Flow-based approaches for lattice gauge theory: scaling properties of a non-equilibrium approach

Alessandro Nada

Università degli Studi di Torino and INFN, Sezione di Torino

2nd LatticeNET workshop on challenges in Lattice field theory

Benasque, 30th March - 5th April 2025

Critical slowing down

(Thermalized) Markov Chain: elegant and *scalable* numerical solution to generate U according to $p(U)$

$$\underbrace{U^{(0)} \xrightarrow{P_p} U^{(1)} \xrightarrow{P_p} \dots \xrightarrow{P_p}}_{\text{thermalization}} \underbrace{U^{(t)} \xrightarrow{P_p} U^{(t+1)} \xrightarrow{P_p} \dots \rightarrow U^{(t+n)}}_{\text{equilibrium}}$$

Critical slowing down

(Thermalized) Markov Chain: elegant and *scalable* numerical solution to generate U according to $p(U)$

$$\underbrace{U^{(0)} \xrightarrow{P_p} U^{(1)} \xrightarrow{P_p} \dots \xrightarrow{P_p}}_{\text{thermalization}} \underbrace{U^{(t)} \xrightarrow{P_p} U^{(t+1)} \xrightarrow{P_p} \dots \to U^{(t+n)}}_{\text{equilibrium}}$$

Configurations sampled sequentially are **autocorrelated**

$$\dots \to U^{(t)} \to U^{(t+1)} \to \dots \to U^{(t+n)}$$

Autocorrelation measured by $\tau_{\mathrm{int}}(\mathcal{O})$ $\to$ effective independent configurations $= n/2\tau_{\mathrm{int}}(\mathcal{O})$

(Thermalized) Markov Chain: elegant and *scalable* numerical solution to generate U according to $p(U)$

$$\underbrace{U^{(0)} \xrightarrow{P_p} U^{(1)} \xrightarrow{P_p} \dots \xrightarrow{P_p}}_{\text{thermalization}} \underbrace{U^{(t)} \xrightarrow{P_p} U^{(t+1)} \xrightarrow{P_p} \dots \to U^{(t+n)}}_{\text{equilibrium}}$$

Configurations sampled sequentially are **autocorrelated**

$$\dots \to U^{(t)} \to U^{(t+1)} \to \dots \to U^{(t+n)}$$

Autocorrelation measured by $\tau_{\text{int}}(\mathcal{O})$ $\to$ effective independent configurations $= n/2\tau_{\text{int}}(\mathcal{O})$

Critical slowing down

When a critical point is approached τ_{int} diverges

E.g. in the continuum limit $a \to 0$

$$\tau_{\text{int}}(\mathcal{O}) \sim a^{-z}$$

where z depends on the algorithm and on the observable under study

What if every new configuration is sampled independently from the previous one **by construction**?

Flow-based approach

find an **exact** mapping between some well-behaving distribution q_0 and the target p

$\rightarrow$ promising approach to fight critical slowing down?

Re-framing lattice simulations: flowing from one distribution to the other

What if every new configuration is sampled independently from the previous one **by construction**?

Flow-based approach

find an **exact** mapping between some well-behaving distribution q_0 and the target p

$\rightarrow$ promising approach to fight critical slowing down?

It works if (a bit roughly):

- the map can be constructed with reasonable computational effort
- the map itself is not too expensive to sample from

Trivializing flows

Original approach **[Lüscher; 0907.5491]**: take standard path integral

$$\langle O \rangle = \frac{1}{Z} \int \mathcal{D}\phi \; \mathcal{O}(\phi) \; e^{-S(\phi)}$$

and perform an invertible field transformation $\tilde{\phi} = \mathcal{F}^{-1}(\phi)$

$$\langle O \rangle = \frac{1}{Z} \int \mathcal{D}\tilde{\phi} \; \mathcal{O}(\mathcal{F}(\tilde{\phi})) \; e^{-S(\mathcal{F}(\tilde{\phi})) + \log \det J(\mathcal{F}(\tilde{\phi}))}$$

$\mathcal{F}$ is a true trivializing map if

$$S(\mathcal{F}(\tilde{\phi})) - \log \det J(\mathcal{F}(\tilde{\phi})) = \text{const}$$

Trivializing flows

Original approach **[Lüscher; 0907.5491]**: take standard path integral

$$\langle O \rangle = \frac{1}{Z} \int \mathcal{D}\phi \; \mathcal{O}(\phi) \; e^{-S(\phi)}$$

and perform an invertible field transformation $\tilde{\phi} = \mathcal{F}^{-1}(\phi)$

$$\langle O \rangle = \frac{1}{Z} \int \mathcal{D}\tilde{\phi} \; \mathcal{O}(\mathcal{F}(\tilde{\phi})) \; e^{-S(\mathcal{F}(\tilde{\phi})) + \log \det J(\mathcal{F}(\tilde{\phi}))}$$

$\mathcal{F}$ is a true trivializing map if

$$S(\mathcal{F}(\tilde{\phi})) - \log \det J(\mathcal{F}(\tilde{\phi})) = \text{const}$$

- trivializing map can be constructed with a flow equation
- approximate maps built expressing the action in the flow equation as a power series in the flow time and truncated
- however: no big improvement in the scaling of the computational cost **[Engel and Schaefer; 1102.1852]**

Normalizing flows: the basics

A normalizing flow is an invertible mapping f_θ constructed as

$$\phi = f_\theta(z) = (f_N \circ \cdots \circ f_1)(z) \qquad\qquad z \sim q_0$$

and the transformed variable follows the distribution

$$q(\phi) = q_0(f_\theta^{-1}(\phi))\,|\det J_\theta|^{-1}$$

Note that in the notation of trivializing maps $\mathcal{F}^{-1} = f_\theta$

Normalizing flows: the basics

A normalizing flow is an invertible mapping f_θ constructed as

$$\phi = f_\theta(z) = (f_N \circ \cdots \circ f_1)(z) \qquad\qquad z \sim q_0$$

and the transformed variable follows the distribution

$$q(\phi) = q_0(f_\theta^{-1}(\phi))\, |\det J_\theta|^{-1}$$

Note that in the notation of trivializing maps $\mathcal{F}^{-1} = f_\theta$

Inserting it in the path integral **[Albergo, Kanwar, Shanahan; 1904.12072]** we get

reweighting-like formula

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \mathcal{D}\phi \, \underbrace{q(\phi)}_{\text{sample}} \, \underbrace{\overbrace{\frac{e^{-S(\phi)}}{q(\phi)}}^{\tilde{w}} \mathcal{O}(\phi)}_{\text{measure}} = \frac{\langle \mathcal{O}(\phi)\tilde{w}(\phi)\rangle_{\phi\sim q}}{\langle \tilde{w}(\phi)\rangle_{\phi\sim q}}$$

independent Metropolis-Hastings

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \mathcal{D}\phi \, \underbrace{q(\phi) \frac{e^{-S(\phi)}}{q(\phi)}}_{\text{sample+MH}} \, \underbrace{\mathcal{O}(\phi)}_{\text{measure}}$$

Normalizing flows: the basics

A normalizing flow is an invertible mapping f_θ constructed as

$$\phi = f_\theta(z) = (f_N \circ \cdots \circ f_1)(z) \qquad\qquad z \sim q_0$$

and the transformed variable follows the distribution

$$q(\phi) = q_0(f_\theta^{-1}(\phi))\,|\det J_\theta|^{-1}$$

Note that in the notation of trivializing maps $\mathcal{F}^{-1} = f_\theta$

Inserting it in the path integral **[Albergo, Kanwar, Shanahan; 1904.12072]** we get

reweighting-like formula

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \mathcal{D}\phi \, \underbrace{q(\phi)}_{\text{sample}} \underbrace{\overbrace{\frac{e^{-S(\phi)}}{q(\phi)}}^{\tilde{w}} \mathcal{O}(\phi)}_{\text{measure}} = \frac{\langle \mathcal{O}(\phi)\tilde{w}(\phi)\rangle_{\phi\sim q}}{\langle \tilde{w}(\phi)\rangle_{\phi\sim q}}$$

independent Metropolis-Hastings

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \mathcal{D}\phi \, \underbrace{q(\phi) \frac{e^{-S(\phi)}}{q(\phi)}}_{\text{sample+MH}} \underbrace{\mathcal{O}(\phi)}_{\text{measure}}$$

