Stochastic normalizing flows as non-equilibrium transformations

Berlin AQTIVATE workshop

01/03/2024

Elia Cellini

In collaboration with:

Andrea Bulgarelli, Michele Caselle, Alessandro Nada, Marco Panaro

Based on:

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

M. Caselle, E.C., A. Nada
- arxiv: 2309.14983 (Lattice 2023 PoS)
- JHEP 02 (2024) 048, arxiv:2307.01107
- arXiv: 24??.???? (in prep.)

A. Bulgarelli, E.C. , A. Nada
- arXiv: 24??.???? (in prep.)

UNIVERSITÀ DI TORINO

INFN

Outline

1. Lattice Field Theory
2. Jarzynski's Equality & Non-Equilibrium MCMC
3. Normalizing Flows
4. Stochastic Normalizing Flows
5. Numerical Results

Lattice Field Theory

Lattice Field Theory

In **Lattice Field Theory** (LFT) the **space-time** **is discretized on a lattice**

$$\Downarrow$$

Statistical mechanics formalization of **quantum field theory**:

$$\langle \mathcal{O} \rangle_{\phi \sim p} = \int D\phi p(\phi) \mathcal{O}(\phi) \qquad p(\phi) = \frac{1}{Z} e^{-S_E[\phi]} \qquad Z \equiv \int D\phi e^{-S_E[\phi]}$$

Markov Chain Monte Carlo (MCMC)

$$\phi_1 \xrightarrow{P} \phi_2 \xrightarrow{P} \phi_3 \xrightarrow{P} \cdots \xrightarrow{P} \phi_M$$

Standard MCMC method:

1. **Transition probability:** $P \propto \exp(-S)$; $S \rightarrow$ target action
2. Observables computed with Monte Carlo

$$\langle \mathcal{O} \rangle_{\phi \sim p} \simeq \mathbb{E}_{\phi \sim p}[\mathcal{O}] = \frac{1}{M} \sum_{i=1}^{M} \mathcal{O}(\phi_i)$$

Problems:

1. **Critical Slowing Down**
2. Direct estimation of **partition functions**

Jarzynski's Equality & 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)

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

$$\phi_0 \xrightarrow{P_1} \phi_1 \xrightarrow{P_2} \phi_2 \xrightarrow{P_3} \cdots \xrightarrow{P_N} \phi_N$$

1. Thermalized q_0 "**prior**"
2. $P_i \propto \exp(-S_i)$ change 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:

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

Where:

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

Crooks Fluctuation Theorem

Thus:

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

$\Rightarrow$ **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]

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

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

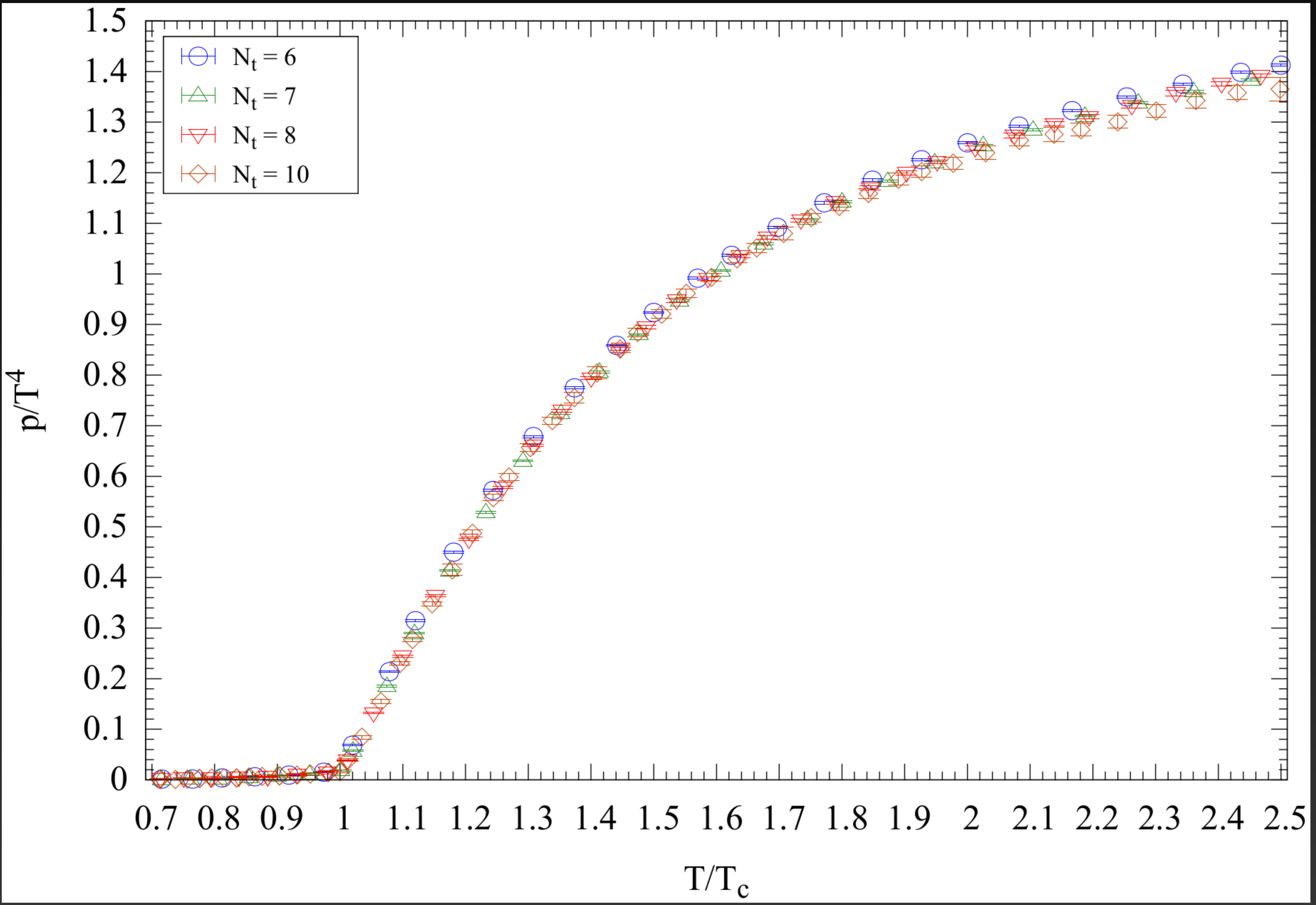
Non-Equilibrium Ensemble

Equilibrium Quantity

NE-MCMC for LFT

Jarzynksi's equality has been exploited to obtain **state-of-the-arts results** in LFT:

- 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]
- Topological freezing
  [Bonanno+; 2402.06561]



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

Equivalent to:

Annealed Importance Sampling
[Neal; physics/9803008]

Numerical Problem

The identities derived before are exact, however, the **exponential** **average** has an high **signal-to-noise** ratio.

