Fighting Topological Freezing with Stochastic Normalizing Flows

Lorenzo Verzichelli

based on [Bonanno *et al.* 2510.25704] + in preparation

In collaboration with Claudio Bonanno, Andrea Bulgarelli, Elia Cellini, Alessandro Nada, Dario Panfalone, Davide Vadacchino

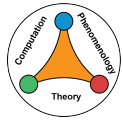
Università di Torino & INFN sezione di Torino



UNIVERSITÀ
DI TORINO



Istituto Nazionale di Fisica Nucleare





Trento ECT*, 8 / 9 / 2026

Outline

1. The Setting and the Problem
2. A Flow-based Methodology
3. Scaling and Continuum Limit
4. Conclusions

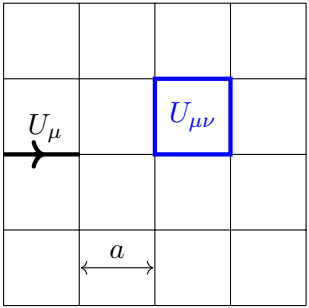
Introduction: Yang–Mills in $D = 3 + 1$

We want to study the **gauge sector** of QCD:
$\mathrm{SU}(3)$ gauge theory in $3 + 1$ space-time dimensions

We consider a discretization on a lattice of spacing a:

$$S = -\beta \sum_{x,\, \mu<\nu} U_{\mu\nu}(x)$$

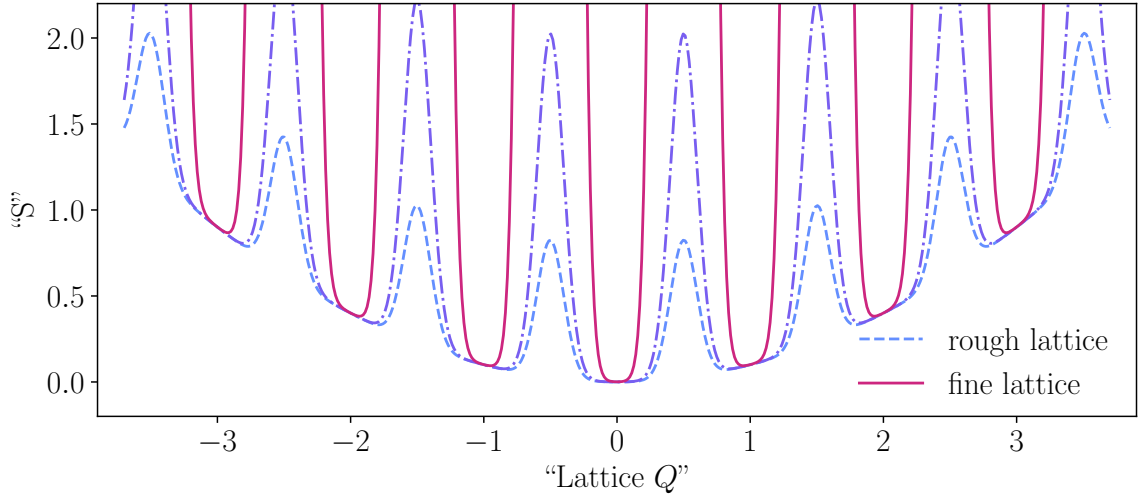
but we are interested in the **continuum limit** $a \to 0$



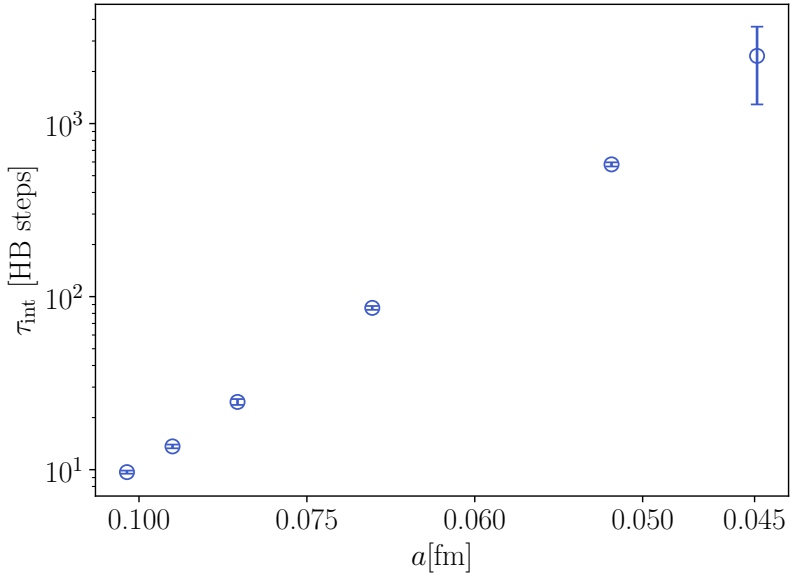
Continuum: (finite-action) configurations are labelled by topological charge $Q \in \mathbb{Z}$
On the lattice: **"potential barriers" separate different topological sectors**

