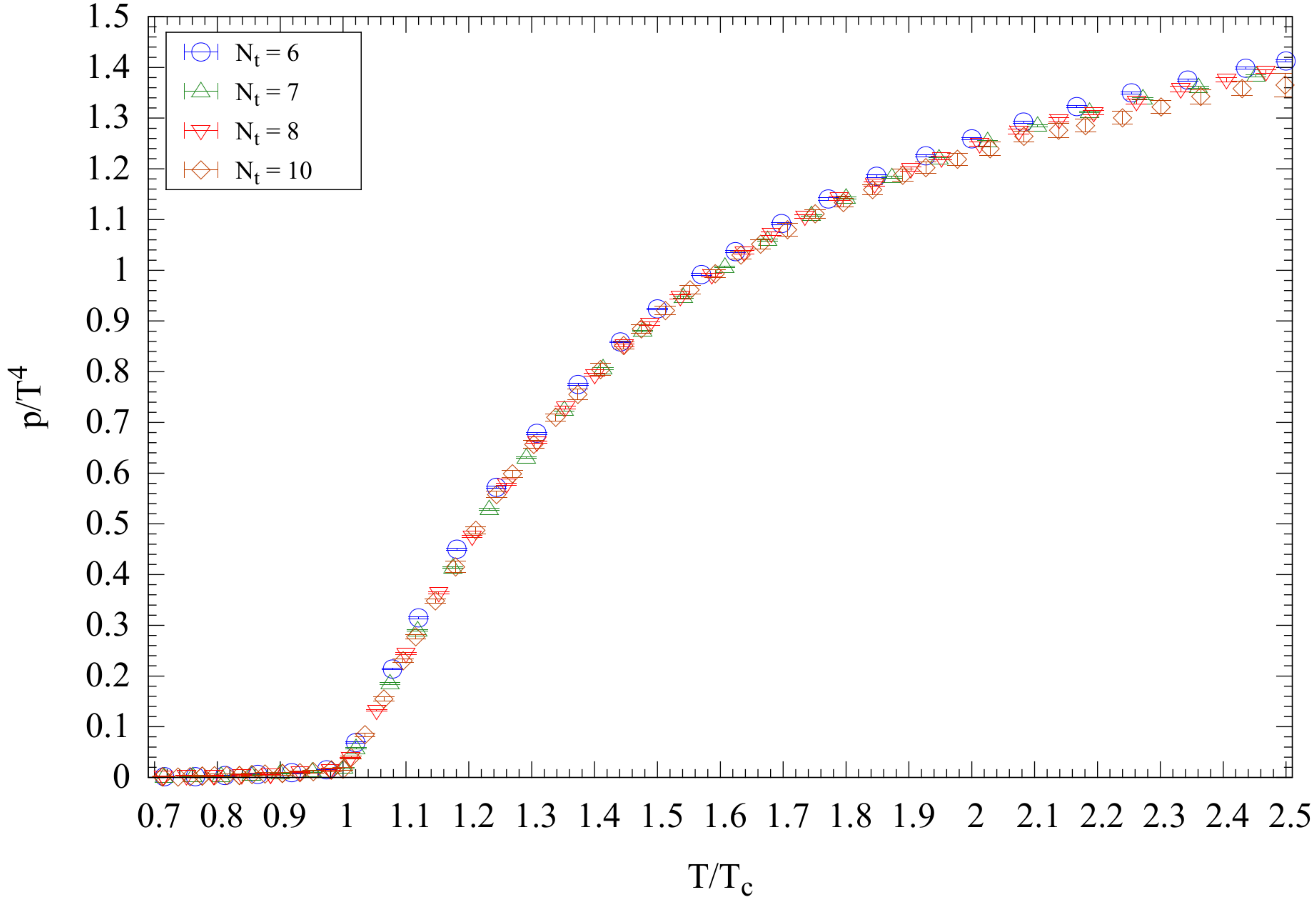
Equilibrium Quantity

NE-MCMC FOR LFT

Jarzynksi's equality has been exploited to obtain **state-of-the-arts results** in LFT: See talk by: Alessandro Nada

- Interface free energy.
  [Caselle+; 1604.05544]
- $SU(3)$ e.o.s.
  [Caselle+; 1801.03110]
- Running coupling
  [Francesconi+; 2003.13734]
- Entanglement entropy
  [Bulgarelli and Panero; 2304.03311, 2404.01987]
- Topological freezing
  [Bonanno+; 2402.06561]

Equivalent to:
Annealed Importance Sampling
[Neal; physics/9803008]



Taken from: [Caselle+; 1801.03110] with authors permission

NUMERICAL PROBLEM

The identities derived before are exact, however, the **exponential average:**

$$\langle e^{-W_d} \rangle_f$$

can be highly inefficient when W_d is large and the statistic is finite.

In order to fight this problem, we want W_d to be "small"

Solution 1) Infinite MCMC steps → quasi-static transformations → "small" W_d

Solution 2) use Machine Learning to minimize W_d

NORMALIZING FLOWS

NORMALIZING FLOWS

A Normalizing Flow (**NF**) g_θ is a **parametric**, **invertible** and **differentiable** function:

[Rezende+; 1505.05770]

$$g_\theta : \; q_0 \to q_\theta \simeq p \qquad \phi = g_\theta(z) \qquad q_\theta(\phi) = q_0(g^{-1}(\phi)) \, | \det J_g |^{-1}$$

NFS: LEARNING BOLTZMANN DISTRIBUTIONS

NFs can be trained to $q_\theta \simeq p(\phi)$ with $p(\phi) = \exp(-S[\phi])/Z$ by minimizing the reverse **Kullback-Leibler divergence**:

[Albergo+; 1904.12072][Noé+; 1812.01729]

$$D_{KL}(q_\theta || p) = \int d\phi q_\theta(\phi) \log \frac{q_\theta(\phi)}{p(\phi)} \geq 0.$$

NFS: SAMPLING BOLTZMANN DISTRIBUTIONS

Partition functions and **observables** can be computed using a re-weighting procedure also called **Importance Sampling**:

[Nicoli+; 1910.13496, 2007.07115]

$$\langle \mathcal{O} \rangle_{\phi \sim p} = \frac{1}{Z} \langle \mathcal{O} \tilde{w} \rangle_{\phi \sim q_\theta} \qquad Z = \langle \tilde{w} \rangle_{\phi \sim q_\theta} \qquad \tilde{w} = \frac{e^{-S[\phi]}}{q_\theta(\phi)}$$

Alternative/equivalent to Jarzynksi's equality!

See talks by:

- Simran Singh
- Ryan Abbott
- Fernando Romero Lopez
- Mathis Gerdes

STOCHASTIC NORMALIZING FLOWS

STOCHASTIC NORMALIZING FLOWS

Stochastic Normalizing Flows (SNFs) combine NE-MCMC update and NF layers:

$$\phi_0 \longrightarrow g_\theta^1(\phi_0) \xrightarrow{P_1} \phi_1 \longrightarrow g_\theta^2(\phi_1) \xrightarrow{P_2} \dots \xrightarrow{P_N} \phi_N = \phi$$

Where g_θ^i are NF layers and P_i are MCMC update

[Wu+; 2002.06707],[Caselle, E.C., Nada, Panero; 2201.08862]

STOCHASTIC NORMALIZING FLOWS

Forward and **Reverse** transition probabilities of NF layers can be written as:

$$P[\phi_n \to \phi_{n+1}] = \delta(\phi_{n+1} - g_\theta^n(\phi_n)) \qquad P[\phi_{n+1} \to \phi_n] = \delta(\phi_n - (g_\theta^n)^{-1}(\phi_{n+1}))$$

And satisfies:

$$q_n(\phi_n)P[\phi_n \to \phi_{n+1}] = q_{n+1}(\phi_{n+1})P[\phi_{n+1} \to \phi_n]$$

$$\Downarrow$$

$$\ln\left(P[\phi_{n+1} \to \phi_n]/P[\phi_n \to \phi_{n+1}]\right) = \ln\left(q_n(\phi_n)/q_{n+1}(\phi_{n+1})\right) = \ln|\det J_{g^n}(\phi_n)|$$

[Wu+; 2002.06707],[Caselle, E.C., Nada, Panero; 2201.08862]

SNFS: DISSIPATED WORK

We have now:

$$W_d^\theta = W_\theta(\phi_0, \dots, \phi_N) - \Delta F = S_N(\phi_N) - S_0(\phi_0) - Q_\theta - \Delta F$$

Where:

$$Q_\theta = \sum_{n=0}^{N-1} \left(S_{n+1}(\phi_{n+1}) - S_{n+1}(\phi_n) + \ln|\det J_{g_\theta^n}| \right)$$

SNFS: TRAINING

We can now train a SNF by minimizing:

$$\mathscr{L}(\theta) = \langle W_d^\theta \rangle_f = D_{KL}(q_0 P_f || p P_r) \geq 0$$

$$\langle W^\theta \rangle_f \geq \Delta F$$

Second Law!

Measure how reversible the process is.

EFFECTIVE STRING THEORY

EFFECTIVE STRING THEORY

Correlators of Polyakov loops modelled in terms of **string partition functions**:

$$\langle P(0)P^{\dagger}(R)\rangle \sim \int DXe^{-S_{EST}[X]} \equiv Z_{EST}$$

The main choice for S_{EST} is the **Nambu-Goto** **(NG)** action:

$$S_{NG}[\phi] = \sigma \int d\xi^2 \sqrt{g}$$

- **Anomalous** at quantum level → **effective, large-distance description** of Yang-Mills theories (**low-energy universality theorem)**.
- Works only up to order $1/R^5$ → first order approximation of a more general theory → **Beyond Nambu-Goto** (BNG)
  [Aharony and Komargodski; 1302.6257], [Brandt and Meineri; 1603.06969],[Caselle; 2104.10486]

NAMBU-GOTO STRING

Main method: **zeta-function regularization**

Main observables:

- **Partition function** → directly associated with the interquark potential. Well known at all the order.
- **Correlation functions** (e.g. width σw^2) → measure of the density of the chromoelectric flux tube

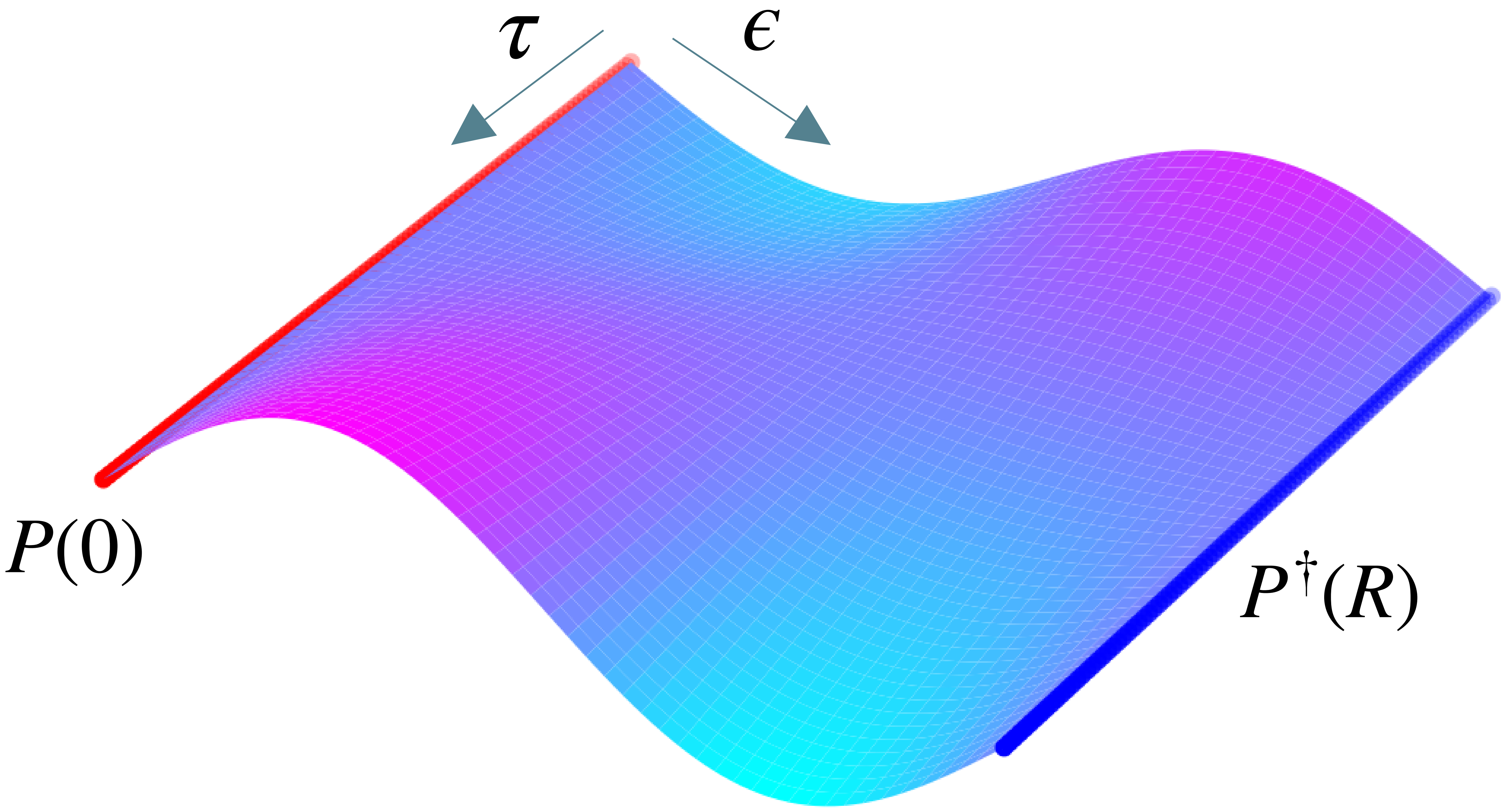
Analytical limits:

