Normalizing Flows for Effective String Theory

Elia Cellini

Università degli Studi di Torino/Istituto Nazionale Fisica Nucleare

June 26th 2023

Machine learning for lattice field theory and beyond
Trento ECT*

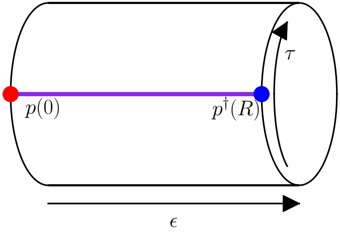
arXiv:2307.?????
M. Caselle, E.C. and A. Nada

Outline

Effective String Theory

Effective string theory (EST) is a non-perturbative framework that provide an effective description of the confining flux tube in term of vibrating string. In particular, the correlator between two Polyakov loops is related to the full partition function of an EST.

$$\langle P(0)P^{\dagger}(R)\rangle = \int D\phi e^{-S_{eff}} \equiv Z(L, R, \sigma).$$



The EST is anomalous at the quantum level and thus must be considered only as an effective, large-distance, description of Yang-Mills theories **[Aharony and Komargodski; 2013]**.

Nambu-Goto String

The most natural choice for S_{eff} is the Nambu-Goto (NG) string **[Nambu; 1974]**,**[Goto; 1971]**.
The main observables we want to compute are:

- The free energy $-logZ$: directly associated with the interquark potential.
- The "width" σw^2: measures the density of chromoelectric flux tube.

Several analytical studies of the free energy together with Monte Carlo simulations of the interquark potential have proven that the NG theory is universal up to terms of order R^{-5}. Nowadays, the community has focused the research of theories "beyond" the Nambu-Goto string.

For recent reviews see **[Aharony and Komargodski; 1302.6257][Brandt and Meineri; 1603.06969][Caselle; 2104.10486]**.

Lack of numerical methods

However, few analytical studies have been provided for the NG width and the theories beyond the NG string. Moreover, it still lacks an efficient numerical method that can be used to study EST where analytical studies are not possible.

Problems:

- Non-linearity of the actions.
- Direct estimation of the partition functions.

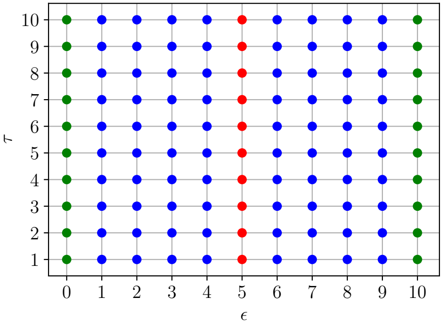
$\rightarrow$ Our proposal: Machine Learning (Normalizing Flows) + Lattice regularization of EST.

Nambu-Goto String on the Lattice

In the $d = 2 + 1$ case, using a "physical gauge" the NG action can be regularized on the lattice as:

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

where Λ is a square lattice of size $L \times R$ with step $a = 1$, $\phi(x) = \phi(\tau, \epsilon) \in \mathbb{R}$ and boundary conditions: $\phi(\tau + L, \epsilon) = \phi(\tau, \epsilon)$ and $\phi(\tau, 0) = \phi(\tau, R) = 0$.



The width of the string can be computed as: $\sigma w^2 = \langle \phi^2(\tau, R/2) \rangle_{\tau,\phi}$

Free Boson expansion

In the limit $\sigma \to \infty$ the Nambu-Goto action can be expanded in series:

$$S_{NG} \sim S_{FB} + O(\sigma^{-1})$$

where:

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

The finite size analytical solutions of $-\log Z_{FB}$ can be found using a gaussian integration:

$$-\log Z_{FB} = A_{FB} RL + C_{FB} L + \log \eta(\xi)$$

where:

$$A_{FB} = -0.3358177... \quad C_{FB} = 0.478252....$$

and:

$$\eta(\xi) = q^{\frac{1}{24}} \prod_{n=1}^{\infty} (1 - q^n) \;\; ; \;\; q = e^{2\pi i \xi} \;\; ; \;\; \xi = i \frac{L}{2R}$$