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

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!

$$W \longleftrightarrow -\ln \tilde{w}$$

Stochastic Normalizing Flows

Stochastic Normalizing Flows

<u>Stochastic Normalizing Flows (SNFs)</u> 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:

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

$$\Downarrow$$

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

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

SNFs: Dissipated Work

We have now:

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

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:

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

$$\Downarrow$$

$$\langle W^\theta \rangle_f \geq \Delta F \quad \Rightarrow \quad \text{Second Law!}$$

Measure how reversible the process is.

Numerical Results

ϕ^4

ϕ^4 in two dimensions → Simplest, non-trivial, continuous LFT

Displays **Spontaneous Symmetry Breaking** → crucial to describe **phase transitions** and **Higgs mechanism**

$$S_{\phi^4}[\phi] = \sum_{x \in \Lambda} - 2\kappa \sum_{\mu=0,1} \phi(x)\phi(x+\hat{\mu}) + (1-2\lambda)\phi(x)^2 + \lambda\phi(x)^4$$

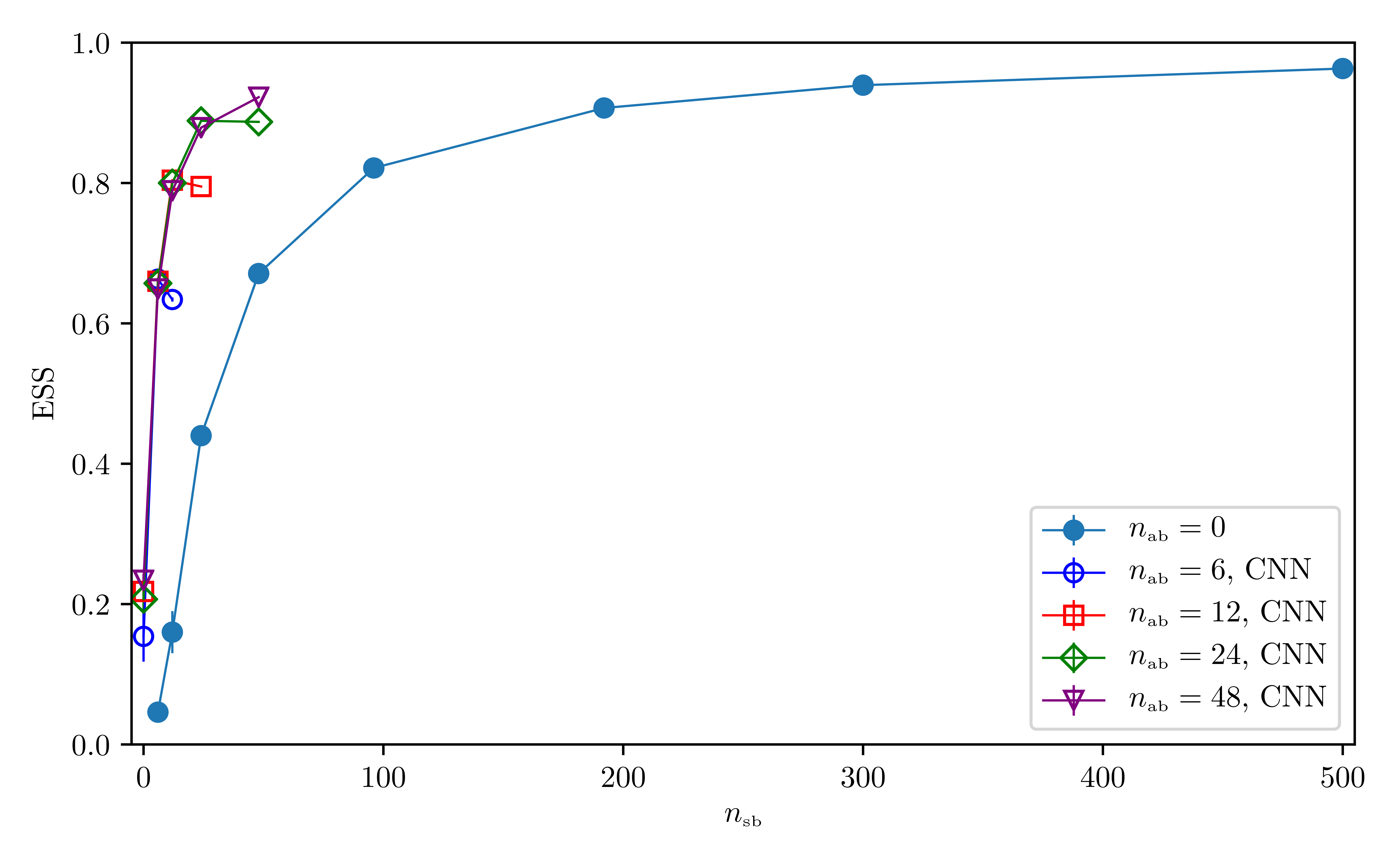
Λ → square lattice of size $N_t \times N_s$ and $\phi(x) \in \mathbb{R}$

$$p(\phi) = \frac{1}{Z} e^{-S[\phi]}$$

ϕ^4: Comparison

$N_t \times N_s = 16 \times 8$ lattice with $\kappa = 0.2$ and $\lambda = 0.022$: unbroken symmetry phase