The partition function is directly accessible! **[Nicoli et al.; 2007.07115]**

$$Z = \langle \tilde{w}(\phi) \rangle_{\phi\sim q}$$

or the ratio Z/Z_0 if you don't know q_0

Normalizing flows: structure and training

Coupling layers

NFs organized as combination of discrete transformations with suitable masking $\rightarrow$ invertibility and triangular Jacobian

$$f_n : \begin{cases} \phi^{n+1}_{\mathrm{frozen}} = \phi^{n}_{\mathrm{frozen}} \\ \phi^{n+1}_{\mathrm{active}} = e^{-s(\phi^{n}_{\mathrm{frozen}})} \phi^{n}_{\mathrm{active}} + t(\phi^{n}_{\mathrm{frozen}}) \end{cases}$$

Coupling layers

NFs organized as combination of discrete transformations with suitable masking → invertibility and triangular Jacobian

$$f_n : \begin{cases} \phi^{n+1}_{\mathrm{frozen}} = \phi^{n}_{\mathrm{frozen}} \\ \phi^{n+1}_{\mathrm{active}} = e^{-s(\phi^{n}_{\mathrm{frozen}})} \phi^{n}_{\mathrm{active}} + t(\phi^{n}_{\mathrm{frozen}}) \end{cases}$$

s and t are the output of neural networks

Coupling layers

NFs organized as combination of discrete transformations with suitable masking → invertibility and triangular Jacobian

$$f_n : \begin{cases} \phi^{n+1}_{\text{frozen}} = \phi^{n}_{\text{frozen}} \\ \phi^{n+1}_{\text{active}} = e^{-s(\phi^{n}_{\text{frozen}})} \phi^{n}_{\text{active}} + t(\phi^{n}_{\text{frozen}}) \end{cases}$$

s and t are the output of neural networks

Training

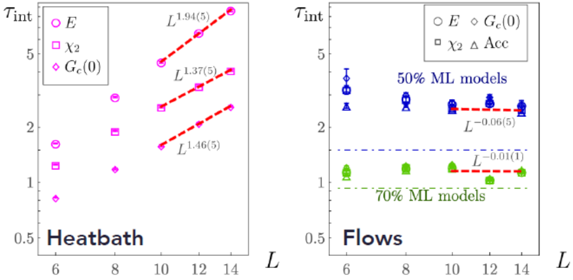
Parameters "trained" with a minimization procedure of a "loss", usually taken to be the Kullback-Leibler divergence

$$\tilde{D}_{\text{KL}}(q\|p) = \int \mathcal{D}\phi \, q(\phi) \log \frac{q(\phi)}{p(\phi)} = -\langle \log \tilde{w} \rangle_{\phi \sim q} + \log Z$$

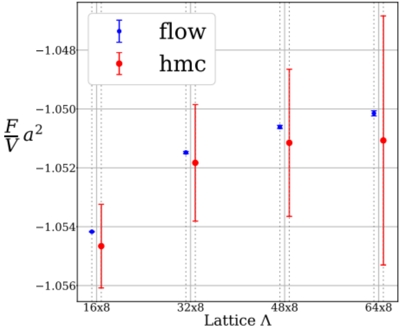
It is a self-learning procedure, no data from target required

Natural toy model: ϕ^4 scalar field theory in 2 dimensions

[Albergo, Kanwar, Shanahan; 1904.12072]



[Nicoli et al.; 2007.07115]



but also **[Del Debbio et al.; 2105.12481] [Caselle, Cellini, AN; 2201.08862] [Gerdes et al.; 2207.00283] [Albandea et al.; 2302.08408]** + . . .

Normalizing flows for lattice gauge theory

Incorporating the symmetries in the flow crucial for an efficient training

Construct equivariant flows:

$$f_\theta(t \cdot \phi) = t \cdot f_\theta(\phi)$$

Incorporating the symmetries in the flow crucial for an efficient training

Construct equivariant flows:

$$f_\theta(t \cdot \phi) = t \cdot f_\theta(\phi)$$

What about gauge theories?

Incorporating the symmetries in the flow crucial for an efficient training

Construct equivariant flows:

$$f_\theta(t \cdot \phi) = t \cdot f_\theta(\phi)$$

What about gauge theories?

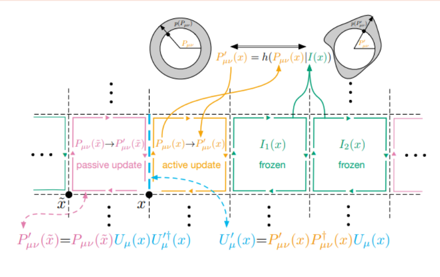
Gauge-equivariant flows

Some architectures developed in the last few years:

- loop-level flow: coupling layers act on (eigenvalues of) untraced loops **[Kanwar et al.; 2003.06413] [Boyda et al.; 2008.05456]**

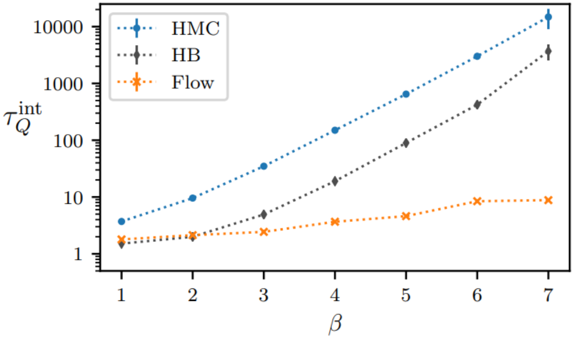
$$P'_{\mu\nu}(x) = h(P_{\mu\nu}(x)|I(x)) \qquad U'_\mu(x) = P'_{\mu\nu}(x)P^\dagger_{\mu\nu}(x)U_\mu(x)$$

- link-level flow (more on that later in the talk)
- also Continuous NFs (next slide)

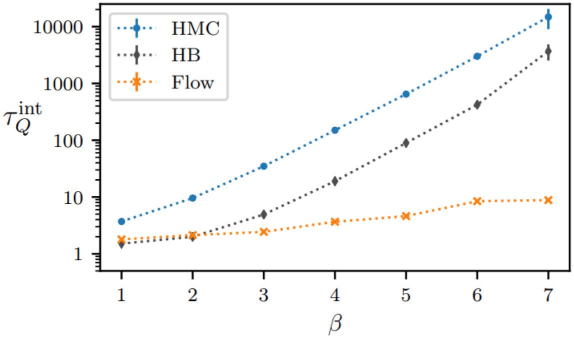


The path until now - 2

U(1) gauge theory in 2d **[Kanwar et al.; 2003.06413]**



U(1) gauge theory in 2d **[Kanwar et al.; 2003.06413]**



SU(N) in 2 dimensions **[Boyda et al.; 2008.05456] [Bacchio et al.; 2212.08469] [Gerdes et al.; 2410.13161]**

but also fermionic theories **[Albergo et al.; 2106.05934]**

Schwinger model **[Finkenrath et al.; 2201.02216] [Albergo et al.; 2202.11712]**

SU(N) with fermions in 2 dimensions **[Abbott et al.; 2207.08945] [Abbott et al.; 2211.07541]**

Wilson flow approach **[Bacchio et al.; 2212.08469]**

no discrete NF, learn the ODE instead

$$\dot{U}_t = Z_t(U_t)U_t$$

with Z_t given by the force of

$$\tilde{S}(U_t, t) = \sum c_i(t)\mathcal{W}_i(U_t)$$

Wilson flow approach **[Bacchio et al.; 2212.08469]**

no discrete NF, learn the ODE instead

$$\dot{U}_t = Z_t(U_t)U_t$$

with Z_t given by the force of

$$\tilde{S}(U_t, t) = \sum c_i(t)\mathcal{W}_i(U_t)$$

Machine-learn the c_i coefficients!