Width of the flux tube

The analytical solution for σw^2 is well known only up to the order σ^{-1} **[Lüscher et al.; 1981][Caselle and Allais; 2009][Gliozzi, Pepe and Wiese; 2010]**:

- $L \gg R$: $\sigma w^2 = \frac{1}{2\pi} \log \frac{R}{Rc} \left(1 - \frac{\pi}{4\sigma R^2}\right) + \frac{5}{96} \frac{1}{\sigma R^2} + \dots$
- $R \gg L$: $\sigma w^2 = \frac{1}{2\pi} \log \frac{L}{Lc} + \frac{R}{4L} + \frac{\pi}{24} \frac{R}{\sigma L^3} + \dots$

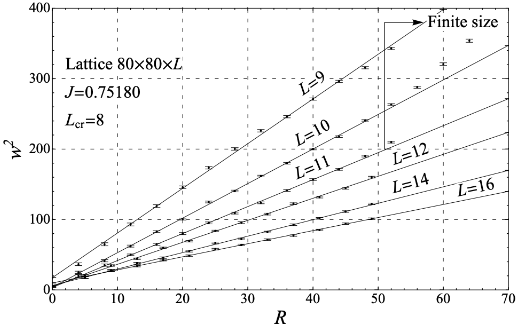
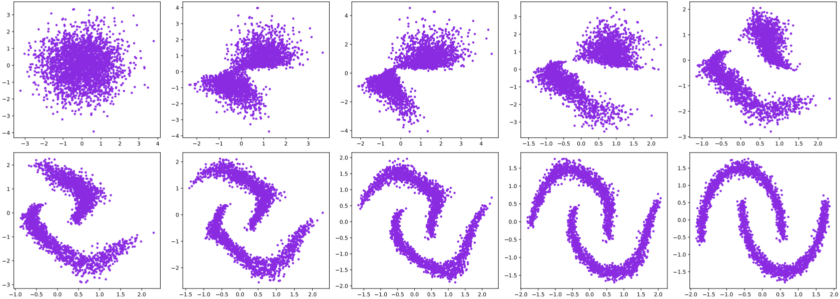


Figure 2: *Flux tube thickness as a function of the interquark distance for various values of the inverse temperature L.*

Normalizing Flows



Normalizing Flows

Normalizing flows (NFs) **[Rezende and Mohamed; 2015]** are a class of deep learning algorithm recently proposed to sample from Boltzmann distributions.

A NF g_θ is a parametric, invertible and differentiable function that maps an easy-to-model prior distribution $q_0(z)$, $z \in \mathbb{R}^n$, to an inferred distribution q_θ which approximate the target $p(\phi)$, $\phi \in \mathbb{R}^n$.

$$g_\theta : \; q_0 \rightarrow q_\theta \simeq p$$

$$\phi = g_\theta(z)$$

$$q_\theta(\phi) = q_0(g^{-1}(\phi))|J_g|^{-1}$$

Targetting Boltzmann distributions

NFs can be trained to $q_\theta \simeq p(\phi)$ with $p(\phi) = \frac{1}{Z}\exp(-S[\phi])$ **[Albergo et al.; 2019]**,**[Noé et al.; 2019]** by minimizing the reverse Kullback-Leibler divergence:

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

Observables can be computed using a re-weighting procedure also called Importance Sampling (IS) in machine learning field **[Nicoli et al.; 2020]**:

$$\langle \mathcal{O} \rangle_{\phi \sim p} = \int D\phi p(\phi) \mathcal{O}(\phi) = \int D\phi q_\theta(\phi) \frac{p(\phi)}{q_\theta(\phi)} \mathcal{O}(\phi) \simeq \frac{1}{\hat{Z}} \langle \mathcal{O} \tilde{w} \rangle_{\phi \sim q_\theta}$$

where

$$\tilde{w} = \frac{e^{-S[\phi]}}{q_\theta(\phi)}$$

and

$$\hat{Z} = \langle \tilde{w} \rangle_{\phi \sim q_\theta}$$

State-of-the-art

Main focus in the lattice community: improve on MCMC simulations by treating critical slowing down