Topological Freezing

In the $a \to 0$ limit: the potential barriers become large
Standard Markov-chain Monte Carlo (MCMC) algorithms get stuck in a sector



Measuring the freezing



τ_{int}: integrated autocorrelation time of Q^2

How many MCMC steps we need to update Q

The error on $\chi \propto \langle Q^2 \rangle$ is proportional to $\sqrt{\tau_{\text{int}}}$

Generative Models

Possible solution: avoid the Markov-chain at least where it freezes

The $D = 1 + 1$, $\mathrm{U}(1)$ **gauge theory** also suffers from topological freezing
Different machine learning approaches have been succesfully applied (examples):

- Diffusion models [Zhu *et al.* 2502.05504],
- Normalizing flows (NF) [Kanwar *et al.* 2003.06413]

Idea: transforming configurations sampled from trivial/tractable distributions

Training similar models for theories in $D = 3 + 1$ is notoriously difficult.

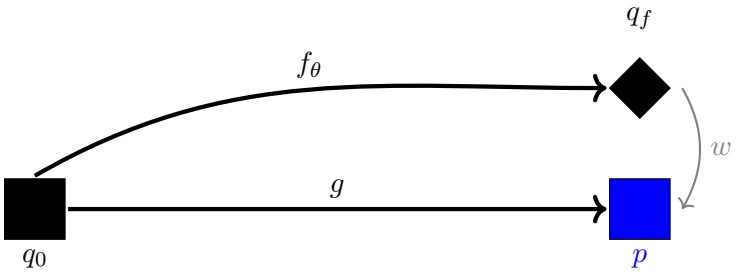
Normalizing Flows

Perfect flow: transformation g mapping atractable q_0 into the target p:

$$p(g(U)) = q_0(U)\,|\mathrm{Det}J_g|^{-1}$$

We use f depending on some parameters θ such that $q_f = q_0\,|\mathrm{Det}J_{f_\theta}|^{-1}$ approximates p and **reweight** by $w = p/q_f$, which is as hard as $\mathrm{Det}J_{f_\theta}$ to compute:

$$\langle \mathcal{O} \rangle_p = \frac{\langle \mathcal{O}\, w \rangle_{q_f}}{\langle w \rangle_{q_f}}$$

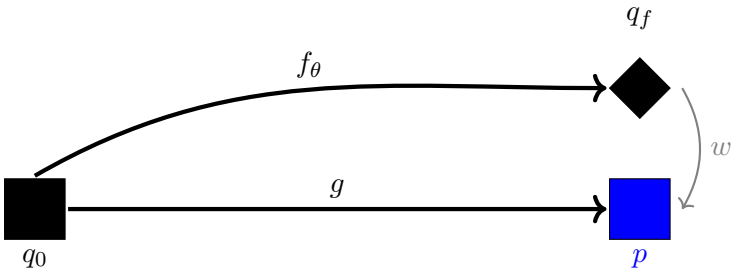


Normalizing Flows

The training optimizes the θ parameters to reduce the noise in the reweighting procedure
Measure the quality of the reweighting:

$$\mathrm{ESS} = \frac{\langle w^2 \rangle_{q_f}}{\langle w \rangle_{q_f}^2} \in [0, 1]$$

Usually the reverse Kullback-Leibler divergence is used as a loss
$\tilde{D}_{\mathrm{KL}}(q_f||p) = \langle \log q_f - \log p \rangle_{q_f}$

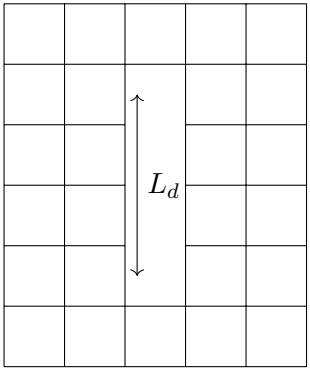


Open Boundary Conditions

Open boundary condition (OBC) solve the problem [Lüscher and Schaefer 1105.4749]
However, **translation invariance is lost**.

Even a "defect" of size ${L_d}^3$ with OBC allows fluctuations of Q.