Very few parameters, efficient in 2d $\mathrm{SU}(3)$ gauge theory

related work: Continuous NFs **[Gerdes et al.; 2207.00283]**, **[Caselle et al.; 2307.01107]**, **[Gerdes et al.; 2410.13161]**

FlowHMC **[Albandea et al.; 2302.08408]**

very close to the original algorithm by Lüscher

$$\langle O \rangle = \frac{1}{Z} \int \mathcal{D}\tilde{\phi}\, \mathcal{O}(f_\theta^{-1}(\tilde{\phi}))\, \underbrace{e^{-S(f_\theta^{-1}(\tilde{\phi})) + \log \det J(f_\theta^{-1}(\tilde{\phi}))}}_{\text{sample}}$$

→ first run an HMC using $S(f_\theta^{-1}(\tilde{\phi})) - \log \det J(f_\theta^{-1}(\tilde{\phi}))$

→ then flow back to the target distribution using f_θ

No MH step is needed + decorrelation obtained (in scalar field theory) even with little overlap with target distribution

However sampling is expensive with deep networks

The path until now - 3

ϕ^4 scalar field theory in 2 dimensions: **[Albergo, Kanwar, Shanahan; 1904.12072] [Nicoli et al.; 2007.07115] [Del Debbio et al.; 2105.12481] [Caselle, Cellini, AN; 2201.08862] [Gerdes et al.; 2207.00283] [Albandea et al.; 2302.08408]**

$\mathrm{U}(1)$ in 2 dimensions: **[Kanwar et al.; 2003.06413] [Singha et al.; 2306.00581]**

$\mathrm{SU}(N)$ in 2 dimensions **[Boyda et al.; 2008.05456] [Bacchio et al.; 2212.08469] [Gerdes et al.; 2410.13161]**

fermionic theories **[Albergo et al.; 2106.05934]**

Schwinger model **[Finkenrath et al.; 2201.02216] [Albergo et al.; 2202.11712]** and $\mathrm{SU}(N)$ gauge theories with fermions in 2 dimensions **[Abbott et al.; 2207.08945]**

first attempts in 4d **[Abbott et al.; 2305.02402]** with interesting applications **[Abbott et al.; 2401.10874]**

The path until now - 3

ϕ^4 scalar field theory in 2 dimensions: **[Albergo, Kanwar, Shanahan; 1904.12072] [Nicoli et al.; 2007.07115] [Del Debbio et al.; 2105.12481] [Caselle, Cellini, AN; 2201.08862] [Gerdes et al.; 2207.00283] [Albandea et al.; 2302.08408]**

$\mathrm{U}(1)$ in 2 dimensions: **[Kanwar et al.; 2003.06413] [Singha et al.; 2306.00581]**

$\mathrm{SU}(N)$ in 2 dimensions **[Boyda et al.; 2008.05456] [Bacchio et al.; 2212.08469] [Gerdes et al.; 2410.13161]**

fermionic theories **[Albergo et al.; 2106.05934]**

Schwinger model **[Finkenrath et al.; 2201.02216] [Albergo et al.; 2202.11712]** and $\mathrm{SU}(N)$ gauge theories with fermions in 2 dimensions **[Abbott et al.; 2207.08945]**

first attempts in 4d **[Abbott et al.; 2305.02402]** with interesting applications **[Abbott et al.; 2401.10874]**

Very efficient samplers in lower-dimensional theories!

Scaling of flow-based samplers?

How do flow-based samplers scale when the d.o.f. of the system are increased? (i.e., in the continuum limit)

Scaling of flow-based samplers?

How do flow-based samplers scale when the d.o.f. of the system are increased? (i.e., in the continuum limit)

We require two features:

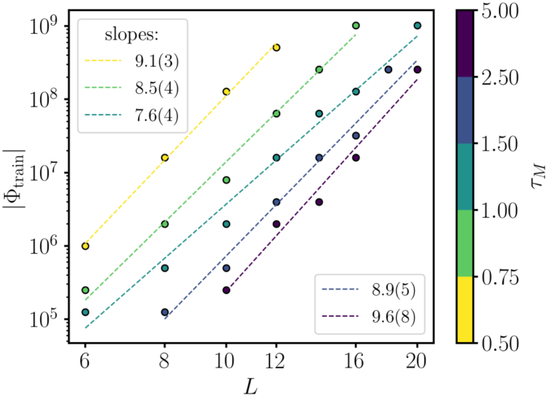
- improved scaling of the sampling with respect to MCMC
- good scaling of the training phase

Scaling of flow-based samplers?

How do flow-based samplers scale when the d.o.f. of the system are increased? (i.e., in the continuum limit)

We require two features:

- improved scaling of the sampling with respect to MCMC
- good scaling of the training phase



Early work: worrying scaling of training cost for fixed models

image from **[Del Debbio et al.; 2105.12481]**

Scaling of flow-based samplers?

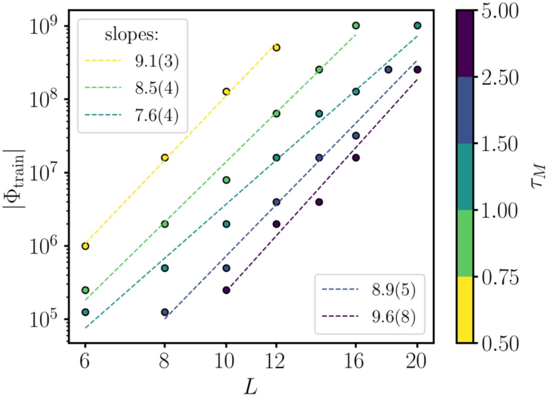
How do flow-based samplers scale when the d.o.f. of the system are increased? (i.e., in the continuum limit)

We require two features:

- improved scaling of the sampling with respect to MCMC
- good scaling of the training phase

Difficult to determine generally! scaling has to be assessed experimentally **[Abbott et al.; 2211.07541]**

see also multiscale NFs **[Abbott et al.; 2404.10819]**



Early work: worrying scaling of training cost for fixed models

image from **[Del Debbio et al.; 2105.12481]**

The rest of this talk: same framework, different approach

As training can be very expensive, can we use a flow-based sampler that requires none + has a clear scaling?

The rest of this talk: same framework, different approach

As training can be very expensive, can we use a flow-based sampler that requires none + has a clear scaling?

Non-equilibrium MCMC

Works like a flow, but purely stochastic

- No training required
- Clear scaling with d.o.f.

The rest of this talk: same framework, different approach

As training can be very expensive, can we use a flow-based sampler that requires none + has a clear scaling?

Non-equilibrium MCMC

Works like a flow, but purely stochastic

- No training required
- Clear scaling with d.o.f.

Stochastic Normalizing Flow

Systematic improvement of NE-MCMC with ML

- Some training required
- Same scaling with d.o.f.

