STOCHASTIC NORMALIZING FLOWS FOR GAUGE THEORIES AND DEFECTS

ELIA CELLINI

Based on:

- Caselle, EC, Nada, Panero; *JHEP* 07 (2022) 01, 2201.08862
- Bulgarelli, EC, Nada; *Phys. Rev. D* 111 (2025) 7, 7, 2412.00200
- Bulgarelli, EC, Jansen, Kühn, Nada, Nakajima, Nicoli, Panero; *Phys. Rev. Lett.* 134 (2025) 15, 151601, 2410.14466
- Bonanno, Bulgarelli, EC, Nada, Panfalone, Vadacchino, Verzichelli 25XX.XXXX

26/08/2025

Bridging analytical and numerical methods for quantum field theory

Trento - ECT*

UNIVERSITÀ DI TORINO

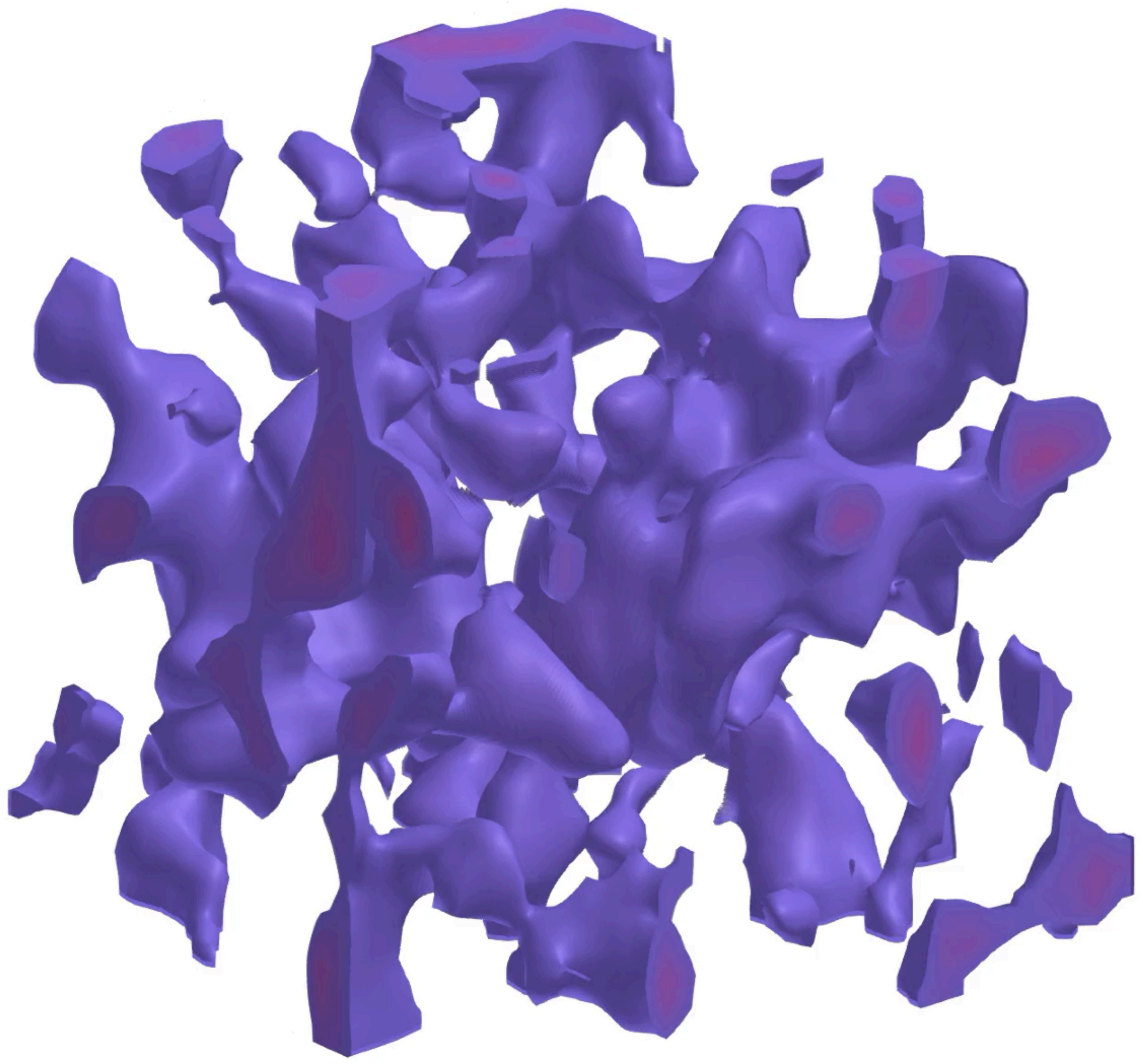
INTRODUCTION AND MOTIVATIONS

LATTICE GAUGE THEORY

In **Lattice Gauge Theory** (LGT) the gauge fields U are defined on links of a N-dimensional lattice representing the discretized space-time

$$\langle \mathcal{O} \rangle_{U \sim p} = \int DU p(U) \mathcal{O}(U)$$

$$p(U) = \frac{1}{Z} e^{-S_E[U]} \qquad Z \equiv \int D\phi e^{-S_E[U]}$$

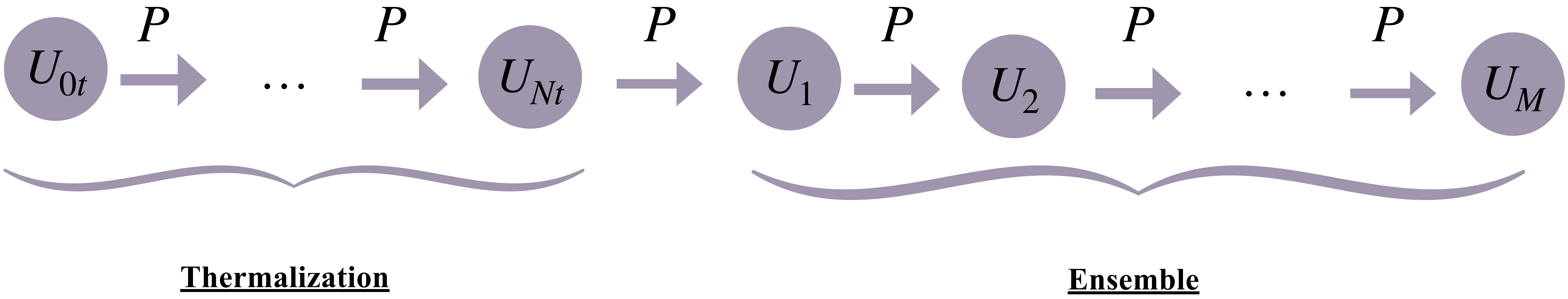


Main method: **Markov chain Monte Carlo (MCMC)**

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

MCMC

In MCMC methods, a stochastic **Markov kernel $P \propto \exp(-S)$** is applied recursively to obtain new configurations



