Flow-based Sampling for Effective String Theory

Turin Lattice Meeting 2023

Elia Cellini

22/12/2023

Based on:

M.Caselle, **EC**, A.Nada, M. Panero

- JHEP 07 (2022) 015, arxiv:2201.08862

M. Caselle, **EC**, A. Nada

- arXiv: 2307.01107 (under review JHEP)
- arxiv: 2309.14983 (Lattice 2023 PoS)

UNIVERSITÀ DI TORINO

INFN

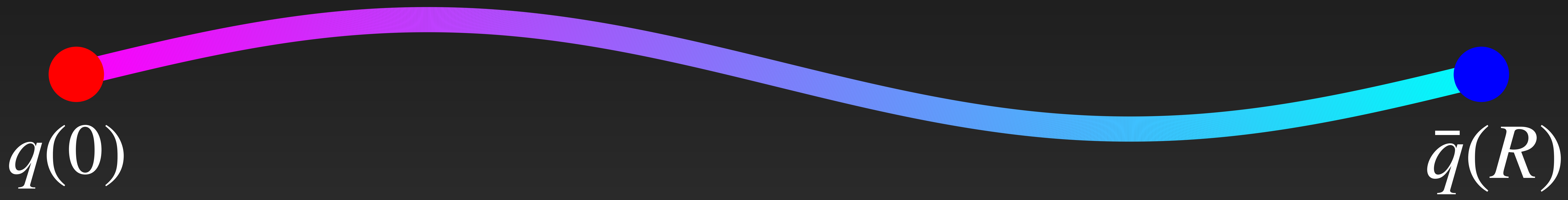
Outline

1. Effective String Theory
2. Normalizing Flows
3. Stochastic Normalizing Flows
4. Conclusion

Effective String Theory

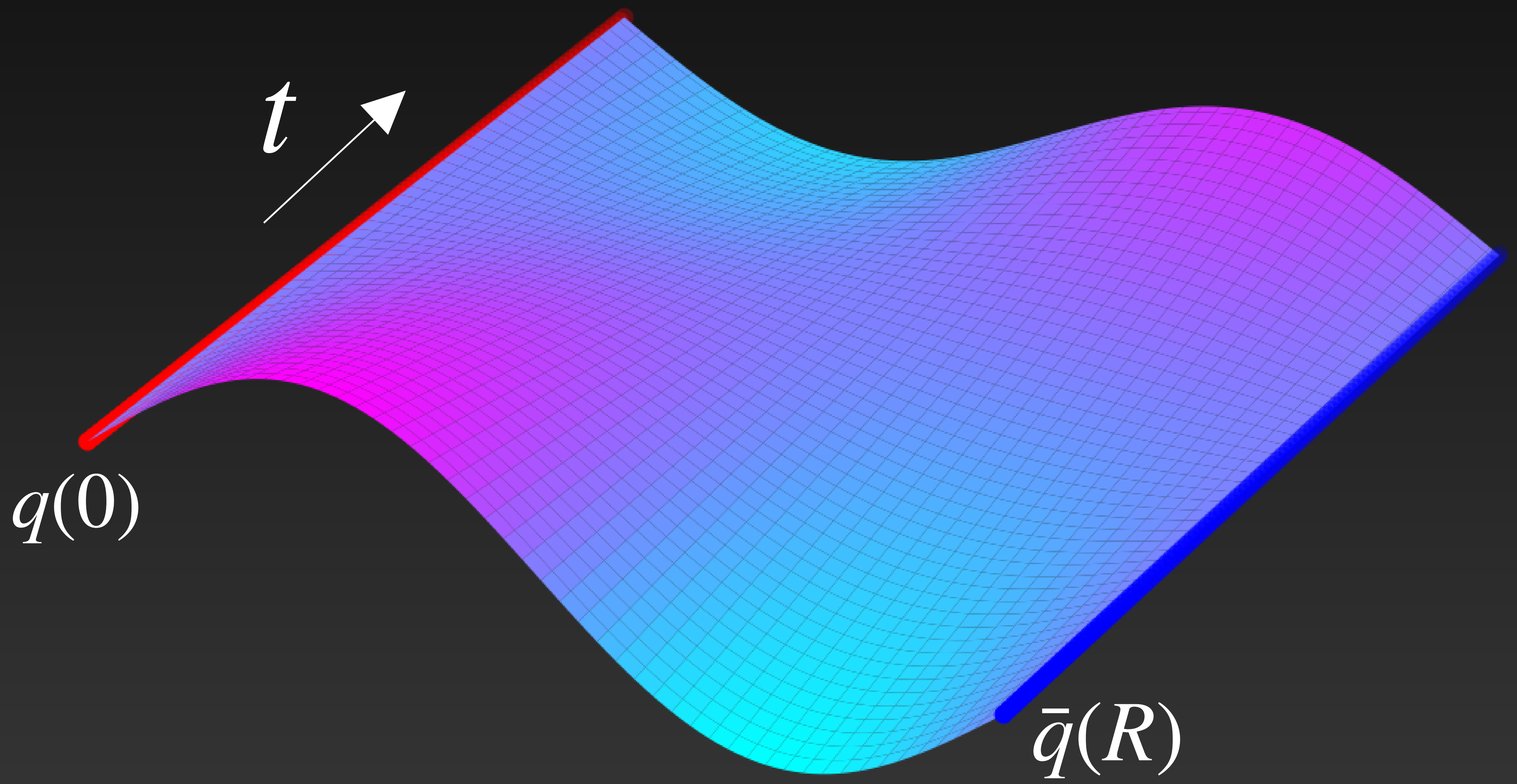
Effective String Theory

Effective String Theory (**EST**) is a powerful tool to study **confinement** in **Yang-Mills** theories