Objective: retain fast sampling of Q from OBC,
but restore periodic boundary conditions (PBC)
and thus translation invariance



Possible approach: parallel tempering:

- in $\mathbb{C}\mathrm{P}^N$ model [Hasenbusch 1706.04443],
- in Yang–Mills [Bonanno *et al.* 2012.14000]
- in full QCD [Bonanno *et al.* 2404.14151]

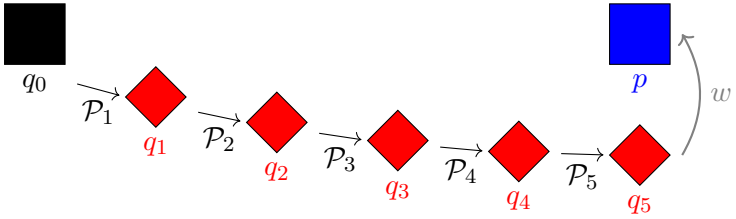
Out of Equilibrium (aka annealing)

Our approach: **out-of-equilibrium trajectories** [Caselle *et al.* 1604.05544]
At each step: we change the update to interpolate between OBC and PBC

Action S changes at each step: **work** done on the system

$$W = \sum_i S_{i+1}\left(U^{(i)}\right) - S_i\left(U^{(i)}\right)$$

Jarzynski equality: $\langle e^{-W} \rangle_{NE} = e^{-\Delta F}$

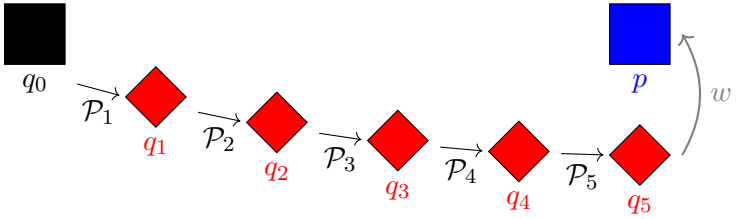


Out of Equilibrium (aka annealing)

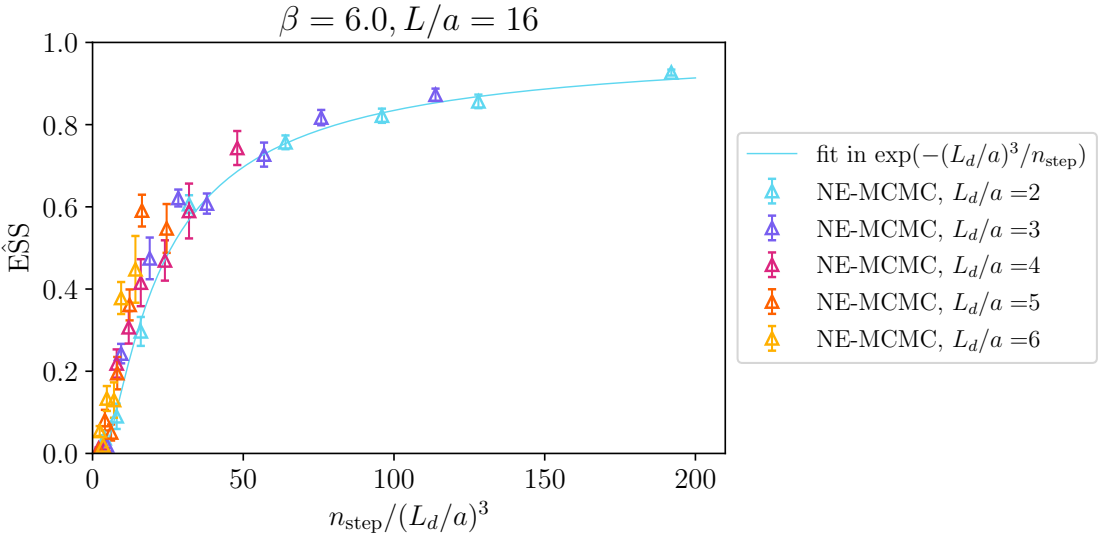
We can compute expectation walues on p (targe, PBC)
as average over out-of-equilibrium trajectories $\langle\cdot\rangle_{NE}$
reweighting by $w = e^{-W}$

$$\langle\mathcal{O}\rangle_p = \frac{\langle\mathcal{O}\, e^{-W}\rangle_{NE}}{\langle e^{-W}\rangle_{NE}}$$

Limit of large number of steps: trajectories at equilibrium, so $W = \Delta F$ has no variance

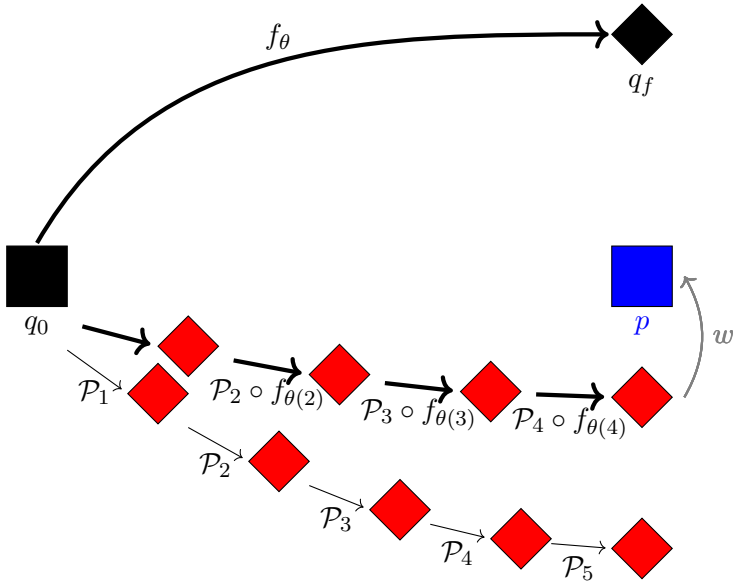


Scaling the Defect Size: NE-MCMC



See also [Bulgarelli *et al.* 2412.00200]