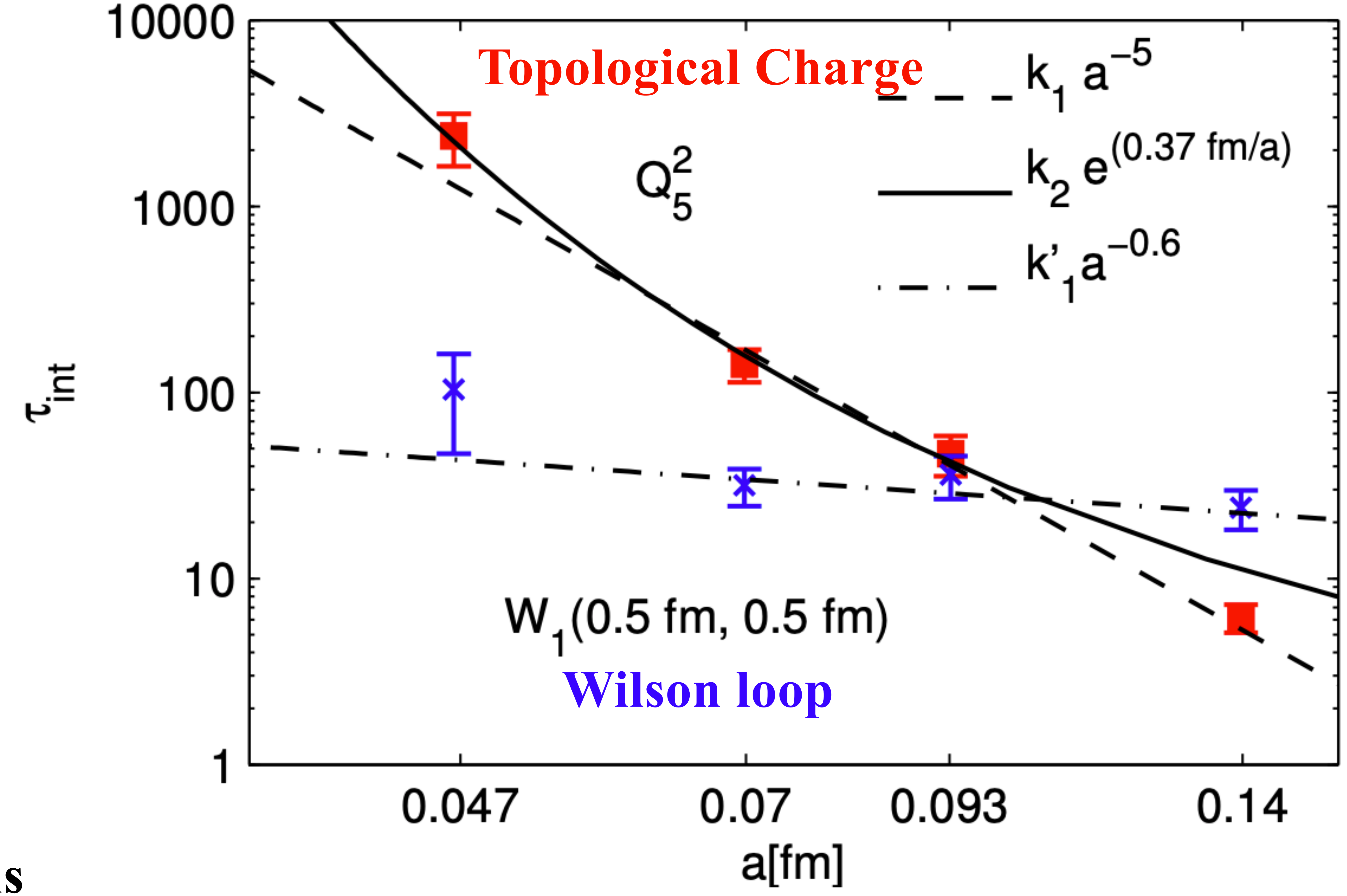
DRAWBACKS OF MCMC

The configurations generated with MCMC are correlated, and the integrated correlation time τ_{int} is related to the lattice step a of the target theory

$$\tau_{int} \sim a^{-z}$$

Approaching the continuum, τ_{int} diverge. This problem is know as the **Critical Slowing Down (CSD).** In particular, a severe CSD is the one on the **topological charge** called **Topological Freezing**

[Schaefer et al.; 1009.5228]



⟵ continuum limit

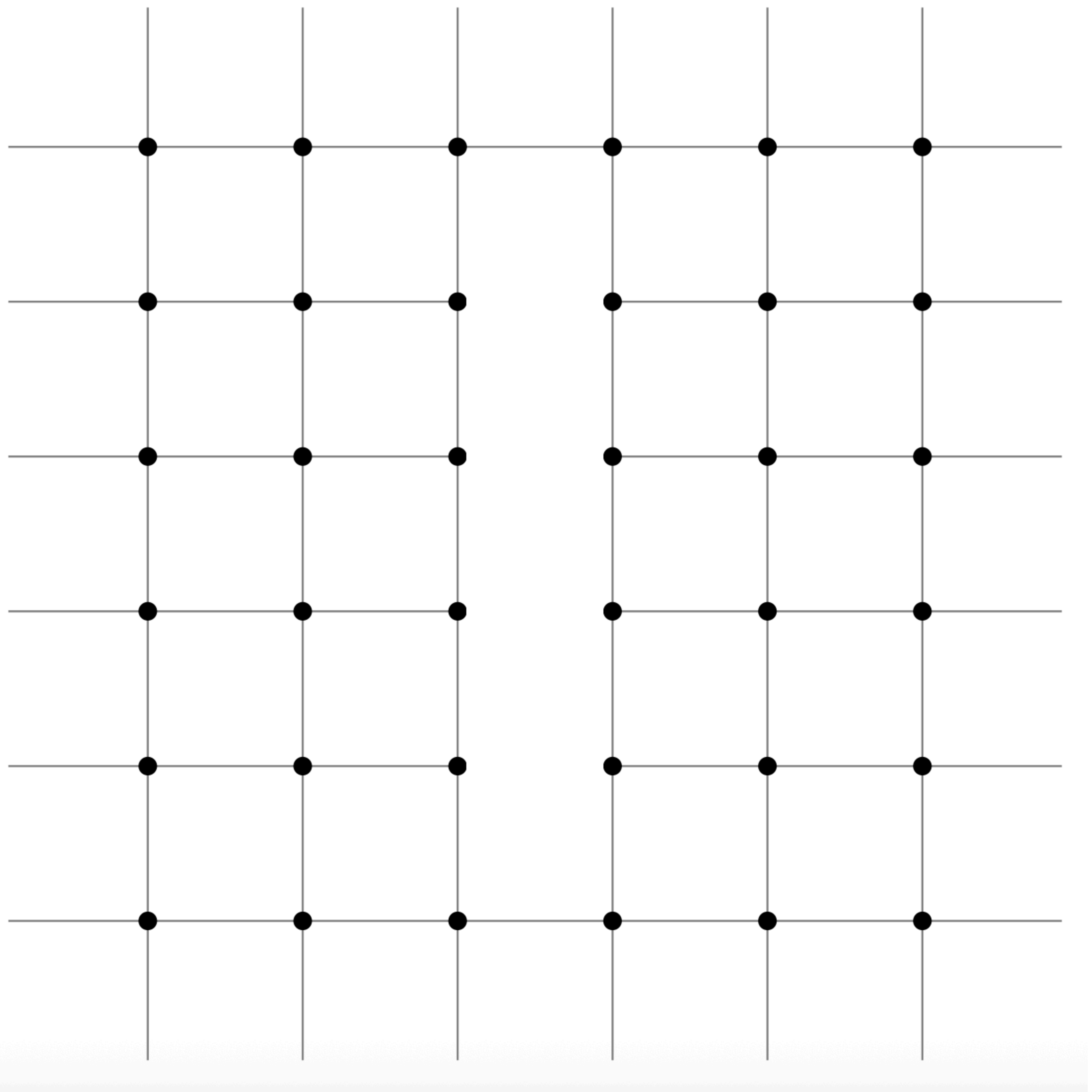
MCMC are also **not efficient in free energy calculations**

OPEN BOUNDARY CONDITIONS

One way to deal with the topological freezing is by adding a **defect** on the theory: the **Open Boundary Conditions (OBC)**

OBC are usually mapped to physical, periodic boundary conditions theories, using **Parallel Tempering (PT)**.

[Lüscher and Schaefer; 1105.4749] [Hasenbusch; 1706.04443]



However, in recent years, two promising new sampling approaches have been introduced in lattice field theory:

- **Non-Equilibrium MCMC (NE-MCMC) → Similar perfomances to PT on OBC!**
- **Normalizing Flows (NFs)**

Interestingly, **both algorithms found successful applications to defects**

Goal: outperform NE-MCMC on OBC using Deep Learning!