$\rightarrow$ recent reviews **[Abbott et al.; 2211.07541][Zhou et al.; 2303.15136]** on the current status of NFs for LFTs

$\rightarrow$ successfully applied in LFTs in 2d: ϕ^4 scalar field theory **[Albergo et al.; 2019]**, **[Kanwar et al.; 2020]**, **[Nicoli et al.; 2020]**, **[Del Debbio et al.; 2021]**, $\mathrm{SU}(N)$ **[Boyda et al.; 2020]**, fermionic theories **[Albergo et al.; 2021]**, $\mathrm{U}(1)$ and $\mathrm{SU}(N)$ with fermions **[Abbott et al.; 2022]**, Schwinger model **[Finkenrath et al.; 2022]**, **[Albergo et al.; 2022]** . . .

$\rightarrow$ strongly related to the idea of trivializing maps **[Lüscher; 2009]**, **[Bacchio et al.; 2022]**, **[Albandea et al.; 2023]**

$\rightarrow$ can be connected to non-equilibrium thermodynamics: Stochastic NFs **[Caselle et al.; 2022]**

$\rightarrow$ new architectures such as Continuous Normalizing Flows **[Gerdes et al.; 2022]**, **<u>used in this work!</u>**

Still several open issues (volume dependence, scalability, mode-collapse ...)

Continuous normalizing flows

Exploiting Neural Ordinary Differential Equations (NODE) **[Chen et al.; 2018]** is possible to build Continuous NFs (CNFs) in which g_θ is the solution of an ODE parameterized by a neural network V_θ:

$$\frac{d\phi(t)}{dt} = V_\theta(\phi(t), t)$$

with

$$\phi(t = 0) = z \sim \mathcal{N}(0, \mathbb{1}/2) \;\; and \; \phi(t = T) = \phi$$

Thus:

$$\phi(T) = ODESOLVER(V_\theta, \phi(0), [0, T])$$

The density of the generated samples can be computed through the ODE:

$$\frac{d \log q_\theta(\phi(t))}{dt} = -(\nabla \cdot V_\theta)(\phi(t), t)$$

CNF: architectures

The architecture used is a continuous in time linear model inspired by **[Gerdes et al.; 2022]**:

$$V_\theta(\phi(t), t) = \sum_{y,d} W_{x,y,d} K(t)_d \phi(t)_y$$

$$(\nabla \cdot V_\theta)(\phi(t), t) = \mathrm{Tr}\left[\sum_d W_d K(t)_d\right]$$

Where $K(t) \in \mathbb{R}^D$ is a temporal kernel of D Fourier coefficients, $W \in \mathbb{R}^{A\times A\times D}$ is a linear neuron with $A = L \times (R - 1)$.

Numerical results

Goals of the numerical studies

The goal of our studies is to provide a proof of concept of the feasibility of the application of NFs as sampler for EST.
To validate our results, we checked:

- Test of the performance of the algorithms using the Effective Sample Size:

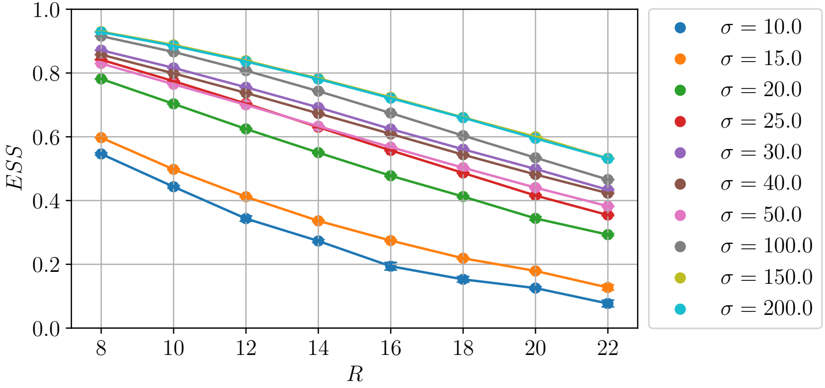
$$ESS = \frac{\langle \tilde{w} \rangle^2}{\langle \tilde{w}^2 \rangle}$$

  We accepted all the models with $ESS > 0.1$

- Benchmark of the Nambu-Goto numerical free energy using $-\log Z_{FB}$.
- Study of the width σw^2.
- Direct comparison with Hybrid Monte Carlo methods (backup slides).