- **Correlation functions**
- **Higher order corrections** (Beyond NG)

LATTICE NAMBU-GOTO STRING

$$S_{NG}(\phi) = \sigma \sum_{x \in \Lambda} \left[\sqrt{1 + (\partial_\mu \phi(x))^2/\sigma} - 1 \right]$$

- $d = 2 + 1$ target Yang-Mills
- σ string tension
- Λ : square lattice of size $L \times R$, $a = 1$
- $\phi(x) = \phi(\tau, \epsilon) \in \mathbb{R}$
- $\phi(\tau + L, \epsilon) = \phi(\tau, \epsilon)$
- $\phi(\tau,0) = \phi(\tau, R) = 0$
- $\sigma w^2 = \langle \phi(\tau, R/2)^2 \rangle_\tau$

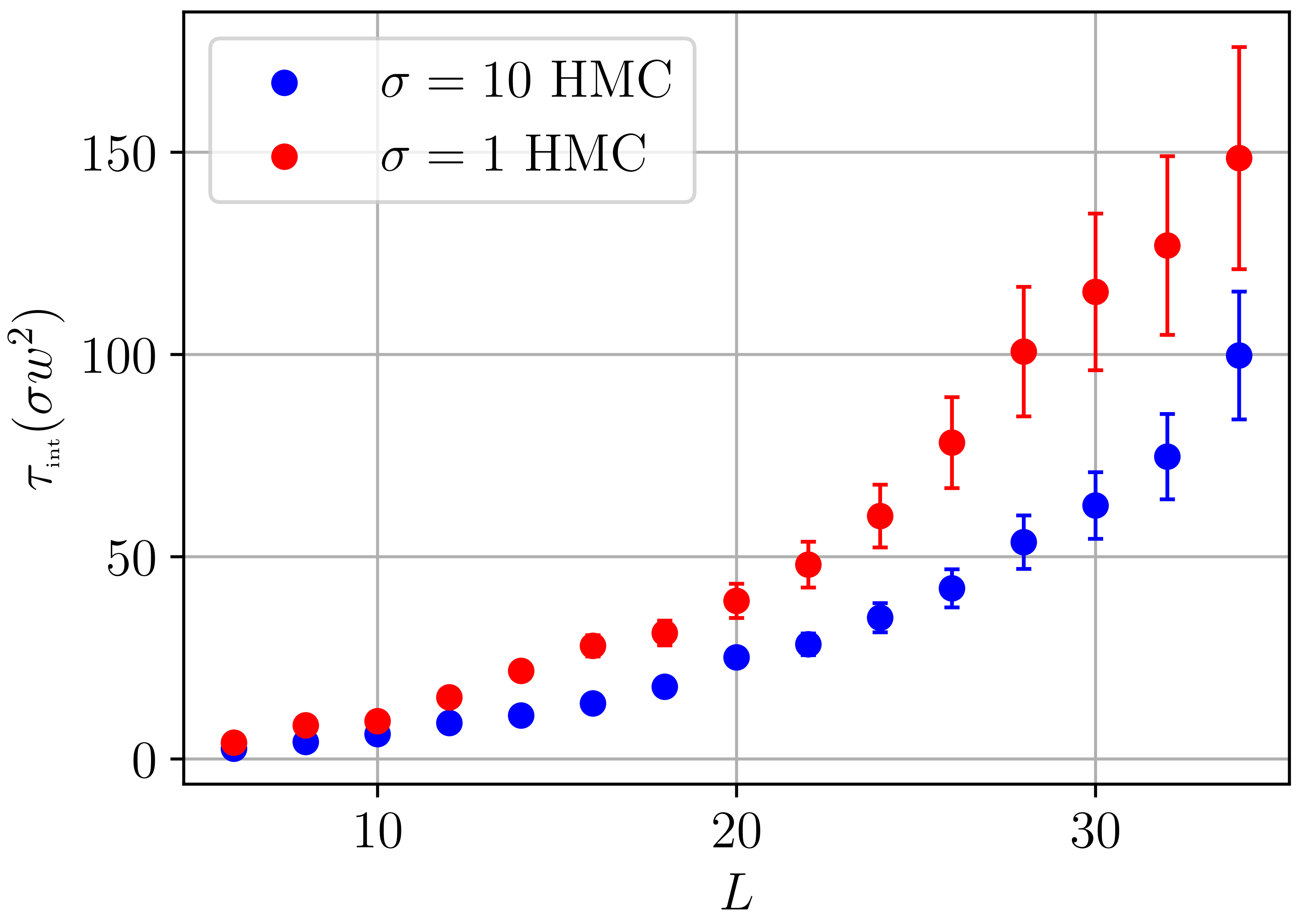


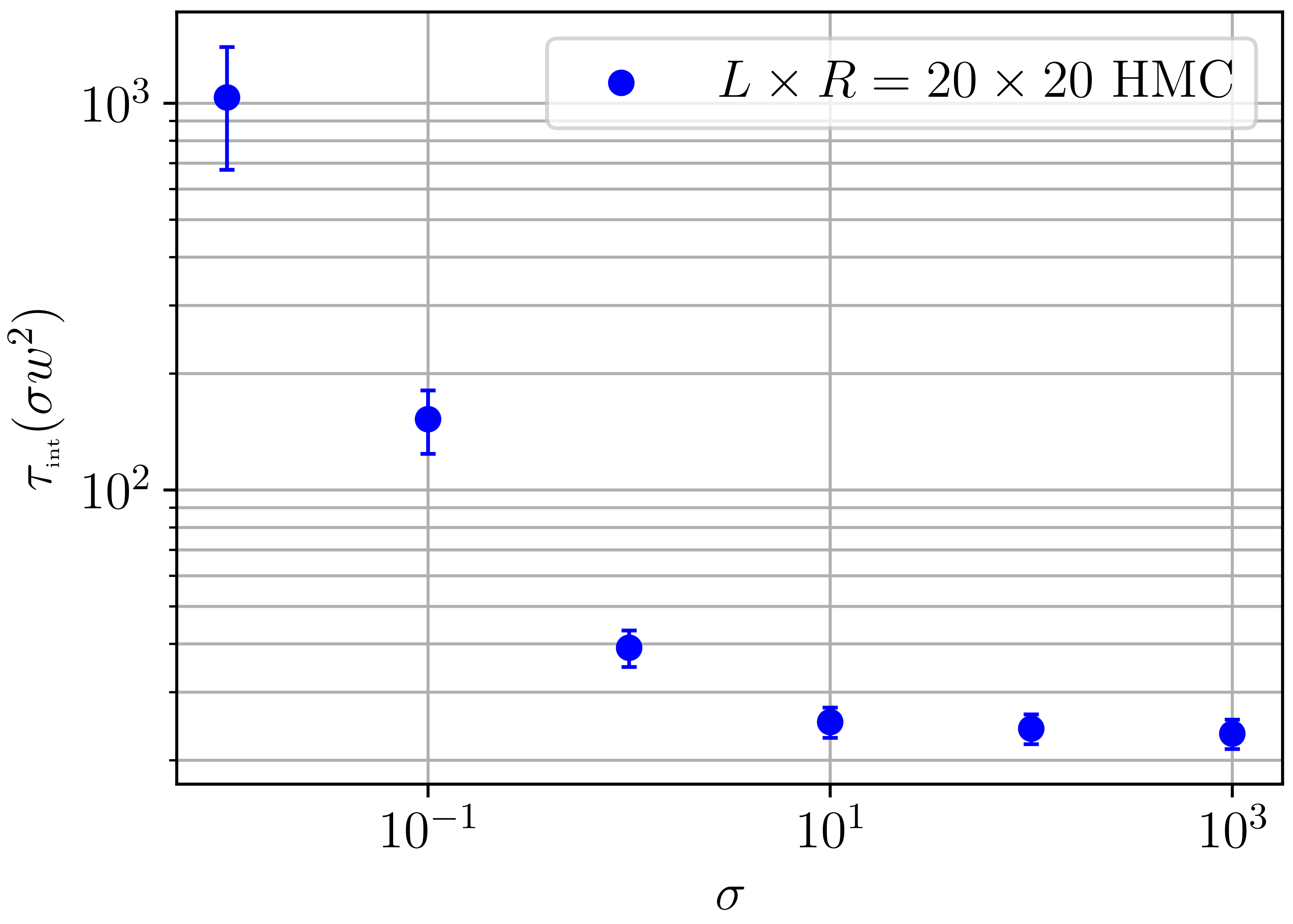
[Caselle, EC, Nada; 2307.01107]

LACKS OF NUMERICAL METHODS

Numerical problems:

- Strong **non-linearity** → **critical** theory (**Critical Slowing Down**)
- Estimation of **partition functions**

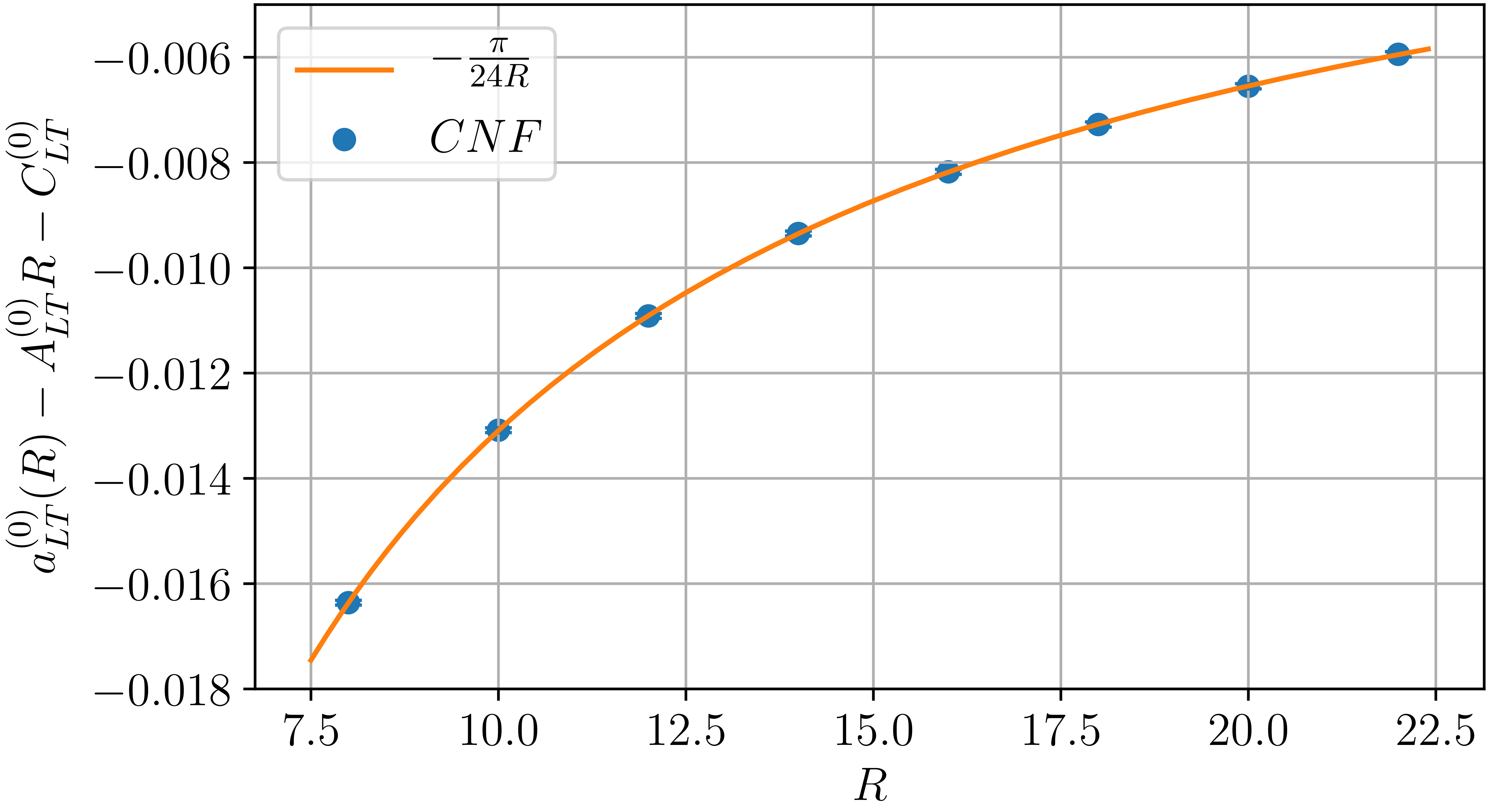




A PROOF-OF-PRINCIPLE

Using **Continuous Normalizing Flows** (CNFs), we proved that flow-based sampling can be successful applied to the NG EST. However, CNFs suffer from poor scaling in σ

$$-\log Z = -\frac{\pi}{24}\frac{L}{R} + \cdots$$



Large σ region ($\sigma \geq 40$),
Fitted coefficient: $-0.1309(2)$,
target: $-0.1308996...$

[Caselle, E.C., Nada; 2307.01107]