NON-EQUILIBRIUM MCMC

CROOKS' FLUCTUATION THEOREM

Given an **arbitrary transformation between two thermodynamics states**, **Crooks' fluctuation theorem relates the exponential of the dissipated work to the forward and reverse transition probability density** of the trajectory:

[Crooks; cond-mat/9901352]

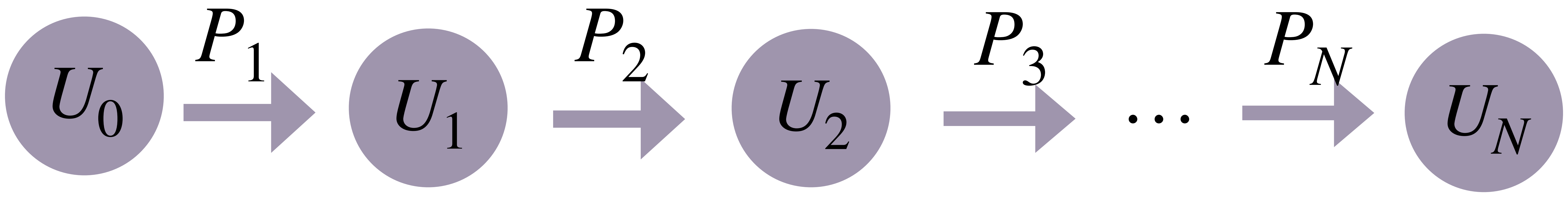
$$\frac{\mathcal{P}_f(W_d)}{\mathcal{P}_r(-W_d)} = e^{W_d}$$

Where the dimensionless dissipated work $W_d = W - \Delta F$

$$W(U, \ldots, U_N) = \sum_{n=0}^{N-1} \left\{ S_{(n+1)}\left(U_n\right) - S_{(n)}\left(U_n\right) \right\}$$

NON-EQUILIBRIUM MCMC

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



[Jarzynski; cond-mat/9610209]

1. **Thermalized** q_0 "**prior**"
2. $P_i \propto \exp(-S_i)$ **change along the processes** (satisfy detailed balance) following a **protocol**
3. $p = \exp(-S_N)/Z_N \to$ "**target**" distribution

Remark: no thermalization during the processes.

$$\langle \mathcal{O} \rangle_p = \langle \mathcal{O} e^{-W_d} \rangle_f$$

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

The average is done over all possible chains

Autocorrelation on the generated samples cannot be larger than the one of the prior configurations!

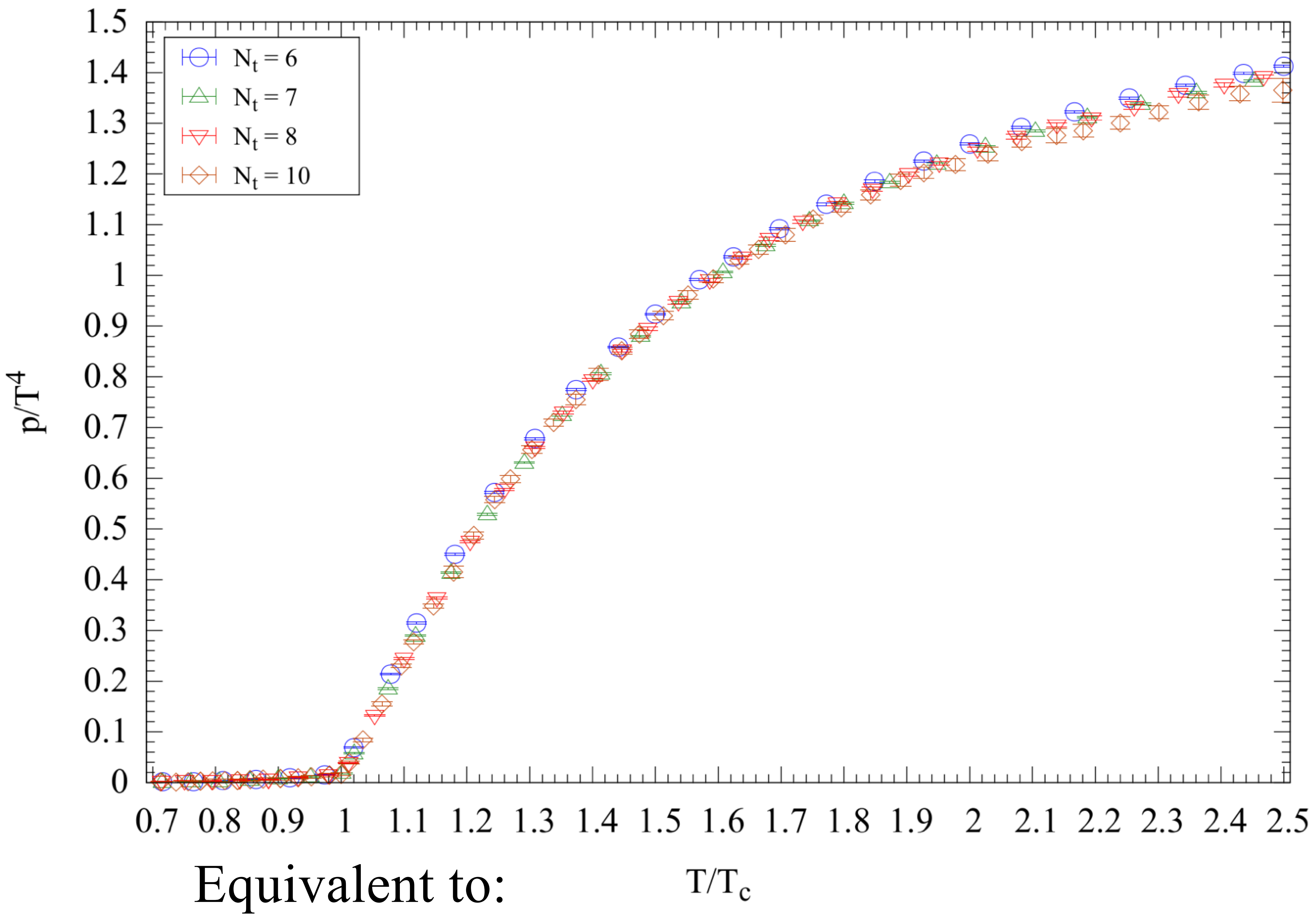
NE-MCMC FOR LFT

NE-MCMC have been exploited to obtain **state-of-the-arts results** in LFT:

- Interface free energy.
  [Caselle et al.; 1604.05544]
- $SU(3)$ e.o.s.
  [Caselle et al.; 1801.03110]
- Running coupling
  [Francesconi et al.; 2003.13734]
- Entanglement entropy
  [Bulgarelli and Panero; 2304.03311, 2404.01987]
- Topological freezing
  [Bonanno et al.; 2402.06561, 2411.00620]
- Casimir energy
  [Bulgarelli et al.; 2505.20403]

Defects

[Caselle et al.; 1801.03110]



Equivalent to:

Annealed Importance Sampling

[Neal; physics/9803008]

DRAWBACKS

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”

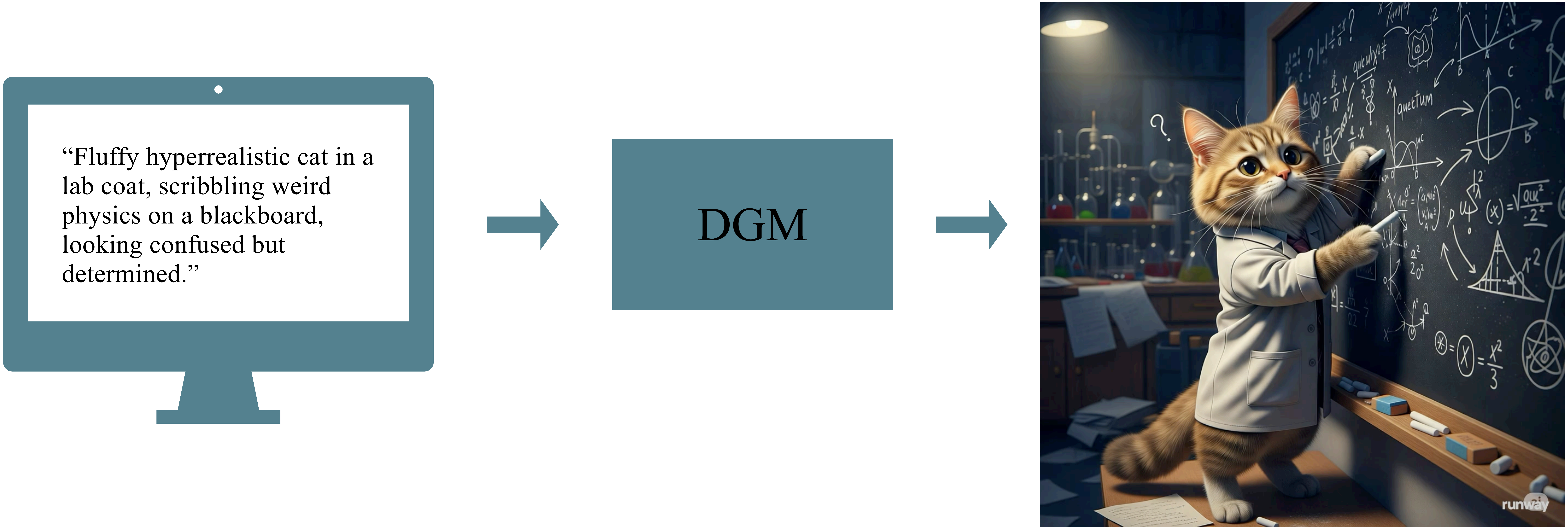
Solution 1) Infinite MCMC steps → quasi-static transformations → “small” W_d

Solution 2) use Machine Learning to minimize W_d

NORMALIZING FLOWS

DEEP GENERATIVE MODELS

Deep generative models (DGM) are **neural networks** trained to **learn and sample from complex data distributions**, enabling them to generate new, realistic data such as images or texts.