Stochastic Normalizing Flow



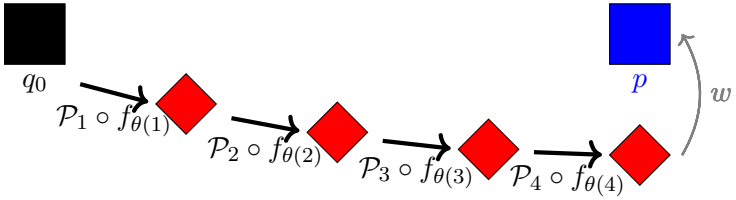
Stochastic Normalizing Flow

Deterministic transformations (flows) inserted to improve the reweighting
Now the work is defined by

$$\sum_i \left[S_{i+1}\left(U^{(i)}\right) - S_i\left(U^{(i)}\right) + \mathrm{Det}J_i \right]$$

so that the reweighting is still exact:

$$\langle \mathcal{O} \rangle_p = \frac{\langle \mathcal{O}\, e^{-W} \rangle_{NE}}{\langle e^{-W} \rangle_{NE}}$$



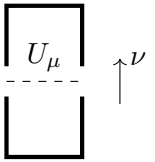
Overview of the Architecture

Our flow use very simple layers **acting on the links of the defect and those adjacent**.
Physically they reduce the fluctuations of the newly formed links with an
"analyitcal smearing" [Morningstar and Peardon hep-lat/0311018]

$$U_\mu \to e^{i\,T(U_\mu\,\Gamma)} U_\mu$$

where

$$\Gamma = \sum_{\nu \neq \mu} \rho(\dots)$$

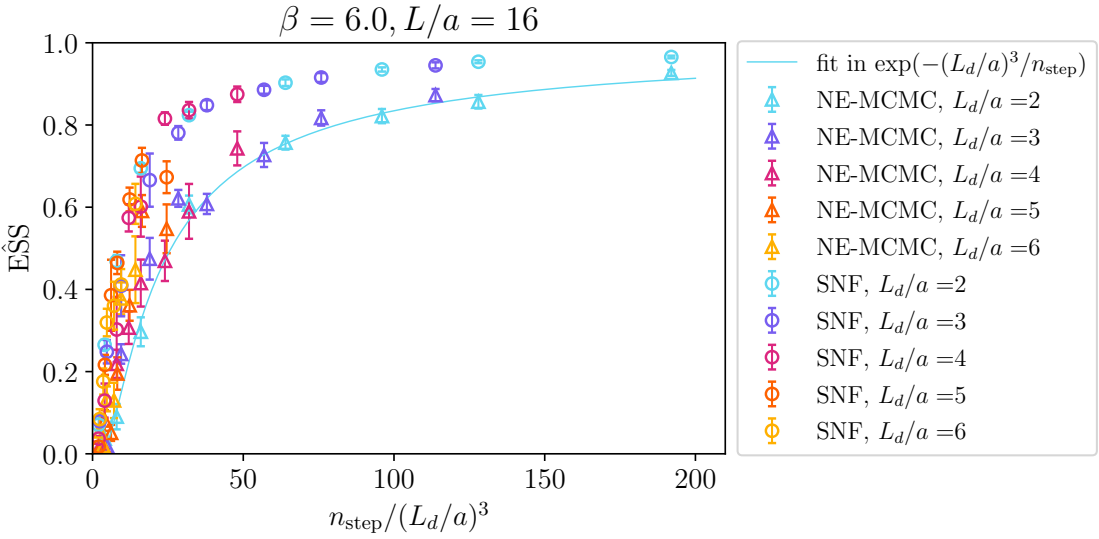


[Nagai and Tomiya 2103.11965] for neural network applications

Probably this function is already implemented in your favourite code!