NG: PHYSICS-INFORMED SNF

In the $\sigma \to \infty$ region:

$$S_{NG}(\phi) \sim S_{FB}(\phi) + \dots$$

$$S_{FB}(\phi) = \frac{1}{2}\sum_x (\partial_\mu \phi(x))^2$$

Prior:

$$q_0(\phi_0) = \frac{1}{Z} e^{-S_{FB}(\phi_0)}$$

MCMC update i:

$$S_i(\phi) = S_{NG}(\phi_i, \sigma_i); \qquad \sigma_i > \sigma_{i+1}$$

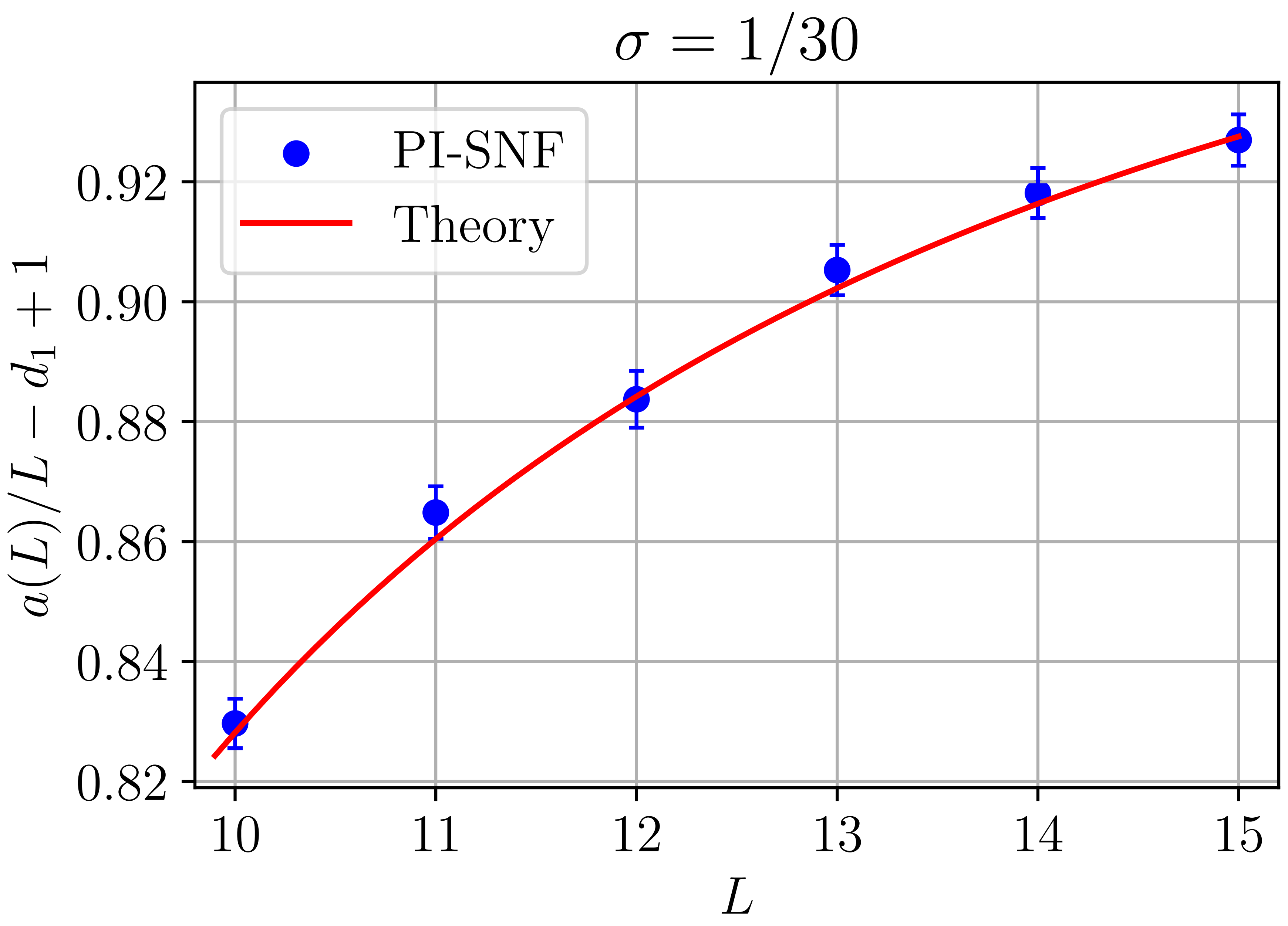
- Design inspired by the $T\bar{T}$ integrable irrelevant perturbation.

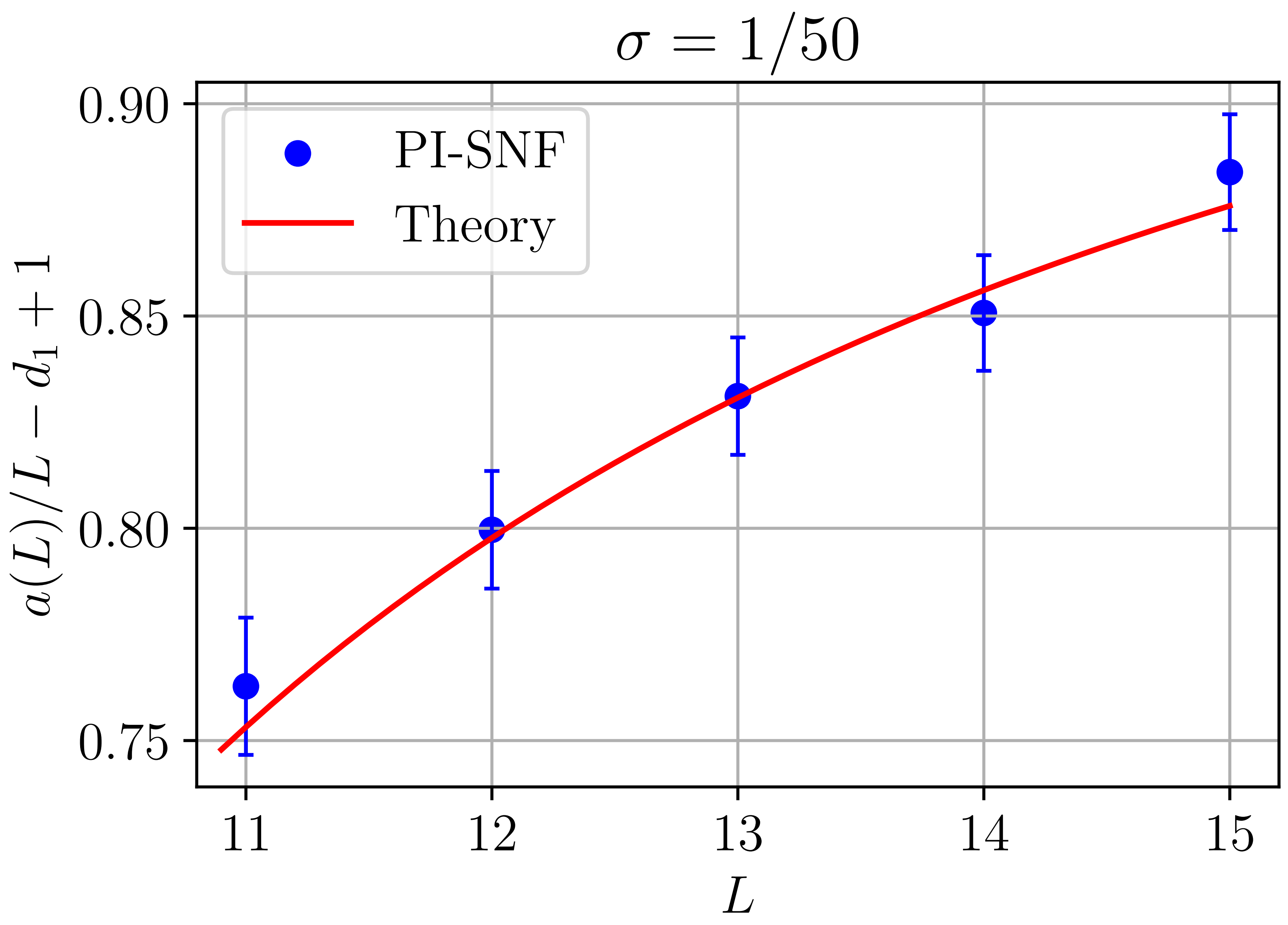
[Cavaglià+; 1608.05534],[Smirnov and Zamolodchikov; 1608.05499],[Caselle, E.C., Nada; 2309.14983]

NG FREE ENERGY $R \gg L$

$$-\log Z = \sigma R L \sqrt{1 - \frac{\pi}{3\sigma L^2}} + \cdots$$

Fitted: $-1.03(2)$,
Target: $-1.047...$





Fitted: $-1.04(7)$,
Target: $-1.047...$

NG WIDTH $R \gg L$

$\sigma = 1/10$

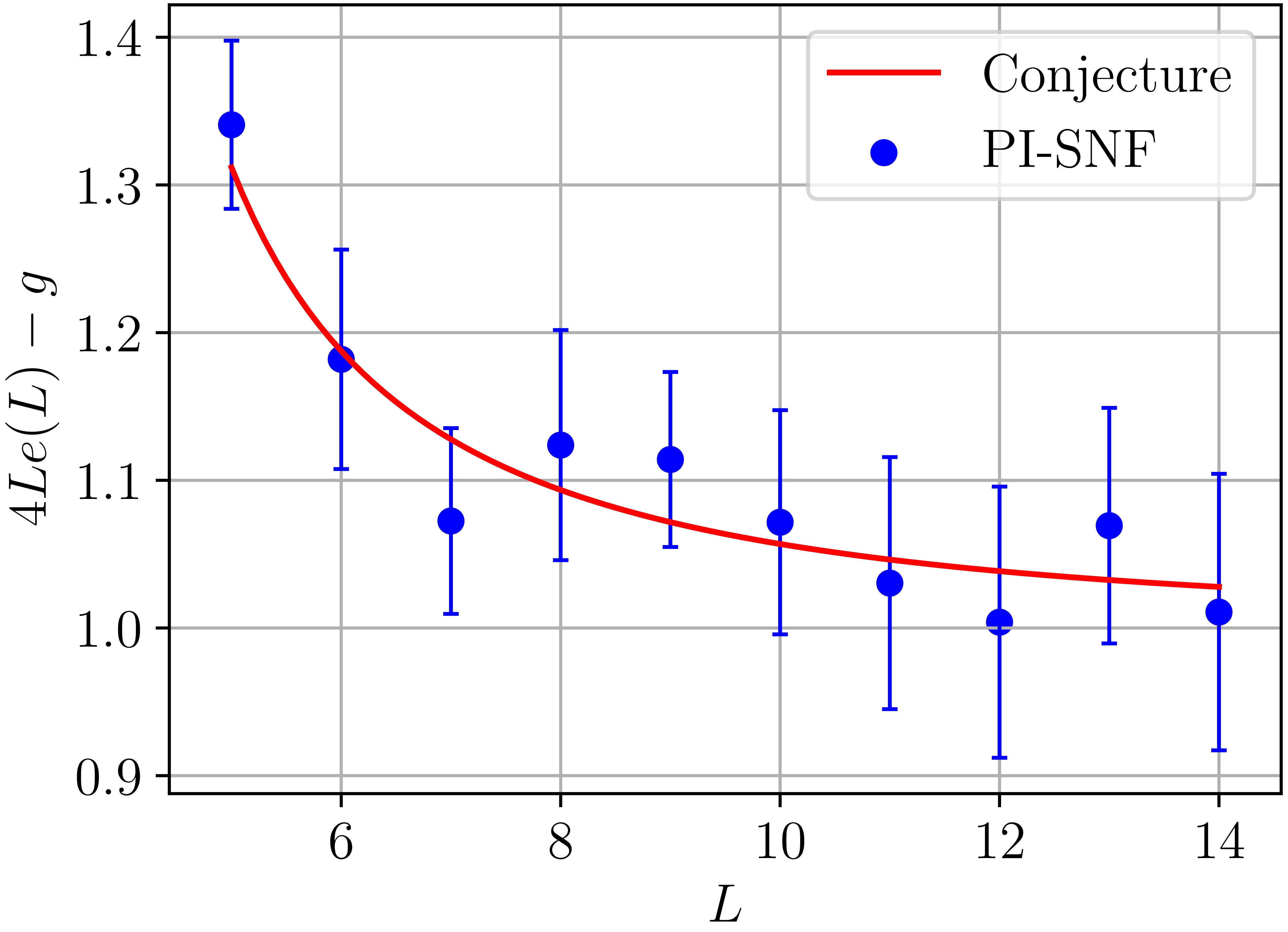
Conjecture:

$$\sigma w^2 = \underbrace{\frac{1}{\sqrt{1 - \frac{\pi}{3\sigma L^2}}}}_{\text{NG: } \sigma(L)/\sigma} \underbrace{\frac{R}{4L}}_{\text{Gaussian solution}} + \cdots$$