Non-equilibrium Monte Carlo

Out-of-equilibrium evolutions

sampling each consecutive step from a sequence of distributions

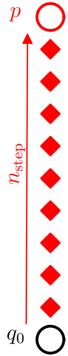
$$q_0 \simeq e^{-S_{c(0)}} \to e^{-S_{c(1)}} \to \cdots \to p \simeq e^{-S_{c(n_{\mathrm{step}})}}$$

Out-of-equilibrium evolutions

sampling each consecutive step from a sequence of distributions

$$q_0 \simeq e^{-S_{c(0)}} \rightarrow e^{-S_{c(1)}} \rightarrow \dots \rightarrow p \simeq e^{-S_{c(n_{\mathrm{step}})}}$$

start **at equilibrium** from the **prior** distribution $q_0 = e^{-S_{c(0)}}/Z_0$



Out-of-equilibrium evolutions

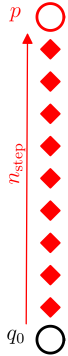
sampling each consecutive step from a sequence of distributions

$$q_0 \simeq e^{-S_{c(0)}} \to e^{-S_{c(1)}} \to \cdots \to p \simeq e^{-S_{c(n_{\mathrm{step}})}}$$

start **at equilibrium** from the **prior** distribution $q_0 = e^{-S_{c(0)}}/Z_0$

at each step:

- change protocol parameter $c(n-1) \to c(n)$
- MC update with transition probability $P_{c(n)}(U_n \to U_{n+1})$
- repeat n_{step} times until the **target** $p = e^{-S}/Z$



Out-of-equilibrium evolutions

sampling each consecutive step from a sequence of distributions

$$q_0 \simeq e^{-S_{c(0)}} \to e^{-S_{c(1)}} \to \dots \to p \simeq e^{-S_{c(n_{\rm step})}}$$

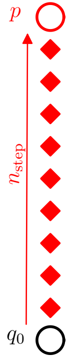
start **at equilibrium** from the **prior** distribution $q_0 = e^{-S_{c(0)}}/Z_0$

at each step:

- change protocol parameter $c(n-1) \to c(n)$
- MC update with transition probability $P_{c(n)}(U_n \to U_{n+1})$
- repeat $n_{\rm step}$ times until the **target** $p = e^{-S}/Z$

"forward" transition probability

$$\mathcal{P}_{\rm f}[U_0, \dots, U] = \prod_{n=1}^{n_{\rm step}} P_{c(n)}(U_{n-1} \to U_n)$$



Crooks' theorem

Look at the ratio of the forward evolution and its reverse

$$\frac{q_0(U_0)\mathcal{P}_{\mathrm{f}}[U_0, \dots, U_{n_{\mathrm{step}}}]}{p(U)\mathcal{P}_{\mathrm{r}}[U_{n_{\mathrm{step}}}, \dots, U_0]} = \frac{q_0(U_0)\prod_{n=1}^{n_{\mathrm{step}}} P_{c(n)}(U_{n-1} \to U_n)}{p(U_{n_{\mathrm{step}}})\prod_{n=1}^{n_{\mathrm{step}}} P_{c(n)}(U_n \to U_{n-1})}$$

Crooks' theorem

Look at the ratio of the forward evolution and its reverse

$$\frac{q_0(U_0)\mathcal{P}_\mathrm{f}[U_0,\dots,U_{n_\mathrm{step}}]}{p(U)\mathcal{P}_\mathrm{r}[U_{n_\mathrm{step}},\dots,U_0]} = \frac{q_0(U_0)\prod_{n=1}^{n_\mathrm{step}} P_{c(n)}(U_{n-1}\to U_n)}{p(U_{n_\mathrm{step}})\prod_{n=1}^{n_\mathrm{step}} P_{c(n)}(U_n\to U_{n-1})}$$

→ **Crooks' theorem** for MCMC **[Crooks; 1999]**: if the update algorithm satisfies detailed balance

$$\frac{q_0(U_0)\mathcal{P}_\mathrm{f}[U_0,\dots,U_{n_\mathrm{step}}]}{p(U)\mathcal{P}_\mathrm{r}[U_{n_\mathrm{step}},\dots,U_0]} = \exp(W - \Delta F)$$

with the generalized **work**

$$W = \sum_{n=0}^{n_\mathrm{step}-1} \left\{ S_{c(n+1)}[U_n] - S_{c(n)}[U_n] \right\}$$

and the **free energy** difference

$$\exp(-\Delta F) = \frac{Z_{c(n_\mathrm{step})}}{Z_{c(0)}}$$

Jarzynski's equality for MCMC

Integrating over all paths gives

$$\int [\mathcal{D}U_0 \dots \mathcal{D}U_{n_{\mathrm{step}}}] q_0(U_0) \mathcal{P}_{\mathrm{f}}[U_0, \dots, U_{n_{\mathrm{step}}}] \exp(-(W - \Delta F)) = 1 \quad \rightarrow \quad \langle \exp(-W_d) \rangle_{\mathrm{f}} = 1$$

with the dissipated work $W_d = W - \Delta F$

Jarzynski's equality for MCMC

Integrating over all paths gives

$$\int [\mathcal{D}U_0 \dots \mathcal{D}U_{n_{\mathrm{step}}}] q_0(U_0) \mathcal{P}_{\mathrm{f}}[U_0, \dots, U_{n_{\mathrm{step}}}] \exp(-(W - \Delta F)) = 1 \quad \rightarrow \quad \langle \exp(-W_d) \rangle_{\mathrm{f}} = 1$$

with the dissipated work $W_d = W - \Delta F$

Formal derivation of **Jarzynski's equality [Jarzynski; 1997]** for MCMC

$$\langle \exp(-W) \rangle_{\mathrm{f}} = \exp(-\Delta F) = \frac{Z}{Z_0}$$

Jarzynski's equality for MCMC

Integrating over all paths gives

$$\int [\mathcal{D}U_0 \ldots \mathcal{D}U_{n_{\mathrm{step}}}] q_0(U_0) \mathcal{P}_{\mathrm{f}}[U_0, \ldots, U_{n_{\mathrm{step}}}] \exp(-(W - \Delta F)) = 1 \quad \rightarrow \quad \langle \exp(-W_d) \rangle_{\mathrm{f}} = 1$$

with the dissipated work $W_d = W - \Delta F$

Formal derivation of **Jarzynski's equality [Jarzynski; 1997]** for MCMC

$$\langle \exp(-W) \rangle_{\mathrm{f}} = \exp(-\Delta F) = \frac{Z}{Z_0}$$

Using Jensen's inequality $\langle \exp x \rangle \geq \exp \langle x \rangle$

$$\exp(-\Delta F) = \langle \exp(-W) \rangle_{\mathrm{f}} \geq \exp(-\langle W \rangle_{\mathrm{f}})$$

we get the Second Law of Thermodynamics

$$\langle W \rangle_{\mathrm{f}} \geq \Delta F$$

From the properties of transition probabilities

$$\langle \mathcal{O} \rangle = \int \mathcal{D}U\, \mathcal{O}(U)\, p(U)$$
$$= \int [\mathcal{D}U_0 \dots \mathcal{D}U]\, \mathcal{O}(U)\, p(U) \mathcal{P}_{\mathrm{r}}[U, \dots, U_0]$$

Then insert $q_0 \mathcal{P}_f$

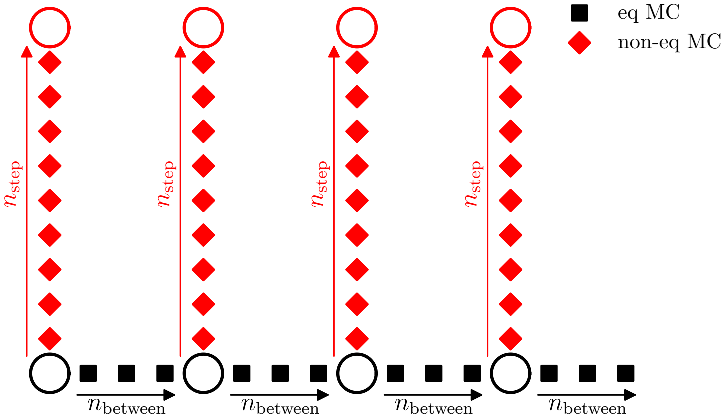
$$\langle \mathcal{O} \rangle = \int [\mathcal{D}U_0 \dots \mathcal{D}U]\, q_0(U_0)\, \mathcal{P}_{\mathrm{f}}[U_0, \dots, U]\, \mathcal{O}(U)\, \frac{p(U)\mathcal{P}_{\mathrm{r}}[U, \dots, U_0]}{q_0(U_0)\mathcal{P}_{\mathrm{f}}[U_0, \dots, U]}$$
$$= \int [\mathcal{D}U_0 \dots \mathcal{D}U]\, q_0(U_0)\, \mathcal{P}_{\mathrm{f}}[U_0, \dots, U]\, \mathcal{O}(U)\, \exp(-(W - \Delta F))$$

NE-MCMC

This goes beyond computing free energy differences! Reweighting-like estimator to compute v.e.v.