The **cost of learning** the ρ parameters is **negligible**
Fully tranferable to different number of steps and lattice volumes

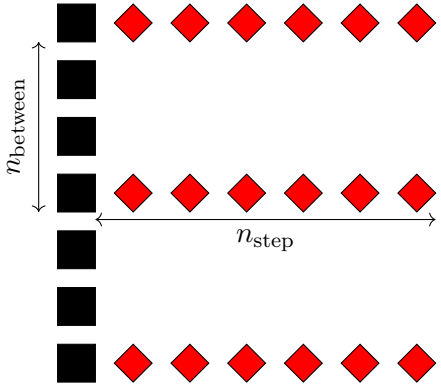
Scaling the Defect Size: SNF



Predicting the cost towords the continuum limit

Number of updates to have a decorrelated configuration:

$$\frac{2\,\tau_{\text{int}}}{\text{ESS}} \times (n_{\text{step}} + n_{\text{between}})$$



Keep τ_{int} constant
with fixed physical L_d:
$n_{\text{between}} \sim a^{-2}$

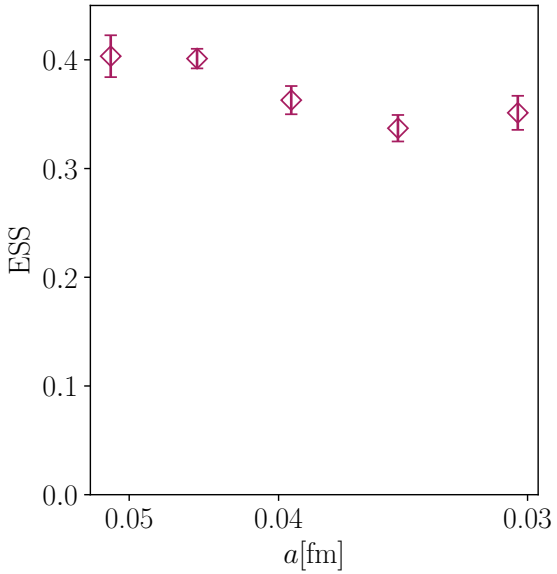
Keep ESS constant:
$n_{\text{step}} \sim a^{-3}$

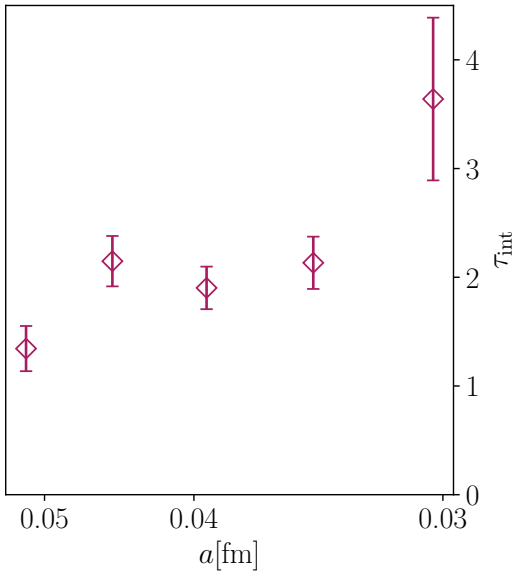
Continuum Limit: simulation parameters

We run simulations at fine lattice spacing and computed the topological susceptibility χ

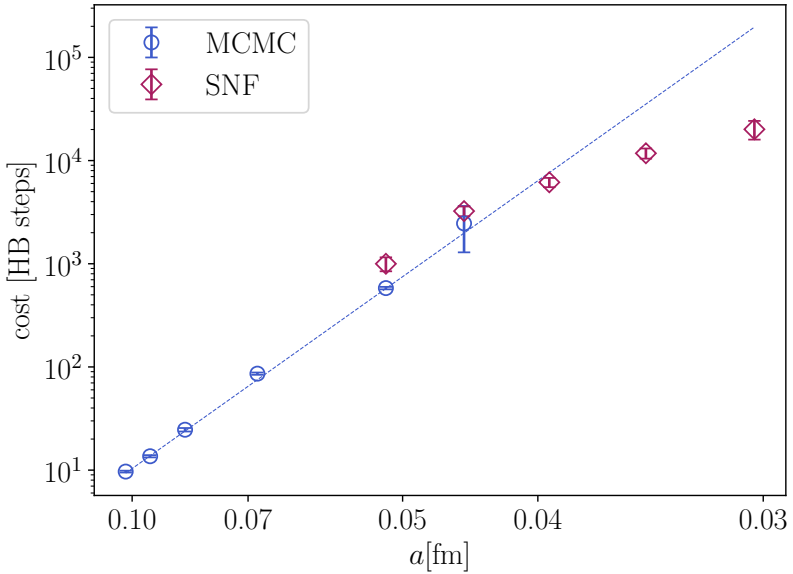
β	a[fm]	V [fm^4]	L_d [fm]	n_{step}	χ [MeV^4]
6.4	0.051594(52)	$(1.55)^4$	0.155	200	$(184.9(2.8))^4$
6.5	0.044900(40)	$(1.53)^4$	0.180	475	$(188.0(2.0))^4$
6.6	0.039333(31)	$(1.49)^4$	0.197	975	$(187.1(2.2))^4$
6.7	0.034489(37)	$(1.52)^4$	0.207	1600	$(184.3(2.1))^4$
6.8	0.030297(36)	$(1.51)^4$	0.182	1600	$(193.8(4.5))^4$