NG: $\sigma(L)/\sigma$

Gaussian solution

Fitted: $-1.09(8)$, target: $-1.047...$

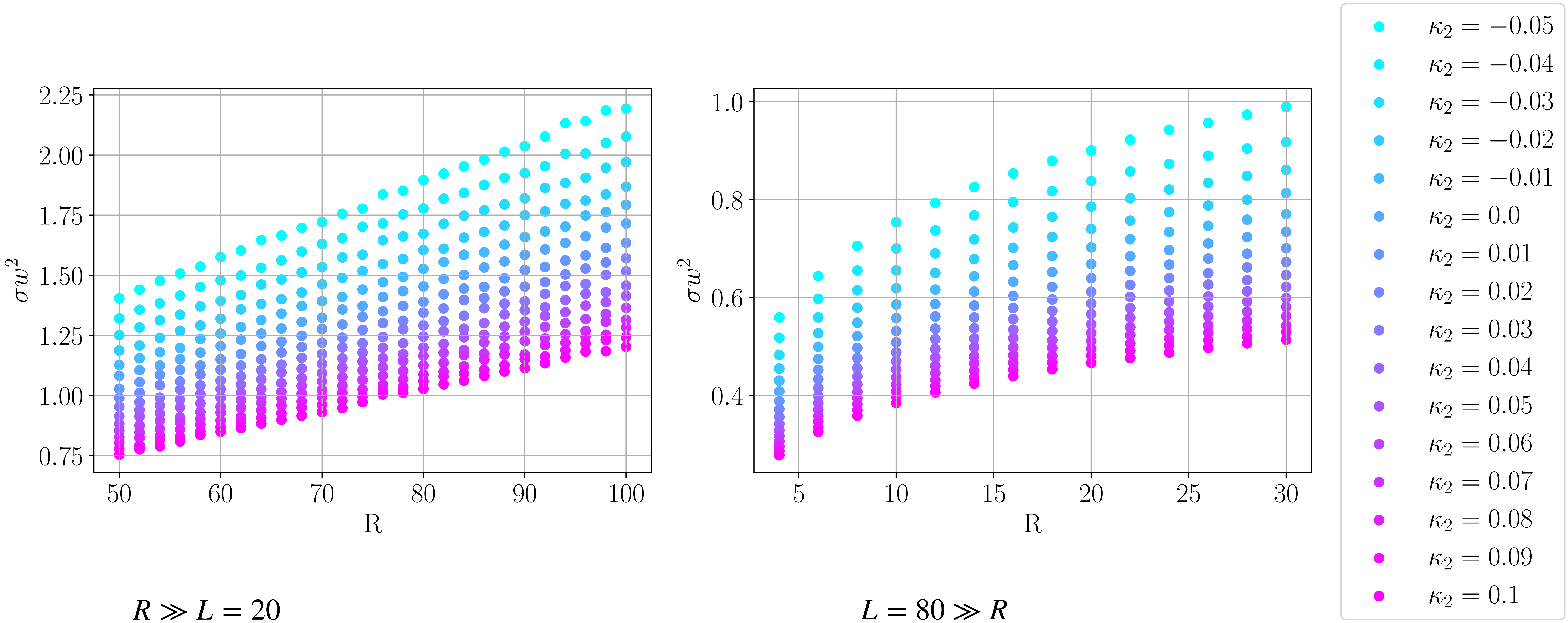


[Caselle, EC, Nada; 2309.14983][Caselle;1004.3875]

BEYOND NG: WIDTH

$$S_{EST}(\phi) = S_{NG}(\phi) + \kappa_2 \mathcal{K}^2(\phi)$$

$$\mathcal{K}^2(\phi) = \sum_{(\tau,\epsilon)\in\Lambda} \mathcal{L}^2(\phi(\tau,\epsilon)) = \sum_{(\tau,\epsilon)\in\Lambda} (\partial_\tau\partial_\tau\phi(\tau,\epsilon))^2 + (\partial_\epsilon\partial_\epsilon\phi(\tau,\epsilon))^2 + 2(\partial_\tau\partial_\epsilon\phi(\tau,\epsilon))^2$$

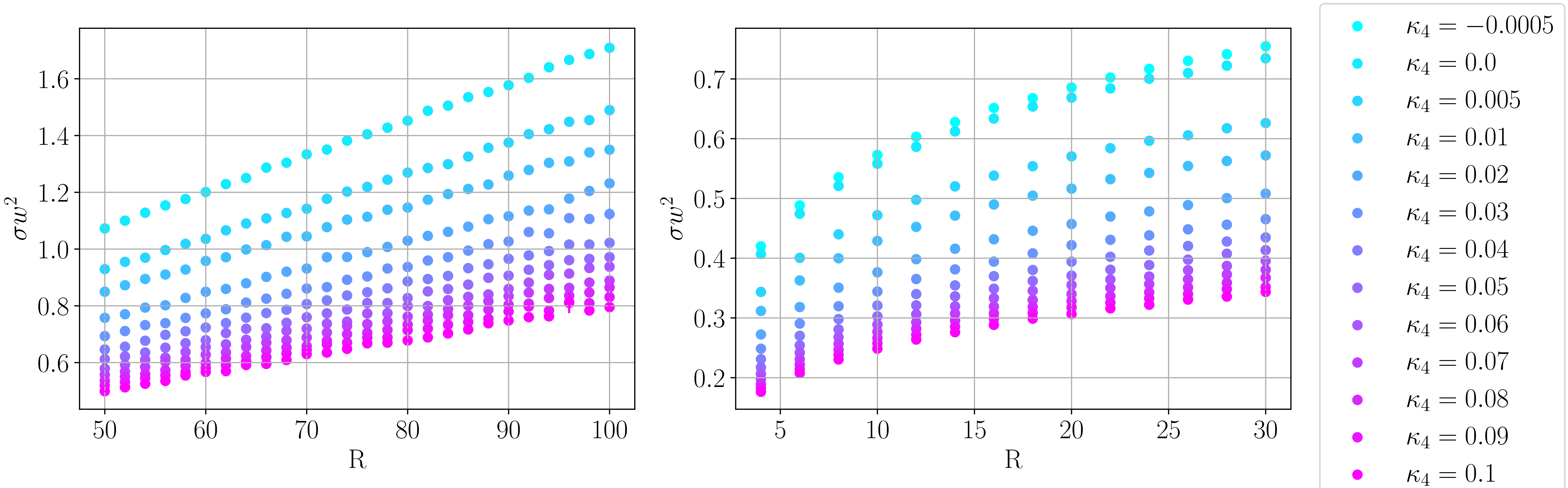


$\sigma = 100$ $R \gg L = 20$ $L = 80 \gg R$

BEYOND NG: WIDTH

$$S_{EST}(\phi) = S_{NG}(\phi) + \kappa_4 \mathcal{K}^4(\phi)$$

$$\mathcal{K}^4(\phi) = \sum_{x\in\Lambda} \left(\mathcal{L}^2(\phi(x))\right)^2$$



$\sigma = 100$ $R \gg L = 20$ $L = 80 \gg R$

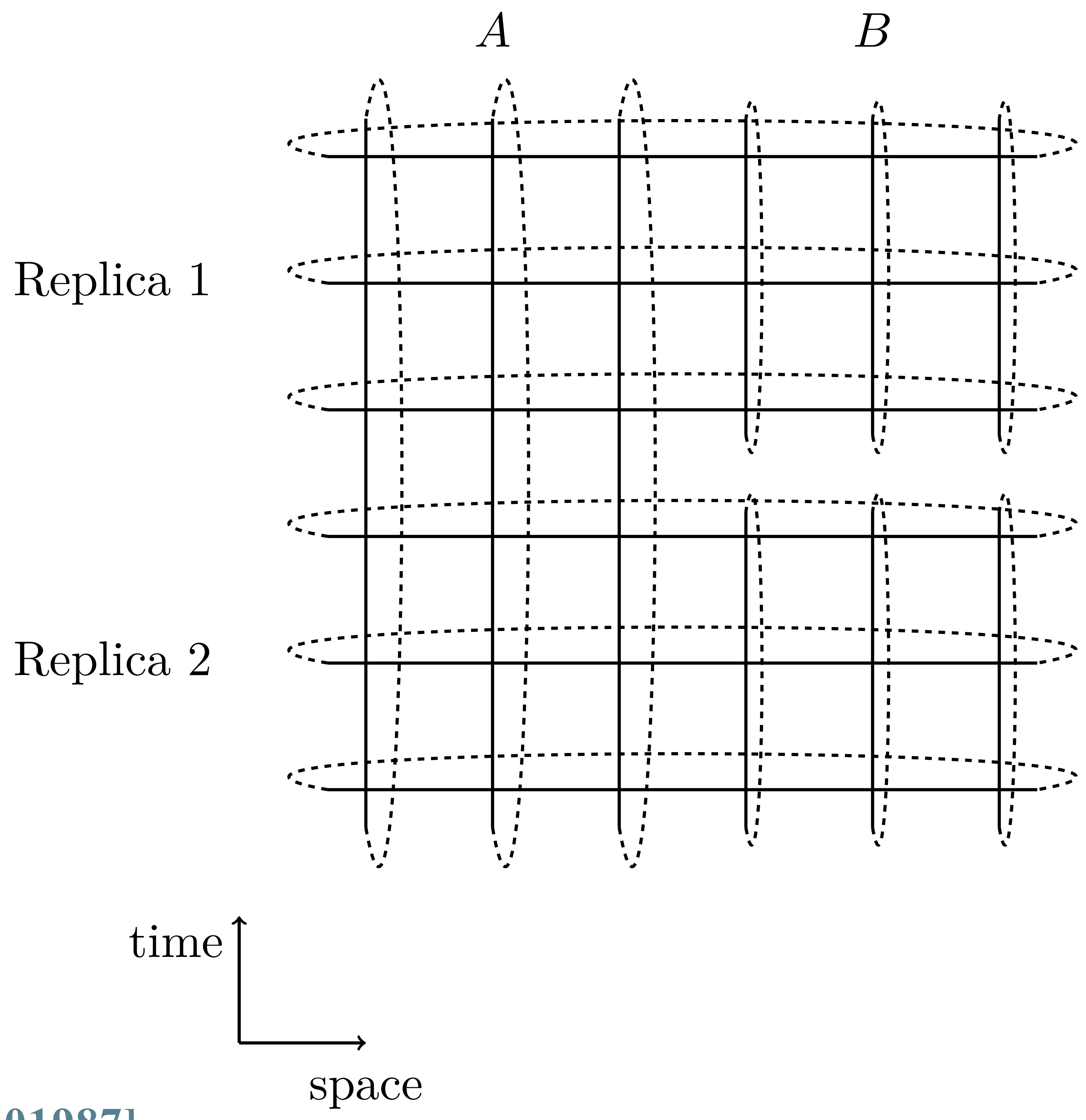
ENTANGLEMENT ENTROPY

ENTANGLEMENT ENTROPY

A way to study the **entanglement in quantum field theory** is by studying the **Rényi entropy** and the **entropic c-functions** using the replica trick:

$$S_n = \frac{1}{1-n} \log \frac{Z_n}{Z^n}$$

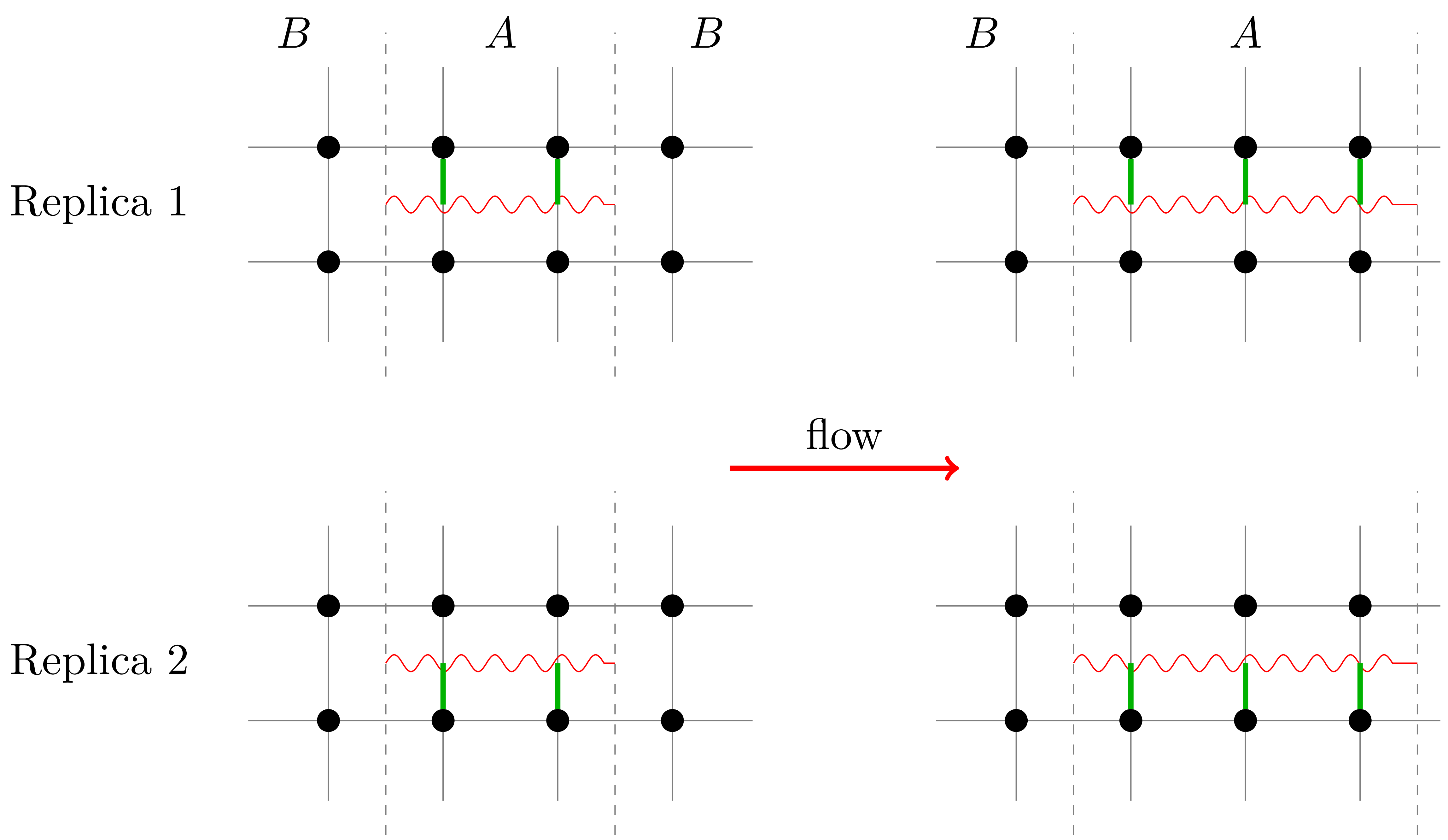
$$C_n = \frac{l^{D-1}}{|\partial A|} \frac{\partial S_n}{\partial l} = \frac{l^{D-1}}{|\partial A|} \frac{1}{n-1} \log \frac{Z_n(l)}{Z_n(l+1)}$$