NEURAL GENERATIVE SAMPLERS (NGS)

Key idea: Leverage a **DGM** to learn a **variational approximation q_θ of the target p** : $q_\theta \simeq p$

[Nicoli et al.; 1910.13496]

$$q_\theta(U) \simeq p = \exp(-S(U))/Z$$

$$U \sim q_\theta(U)$$

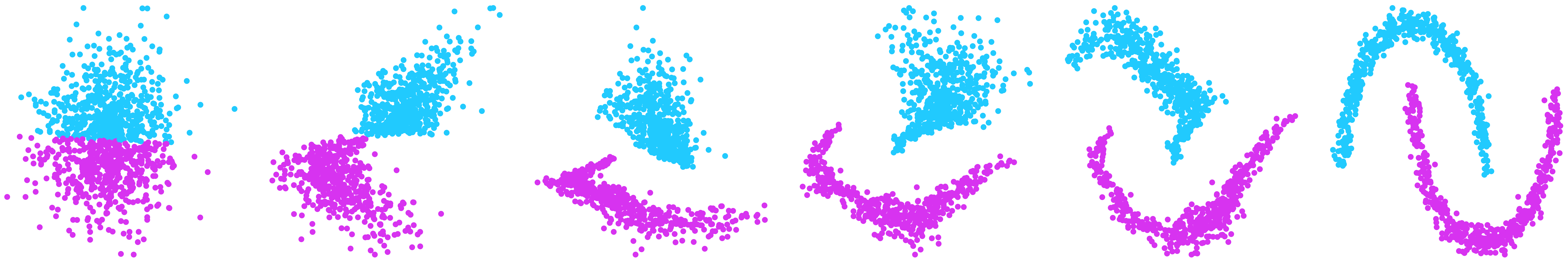
Fundamental aspect: the learned variational density must be computed exactly

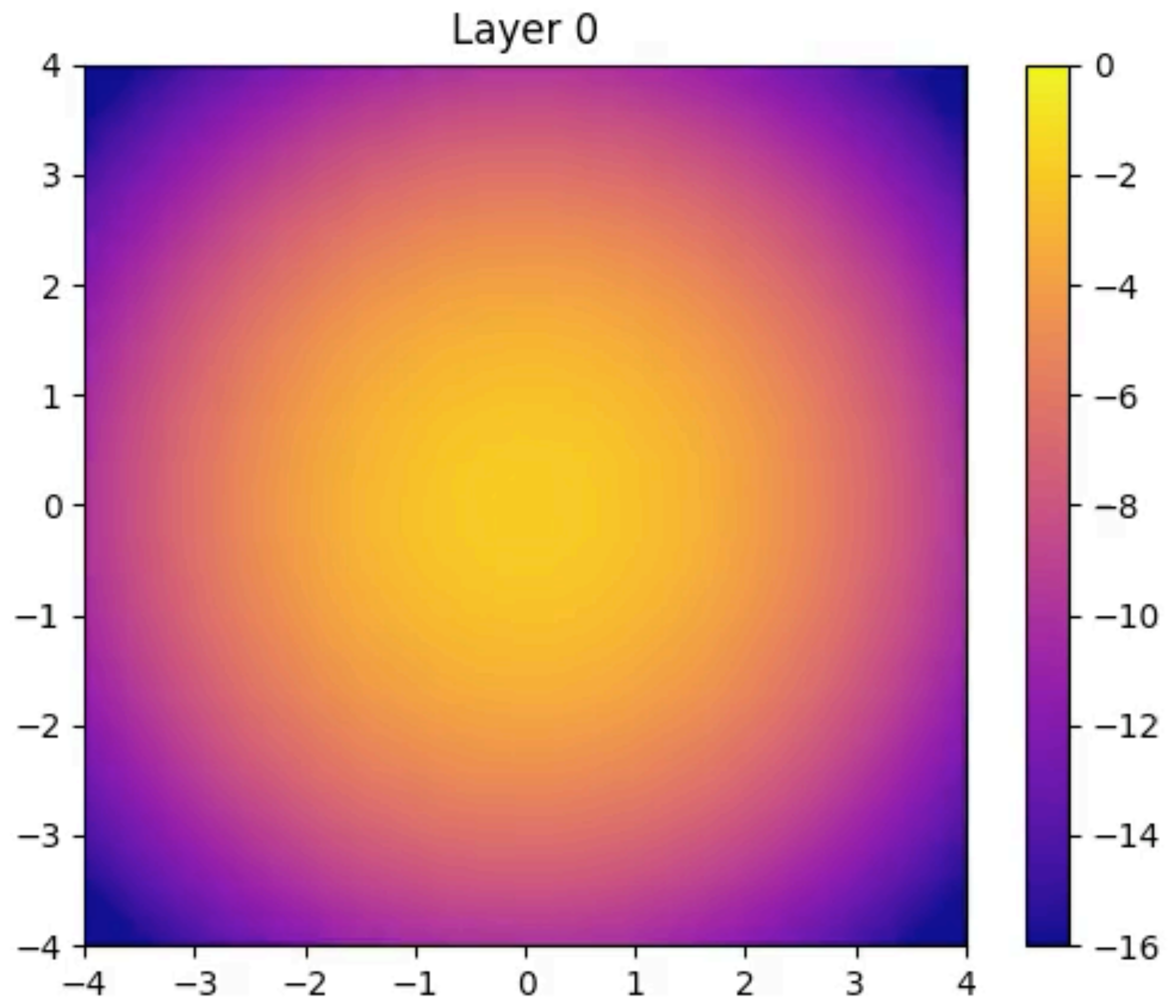
NORMALIZING FLOWS

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

[Tabak and Vanden-Eijnden; 2010], [Tabak and Turner; 2013],[Rezende et al.; 1505.05770]

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





TRAINING AND SAMPLING

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

[Wu et al.; 1809.10606],[Noé et al.; 1812.01729],[Albergo et al.; 1904.12072],[Nicoli et al.; 1910.13496, 2007.07115]

$$D_{KL}(q_\theta || p) = \int dU q_\theta(U) \log \frac{q_\theta(U)}{p(U)} = \int dU q_\theta(U) \big(S(U) + \log q_\theta(U)\big) - F \geq 0.$$