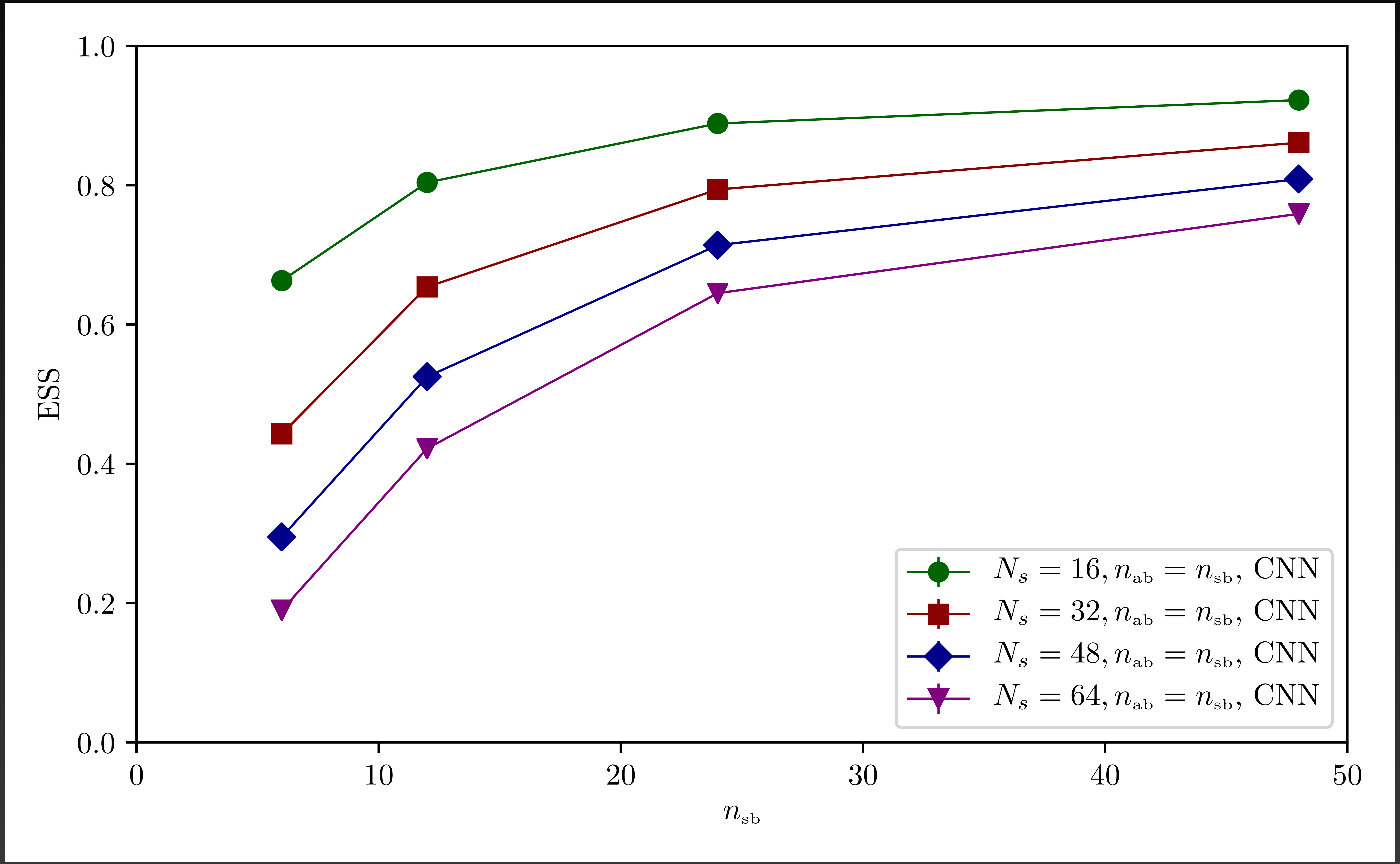
$$ESS = \frac{\langle e^{-W} \rangle_f^2}{\langle e^{-2W} \rangle_f}$$



[Caselle, E.C., Nada, Panero; 2201.08862]

ϕ^4: SNFs Volume Scaling

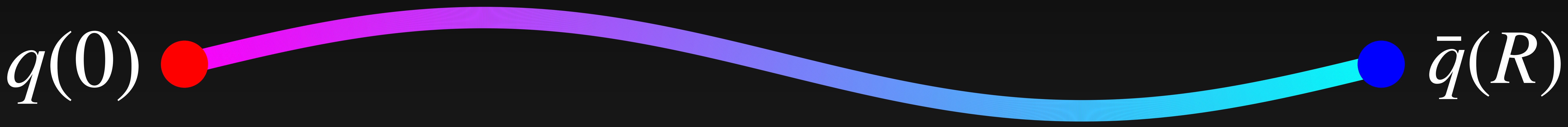
$N_t = 8$, $\kappa = 0.2$ and $\lambda = 0.022$



[Caselle, E.C., Nada, Panero; 2201.08862]

Nambu-Goto Effective String Theory (NG EST)