[Calabrese and Cardy; hep-th/0405152], [A. Bulgarelli and M. Panero; 2304.03311, 2404.01987]

ENTROPIC C-FUNCTION

The entropic c-functions can be computed on the lattice using the Jazynski's equality:



Protocol: $Z_n(l) \rightarrow Z_n(l+1)$

[A. Bulgarelli and M. Panero; 2304.03311, 2404.01987]

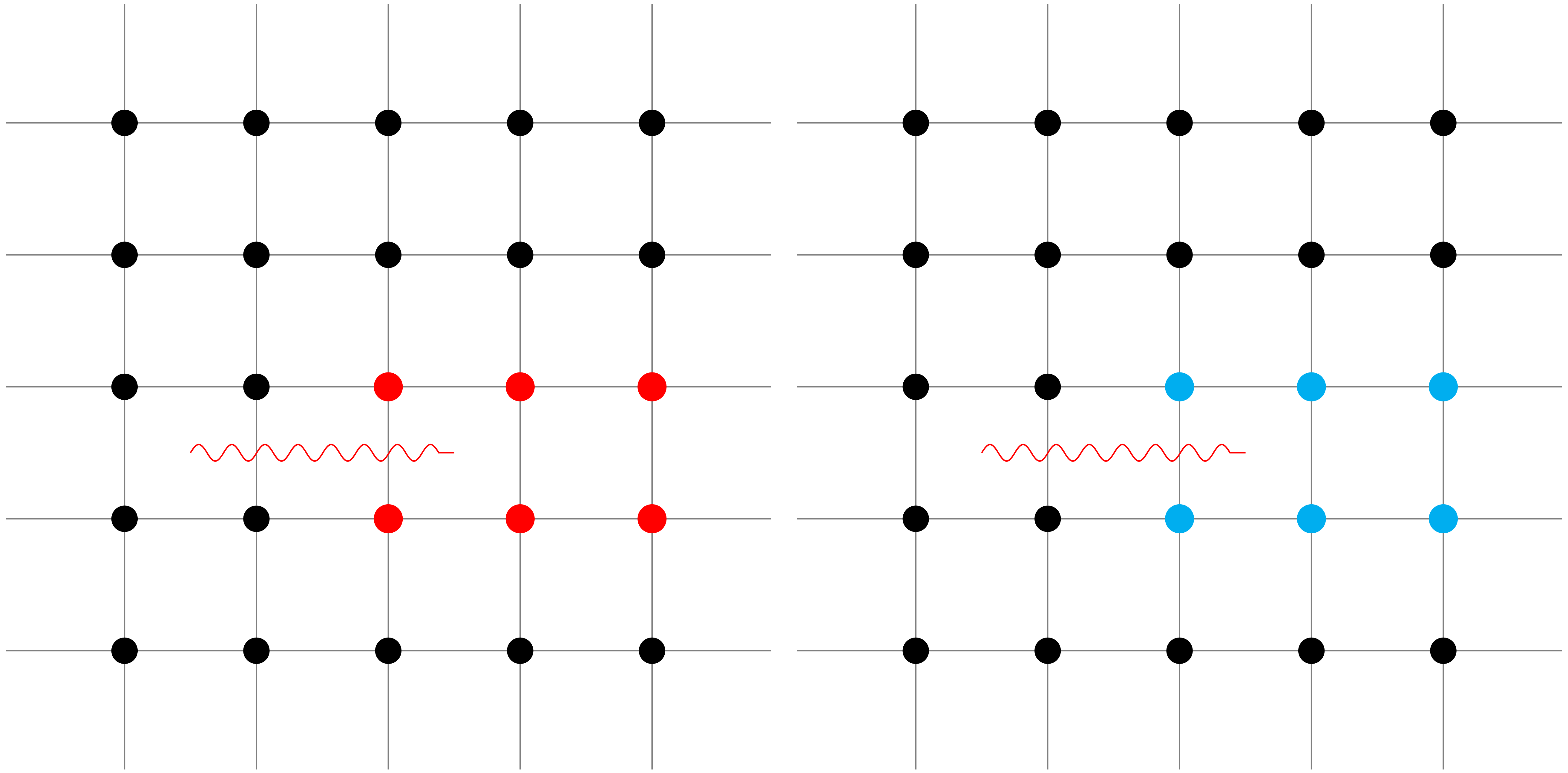
Related work:

[Białas et al.; 2406.06193]

See talk by: Tomasz Stebel

DEFECT COUPLING LAYERS

The coupling layers (CLs) act locally only on a portion of the defect



Frozen Replica

Active Replica

- Environment (always frozen)
- Frozen (input to CL)
- Active (transformed by CL)

Patch used: 4×5

(2×3 for pure CNN NF)

Related work:

[Abbott et al.; 2404.11674]

SOME TECHNICAL DETAILS

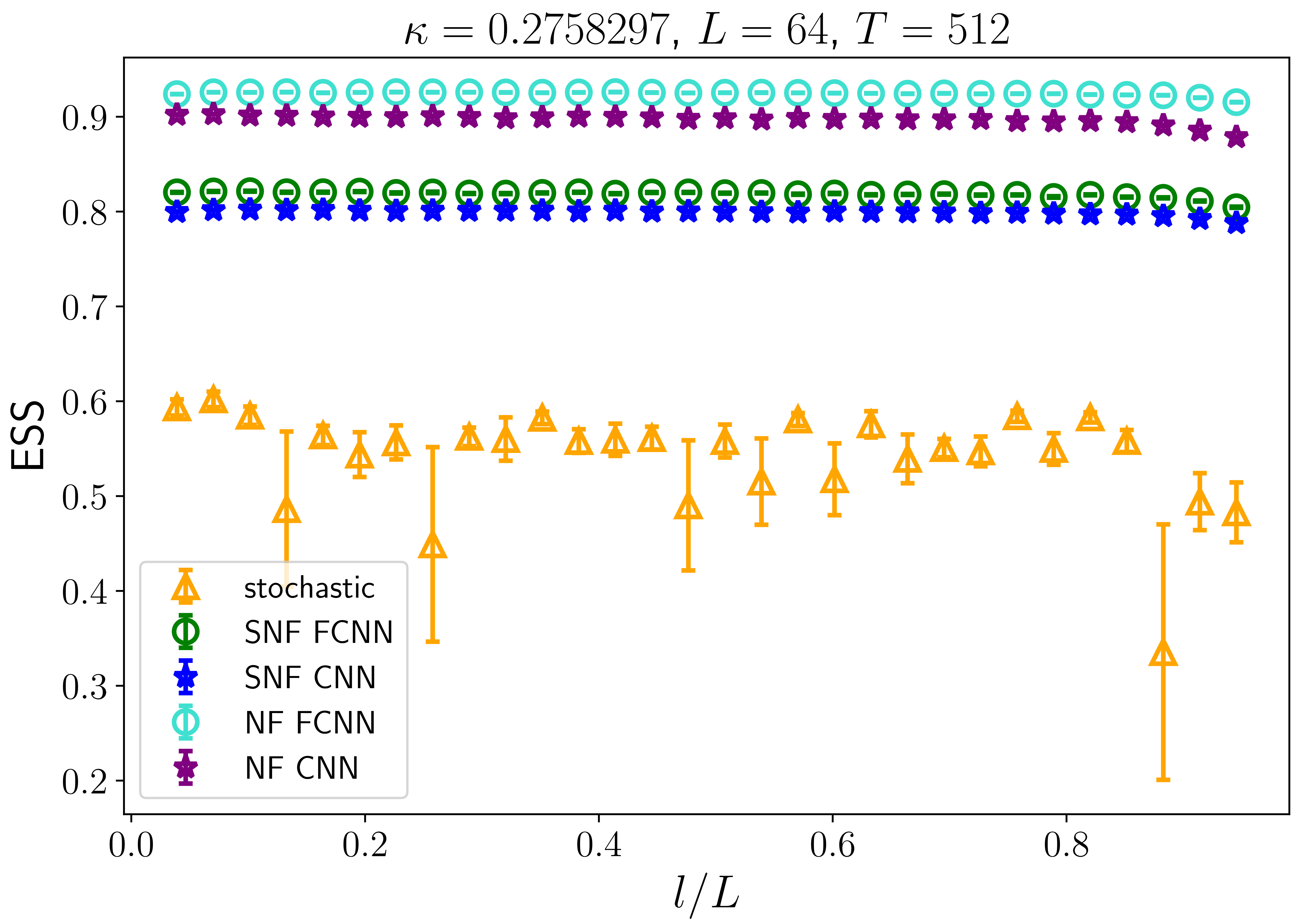
We studied the entropic c-functions C_2 in the **ϕ^4 scalar field theory** with action:

$$S = -2k \sum_{<ij>} \phi_i \phi_j + (1-2\lambda) \sum_i \phi_i^2 + \lambda \sum_i \phi_i^4$$

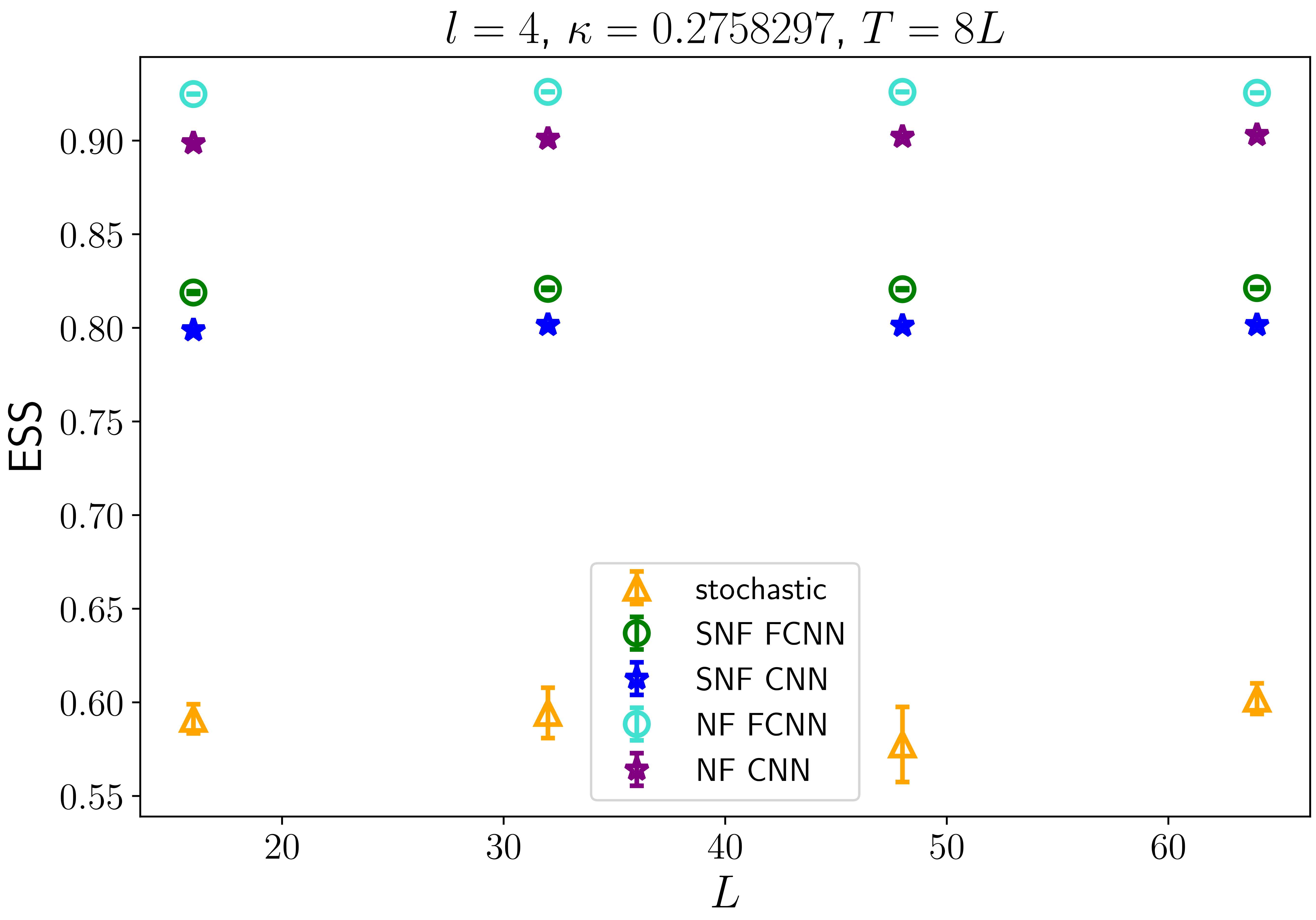
- We generally worked at **criticality**: $\kappa = 0.2758297$, $\lambda = 0.03$
  [Bosetti et al.; 1506.08587]
- In this presentation, we show results only for the $D = 1 + 1$ case with $V = T \times L = 8L \times L$ (zero temperature)

We trained the CLs only for $L = 16$ and $l = 1$, we then evaluate the model for different L and l without any retraining!