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

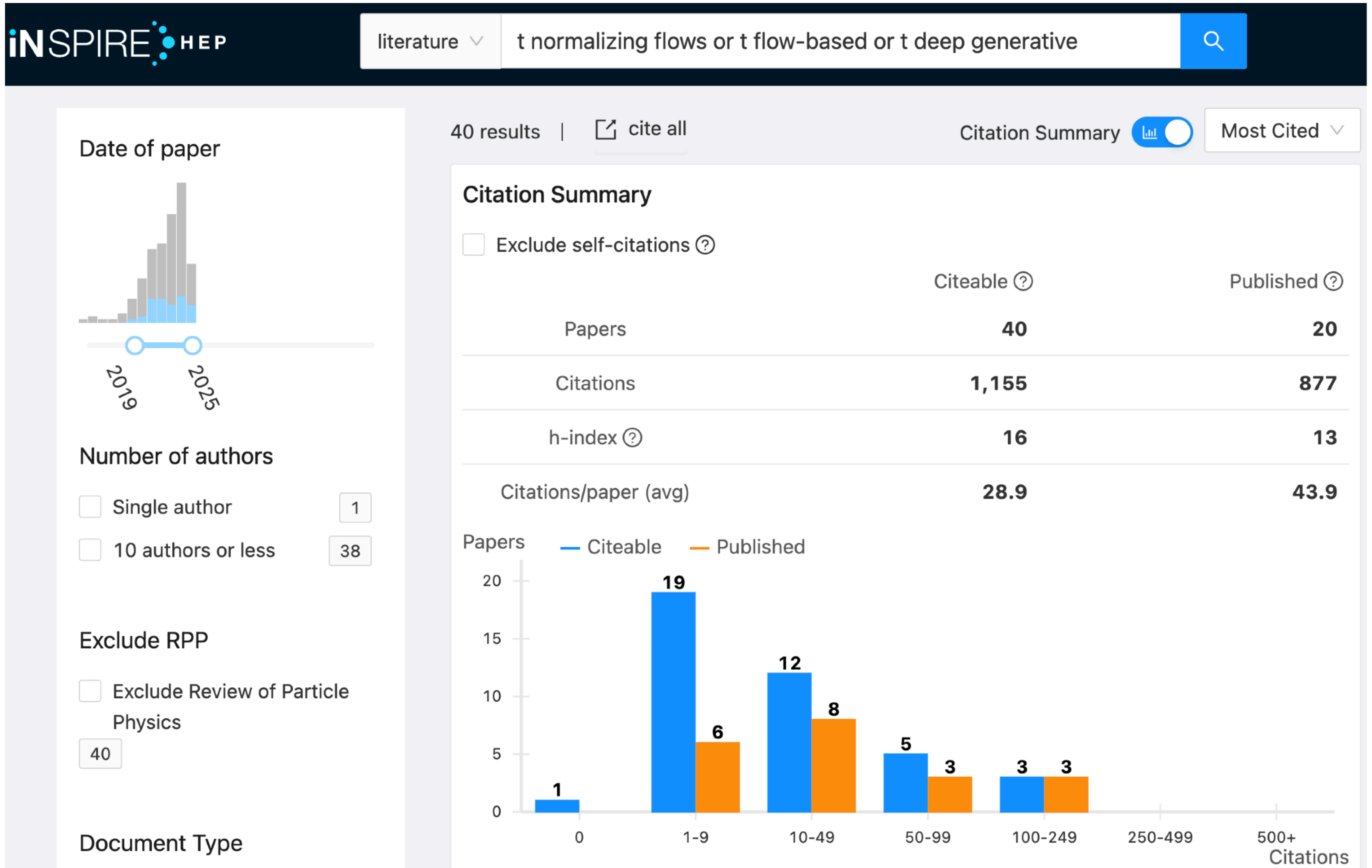
[Nicoli et al.; 1910.13496, 2007.07115]

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

Autocorrelation on the generated samples cannot be larger than the one of the prior configurations!

NF IN LATTICE

Widely investigated in lattice field theory, ranging from scalar field theory, gauge theory and QCD.



However, in general, **NFs suffer from poor scaling!** **General purpose "Lattice GPT" models are still far…**

3 PRACTICAL APPLICATIONS

1. **Derivative of observables with respect to action parameters α using correlated ensemble**

[Bacchio, 2305.07932],[Abbot et al.; 2401.10874]

$$\frac{d\langle O\rangle_\alpha}{d\alpha} \simeq \frac{\langle O\rangle_\alpha - \langle O\rangle_{\alpha+\epsilon}}{\epsilon} = \frac{1}{\epsilon}\langle O(U) - w(f(U))O(f(U))\rangle_\alpha$$

$$f : S_\alpha \to S_{\alpha+\epsilon}$$

$$w = \tilde{w}/\langle\tilde{w}\rangle$$

Observe that:

$$\langle O(t)\rangle_\lambda = \frac{1}{Z}\int DU\, O(t) e^{-S[U]+\lambda\, O(0)}$$

$$C(t) = \langle O(t)O(0)\rangle = \left.\frac{d\langle O(t)\rangle_\lambda}{d\lambda}\right|_{\lambda=0}$$

→ **Computing $n-$point functions by evaluating $(n-1)$ derivatives of one point functions.**

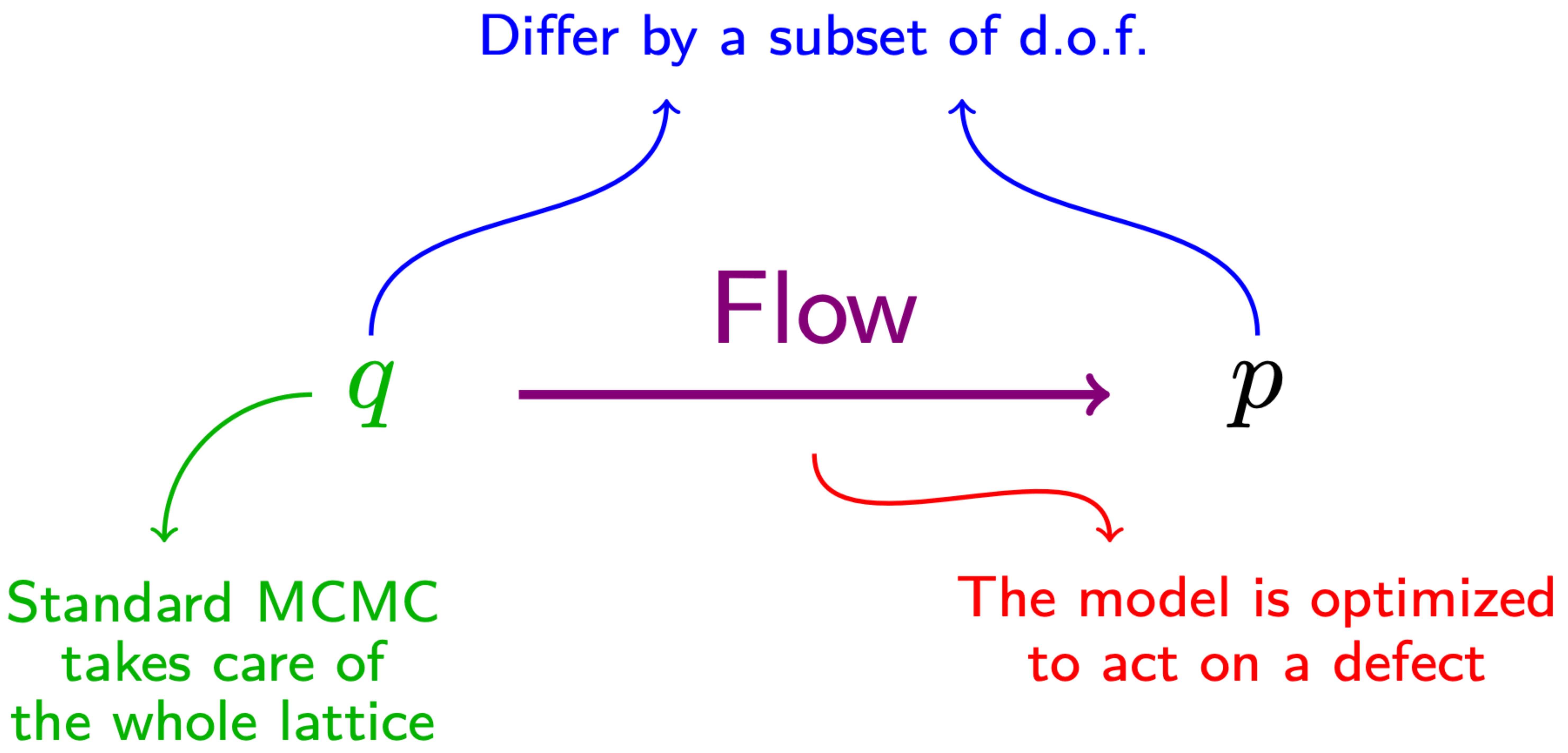
Tackling Signal To Noise in large distance correlators!

[Catumba and Ramos; 2502.15570]

3 PRACTICAL APPLICATIONS

2. <u>Application to Defects</u> → <u>Non trivial non-perturbative physics</u>

[Bulgarelli, EC et al.; 2410.14466]



<u>State of the Arts results in scalar field theory Entanglement Entropy calculations with pure NFs!</u>

3 PRACTICAL APPLICATIONS

3. **Combination with MCMC** → **Scaling up powerful algorithms**

Standard MCMC

↓

Deep learning HMC

[Foreman et al.; 2105.03418]

Non-Equilibrium MCMC

↓

Stochastic NFs

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 et al; 2002.06707],[Caselle, EC, Nada, Panero; 2201.08862],[Bulgarelli, EC, Nada.; 2412.00200]

VARIATIONAL 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:

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

TRAINING OF SNF

We can now train a SNF by minimizing:

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

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

Second Law!