$$\langle \mathcal{O} \rangle = \frac{\langle \mathcal{O}\ \exp(-W) \rangle_{\mathrm{f}}}{\langle \exp(-W) \rangle_{\mathrm{f}}} = \langle \mathcal{O}\ \exp(-W_d) \rangle_{\mathrm{f}}$$

A non-equilibrium paradigm to perform MCMC



Applications of Jarzynski's equality in Lattice Field Theory

Computation of free energies and/or sampling problematic distributions

- Calculation of the interface free-energy in the Z_2 gauge theory **[Caselle et al.; 1604.05544]**
- $\mathrm{SU}(3)$ pure gauge equation of state in 4d from the pressure **[Caselle et al.; 1801.03110]**
- Renormalized coupling for $\mathrm{SU}(N)$ YM theories **[Francesconi et al.; 2003.13734]**
- Connection with Stochastic Normalizing Flows: first test for ϕ^4 scalar field theory **[Caselle et al.; 2201.08862]**
- Entanglement entropy **[Bulgarelli and Panero; 2304.03311]**, also with (S)NFs **[Bulgarelli et al.; 2410.14466]**
- Topological unfreezing for $\mathrm{CP}^{(N-1)}$ model **[Bonanno et al.; 2402.06561]**
- Numerical simulations of Effective String Theory **[Caselle et al.; 2409.15937]**

Applications of Jarzynski's equality in Lattice Field Theory

Computation of free energies and/or sampling problematic distributions

- Calculation of the interface free-energy in the Z_2 gauge theory **[Caselle et al.; 1604.05544]**
- $\mathrm{SU}(3)$ pure gauge equation of state in 4d from the pressure **[Caselle et al.; 1801.03110]**
- Renormalized coupling for $\mathrm{SU}(N)$ YM theories **[Francesconi et al.; 2003.13734]**
- Connection with Stochastic Normalizing Flows: first test for ϕ^4 scalar field theory **[Caselle et al.; 2201.08862]**
- Entanglement entropy **[Bulgarelli and Panero; 2304.03311]**, also with (S)NFs **[Bulgarelli et al.; 2410.14466]**
- Topological unfreezing for $\mathrm{CP}^{(N-1)}$ model **[Bonanno et al.; 2402.06561]**
- Numerical simulations of Effective String Theory **[Caselle et al.; 2409.15937]**

How far are we from equilibrium?

Intuitively we want to be as close to equilibrium as possible!

We can measure the similarity of forward and reverse processes

$$\tilde{D}_{\mathrm{KL}}(q_0\mathcal{P}_{\mathrm{f}}\|p\mathcal{P}_{\mathrm{r}}) = \int [\mathcal{D}U_0 \dots \mathcal{D}U]\, q_0(U_0)\mathcal{P}_{\mathrm{f}}[U_0, \dots, U] \log \frac{q_0(U_0)\mathcal{P}_{\mathrm{f}}[U_0, \dots, U]}{p(U)\mathcal{P}_{\mathrm{r}}[U, U_{n_{\mathrm{step}}-1}, \dots, U_0]}$$

How far are we from equilibrium?

Intuitively we want to be as close to equilibrium as possible!

We can measure the similarity of forward and reverse processes

$$\tilde{D}_{\mathrm{KL}}(q_0\mathcal{P}_{\mathrm{f}}\|p\mathcal{P}_{\mathrm{r}}) = \int [\mathcal{D}U_0 \ldots \mathcal{D}U]\, q_0(U_0)\mathcal{P}_{\mathrm{f}}[U_0,\ldots,U] \log \frac{q_0(U_0)\mathcal{P}_{\mathrm{f}}[U_0,\ldots,U]}{p(U)\mathcal{P}_{\mathrm{r}}[U, U_{n_{\mathrm{step}}-1},\ldots,U_0]}$$

Clear "thermodynamic" interpretation

$$\tilde{D}_{\mathrm{KL}}(q_0\mathcal{P}_{\mathrm{f}}\|p\mathcal{P}_{\mathrm{r}}) = \langle W\rangle_{\mathrm{f}} + \log\frac{Z}{Z_0} = \underbrace{\langle W\rangle_{\mathrm{f}} - \Delta F \geq 0}_{\text{Second Law of thermodynamics!}}$$

→ measure of how reversible the process is!

For NFs we minimize $\tilde{D}_{\mathrm{KL}}(q\|p)$.
But interestingly

$$\tilde{D}_{\mathrm{KL}}(q\|p) \leq \tilde{D}_{\mathrm{KL}}(q_0\mathcal{P}_{\mathrm{f}}\|p\mathcal{P}_{\mathrm{r}})$$

The effective sample size

Effective Sample Size: defined in general as the ratio between the "theoretical" variance and the actual variance of the NE observable

$$\frac{\mathrm{Var}(\mathcal{O})_{\mathrm{NE}}}{n} = \frac{\mathrm{Var}(\mathcal{O})_p}{n\,\mathrm{ESS}}$$

but difficult to compute

We use the (customary) approximate estimator

$$\hat{\mathrm{ESS}} = \frac{\langle \exp(-W)\rangle_f^2}{\langle \exp(-2W)\rangle_f} = \frac{1}{\langle \exp(-2W_d)\rangle_f}$$

Easy to understand in terms of the variance of $\exp(-W)$:

$$\mathrm{Var}(\exp(-W)) = \left(\frac{1}{\hat{\mathrm{ESS}}} - 1\right)\exp(-2\Delta F) \geq 0$$

which leads to

$$0 < \hat{\mathrm{ESS}} \leq 1$$

Non-equilibrium Monte Carlo in β in $\mathrm{SU}(3)$

Most natural example: changing β in $\mathrm{SU}(3)$ pure gauge **[Bulgarelli, Cellini, AN; 2412.00200]**

Prior: thermalized Markov Chain at $\beta_0 < \beta_{\mathrm{target}}$

Protocol: linearly increase β (compress the volume)

d.o.f. involved $\sim (L/a)^4$

Sampling possible at any intermediate β

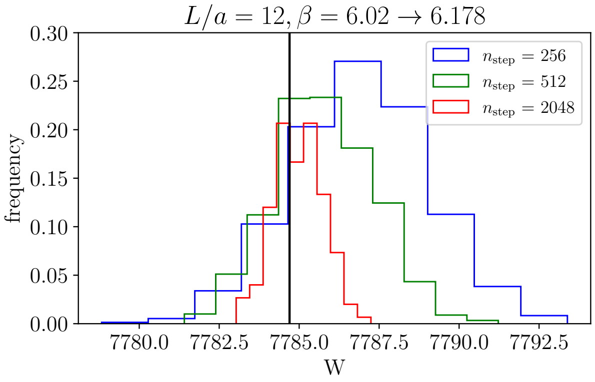
Most natural example: changing β in SU(3) pure gauge **[Bulgarelli, Cellini, AN; 2412.00200]**

Prior: thermalized Markov Chain at $\beta_0 < \beta_{\rm target}$

Protocol: linearly increase β (compress the volume)

d.o.f. involved $\sim (L/a)^4$

Sampling possible at any intermediate β

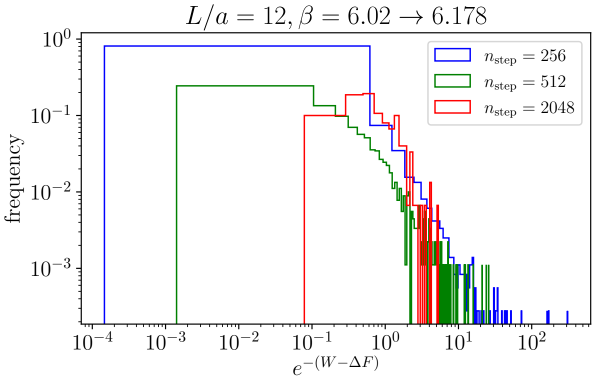


$\tilde{D}_{\rm KL} = \langle W \rangle_{\rm f} - \Delta F$ decreases with increasing $n_{\rm step}$