Low dimensional model for **confinement** in pure **gauge theories.**



Used to model and predict infrared behaviors of lattice gauge theories

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

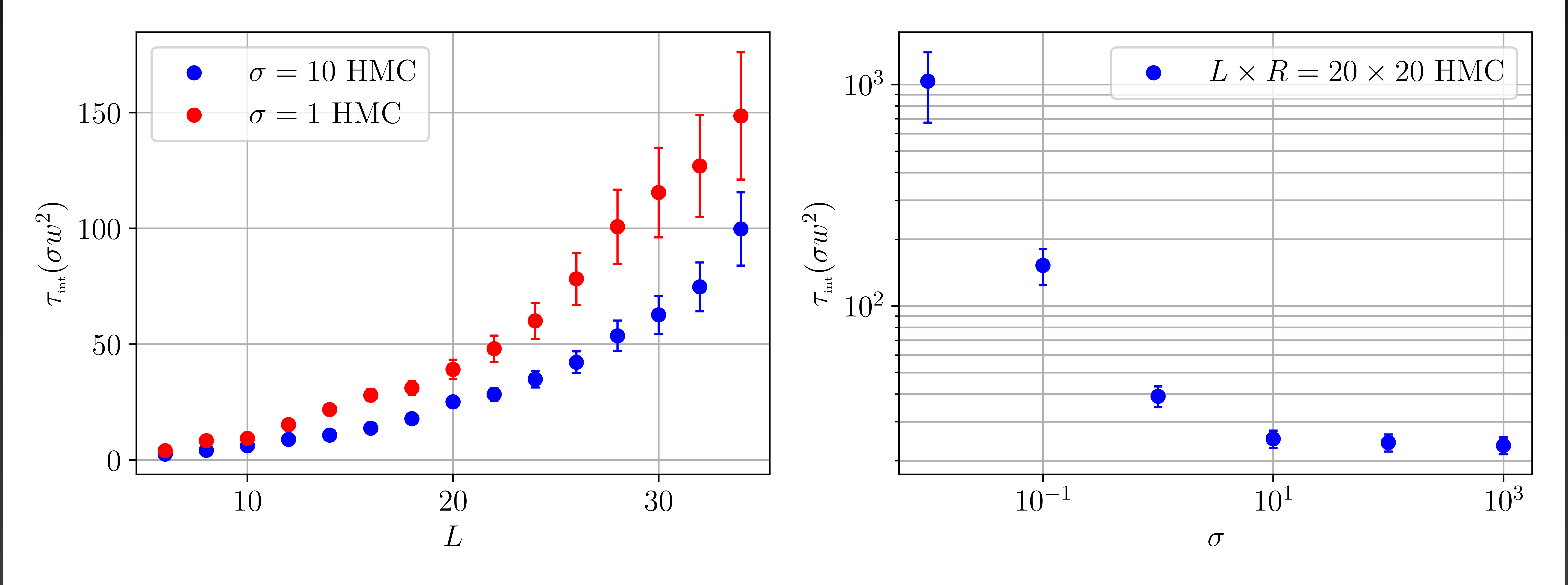
$$S_{NG}[\phi] = \sigma \sum_{x\in\Lambda=L\times R}\left[\sqrt{1+(\partial_{\mu}\phi)^2/\sigma}-1\right]$$

[Aharony and Komargodski; 1302.6257],[Caselle; 2104.10486],[Caselle, E.C., Nada; 2307.01107]

NG: Lacks of numerical methods

Numerical problems:

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

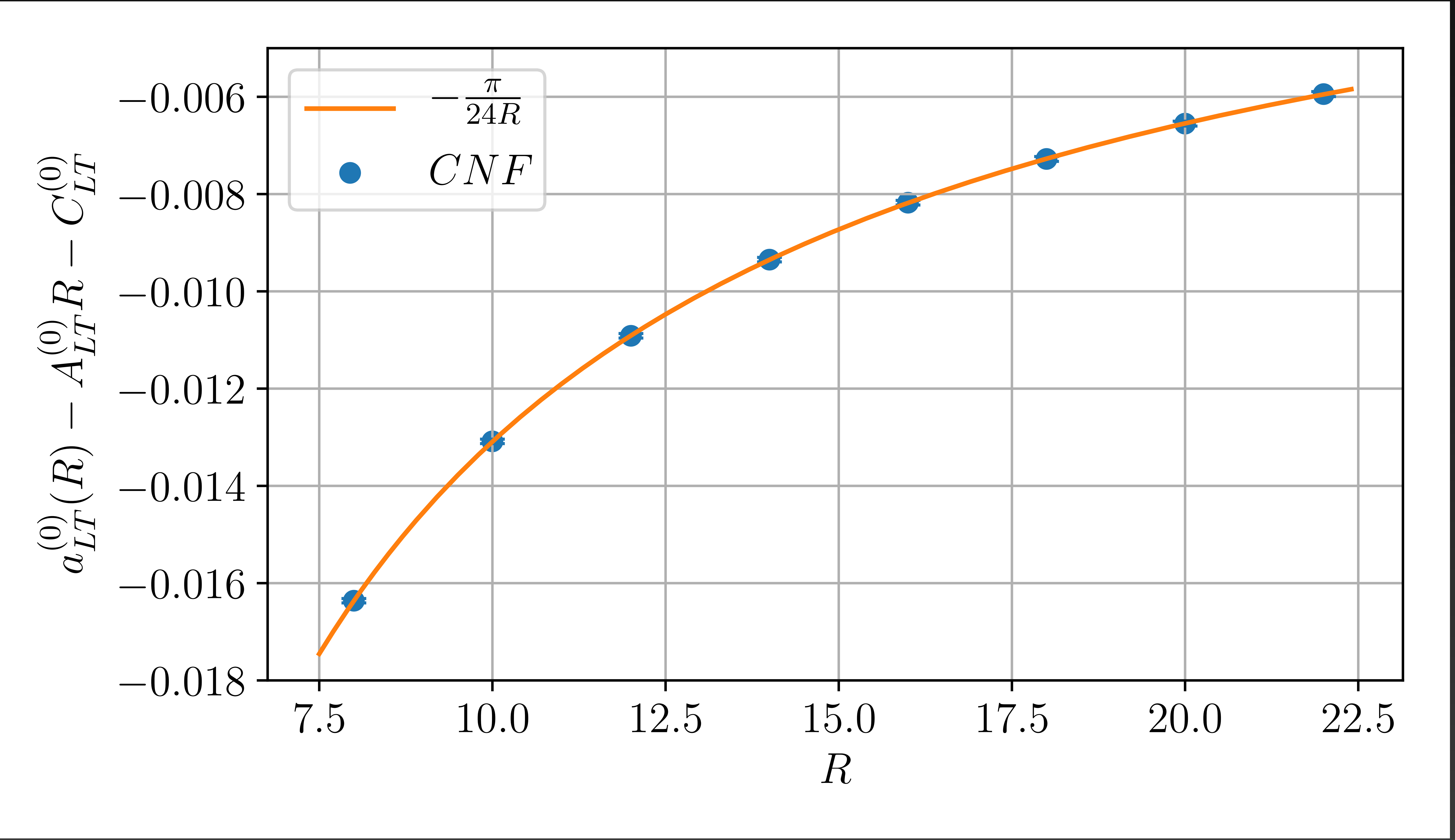


Nambu-Goto: CNF

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: $-0.1308996\ldots$



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

NG: Physics-Informed SNF

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

$$S_{NG}(\phi) \sim S_{FB}(\phi) + \dots \qquad S_{FB}(\phi) = \frac{1}{2}\sum_{x} (\partial_\mu \phi(x))^2$$

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.