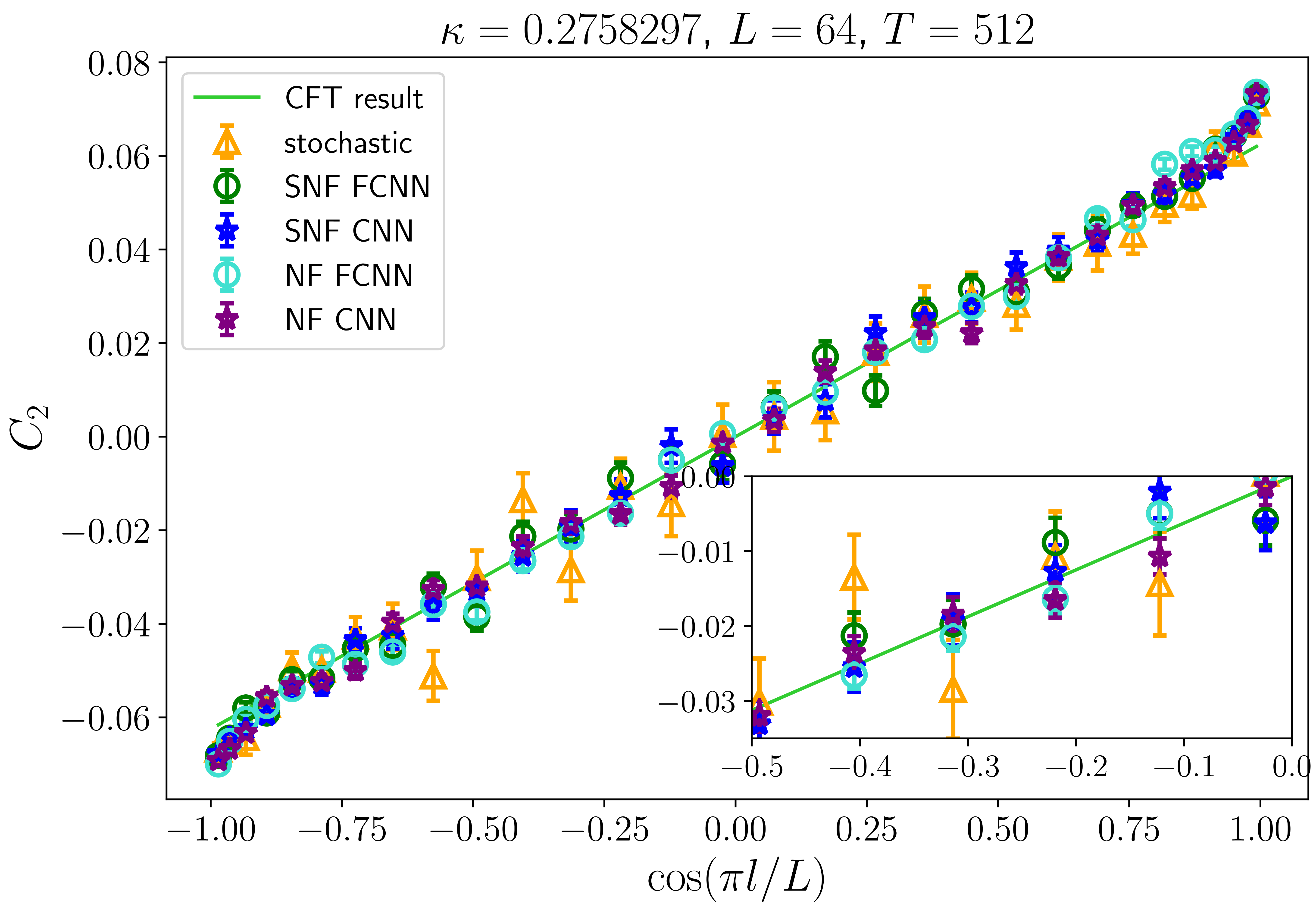
ESS VS CUT LENGTH



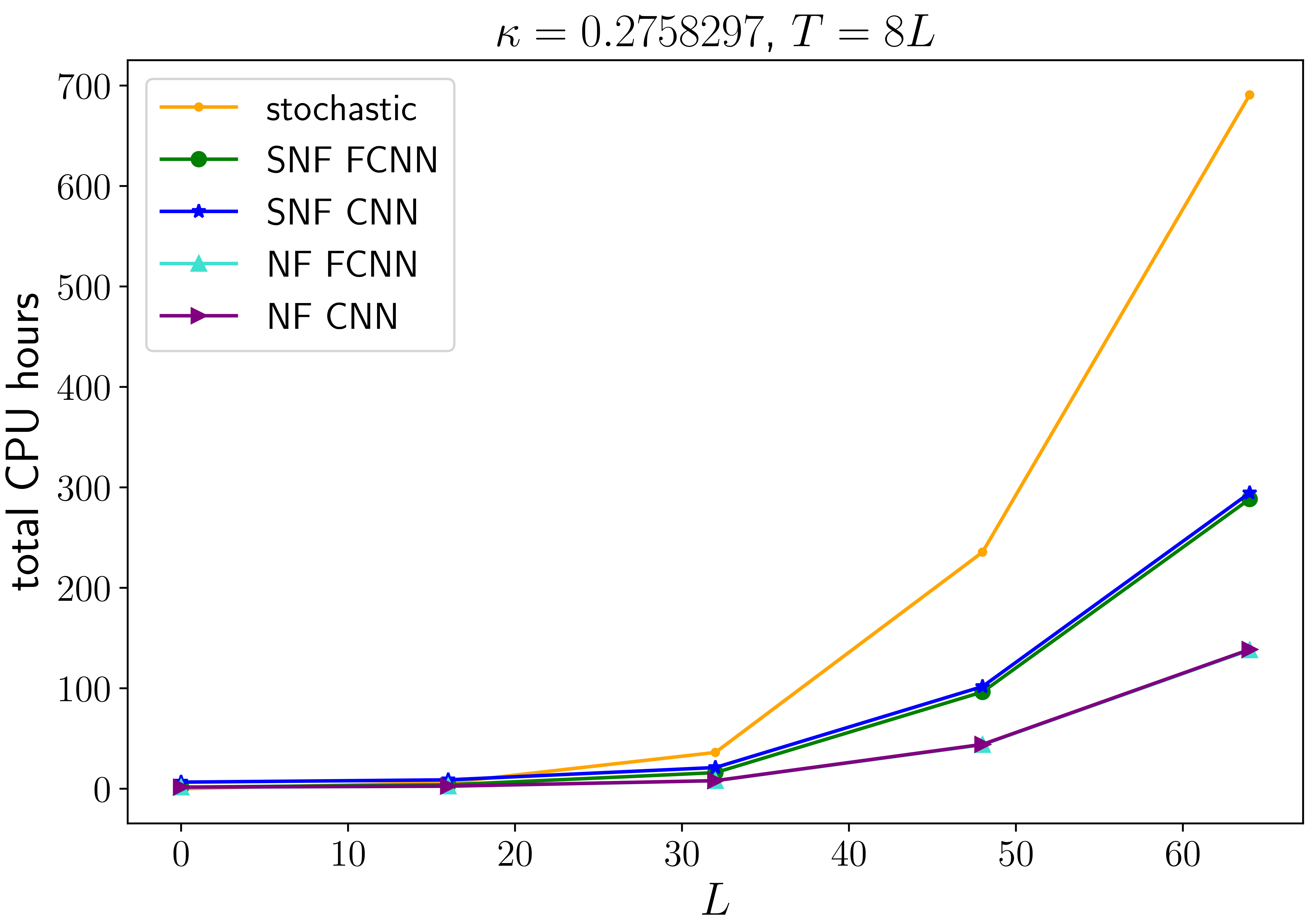
ESS VS VOLUME



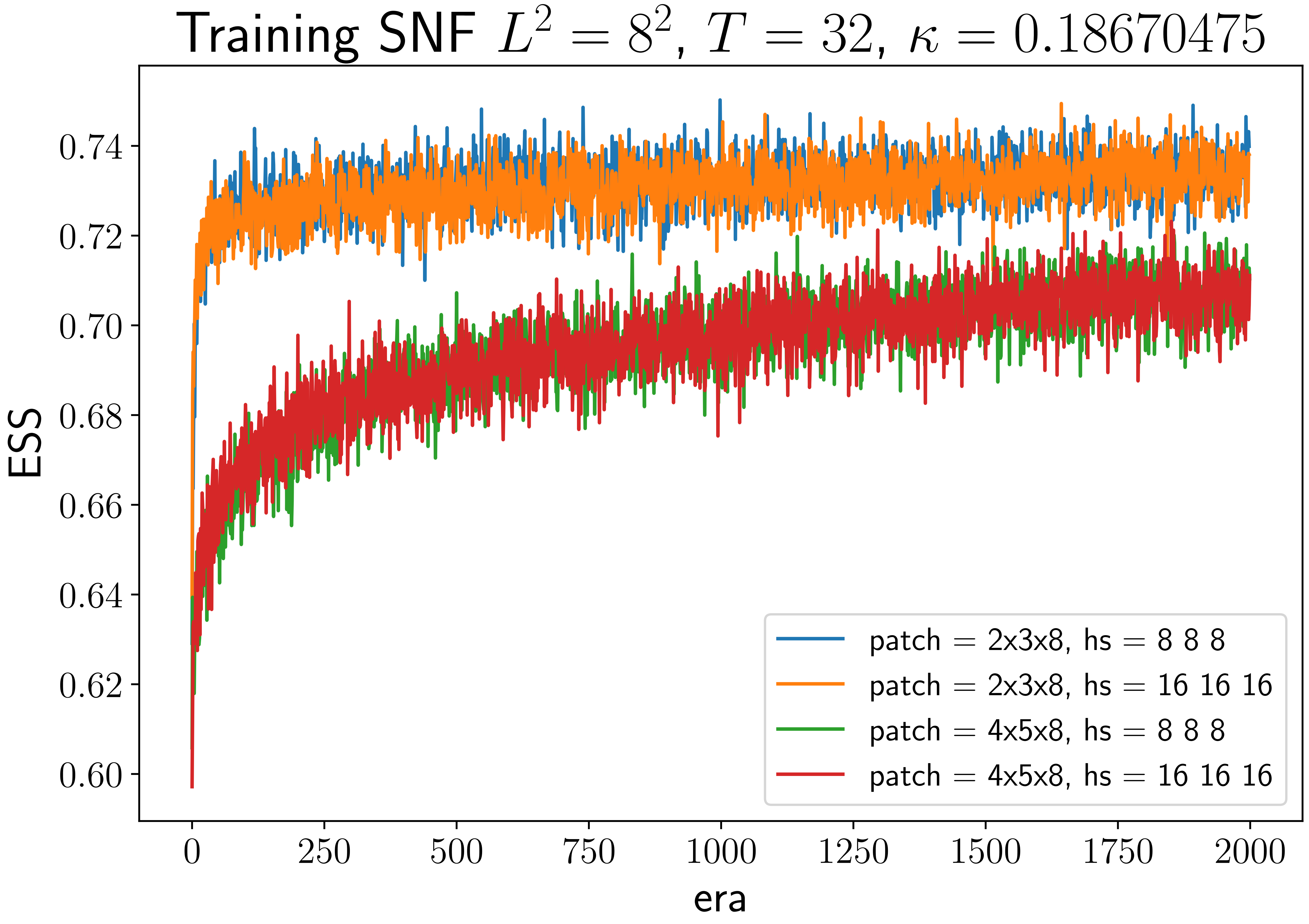
CFT PREDICTION

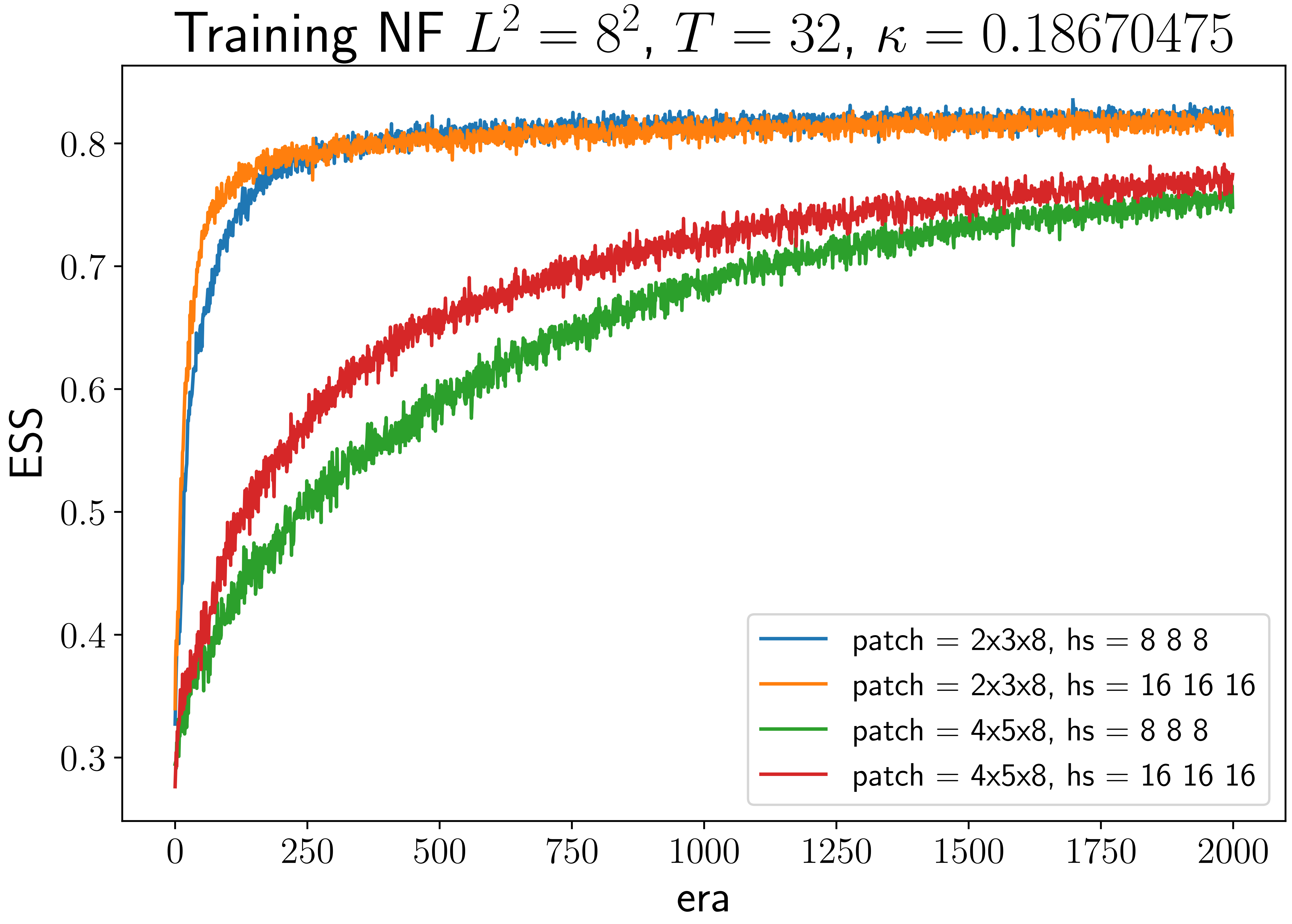


SCALING TIMES



$D = 2 + 1$


Training SNF $L^2 = 8^2$, $T = 32$, $\kappa = 0.18670475$
ESS
era
0.60
0.62
0.64
0.66
0.68
0.70
0.72
0.74
0
250
500
750
1000
1250
1500
1750
2000
patch = 2x3x8, hs = 8 8 8
patch = 2x3x8, hs = 16 16 16
patch = 4x5x8, hs = 8 8 8
patch = 4x5x8, hs = 16 16 16


Training NF $L^2 = 8^2$, $T = 32$, $\kappa = 0.18670475$
ESS
era
0.3
0.4
0.5
0.6
0.7
0.8
0
250
500
750
1000
1250
1500
1750
2000
patch = 2x3x8, hs = 8 8 8
patch = 2x3x8, hs = 16 16 16
patch = 4x5x8, hs = 8 8 8
patch = 4x5x8, hs = 16 16 16

OUTLOOK

- We outlined SNFs pointing out the connections with non-equilibrium thermodynamics
  1. see talk by A. Nada on SNFs for $SU(3)$

- SNFs can be successful applied to study EST:
  1. Beyond Nambu-Goto
  2. EST with “fermions”
  3. Four dimensions and axions