We run the models on NVIDIA Tesla V100 GPU of the Marconi100 (CINECA).

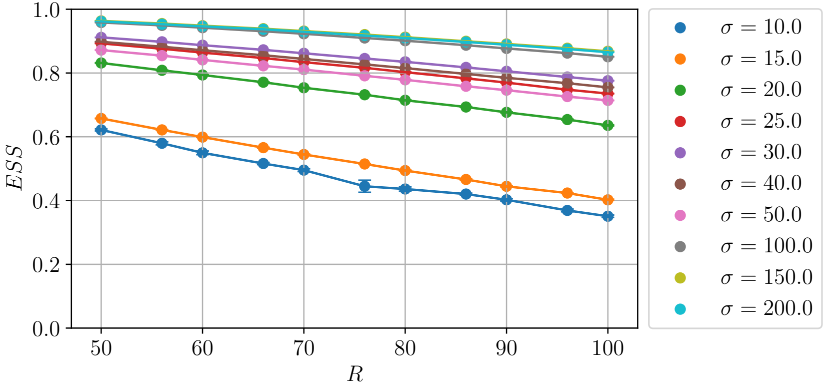
low-temperature, $R \ll L = 90$



Decrease in performance towards smaller $\sigma \rightarrow$ better architecture are being tested

Scaling

high-temperature, $R \gg L = 10$



Decrease in performance towards smaller $\sigma \rightarrow$ better architecture are being tested

Partition function at low-temperature

global fit in (σ, L)

$$-\log Z = \left(a^{(0)}_{LT}(R) + \frac{a^{(1)}_{LT}(R)}{\sigma} + \frac{a^{(2)}_{LT}(R)}{\sigma^2} + \frac{a^{(3)}_{LT}(R)}{\sigma^3} \right) L$$

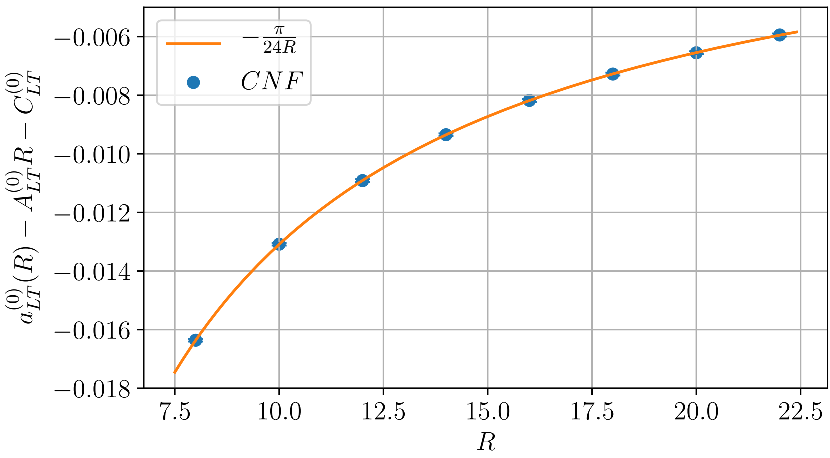
and then

$$a^{(0)}_{LT}(R) = A^{(0)}_{LT} R + \frac{B^{(0)}_{LT}}{R} + C^{(0)}_{LT}$$

and we found:

	$A^{(0)}_{LT}$	$B^{(0)}_{LT}$	$C^{(0)}_{LT}$	$\chi^2/d.o.f.$
CNFs	$-0.335820(2)$	$-0.1309(2)$	$0.47822(4)$	0.93
Theory	$-0.3358177...$	$-0,13089969...$	$0.478252...$	