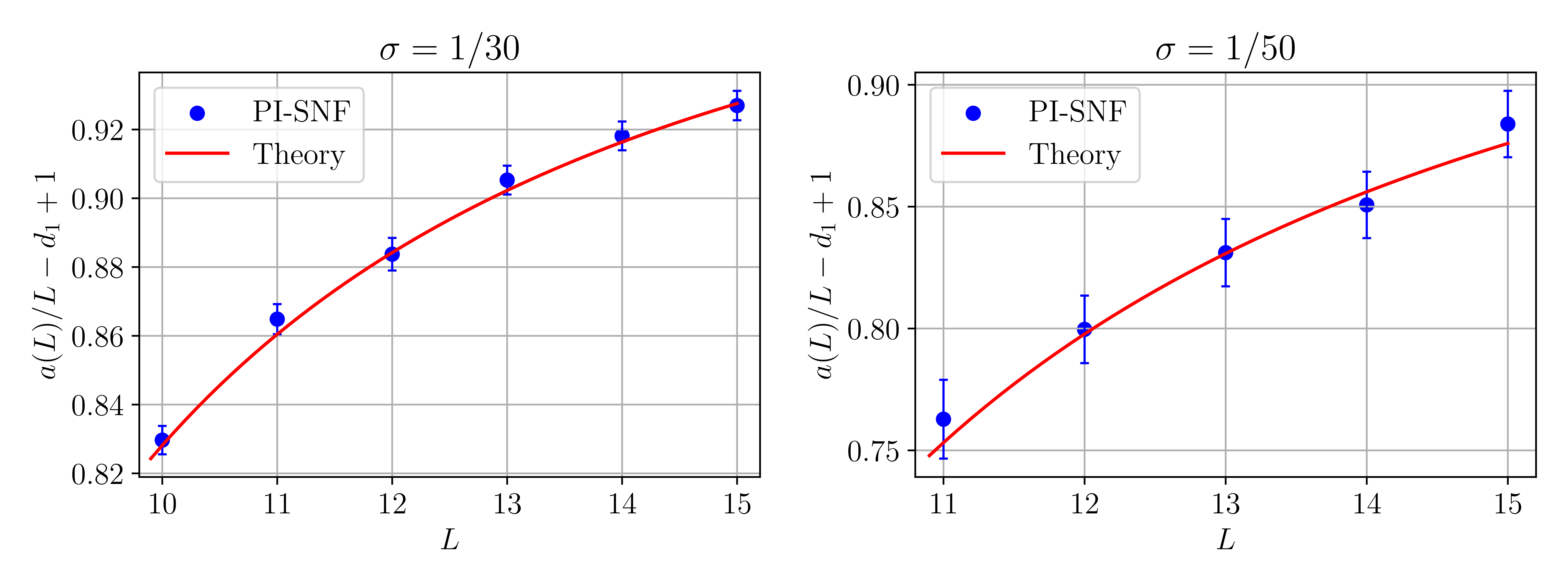
  [Cavaglià+; 1608.05534],[Smirnov and Zamolodchikov; 1608.05499],[Caselle, E.C., Nada; 2309.14983]