  4. Others integrable models

- SNFs can improve the study of entanglement entropy on the lattice
  1. More complex theories
  2. Others defects (open boundary conditions, …)

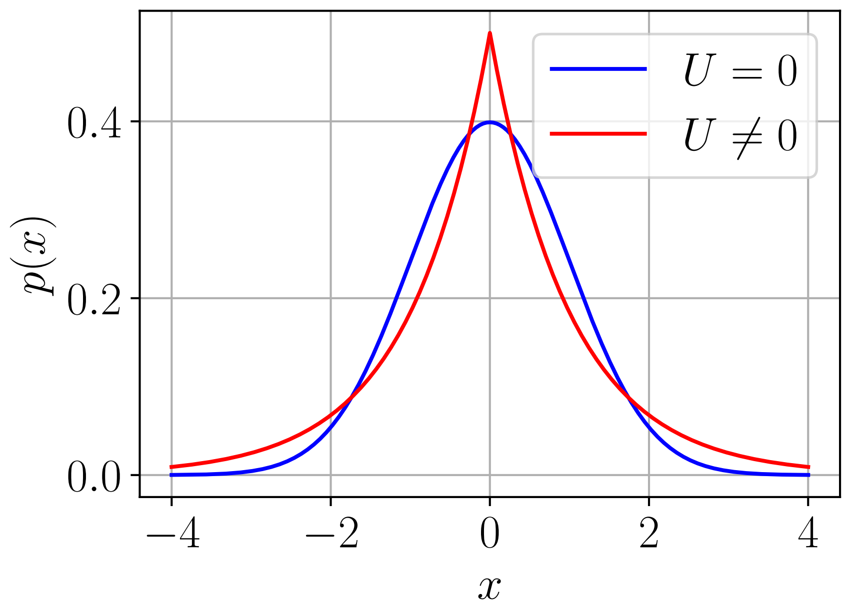
THANK YOU FOR YOUR ATTENTION!

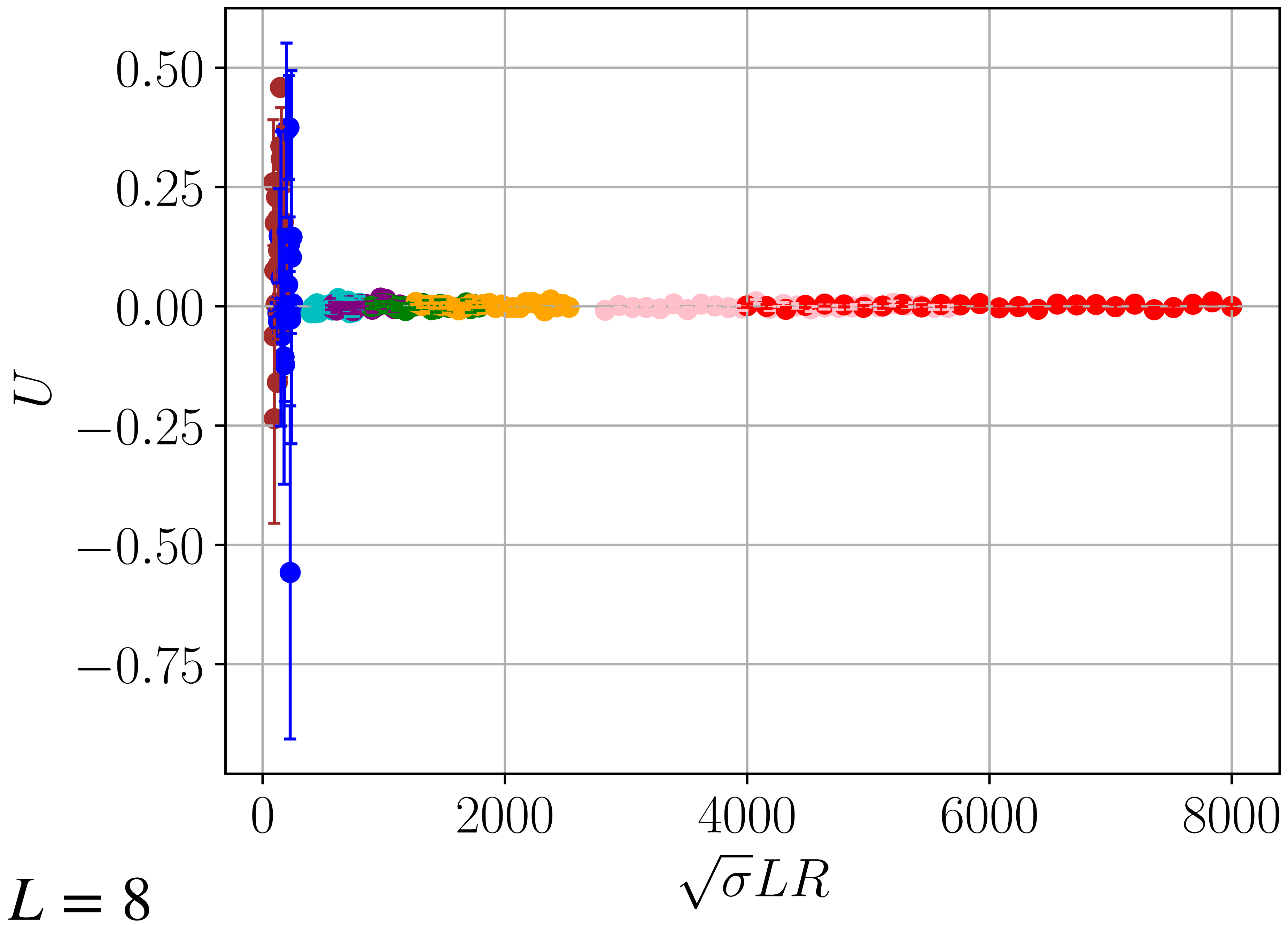
SNFS: RELATED WORKS

- Annealed Importance Sampling: Equivalent to Jarzynski's equality. Used in the original SNF paper
  [Neal; physics/9803008]
- Sequential Monte Carlo: Generalization of AIS.
  [Dai+; 2007.11936]
- SNF idea reworked in CRAFT
  [Matthews+; 2201.13117]
- An hybrid (deterministic/stochastic) approach with no neural networks has been proposed also by Jarzynksi in 2011
  [Vaikuntanathan and Jazynski; 1101.2612]
- FAB: combination of NFs and AIS.
  [Midgley+; 2208.01893]
- Exact work for discretized Langevin dynamics.
  [Sivak+; 1107.2967]

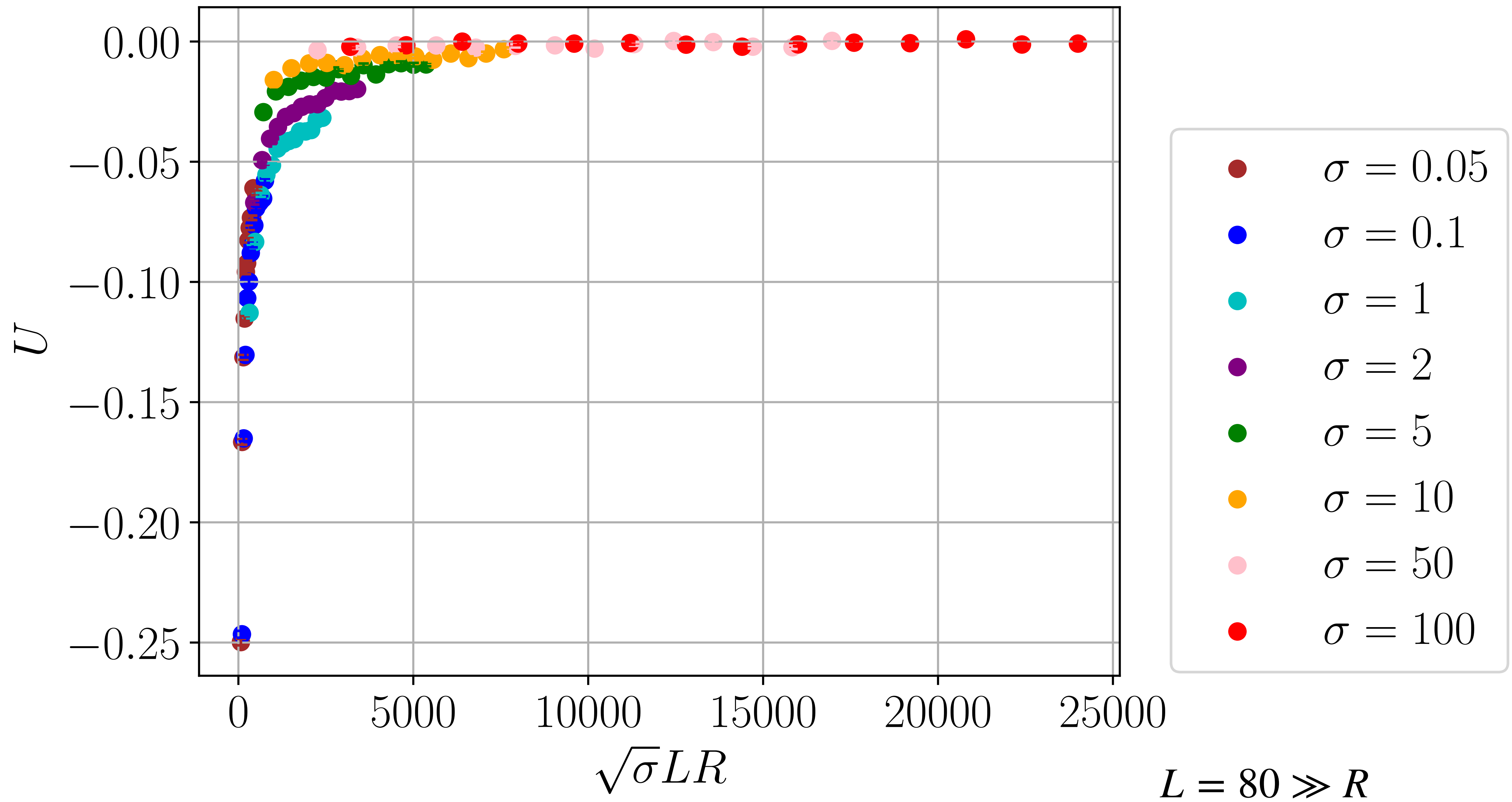
IS THE NG FLUX TUBE SHAPE GAUSSIAN?

$$U = 1 - \frac{\langle \phi^4(\tau, R/2) \rangle_\tau}{3\langle \phi^2(\tau, R/2) \rangle_\tau^2}$$