Non-equilibrium Monte Carlo in β in SU(3)

Most natural example: changing β in SU(3) pure gauge

[Bulgarelli, Cellini, AN; 2412.00200]



Prior: thermalized Markov Chain at $\beta_0 < \beta_{\text{target}}$

Protocol: linearly increase β (compress the volume)

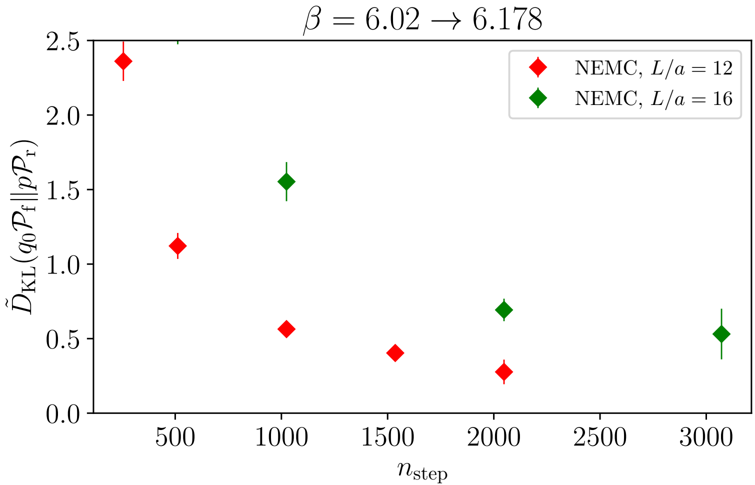
d.o.f. involved $\sim (L/a)^4$

Sampling possible at any intermediate β

$\text{Var}(\exp(-W_d)) = \frac{1}{\hat{\text{ESS}}} - 1$ decreases with increasing n_{step}

Evolutions in β: volume scaling

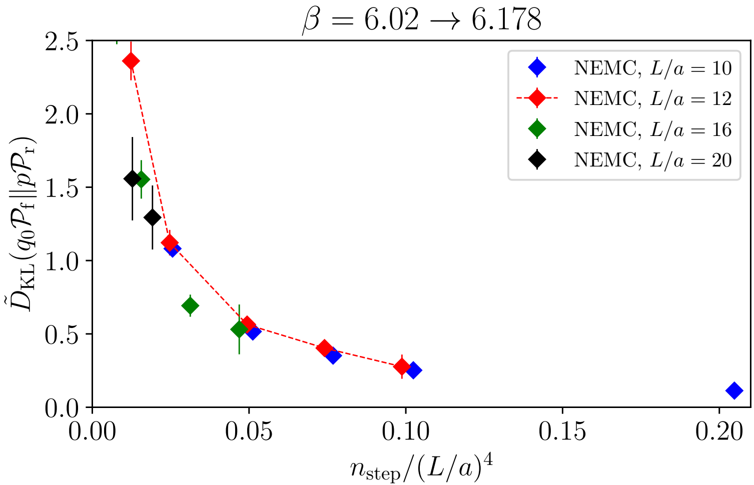
$(1.8\text{fm})^4 \to (1.4\text{fm})^4$ for $L/a = 20$



Evolutions in β: volume scaling

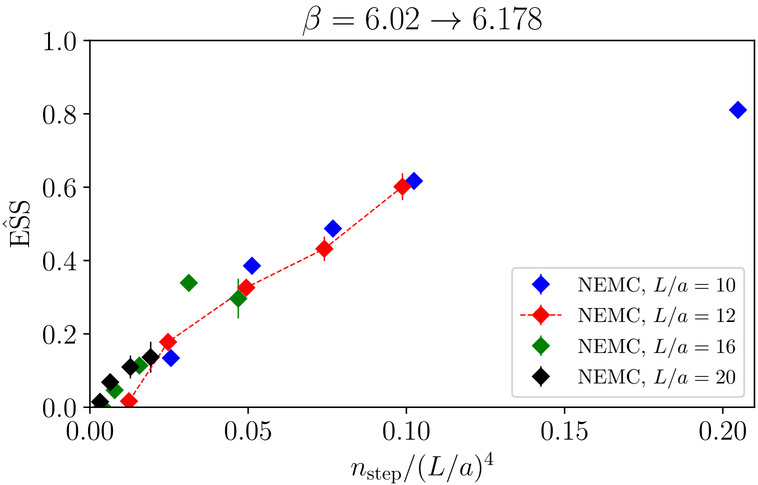
$(1.8\text{fm})^4 \to (1.4\text{fm})^4$ for $L/a = 20$

clear scaling with $n_{\text{step}} \sim (L/a)^4$



$(1.8\mathrm{fm})^4 \to (1.4\mathrm{fm})^4$ for $L/a = 20$

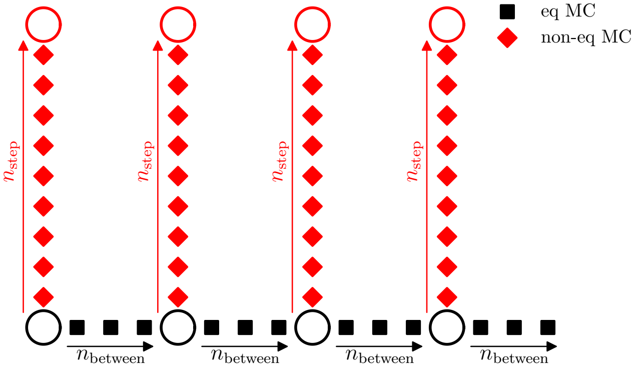
clear scaling with $n_{\mathrm{step}} \sim (L/a)^4$



Stochastic Normalizing Flows

SNFs as systematic improvement of non-equilibrium evolutions

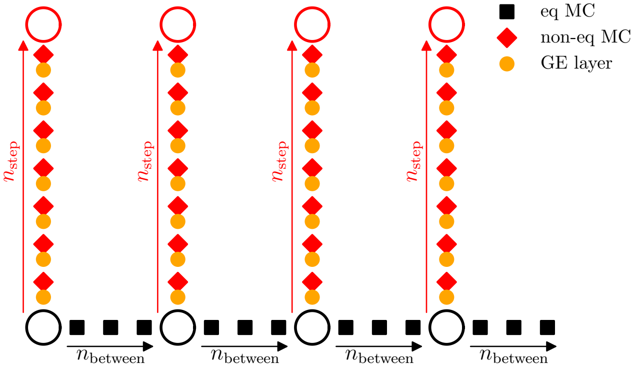
What if you introduce the same transformations used in NFs **between** the non-equilibrium Monte Carlo updates?



What if you introduce the same transformations used in NFs **between** the non-equilibrium Monte Carlo updates?

Stochastic Normalizing Flows (introduced in **[Wu et al.; 2002.06707]**)

$$U_0 \xrightarrow{g_1} g_1(U_0) \xrightarrow{P_{c(1)}} U_1 \xrightarrow{g_2} g_2(U_1) \xrightarrow{P_{c(2)}} U_2 \xrightarrow{g_3} \dots \xrightarrow{P_{c(n_{\text{step}})}} U_{n_{\text{step}}}$$



SNFs as systematic improvement of non-equilibrium evolutions

What if you introduce the same transformations used in NFs **between** the non-equilibrium Monte Carlo updates?

Stochastic Normalizing Flows (introduced in **[Wu et al.; 2002.06707]**)

$$U_0 \xrightarrow{g_1} g_1(U_0) \xrightarrow{P_{c(1)}} U_1 \xrightarrow{g_2} g_2(U_1) \xrightarrow{P_{c(2)}} U_2 \xrightarrow{g_3} \dots \xrightarrow{P_{c(n_{\mathrm{step}})}} U_{n_{\mathrm{step}}}$$

The (generalized) work now is **[Caselle, Cellini, AN, Panero; 2201.08862]**

$$W = \sum_{n=0}^{n_{\mathrm{step}}-1} \underbrace{S_{c(n+1)}(g_n(U_n)) - S_{c(n)}(g_n(U_n))}_{\text{stochastic}} - \underbrace{\log |\det J_n(U_n)|}_{\text{deterministic}}$$

- use gauge-equivariant layers to effectively decrease n_{step}
- how to do training? advantages from the architecture
- scaling with the volume?