NG: Free Energy

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

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

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



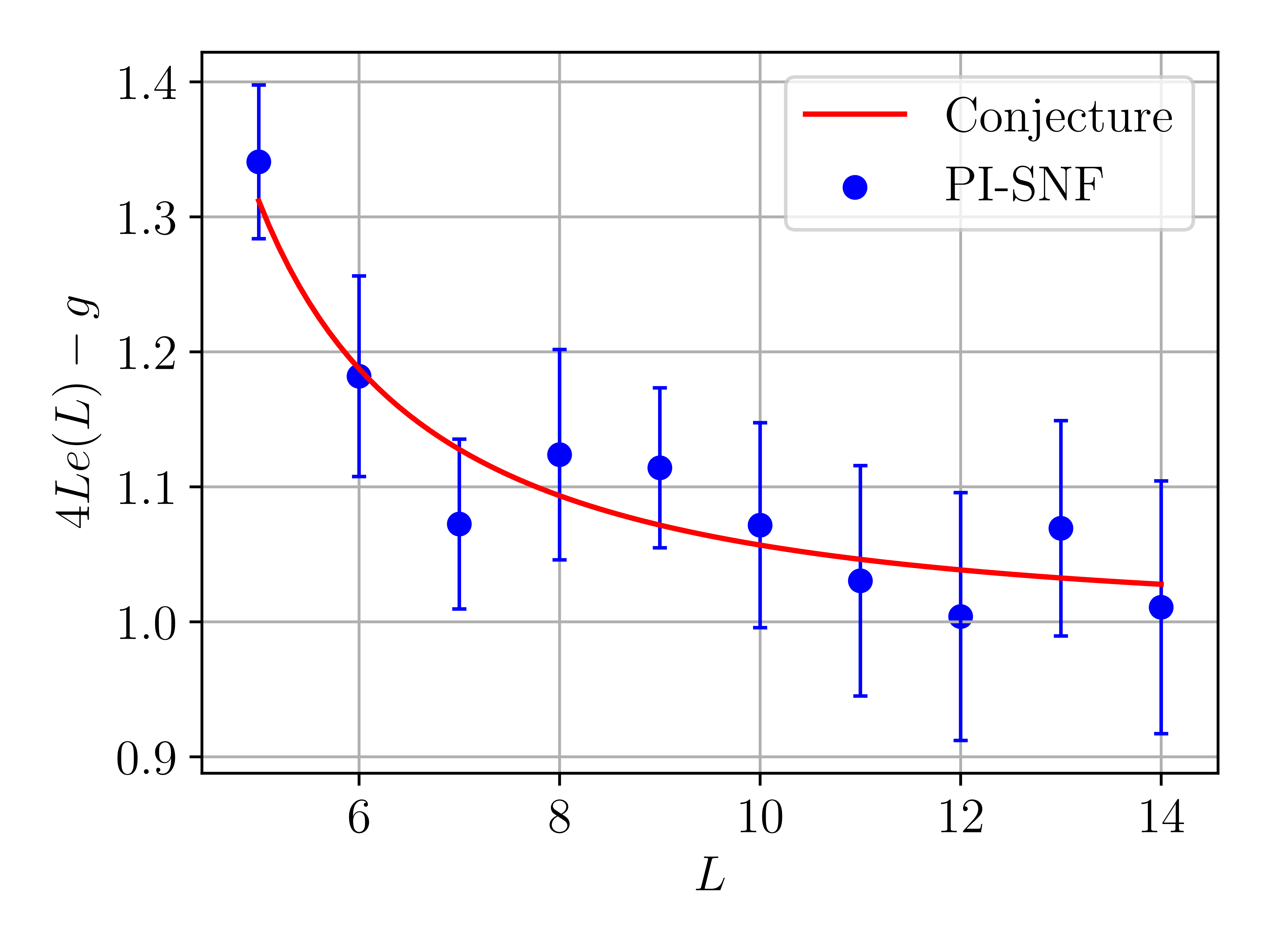
NG Width

Width of the string → measure the **density** of the **chromoelectric flux tube**

$$\sigma = 1/10,$$

$$\sigma w^2 = \frac{1}{\sqrt{1-\frac{\pi}{3\sigma L^2}}}\frac{R}{4L} + \cdots$$

Fitted: $-1.09(8)$, target: $-1.047\ldots$

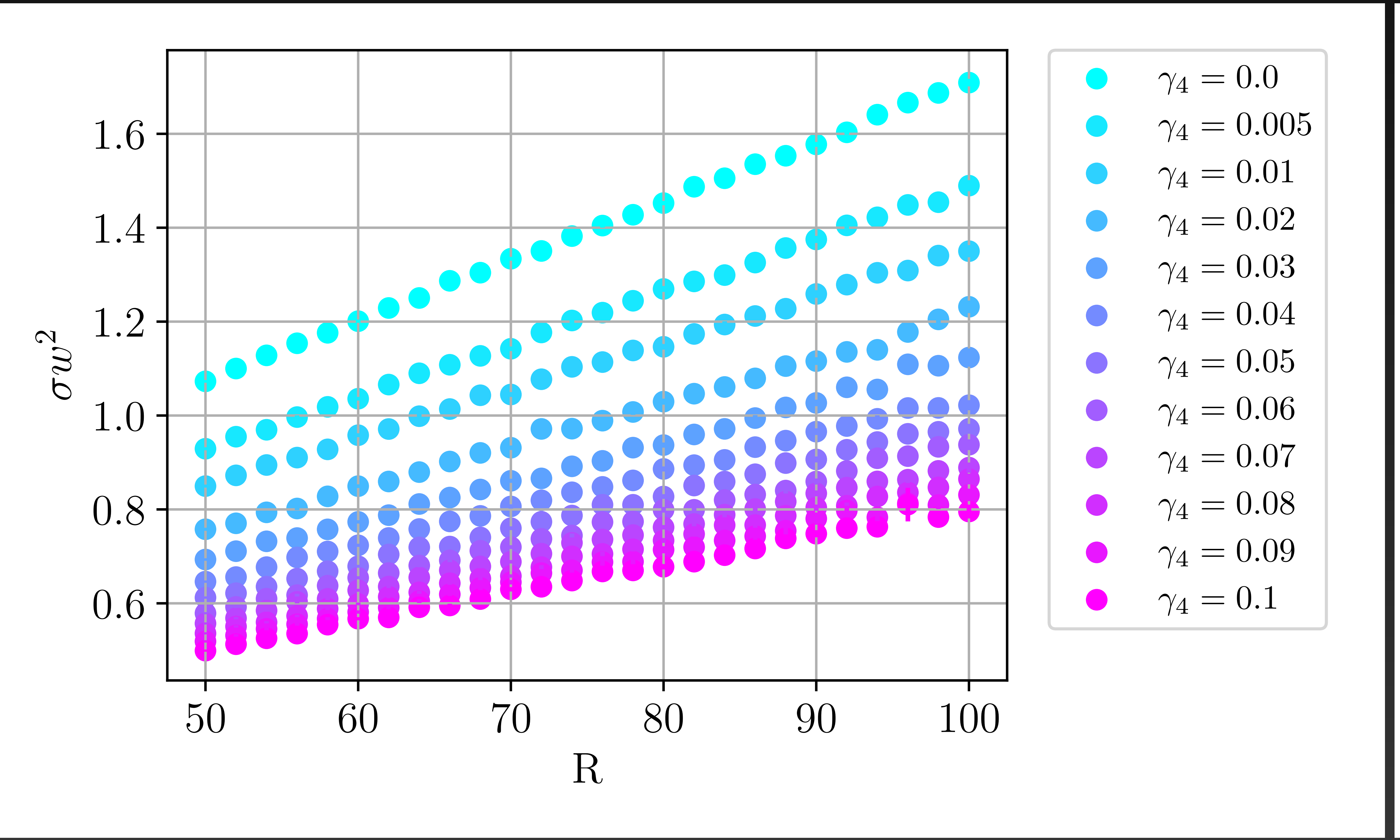


[Caselle, E.C., Nada; 2309.14983][Caselle;1004.3875]