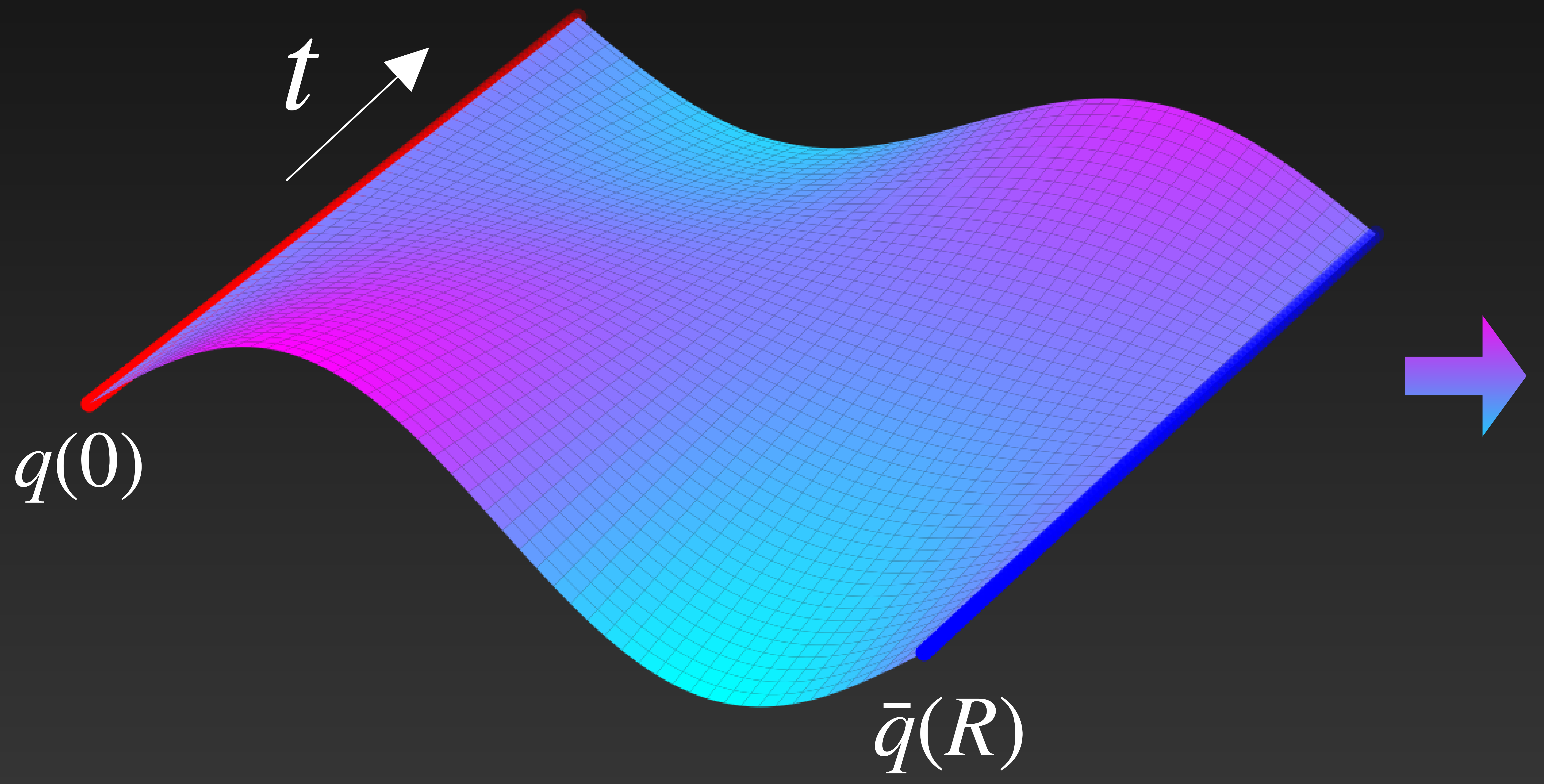
Effective String Theory

Effective String Theory (**EST**) is a powerful tool to study **confinement** in **Yang-Mills** theories



Effective String Theory

Effective String Theory (**EST**) is a powerful tool to study **confinement** in **Yang-Mills** theories



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

Nambu-Goto Theory

The main choice for S_{EST} is the Nambu-Goto action:

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

- **Anomalous** at quantum level $\rightarrow$ **effective**, **large-distance** **description** of Yang-Mills theories.
  [Aharony and Komargodski; 1302.6257], [Brandt and Meineri; 1603.06969],[Caselle; 2104.10486]