More reversible trajectories → smaller work → better error!

Successful application to **Effective String Theory** (scalar theory)

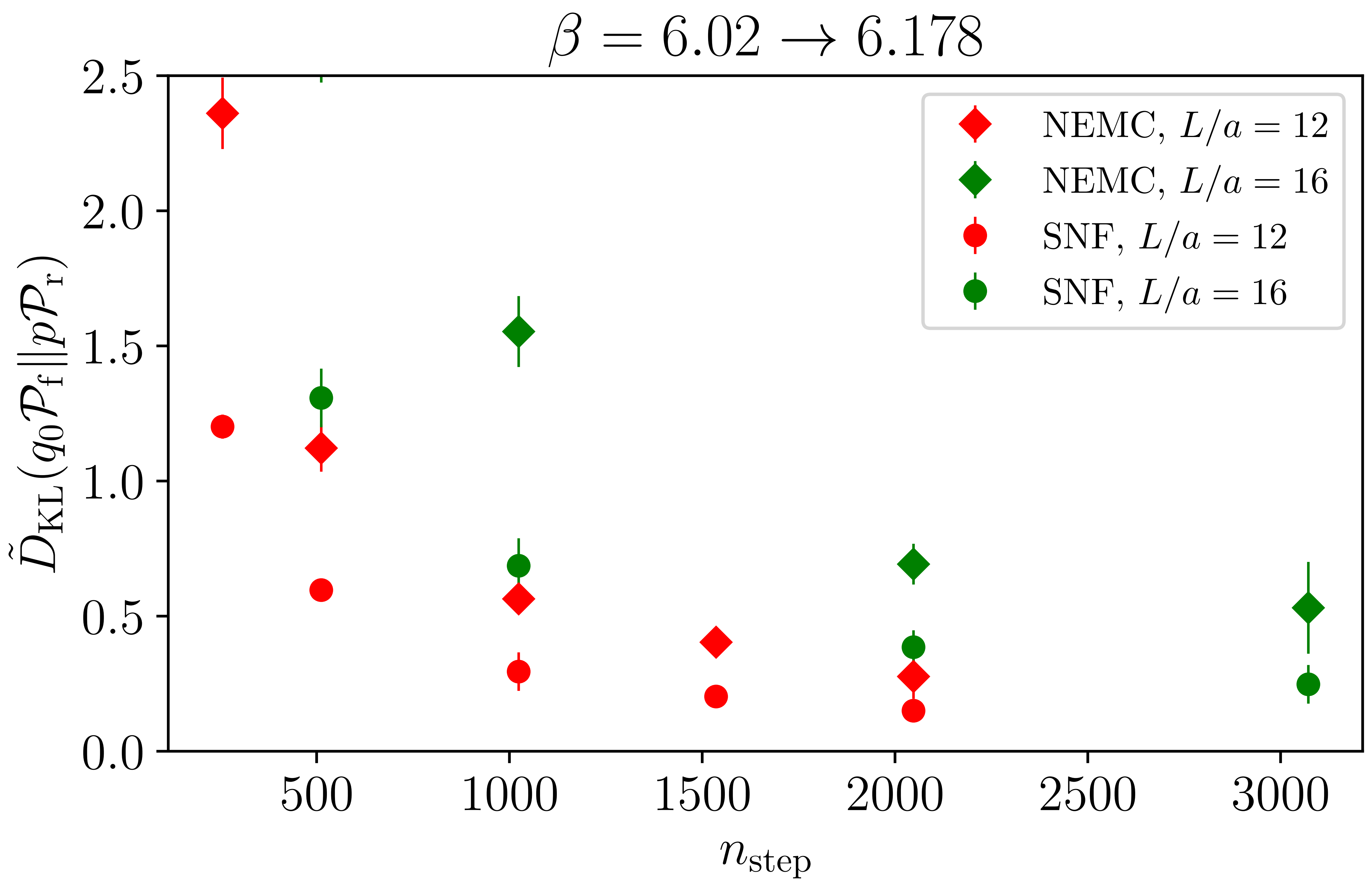
[Caselle, EC, Nada; 2307.01107, 2409.15937]