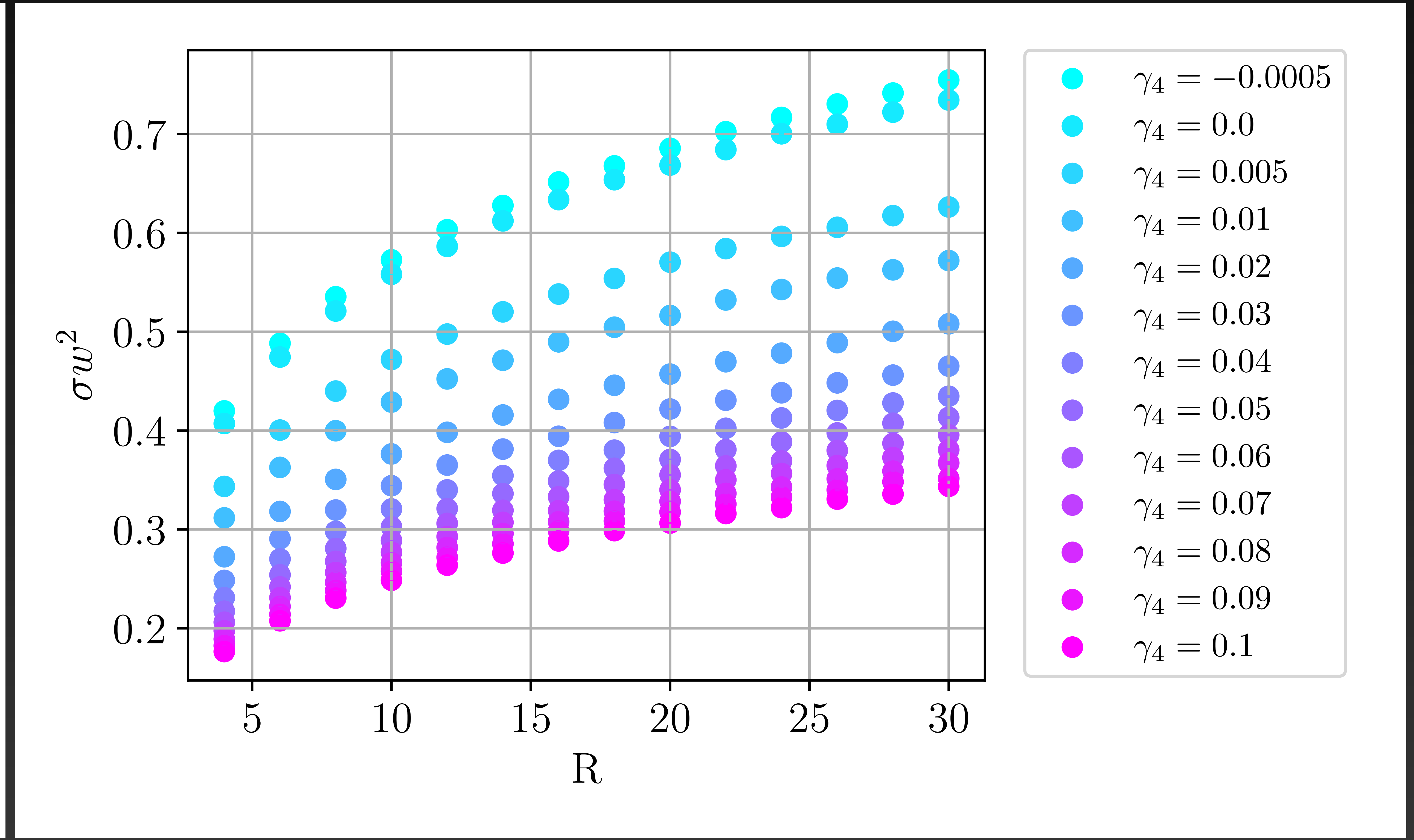
Beyond NG: width

$$S_{BNG} = S_{NG} + \gamma_4 \mathcal{K}^4$$

$R \gg L = 20$



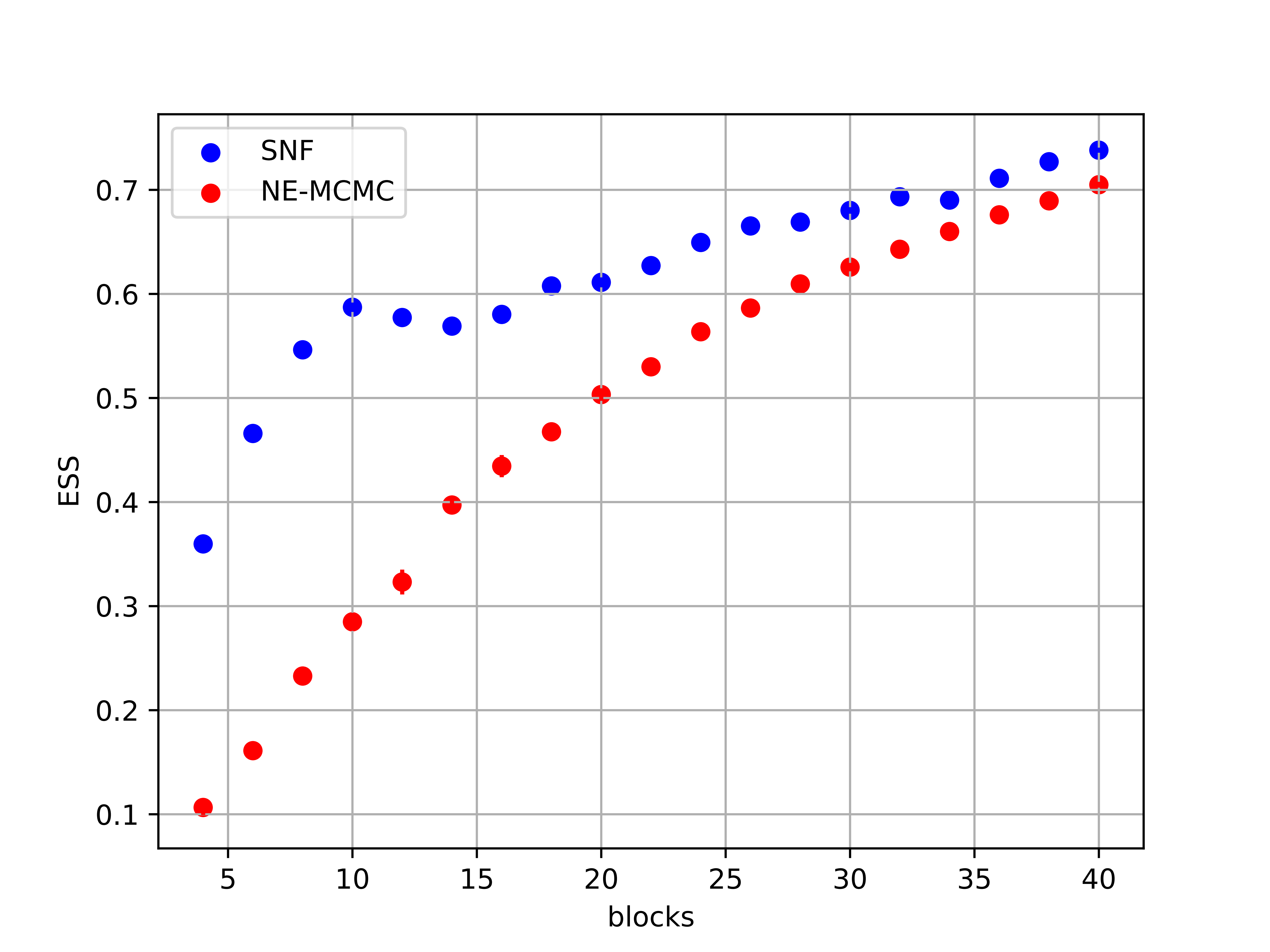
$L = 80 \gg R$



[Caselle, E.C., Nada; in prep.]