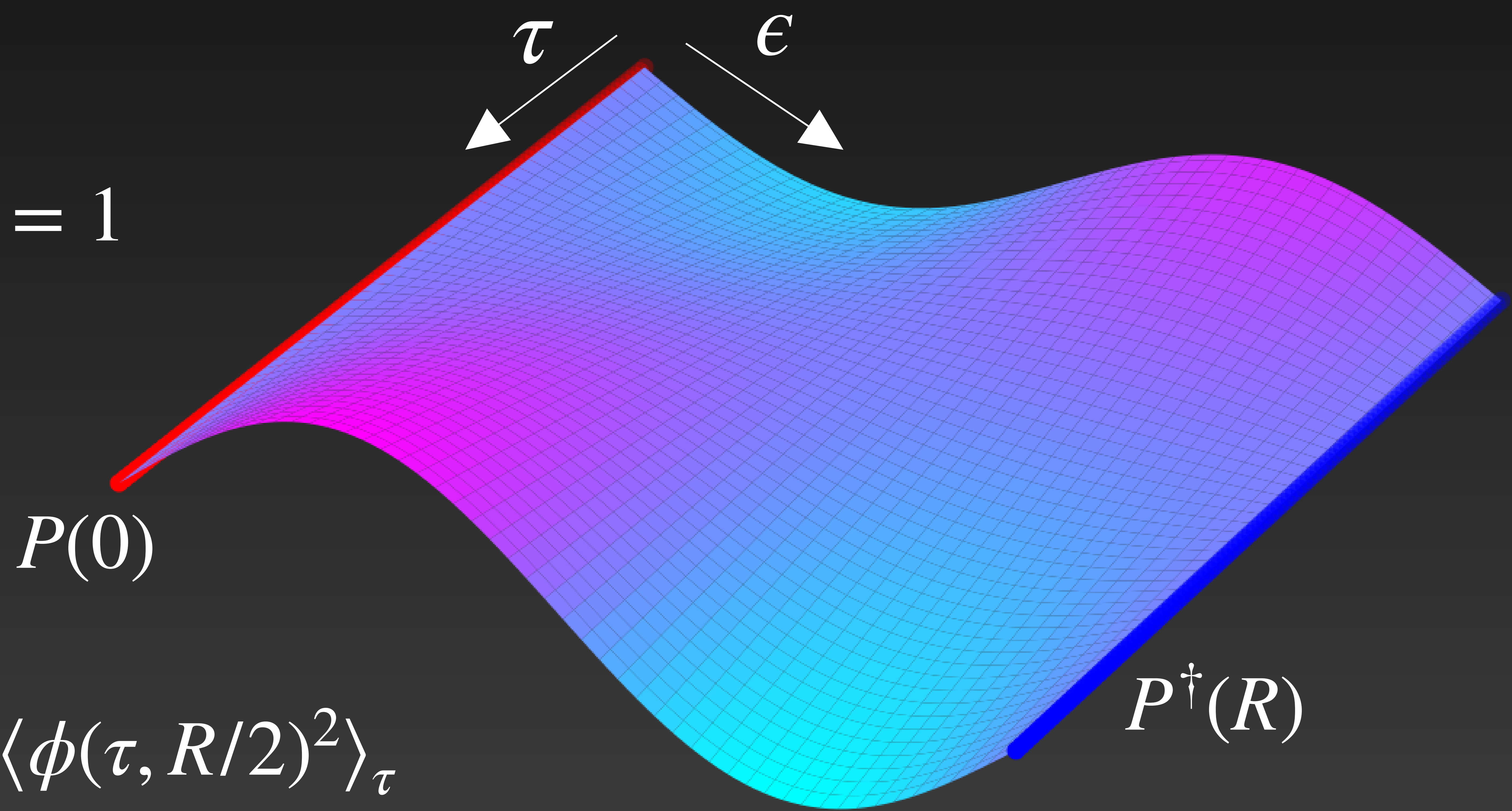
Analytical problems

- **Correlation functions** (e.g. width σw^2)
- **Partition functions** of higher order corrections

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 $L \times R$, $a = 1$
- $\phi(x) = \phi(\tau, \epsilon) \in \mathbb{R}$
- $\phi(\tau + L, \epsilon) = \phi(\tau, \epsilon)$
- $\phi(\tau, 0) = \phi(\tau, R) = 0$



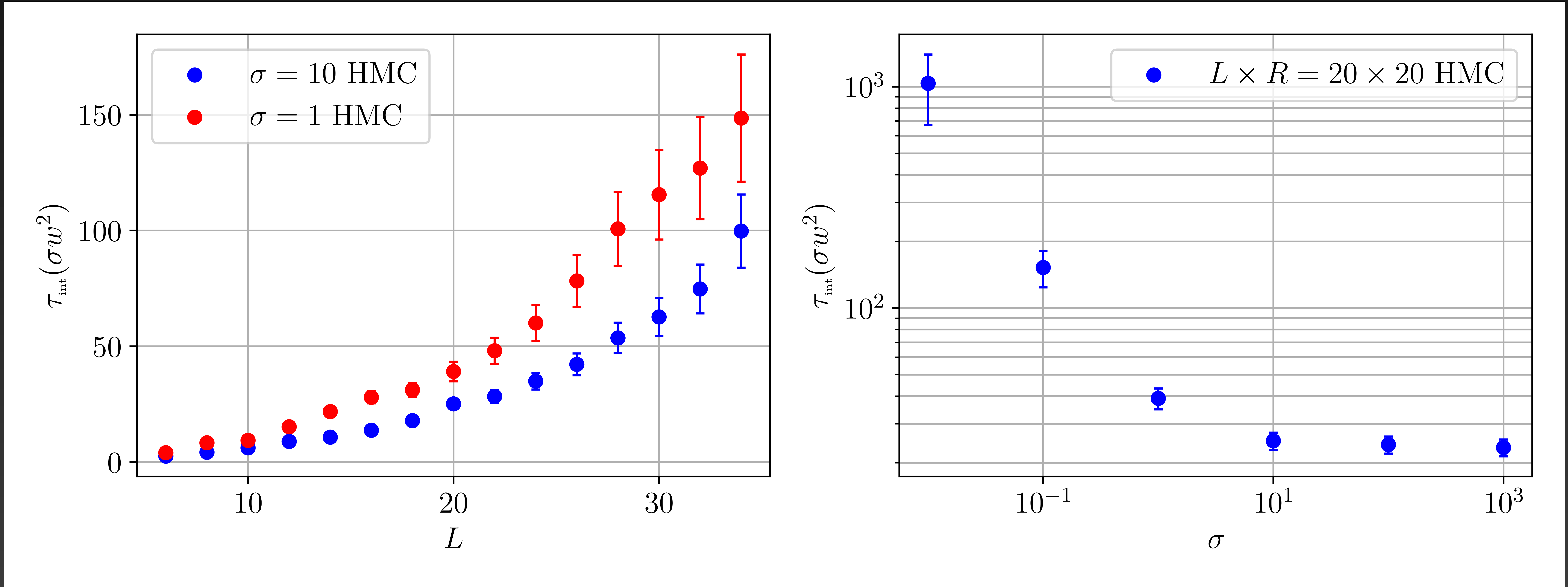
$$\sigma w^2 = \langle \phi(\tau, R/2)^2 \rangle_\tau$$

[Caselle, **EC**, Nada; 2307.01107]

Lacks of numerical methods

Numerical problems:

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



Normalizing Flows

Normalizing Flows

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

[Rezende+; 1505.05770]

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

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]

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

Sampling Boltzmann Distributions

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

[Nicoli+; 2007.07115]

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

↓ Uncorrelated!