Plot of the Dedekind prediction $-\frac{\pi}{24R}$ (Lüscher term) compared to the numerical simulations:



global fit in (σ, R)

$$-\log Z - \frac{1}{2}\log\left(\frac{2R}{L}\right) =$$

$$= \left(a_{HT}^{(0)}(L) + \frac{a_{HT}^{(1)}(L)}{\sigma} + \frac{a_{HT}^{(2)}(L)}{\sigma^2} + \frac{a_{HT}^{(3)}(L)}{\sigma^3}\right) R + c_{HT}^{(0)}(L) + \frac{c_{HT}^{(1)}(L)}{\sigma}$$

then fit in L

$$a_{HT}^{(0)} = A_{HT}^{(0)} L + \frac{B_{HT}^{(0)}}{L} + \frac{B1_{HT}^{(0)}}{L^3} + \frac{B2_{HT}^{(0)}}{L^5}$$

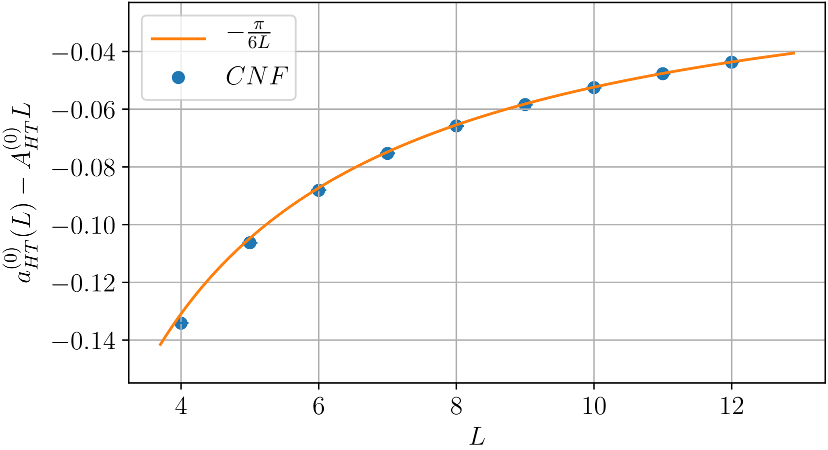
and

$$c_{HT}^{(0)}(L) = C_{HT}^{(0)} L + D_{HT}^{(0)}$$

and we found:

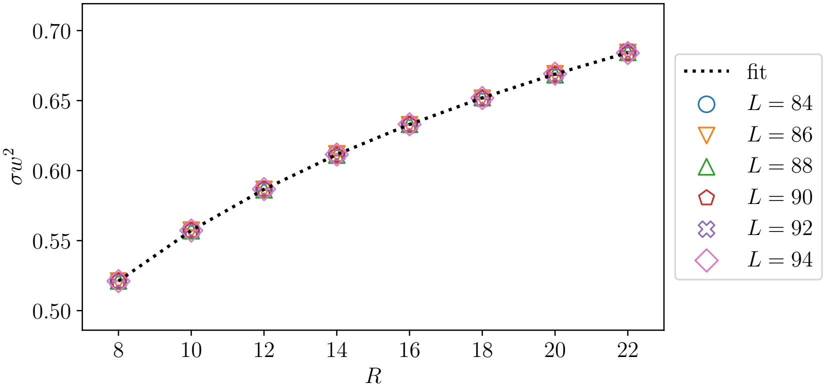
	$A_{HT}^{(0)}$	$B_{HT}^{(0)}$	$C_{HT}^{(0)}$	$\chi^2/d.o.f.$
CNFs	$-0.335823(2)$	$-0.5234(2)$	$0.47827(3)$	$1.89, 1.50$
Theory	$-0.3358177\ldots$	$-0.523598\ldots$	$0.478252\ldots$	