Gauge-equivariant layers

Implementation of the coupling layers introduced in **[Nagai and Tomiya; 2103.11965]** and the link-level flow used in **[Abbott et al.; 2305.02402]**

Gauge-equivariant layers

Implementation of the coupling layers introduced in **[Nagai and Tomiya; 2103.11965]** and the link-level flow used in **[Abbott et al.; 2305.02402]**

Essentially a stout-smearing transformation **[Morningstar and Peardon; 2003]**

$$U'_\mu(x) = g_l(U_\mu(x)) = \exp\left(Q^{(l)}_\mu(x)\right)\, U_\mu(x)$$

with the algebra-valued

$$Q^{(l)}_\mu(x) = 2\left[\Omega^{(l)}_\mu(x)\right]_{\mathrm{TA}} \qquad\qquad \Omega^{(l)}_\mu(x) = \underbrace{C^{(l)}_\mu(x)}_{\text{frozen}}\underbrace{U^\dagger_\mu(x)}_{\text{active}}$$

Invertibility + easy computation of $\log J$ guaranteed by $2 \times D = 8$ masks

Gauge-equivariant layers

Implementation of the coupling layers introduced in **[Nagai and Tomiya; 2103.11965]** and the link-level flow used in **[Abbott et al.; 2305.02402]**

Essentially a stout-smearing transformation **[Morningstar and Peardon; 2003]**

$$U'_\mu(x) = g_l(U_\mu(x)) = \exp\left(Q^{(l)}_\mu(x)\right)\, U_\mu(x)$$

with the algebra-valued

$$Q^{(l)}_\mu(x) = 2\left[\Omega^{(l)}_\mu(x)\right]_{\mathrm{TA}} \qquad \Omega^{(l)}_\mu(x) = \underbrace{C^{(l)}_\mu(x)}_{\text{frozen}}\underbrace{U^\dagger_\mu(x)}_{\text{active}}$$

Invertibility + easy computation of $\log J$ guaranteed by $2 \times D = 8$ masks

Sum of frozen staples

$$C^{(l)}_\mu(x) = \sum_{\nu\neq\mu} \underbrace{\rho^{(l)}_{\mu\nu}(x)}_{\text{train}}\underbrace{S_{\mu\nu}(x)}_{\text{staple}}$$

in this work: $\rho^{(l)}_{\mu\nu}(x) \longrightarrow \rho^{(l)}$, meaning 1 parameter per mask/8 parameters per layer

Learning ρ

[Bulgarelli, Cellini, AN; 2412.00200]

Architecture: (1 gauge-equivariant CL + 1 full MC update) $\times n_{\mathrm{step}}$

Training: minimizing $\tilde{D}_{\mathrm{KL}}(q_0 \mathcal{P}_{\mathrm{f}} \| p \mathcal{P}_{\mathrm{r}}) = \langle W \rangle_{\mathrm{f}} + \mathrm{const}$

[Bulgarelli, Cellini, AN; 2412.00200]

Architecture: (1 gauge-equivariant CL + 1 full MC update) $\times n_{\mathrm{step}}$

Training: minimizing $\tilde{D}_{\mathrm{KL}}(q_0\mathcal{P}_{\mathrm{f}} \| p\mathcal{P}_{\mathrm{r}}) = \langle W \rangle_{\mathrm{f}} + \mathrm{const}$

To avoid memory issues for large n_{step} and large volumes we train each layer separately during the non-equilibrium evolution

It's a feature of SNFs

$$U_0 \xrightarrow{g_1} g_1(U_0) \xrightarrow{P_{c(1)}} U_1 \xrightarrow{g_2} g_2(U_1) \xrightarrow{P_{c(2)}} U_2 \xrightarrow{g_3} \dots \xrightarrow{P_{c(n_{\mathrm{step}})}} U_{n_{\mathrm{step}}}$$

Look at the loss $W = S(U_{n_{\mathrm{step}}}) - S_0(U_0) - Q - \log J$

$$Q + \log J = \sum_{n=0}^{n_{\mathrm{step}}-1} S_{c(n+1)}(U_{n+1}) - S_{c(n+1)}(g_n(U_n)) + \log\det J_n(U_n)$$

the terms in the sum can be trained separately!

→ each layer connects two neighbouring intermediate distributions

→ reminiscent of CRAFT **[Matthews at al.; 2201.13117]**

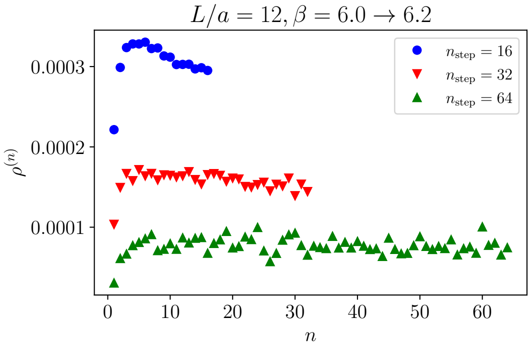
→ memory usage independent of n_{step}!

→ bias in the gradient (no visible effect)

Transferring ρ

Short training: 200-1000 epochs enough to saturate **[Bulgarelli, Cellini, AN; 2412.00200]**

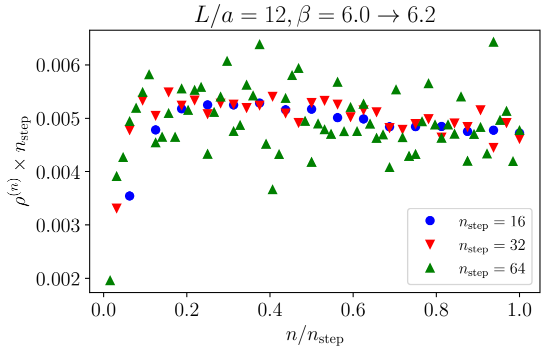
Training only with small n_{step}: clear pattern emerges



Transferring ρ

Short training: 200-1000 epochs enough to saturate **[Bulgarelli, Cellini, AN; 2412.00200]**

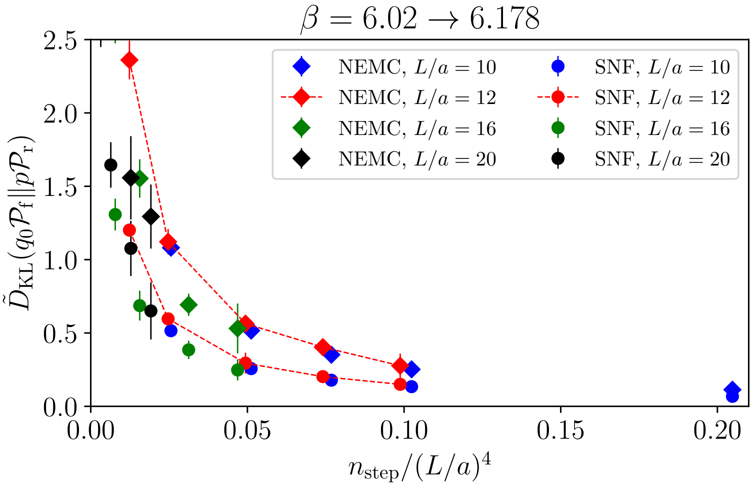
Training only with small n_{step}: clear pattern emerges



- global interpolation of ρ from trainings at $n_{\text{step}} = 16, 32, 64$
- $\rho^{(l)}$ extrapolated to large n_{step} $\to$ no retraining!
- Heavy use of **transfer learning** for each $\beta_0 \to \beta$ evolution
- Transfer learning also possible between different volumes