↓ Partition functions!

Continuous Normalizing Flows

Continuous NFs (CNFs) are NFs in which the flow g_θ is the solution of a Neural Ordinary Differential Equation (NODE):

$$\frac{d\phi(t)_x}{dt} = \sum_{y,f} W_{x,y,f} K(t)_f \phi(t)_y$$

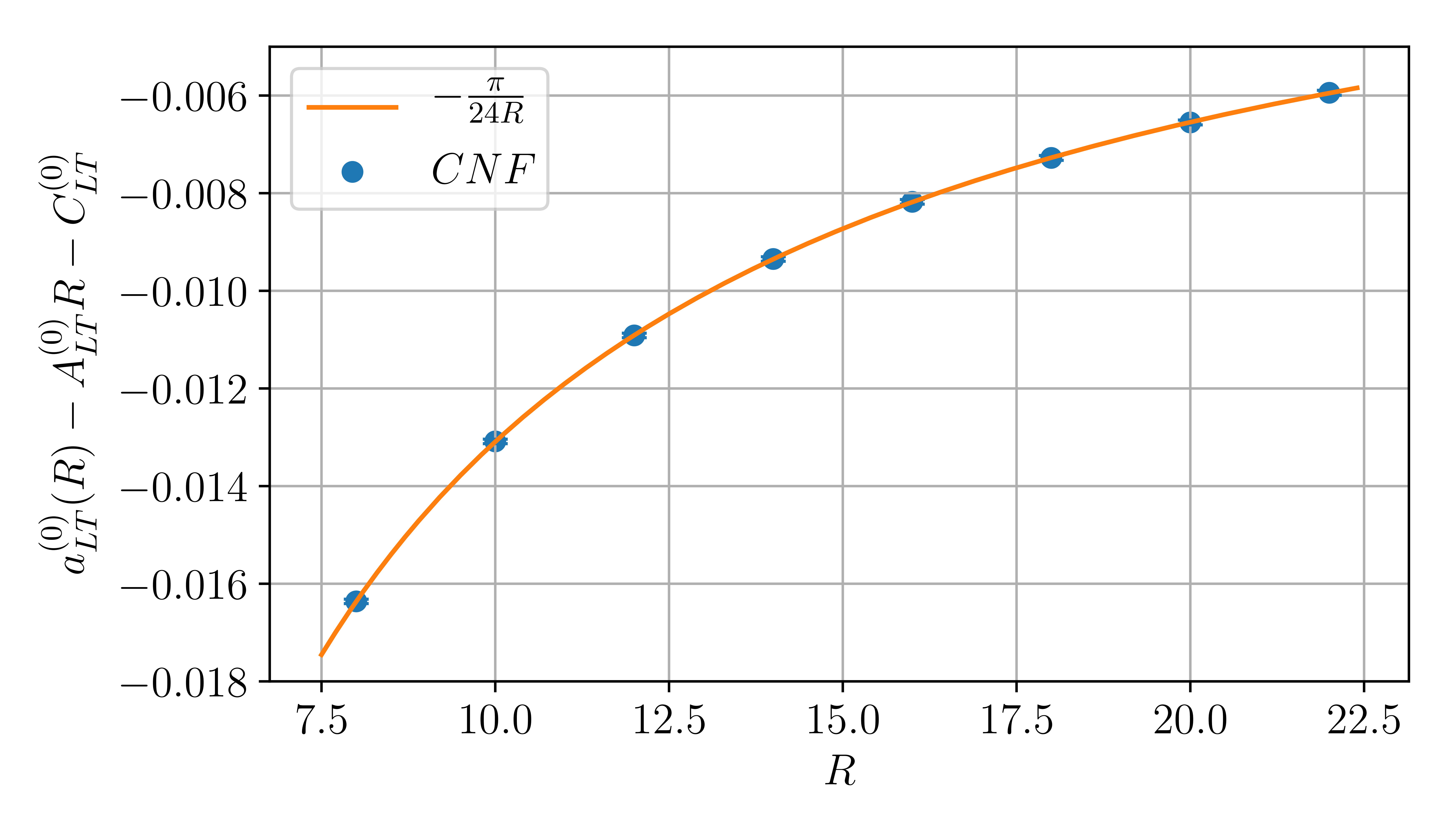
$$\frac{d \log q_\theta(\phi(t))}{dt} = -Tr\left[\sum_f W_f K(t)_f\right]$$

$K(t) \in \mathbb{R}^F$ temporal kernel of F Fourier coefficients

[Chen+;1806.07366],[Gerdes+,2207.00283] [Caselle, **EC**, Nada; 2307.01107]