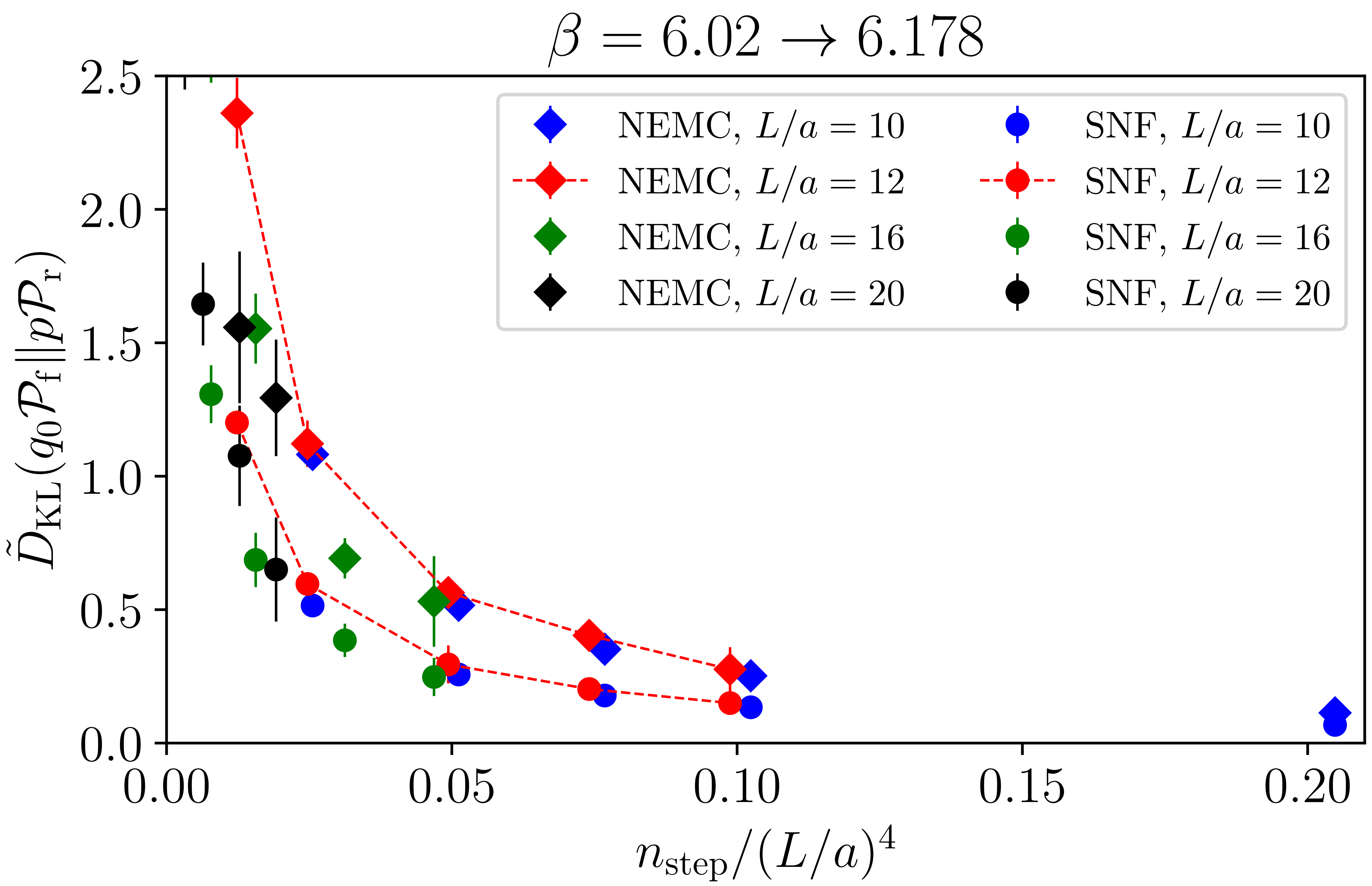
$d = 3 + 1$ SU(3): FLOWS IN β

Bulgarelli, EC, Nada; *Phys. Rev. D* 111 (2025) 7, 7, 2412.00200

Gauge equivariant "Smearing" NFs layer + Heatbath + Over relaxation

VOLUME SCALING DKL



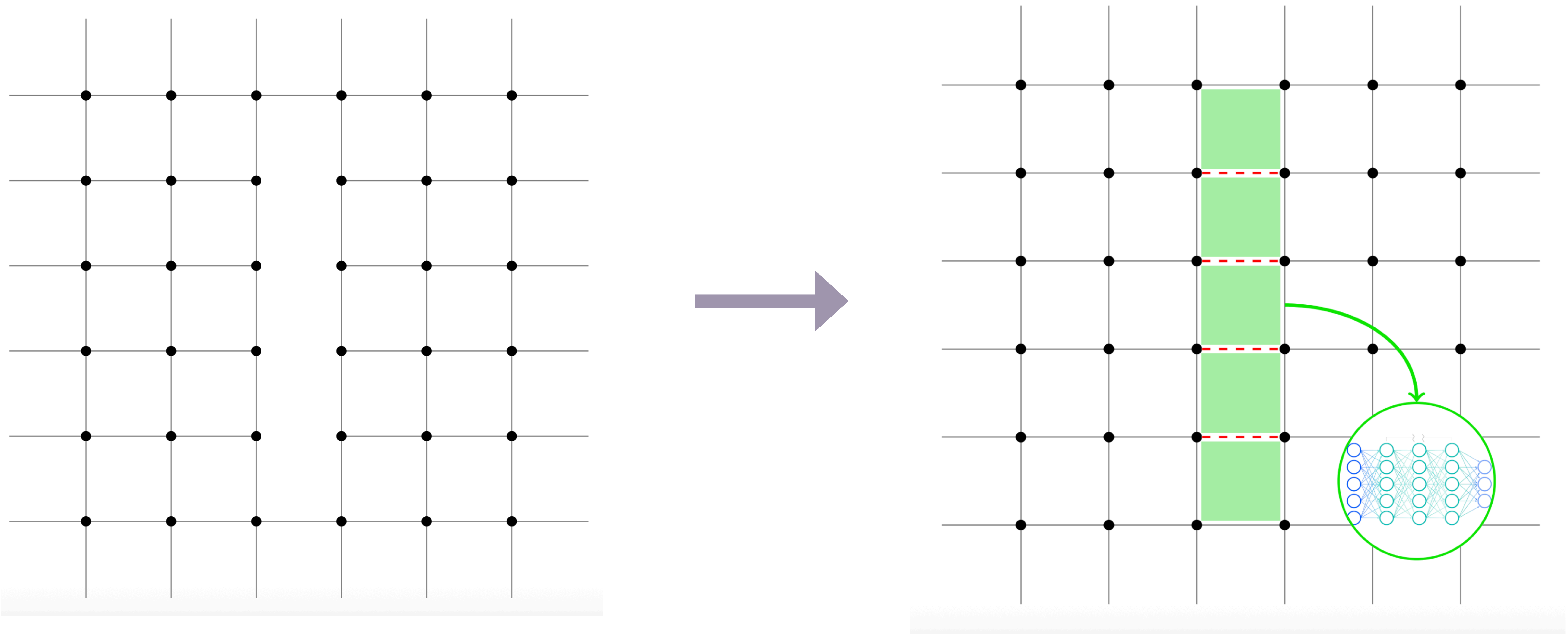


$d = 3 + 1$ SU(3): OPEN BOUNDARY CONDITIONS

Bonanno, Bulgarelli, EC, Nada, Panfalone, Vadacchino, Verzichelli; 25XX.XXXX

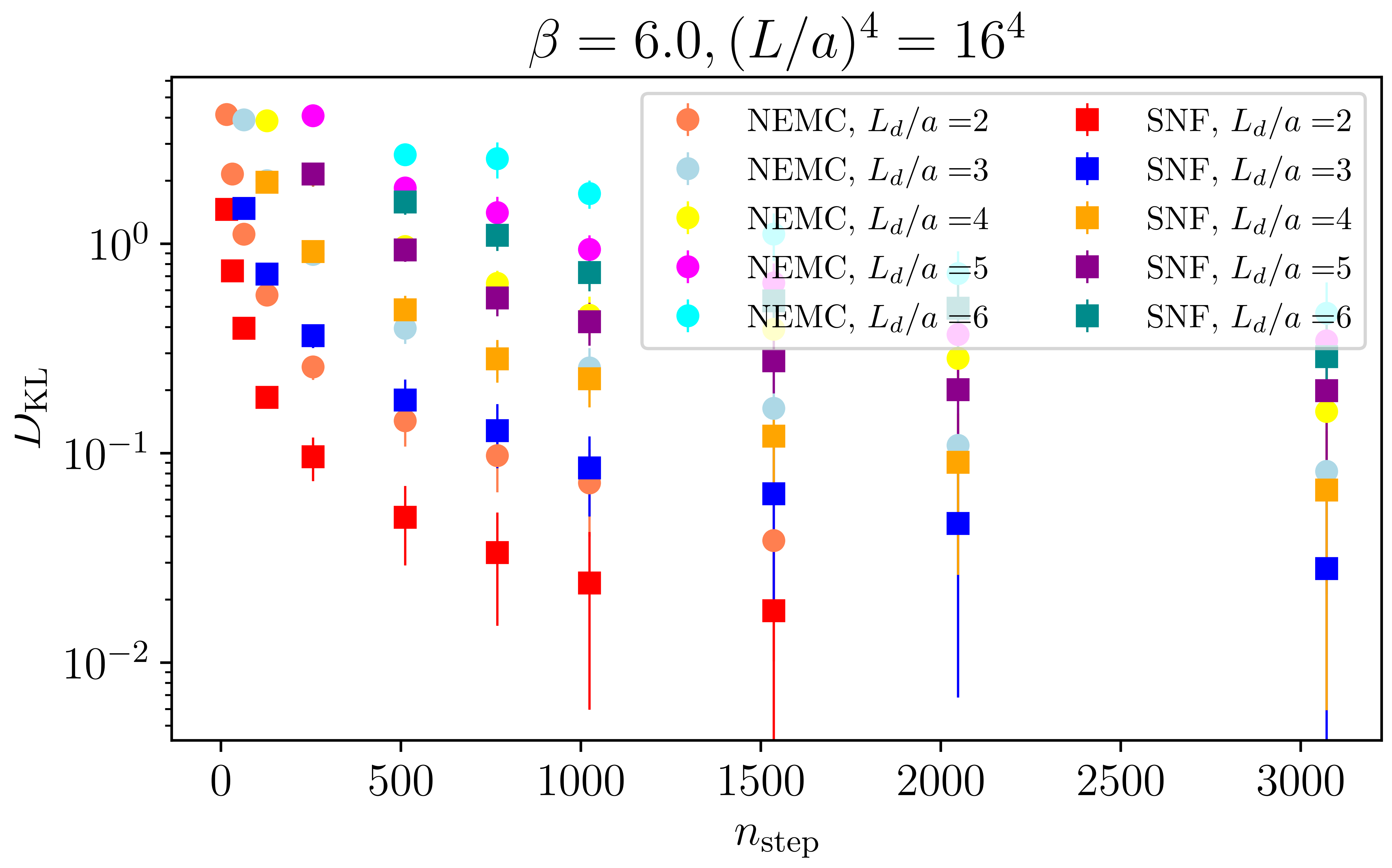
Gauge equivariant “Smearing” NFs layer + Heatbath + Over relaxation