Monitoring the ESS and τ_{int}





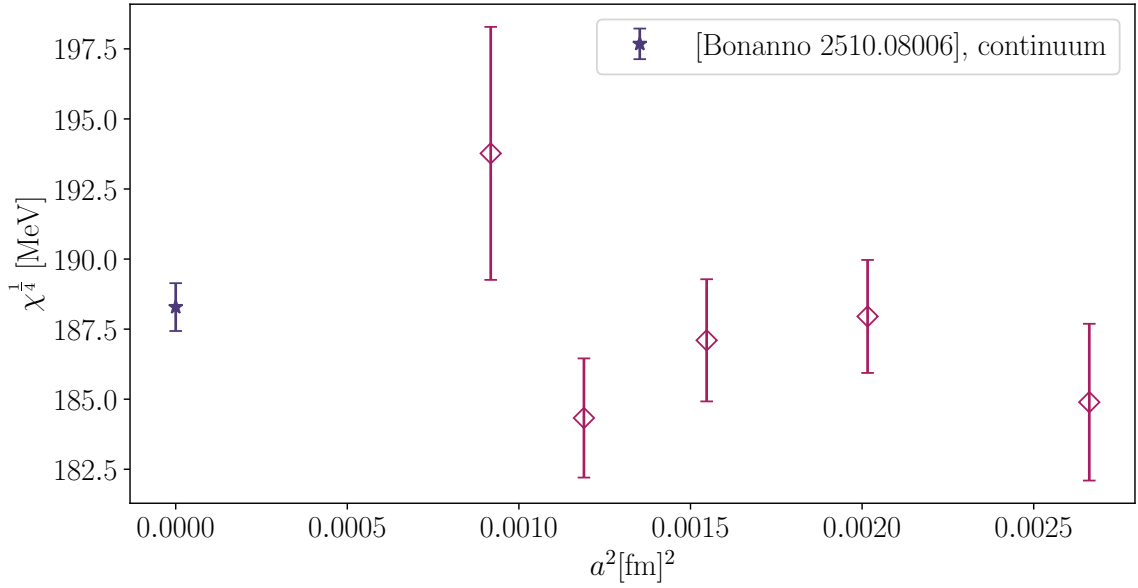
Comparison of Real Cost



Real cost for one decorrelated configuration

- τ_{int} for MCMC
- $\tau_{\text{int}}/\text{ESS}$ times $(n_{\text{between}} + n_{\text{step}})$ for SNF

The Topological Susceptibility: results



Conclusions

- SNF are a suitable tool to fight topological freezing
- They are a combination of out-of-equilibrium Monte Carlo with normalizing flows
- They retain the scalability of the Monte Carlo (stochastic) methods
- The Normalizing Flow (deterministic) can systematically improve the performance
- The precision on χ defines a benchmark for topological freezing

Thank you for your attention!

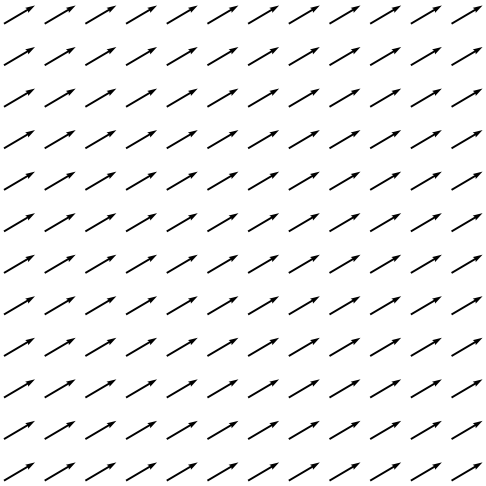


C production code

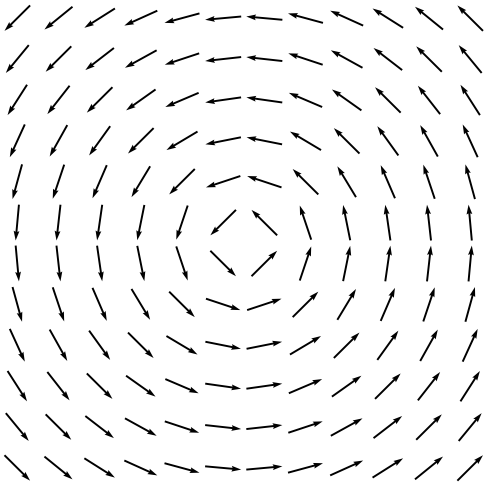


Python prototypes

Instanton (2D cartoons)