Proof of Concept: Partition Function

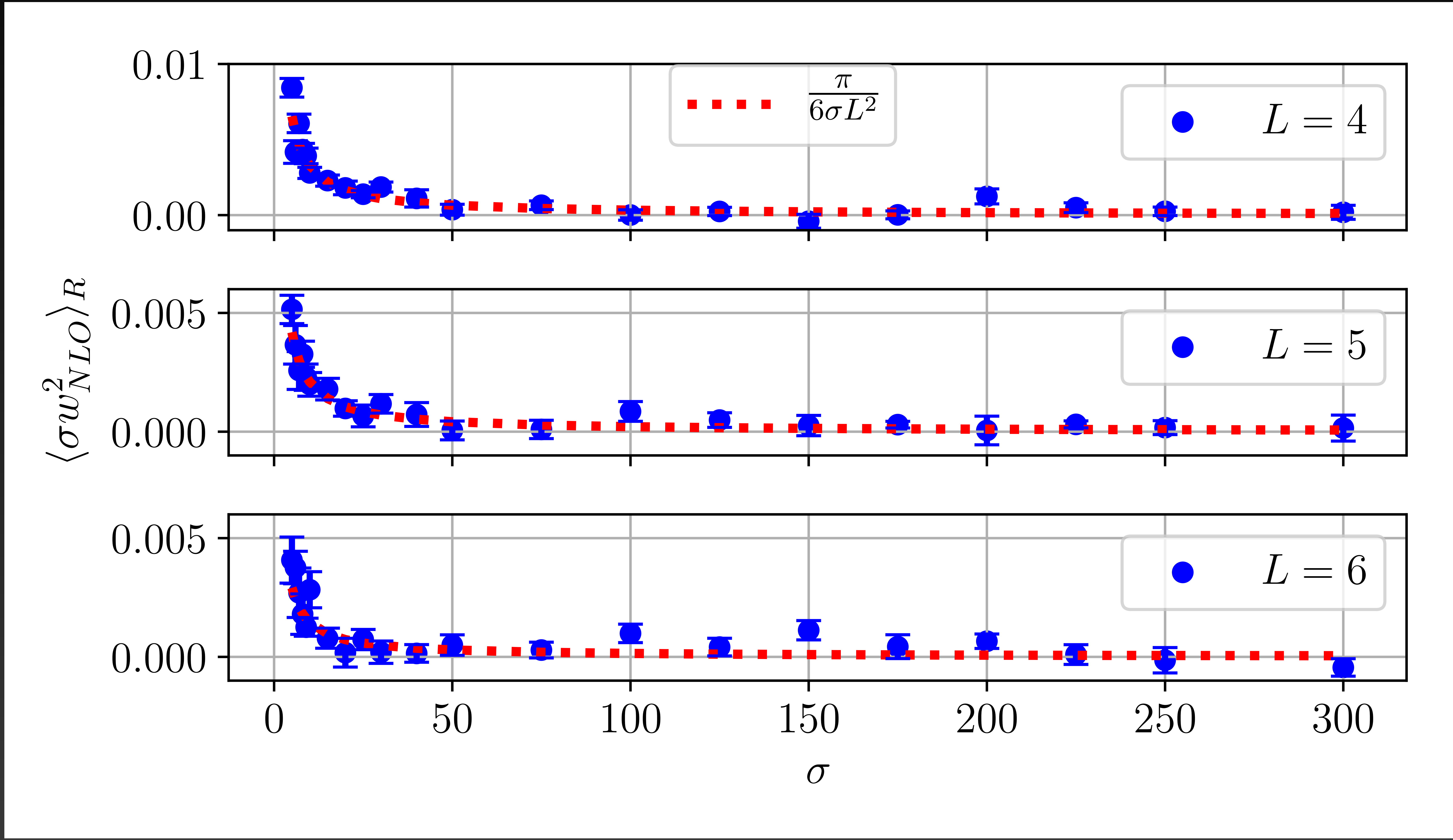
Large σ region ($\sigma \geq 40$), inferred coefficient: $-0.1309(2)$, target: $-0.1308996\ldots$



[Caselle, **EC**, Nada; 2307.01107]

Proof of Concept: Width

Large σ region ($\sigma \geq 5$), inferred coefficient: $0.55(5)$, target: $0.523598...$



[Caselle, **EC**, Nada; 2307.01107][Gliozzi+; 1002.4888]

Stochastic Normalizing Flows

Jarzynski's Equality

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

[Jarzynski; cond-mat/9610209]

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

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

Non-Equilibrium 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. $P_i \propto \exp(-S_i)$ 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:

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

Reverse probability density:

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

Dissipated Work

Observe that:

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

Where:

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

Crooks Fluctuation Theorem

Thus:

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

→ **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

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

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

Jarzynski's equality

[Jarzynski; cond-mat/9610209]