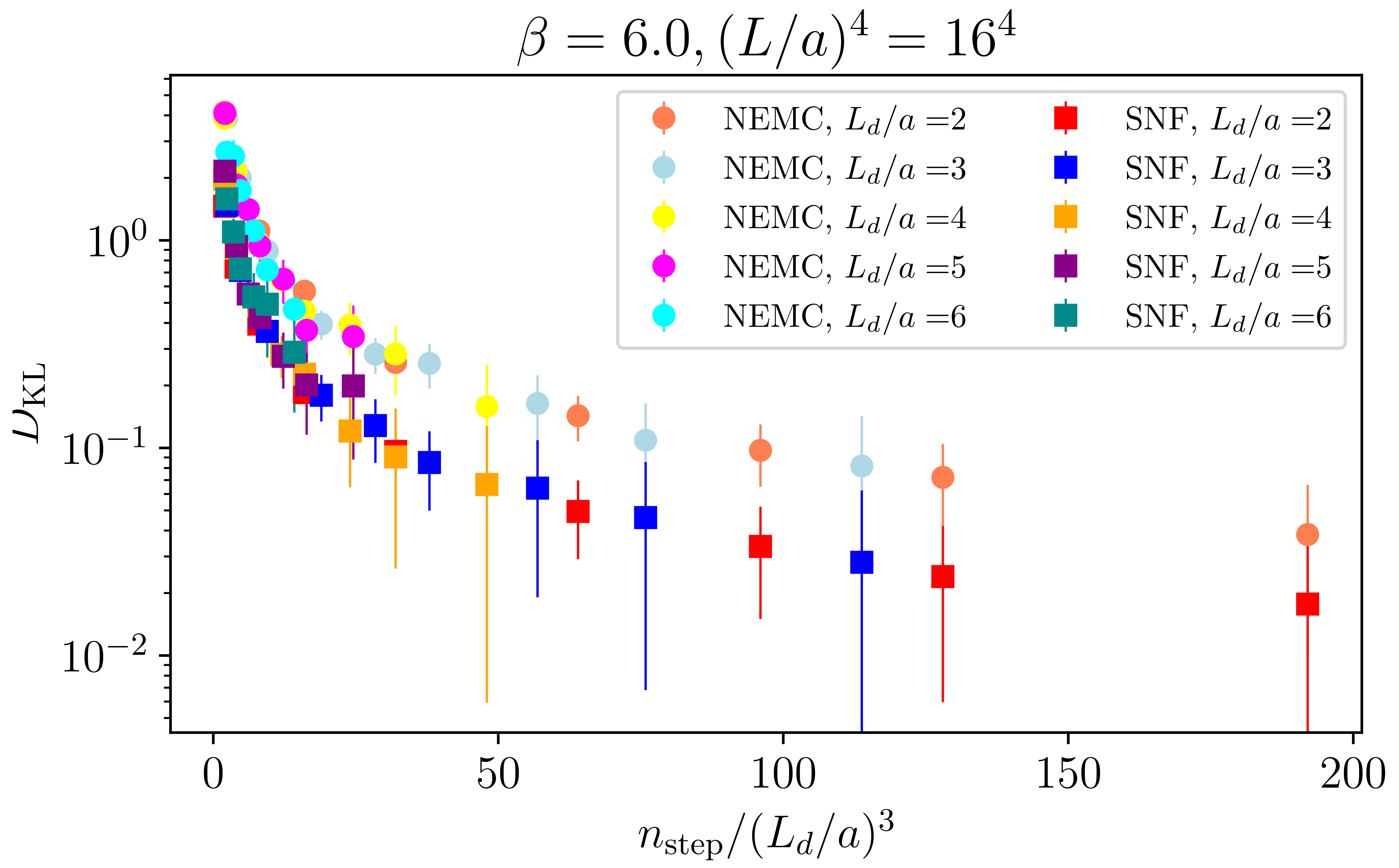
DEFECT COUPLING LAYER



Protocol: $\beta_{defect} : 0 \rightarrow \beta_{bulk}$

DEFECT VOLUME SCALING





OUTLOOK

- **Flow-based and Non-Equilibrium samplers provide new routes to study lattice field theory!**
- **Great scaling of SNFs with the d.o.f!**
- **Toward the continuum extrapolation of the topological susceptibility in SU(3) (and QCD!)**
- **Flow-based samplers fit well with defects:**
  1. MCMC takes care of the main bulk of the calculations
  2. ML work only on the defects:
     low dimensional structure → small number of d.o.f.

QUESTIONS

- **Other defects?**
- **Free-energy related observables?**

THANK YOU FOR YOUR ATTENTION!

Turin Lattice Field Theory group Github

BACKUP SLIDE

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 Jarzynski 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]
- NETS: a Non-Equilibrium Transport Sampler
  [Albergo+; 2410.02711]
- Replica Exchange with NFs:
  [Invernizzi+; 2210.14104], [Abbott+; 2404.11674]

GAUGE EQUIVARIANT LAYER

[Nagai and Tomya.;2103.11965],[Abbott et al.;2305.02402]

(Masked) Stout smearing:

$$U'_\mu(x) = g_n(U_\mu(x)) = \exp\left(iQ^{(n)}_\mu(x)\right) U_\mu(x),$$

Traceless Hermitian

$$Q^{(n)}_\mu(x) = \frac{i}{2}\left((\Omega^{(n)}_\mu(x))^\dagger - \Omega^{(n)}_\mu(x)\right) - \frac{i}{2N}\mathrm{Tr}\left((\Omega^{(n)}_\mu(x))^\dagger - \Omega^{(n)}_\mu(x)\right)$$

Active link transformed
with frozen staples

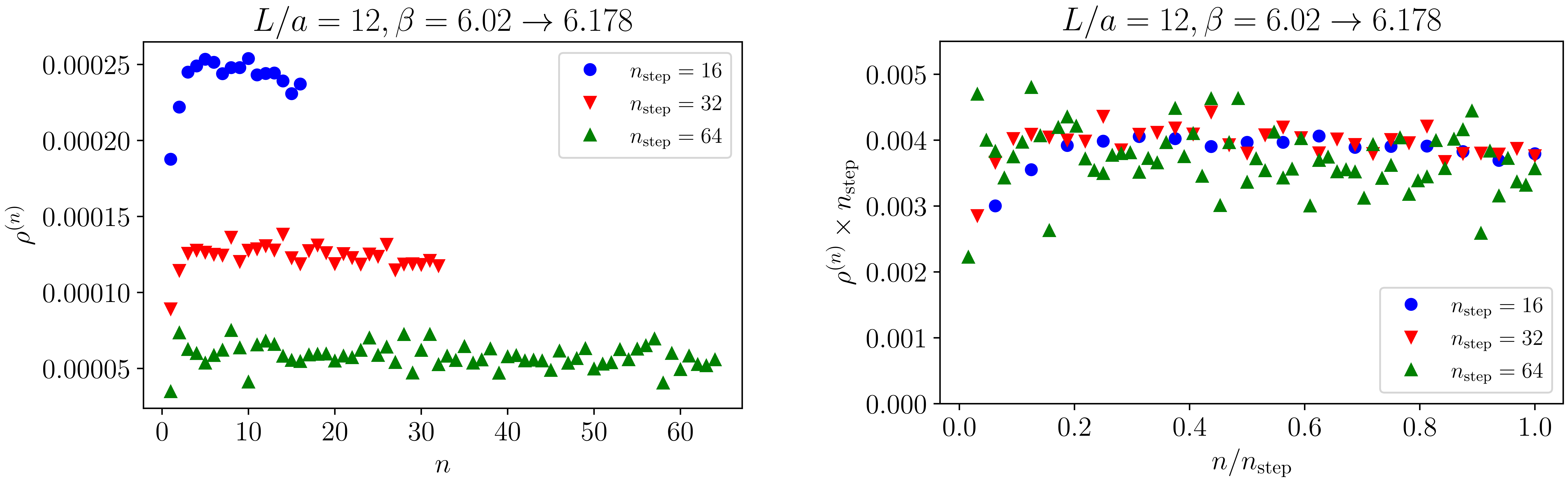
$$\Omega^{(n)}_\mu(x) = C^{(n)}_\mu(x)U^\dagger_\mu(x)$$

$$C^{(n)}_\mu(x) = \sum_{\nu\neq\mu} \rho^{(n)}_{\mu\nu}(x)\Big(U_\nu(x)U_\mu(x+\hat{\nu})U^\dagger_\nu(x+\hat{\mu}) + U^\dagger_\nu(x-\hat{\nu})U_\mu(x-\hat{\nu})U_\nu(x-\hat{\nu}+\hat{\mu})\Big)$$

One parameter for each mask, 8 masks.

EXTRAPOLATION OF THE PARAMETERS

We trained the model at fixed volume ad for few values of n_{step}, then we fitted ρ and extrapolate the values for larger steps and volumes.



VOLUME SCALING ESS

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