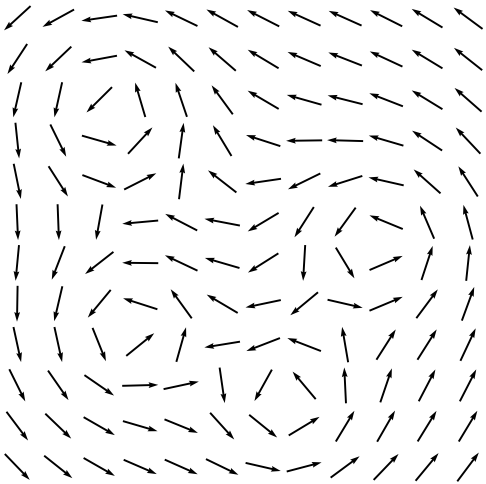


No instantons ($Q = 0$)

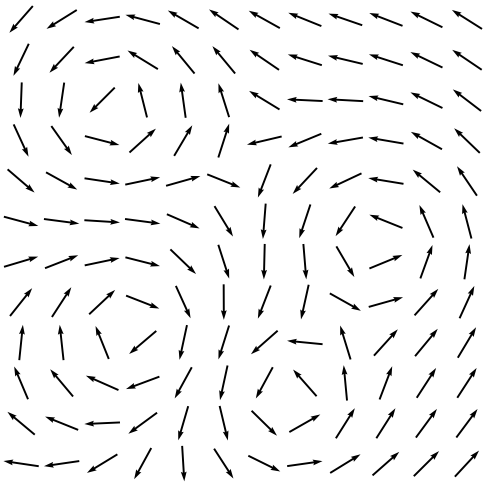


One Instanton ($Q = 1$)

Gas of Instantons (2D cartoons)



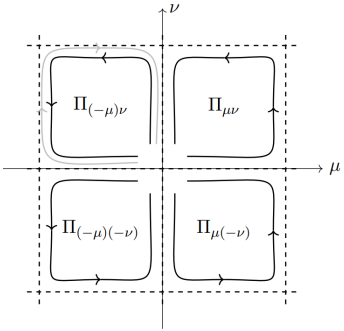
$Q = 4$



$Q = 2$

Discretiztion of Q

Q looks like $\int \epsilon_{\mu\nu\rho\sigma} F_{\mu\nu} F_{\rho\sigma}$
On the lattice: we want it exactly CP odd, thus **clover discretization**



In order to remove the fluctuations we performed 60 to 80 (at the finest a) **cooling steps**
Cooling is a local update that reduces the action
It brings the configuration close to the minimum of the action in the topological sector

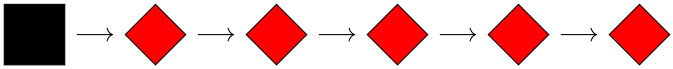
Out of Equilibrium MC

System with action $S_\lambda[\phi]$ explicitly depending on a parameter λ

For each value of λ, can device a transition probability $\mathcal{P}_\lambda(\phi \to \phi')$

We can lead the system **out-of-equilibrium** changing λ during the Markov Chain:

$$\phi_0 \xrightarrow{\mathcal{P}_{\lambda(1)}} \phi_1 \xrightarrow{\mathcal{P}_{\lambda(2)}} \phi_2 \xrightarrow{\mathcal{P}_{\lambda(3)}} \dots \xrightarrow{\mathcal{P}_{\lambda(n_{\text{step}})}} \phi_{n_{\text{step}}}$$



We will assume detailed balance to hold for each $\mathcal{P}_\lambda$

Heat and Work

In general:

$$S_{\lambda(0)}[\phi_0] \neq S_{\lambda(n_{\text{step}})}[\phi_{n_{\text{step}}}]$$

We will call **Work**[1] the change in the action due to change in parameters:

$$\delta W_k = S_k[\phi_{k-1}] - S_{k-1}[\phi_{k-1}]$$

And **Heat** the change in the action due to change of the configuration:

$$\delta Q_k = S_k[\phi_k] - S_k[\phi_{k-1}]$$

So that

$$\Delta S = \sum_{k=1}^{n_{\text{step}}} \left(\delta W_k + \delta Q_k\right) = W + Q$$

$$\phi_k \xrightarrow{\mathcal{P}_{\lambda(k+1)}} \phi_{k+1}$$

[1] done on the system

Jarzynski Equality

$$\langle e^{-W} \rangle_{\mathrm{f}} = e^{-\Delta F}$$

[Jarzynski, 1996]

$\langle \cdot \rangle_{\mathrm{f}}$ is the average over **forward trajectories**

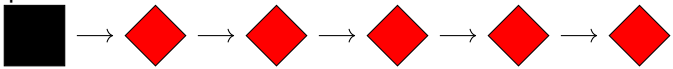
ΔF is the **free energy difference**
between $\lambda(0)$ and $\lambda(n_{\mathrm{step}})$

Let the **prior** be $q_0 \sim e^{-S_{\lambda(0)}}$
And the **target** be $p \sim e^{-S_{\lambda(n_{\mathrm{step}})}}$