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

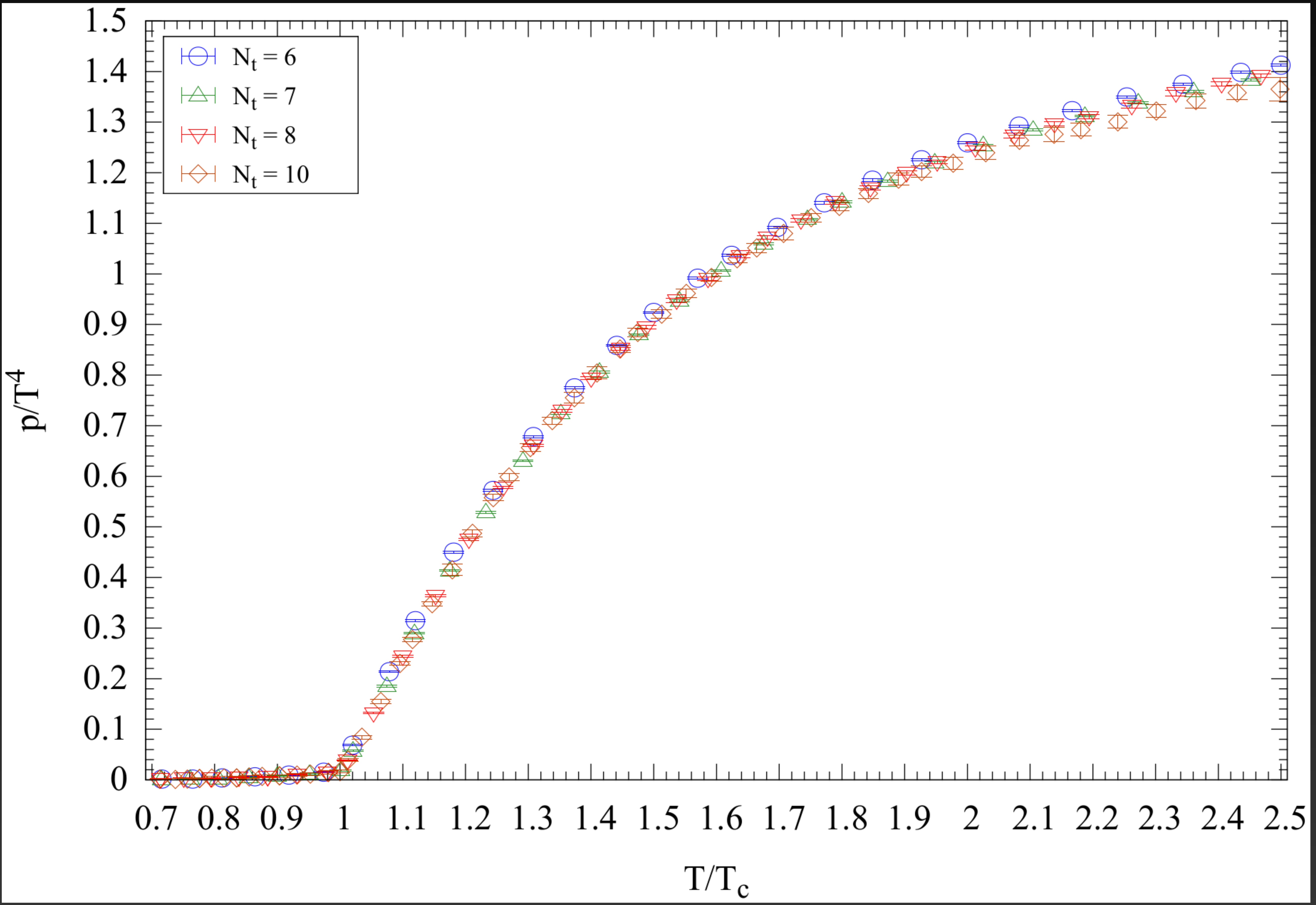
Non-Equilibrium Ensemble

Equilibrium Quantity

Jarzynksi 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+; 2310.11979]



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

Stochastic Normalizing Flows (SNFs)

Solution 2) use Machine Learning to minimize W_d

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

Where g_θ^i are NF layers with forward/reverse transition probability:

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

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

[Wu+; 2002.06707],[Caselle, **EC**, 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:

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

SNFs: Training

We can now train a SNF to minimize W_d^θ

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

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

Second Law!

Measure how reversible the process is.

SNFs: Physics-Informed Design

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:

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

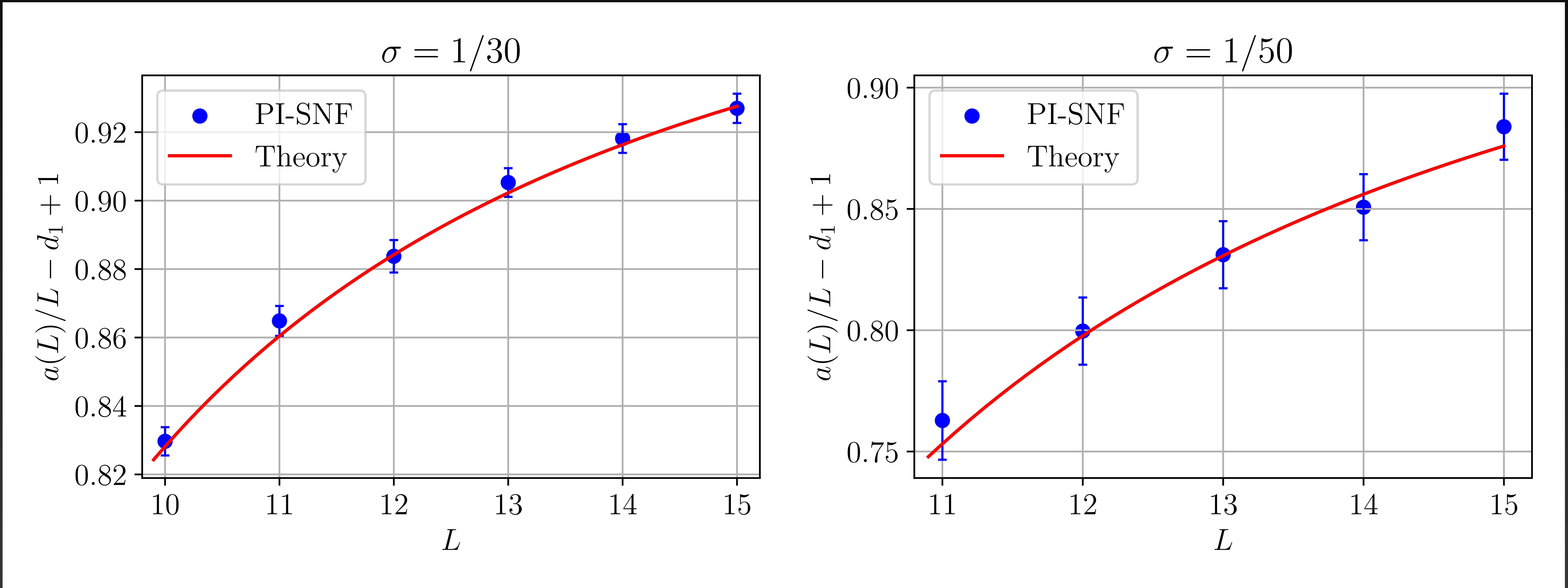
- Design inspired by the $T\bar{T}$ integrable irrelevant perturbation.