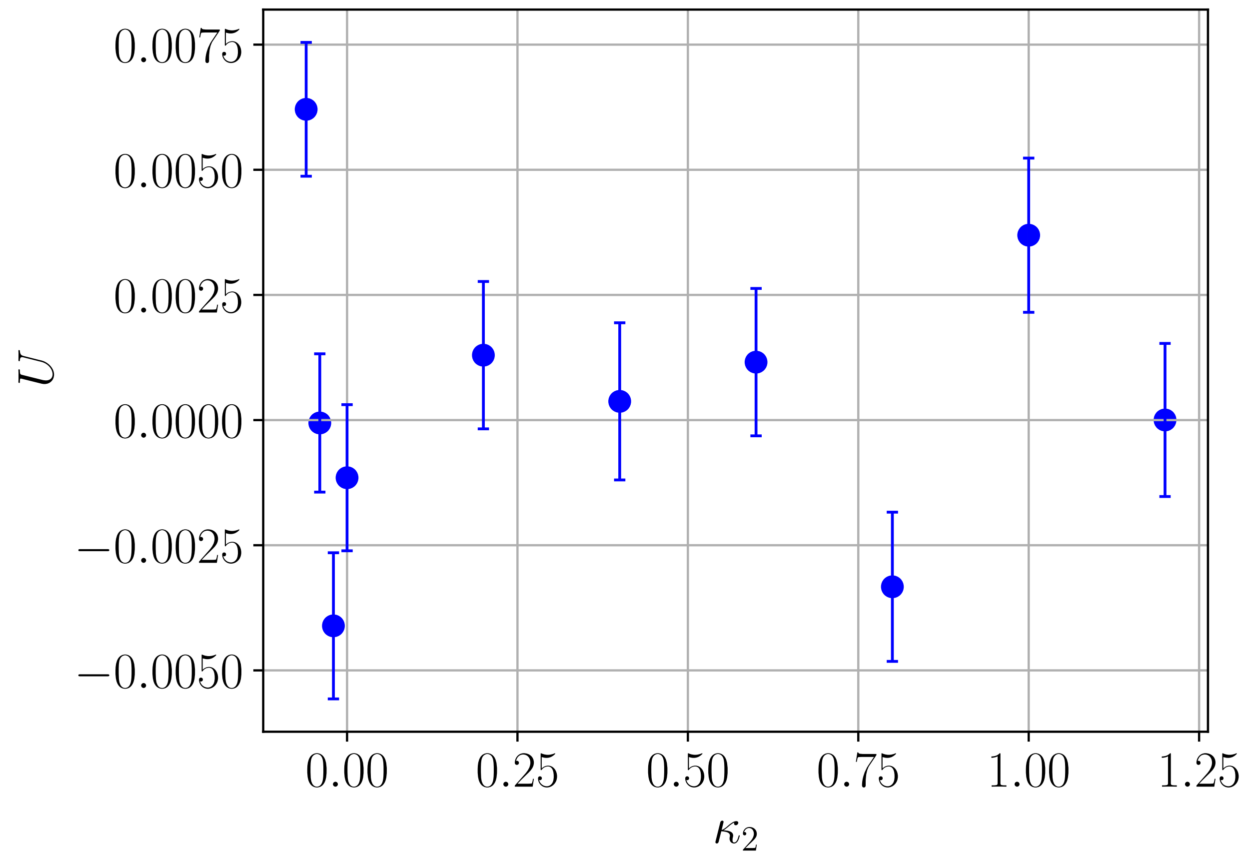
$R \gg L = 8$

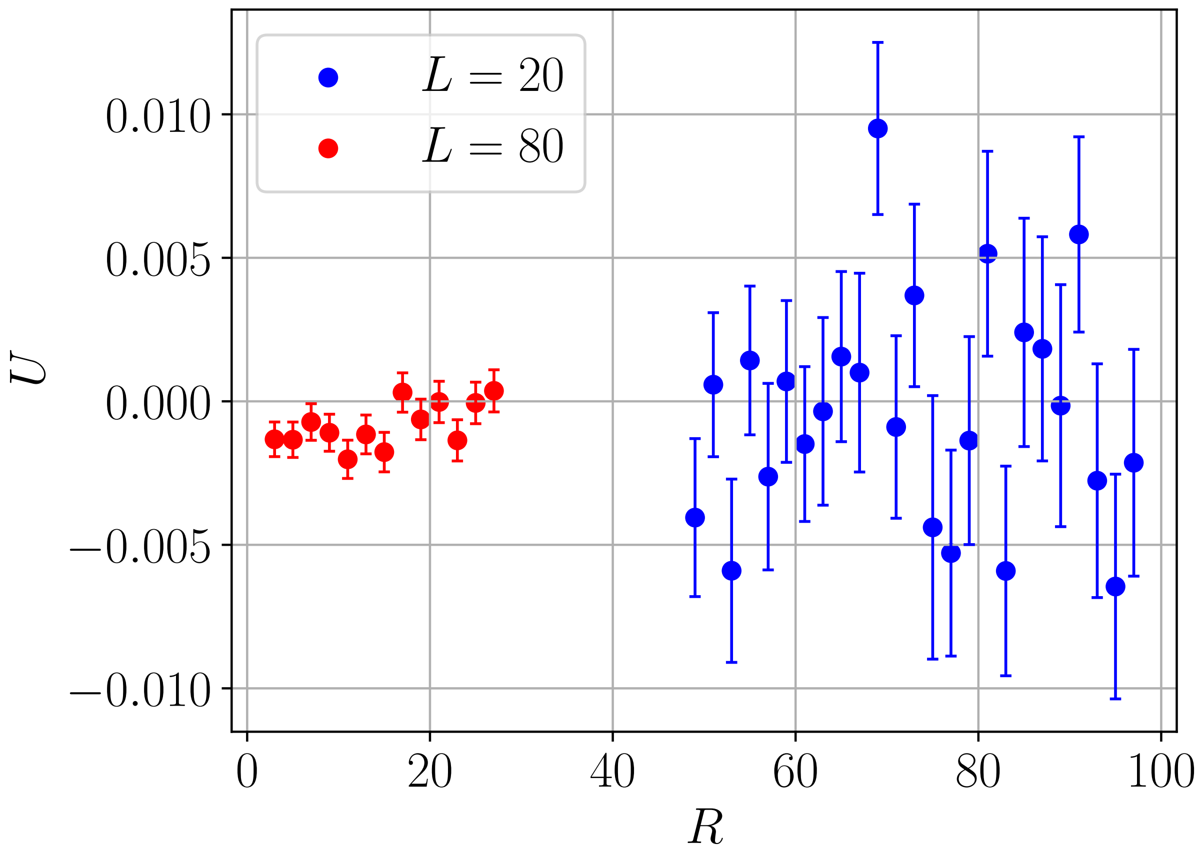


$L = 80 \gg R$

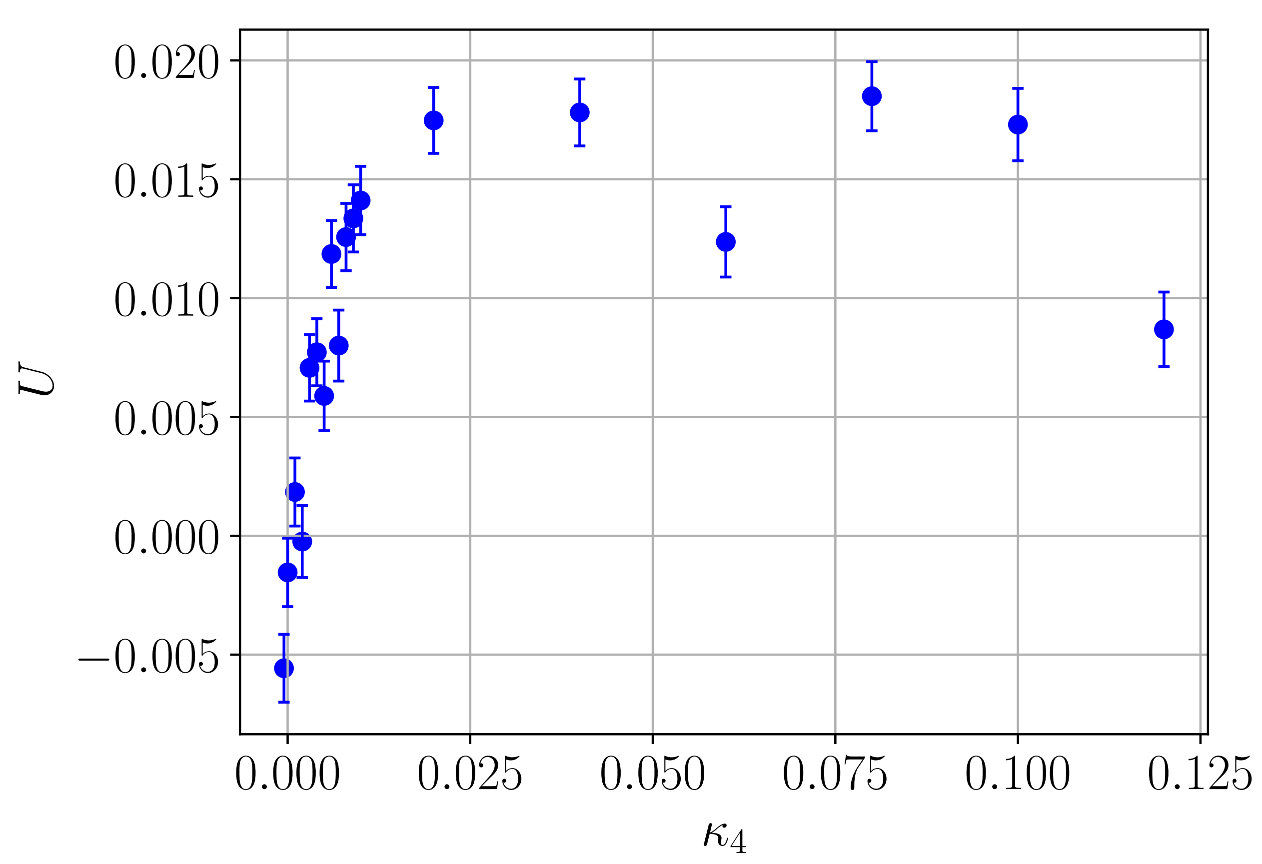
BEYOND NG: GAUSSIANITY

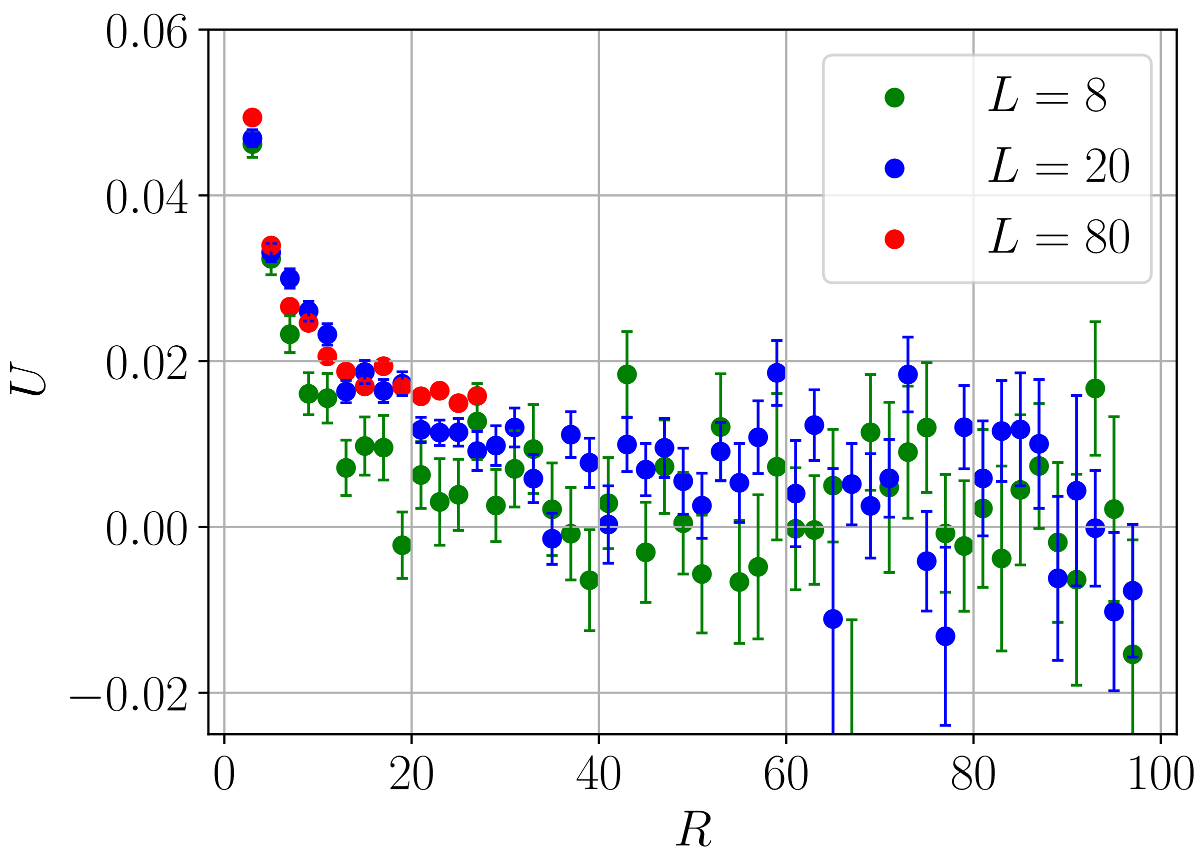
$\sigma = 100$





$$S_{NG} + \kappa_2 \mathcal{K}^2$$





$$S_{NG} + \kappa_4 \mathcal{K}^4$$