ΔF is the difference in free energy between
the q_0 and p **equilibrium distribution**

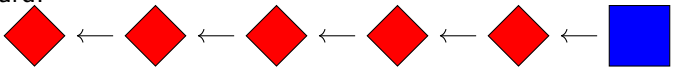
Forward and Reverse Process

Besides the forward process



$$P_{\mathrm{f}} = q_0(\phi_0) \prod_{k=1}^{n_{\mathrm{step}}} \mathcal{P}_{\lambda(k)}(\phi_{k-1} \to \phi_k)$$

Consider the backward:



$$P_{\mathrm{r}} = p(\phi_{n_{\mathrm{step}}}) \prod_{k=1}^{n_{\mathrm{step}}} \mathcal{P}_{\lambda(k)}(\phi_k \to \phi_{k-1})$$

Crooks Fluctuation Theorem

Let us compute

$$\log \frac{P_\mathrm{f}}{P_\mathrm{r}} = \log \frac{q_0(\phi_0)}{p(\phi_{n_\mathrm{step}})} \sum_{k=1}^{n_\mathrm{step}} \log \frac{\mathcal{P}_{\lambda(k)}(\phi_{k-1} \to \phi_k)}{\mathcal{P}_{\lambda(k)}(\phi_k \to \phi_{k-1})} = S_{n_\mathrm{step}}(\phi_{n_\mathrm{step}}) - S_0(\phi_0) - \Delta F - Q$$

since

$$\sum_{k=1}^{n_\mathrm{step}} \log \frac{\mathcal{P}_{\lambda(k)}(\phi_{k-1} \to \phi_k)}{\mathcal{P}_{\lambda(k)}(\phi_k \to \phi_{k-1})} = \sum_{k=1}^{n_\mathrm{step}} (S_k(\phi_k) - S_k(\phi_{k-1})) = Q$$

by detailed balance [Crooks, 1999]

We call the difference $S_{n_\mathrm{step}}(\phi_{n_\mathrm{step}}) - S_0(\phi_0) - Q - \Delta F$ **dissipated work**:

$$\log \frac{P_\mathrm{f}}{P_\mathrm{r}} = W - \Delta F = W_d$$

Unbiased Sampling and Proving Jarzynski

$$P_{\mathrm{r}} = P_{\mathrm{f}}\, e^{-W_d} = P_{\mathrm{f}}\, e^{\Delta F - W}$$

We would like to sample the target p:

$$\langle \mathcal{O} \rangle_p = \int \mathcal{D}\phi\, p(\phi)\, \mathcal{O}(\phi) = \int \prod_{k=0}^{n_{\mathrm{step}}} \mathcal{D}\phi_k\, P_{\mathrm{r}}\, \mathcal{O}(\phi_{n_{\mathrm{step}}}) = \int \prod_{k=0}^{n_{\mathrm{step}}} \mathcal{D}\phi_k\, P_{\mathrm{f}}\, e^{-W_d}\, \mathcal{O}(\phi_{n_{\mathrm{step}}})$$

or shortly

$$\langle \mathcal{O} \rangle_p = \langle e^{-W_d}\, \mathcal{O}(\phi_{n_{\mathrm{step}}}) \rangle_{\mathrm{f}}$$

Take the special case $\mathcal{O} = 1$:

$$1 = \langle e^{-W_d} \rangle_{\mathrm{f}} \qquad \text{or} \qquad e^{-\Delta F} = \langle e^{-W} \rangle_{\mathrm{f}}$$

Metrics of Jarzynski: the Dissipated Work

$$\langle \mathcal{O} \rangle_p = \langle e^{-W_d} \, \mathcal{O}(\phi_{n_{\text{step}}}) \rangle_{\text{f}}$$

looks dangerously like a **reweighting formula**...

We need to make sure W_d does not grow with the "size" of the system

Let us make an observation:

$$0 \leq D_{\text{KL}}(P_{\text{f}} || P_{\text{r}}) = \langle \log \frac{P_{\text{f}}}{P_{\text{r}}} \rangle_{\text{f}} = \langle W_d \rangle_{\text{f}}$$

Where $D_{\text{KL}}(q||q')$ is the **Kullback-Leibler divergence**, greater or equal to 0
It is a measure of how much two distributions differ

This also shows $\langle W \rangle_{\text{f}} \geq \Delta F$
(Second law of thermodynamics)

Calsses of Smearing Parameters

class	$U_\mu(x)$ on defect	staple contains defect link	# other staples containing defect link
t1b	Yes	No	5
t1e	Yes	No	4
t1c	Yes	No	3
t2	Yes	Yes	5
t2b	Yes	Yes	4
t2e	Yes	Yes	3
t2c	Yes	Yes	2
sp	No ($\mu \neq \hat{0}$)	Yes	-
ex	No ($\mu = \hat{0}$)	Yes	-

Classes of Smearing Parameters