  [Cavaglià+; 1608.05534],[Smirnov and Zamolodchikov; 1608.05499]

SNFs: NG Partition Functions

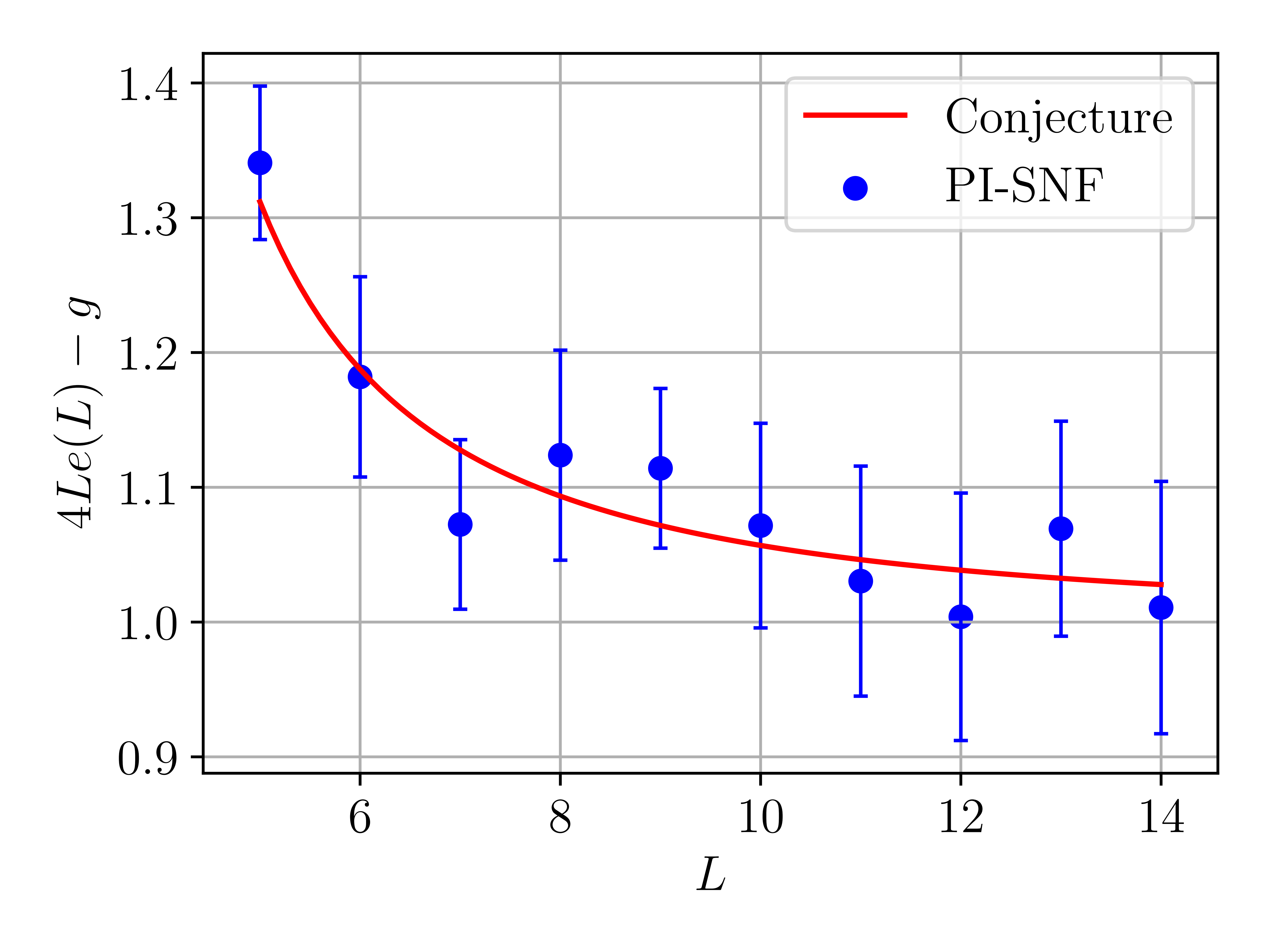
inferred: $-1.03(2)$, target: $-1.047...$

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



SNFs: NG Width

$\sigma = 1/10$, inferred: $1.09(8)$, target: $1.047...$



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

Conclusion

We showed that (S)NFs can be successful applied to sample from the EST probability distribution!

1. Flow-based application to Lattice Field Theory which is not a toy-model.
2. State-of-the-art results for Effective String Theory.
3. Toward the study of the “rigid” string:

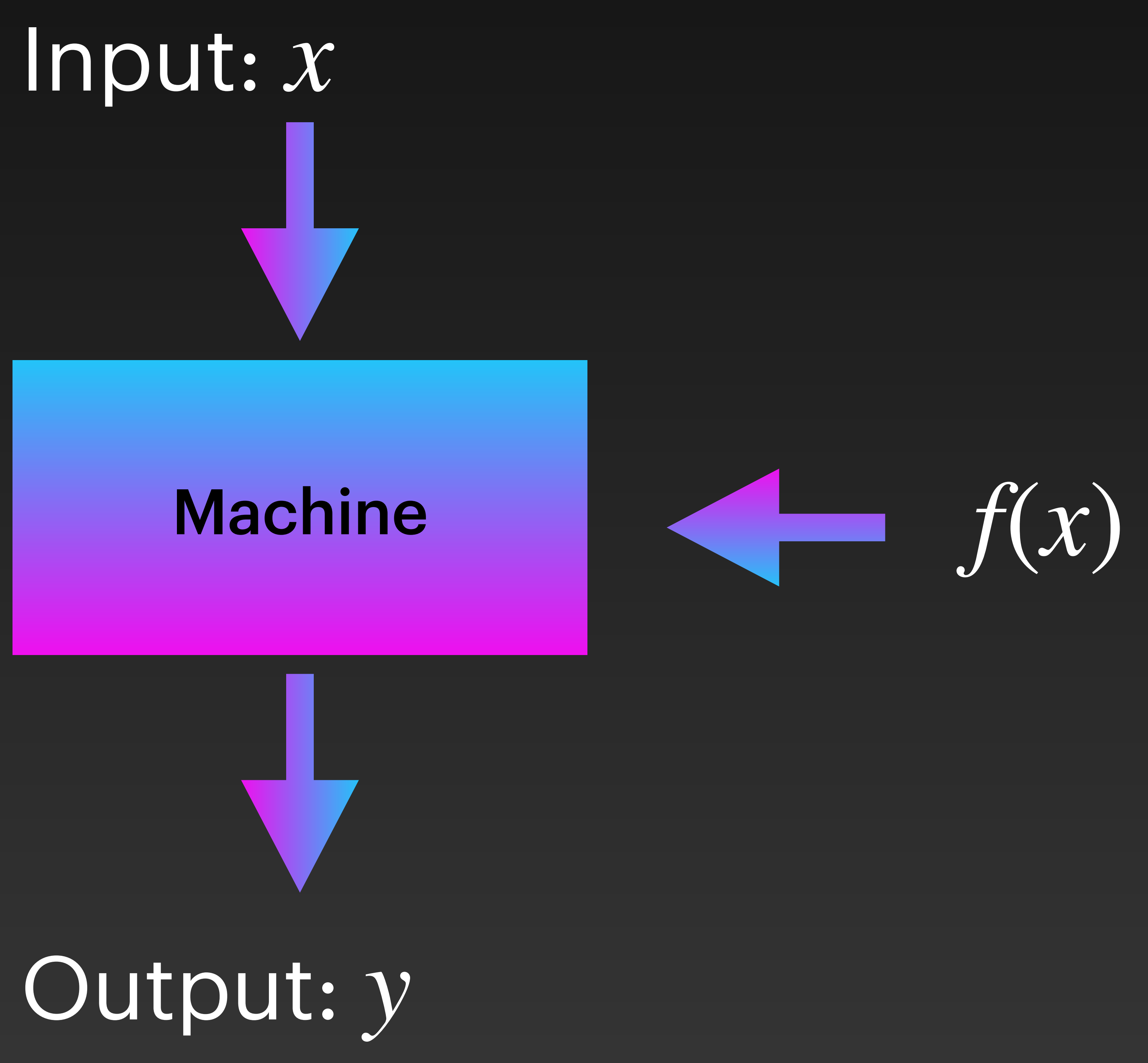
$$S_{BNG} = S_{NG} + \gamma \mathcal{K}^4$$