SU(3) in $d = 3 + 1$

SNF: Neural Smearing + Heatbath; lattice $V = 4^4$; prior: thermalized MCMC with $\beta_0 = 6.0$, target: $\beta = 6.2$



[Morningstar and Peardon; hep-lat/0311018],[Nagai and Tomika; 2103.11965],[Abbott+; 2211.07541],[Bulgarelli, E.C., Nada; in prep.]

Conclusions

1. We showed that SNFs can relay on a strong theoretical background

2. ϕ^4 : Promising volume scaling

3. EST: State-of-the-art results for Effective String Theory $\rightarrow$ leading numerical method

4. $\mathrm{SU}(3)$: Proof-of-principle in $d = 3 + 1$ $\rightarrow$ toward physical lattice volume and fight topological freezing

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.

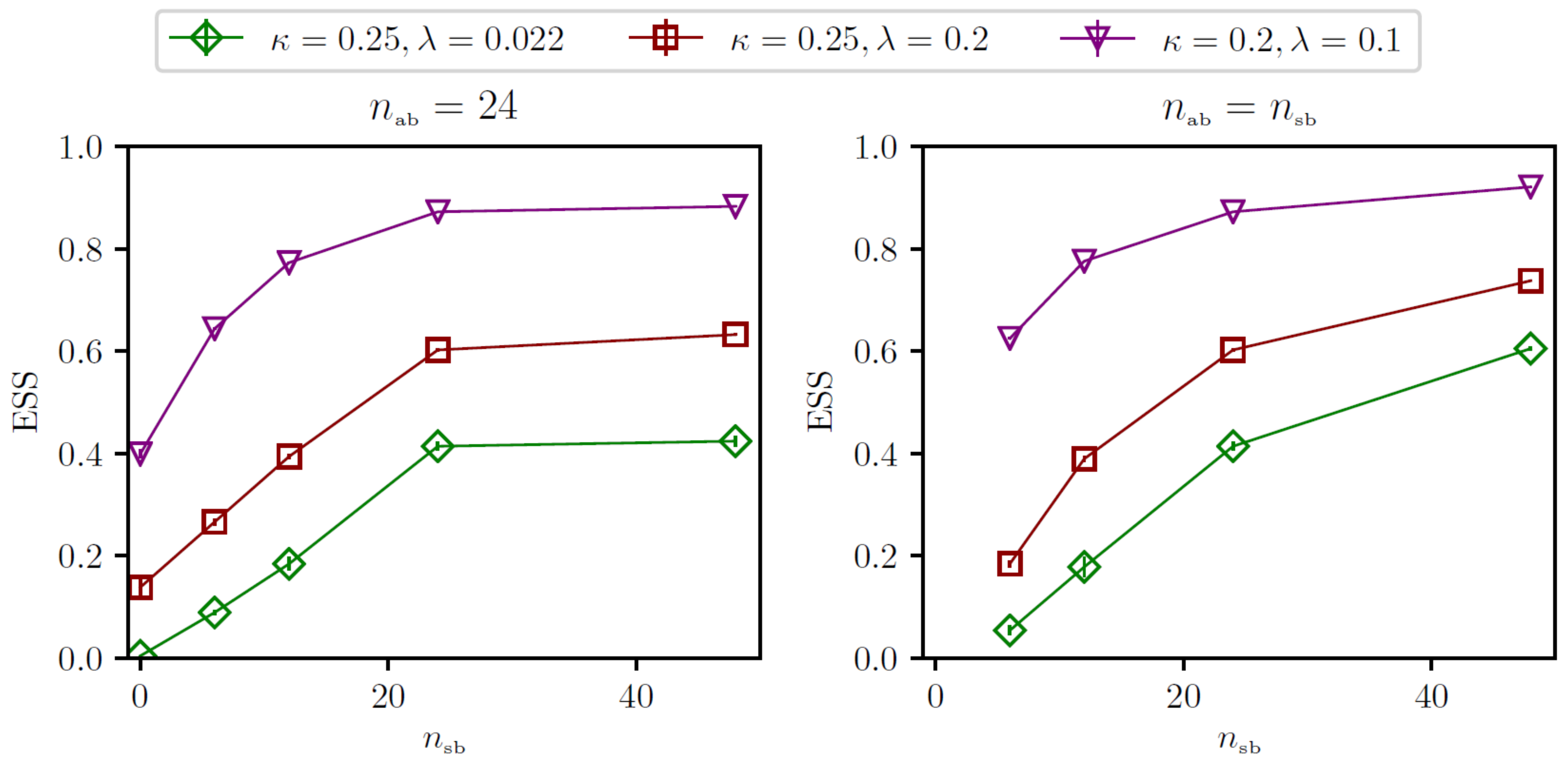
  [Midgley+; 2208.01893]

Technical details: ϕ^4

- Prior: Normal distribution ($\kappa = \lambda = 0$)
- Linear protocol
- Heatbath algorithm for stochastic updates
- Equivariant affine coupling layers, one convolutional layer with 3×3 kernel and two channels

ϕ^4: SNFs Couplings

Test with different action parameters on a $N_t \times N_s = 16 \times 8$ lattice



Technical details: NG

- Prior massless free boson and linear protocol in $t = 1/\sigma \rightarrow$ Inspired by Irrelevant Perturbations
- HMC for stochastic updates
- Affine coupling layers, 3 convolutional layers with $3 \times 3 \times 16$ kernels and a two channels output layer. Each blocks (even-odd) share the same network