Plot of the Dedekind prediction $-\frac{\pi}{6L}$ compared to the numerical simulations:



global fit in (σ, L, R) - plot for $\sigma = 100.0$

$$\sigma w^2 = \left(1 + \frac{e_{LT}^{(0)}}{\sigma} + \underbrace{e_{LT}^{(1)}}_{-\pi/4}\frac{1}{\sigma R^2}\right)\left(\underbrace{f_{LT}}_{1/2\pi}\log(R) + g_{LT}\right) + \frac{h_{LT}^{(0)}}{R^2} + \underbrace{h_{LT}^{(1)}}_{5/96}\frac{1}{\sigma R^2}$$

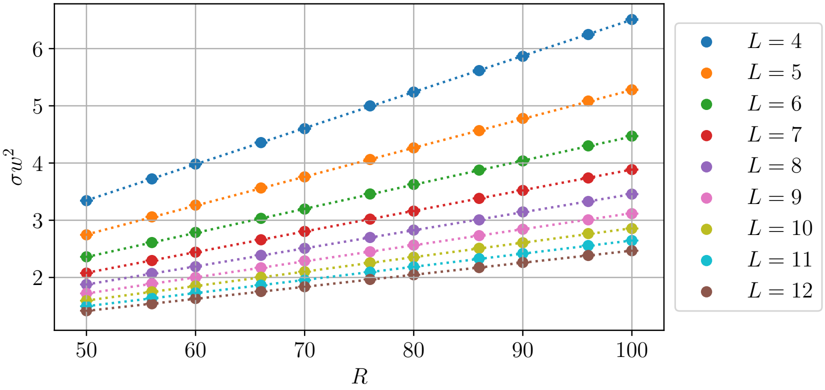


Width at high-temperature

global fit in (σ, L, R)

$$\sigma w^2(\sigma, L, R) = \left(1 + \frac{i_{HT}^{(0)}}{\sigma} + \underbrace{i_{HT}^{(1)}}_{\pi/6} \frac{1}{\sigma L^2}\right)\left(\underbrace{j_{HT}}_{1/4} \frac{R}{L} + k_{HT} + \underbrace{l_{HT}}_{1/2\pi} \log(L)\right)$$

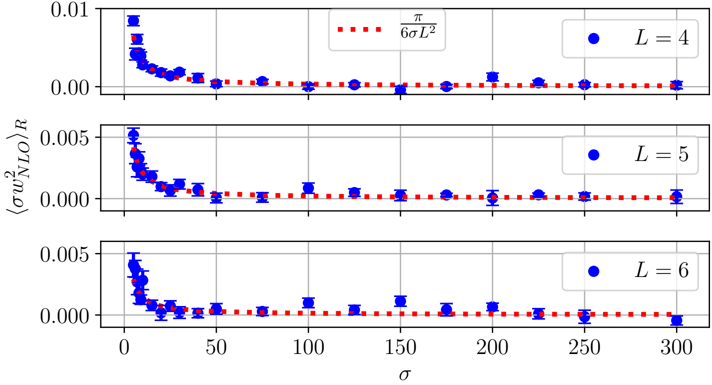
plot for $\sigma = 100.0$



Plot of:

$$\langle \sigma w^2_{2nd} \rangle_R(\sigma, L) = \langle \frac{\sigma w^2(\sigma, L, R)}{\left(j_{HT}\frac{R}{L} + k_{HT} + l_{HT}\log(L)\right)} - 1 - \frac{i^{(0)}_{HT}}{\sigma} \rangle_R,$$

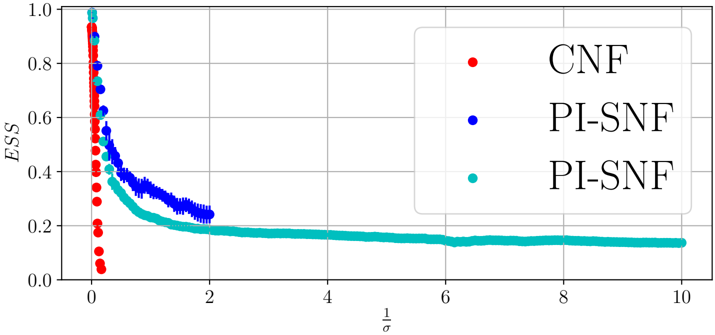
compared to the analytical solution $\frac{\pi}{6\sigma L^2}$. **First numerical observation of this term!**