Thank you for your attention!

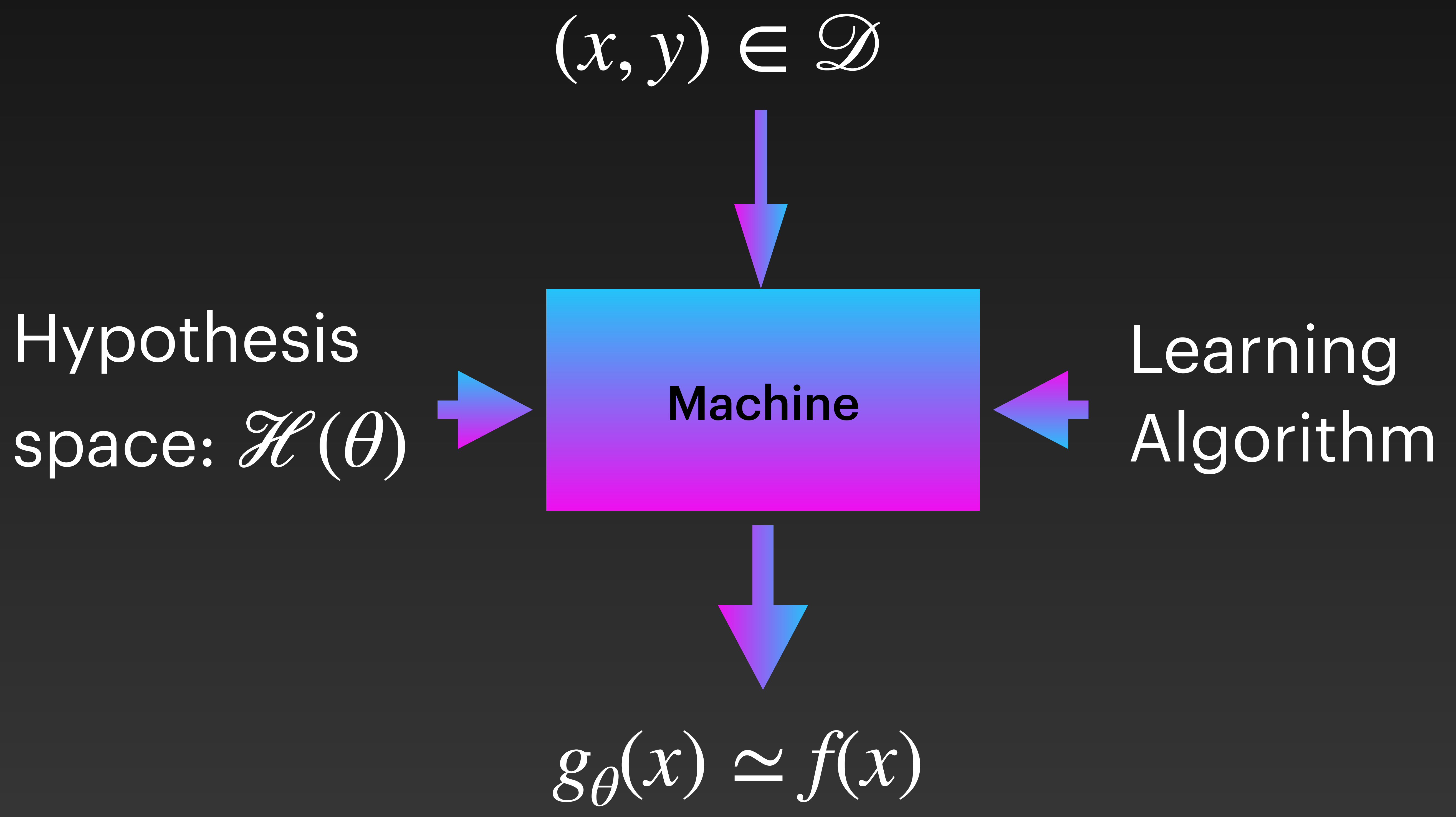
Machine Learning

We want a program ables to compute: $f(x) = y$ for a given input x

Classical programming:



Machine Learning:



SNFs: Technical details

- 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

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]