Outlook

Outlook

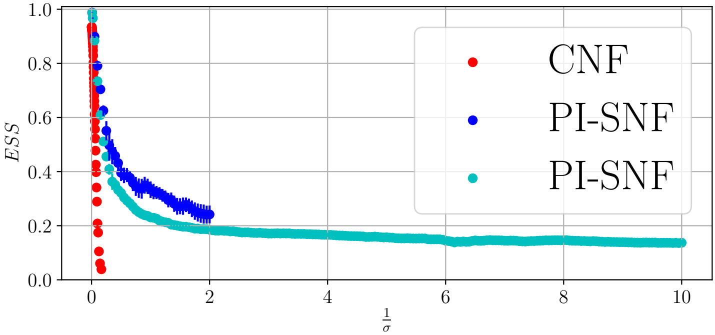
- We showed that CNFs are able to sample efficiently from the probability distribution of the Nambu-Goto EST.
- More work is put into better architectures → "Physics-Informed" Stochastic NFs



- Toward the study of the theories beyond the Nambu-Goto string using flows-based sampler as leading numerical method!

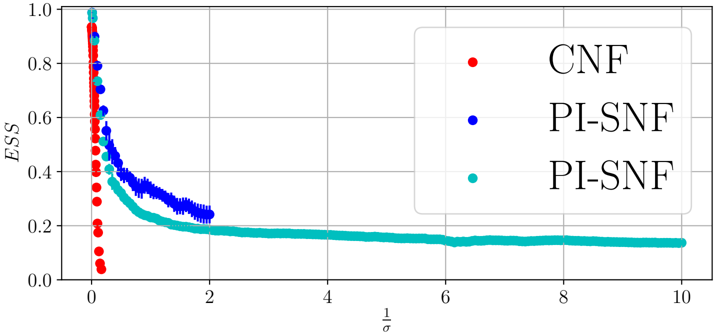
Outlook

- We showed that CNFs are able to sample efficiently from the probability distribution of the Nambu-Goto EST.
- More work is put into better architectures → "Physics-Informed" Stochastic NFs



- Toward the study of the theories beyond the Nambu-Goto string using flows-based sampler as leading numerical method!

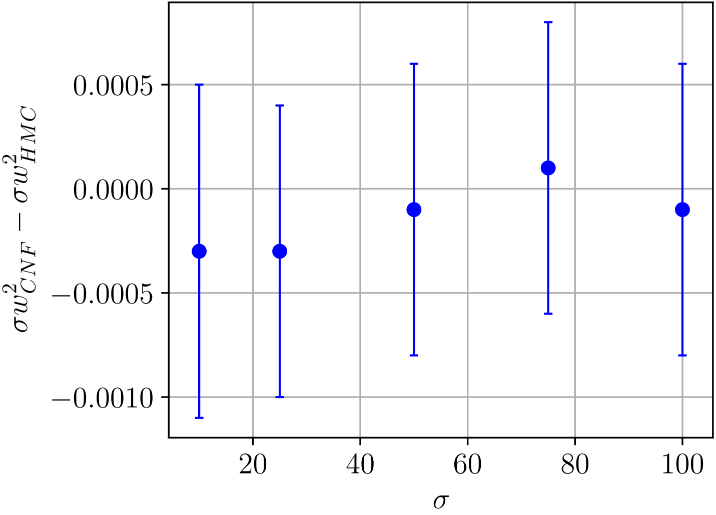
- We showed that CNFs are able to sample efficiently from the probability distribution of the Nambu-Goto EST.
- More work is put into better architectures → "Physics-Informed" Stochastic NFs



- Toward the study of the theories beyond the Nambu-Goto string using flows-based sampler as leading numerical method!

Thank you for your attention!

Comparison with HMC: Bias



Comparison with HMC: Sampling time