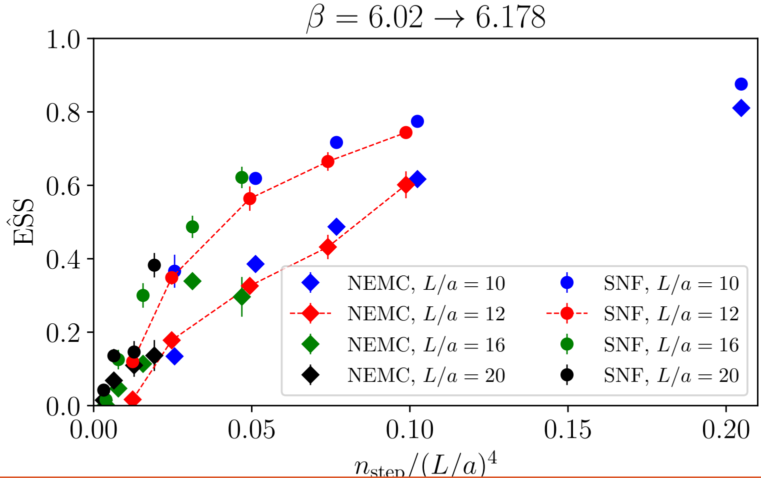
SNF volume scaling

[Bulgarelli, Cellini, AN; 2412.00200]



[Bulgarelli, Cellini, AN; 2412.00200]

$n_{\text{step}} \sim V$ for fixed $\tilde{D}_{\text{KL}}$ or ESS



A strategy to mitigate topological freezing

Topological freezing in lattice gauge theory

On the lattice: topological sectors characterized by integer Q emerge for $a \to 0$

Transition between these sectors is strongly suppressed using standard MCMC

Topological freezing in lattice gauge theory

On the lattice: topological sectors characterized by integer Q emerge for $a \to 0$

Transition between these sectors is strongly suppressed using standard MCMC

- Strong freezing of topology at $\beta \geq 6.5$ ($r_0/a > 11$)
- $\tau_{\text{int}}(Q^2) > 10^3$ with 1 heat-bath step + 4 over-relaxation steps ($z \sim 5$)
- **Open Boundary Conditions [Lüscher and Schaefer; 1105.4749]** remove the sectors and mitigate the issue. But: complications due to boundary effects
- General strategy for mitigation of critical slowing down?

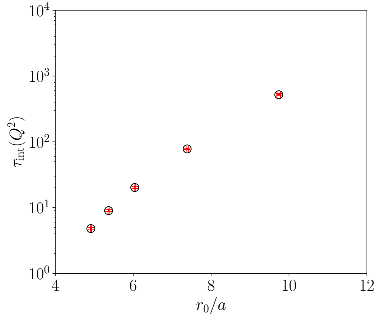


Image courtesy of C. Bonanno

A non-equilibrium strategy for topological freezing in $\mathrm{SU}(3)$

Goal: sample (frozen) topological observables at β_{target} on a L^4 lattice

Goal: sample (frozen) topological observables at β_{target} on a L^4 lattice

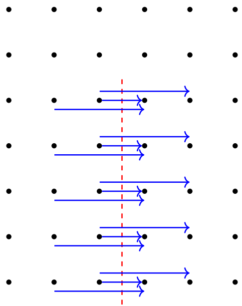
NE-MCMC in the Boundary Conditions

Prior: thermalized Markov Chain at β_{target} with OBC on a L_d^3 defect

Protocol: switch defect BC (# d.o.f. $\sim (L_d/a)^3$) linearly until PBC

Similar strategy used in Parallel Tempering →**see Claudio's talk on Monday**

Can be combined with coupling layers acting around the defect to get a SNF



Goal: sample (frozen) topological observables at β_{target} on a L^4 lattice

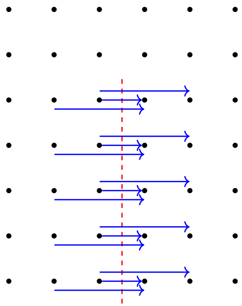
NE-MCMC in the Boundary Conditions

Prior: thermalized Markov Chain at β_{target} with OBC on a L_d^3 defect

Protocol: switch defect BC ($\#$ d.o.f. $\sim (L_d/a)^3$) linearly until PBC

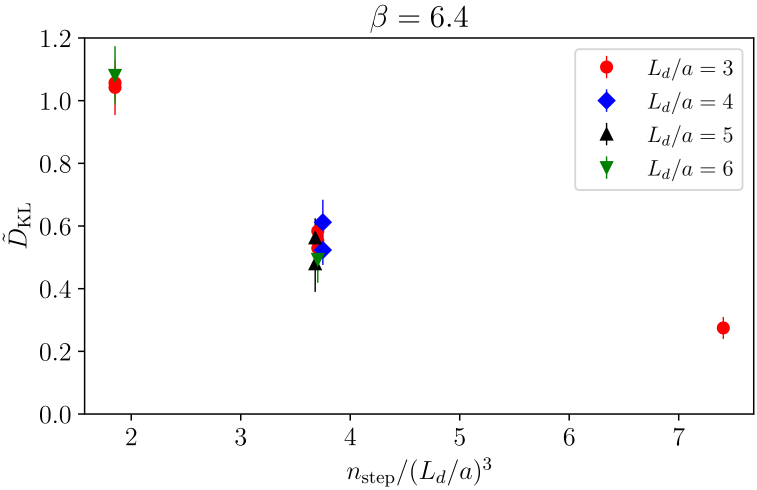
Similar strategy used in Parallel Tempering →**see Claudio's talk on Monday**

Can be combined with coupling layers acting around the defect to get a SNF

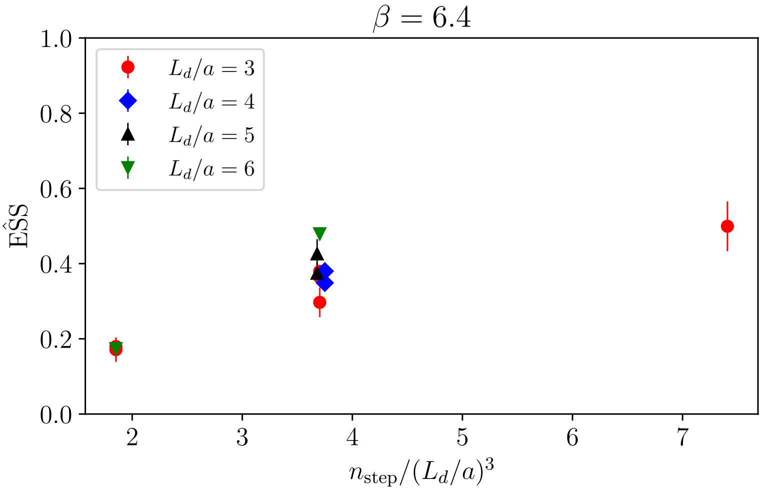


Applied to CP^{N-1} model in 2d **[Bonanno, AN, Vadacchino; 2402.06561]**

Promising results: $\tau_{\mathrm{int}}(Q^2) \sim 10^5$ tamed to effectively a few thousands + length of non-equilibrium evolutions scales with defect size preliminary results



preliminary results

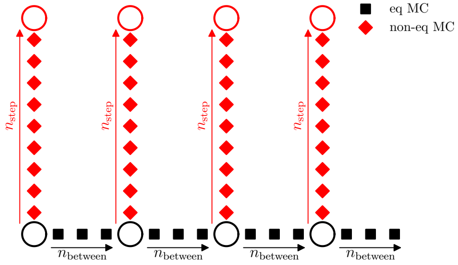


Metric for scaling

$$n_{\mathrm{ev}}^{(\mathrm{eff})} \times (n_{\mathrm{step}} + n_{\mathrm{between}}) \simeq n_{\mathrm{ev}} \frac{2\tau_{\mathrm{int}}}{\mathrm{E\hat{S}S}} \times (n_{\mathrm{step}} + n_{\mathrm{between}})$$

Metric for scaling

$$n_{\rm ev}^{\rm (eff)} \times (n_{\rm step} + n_{\rm between}) \simeq n_{\rm ev} \frac{2\tau_{\rm int}}{\hat{\rm ESS}} \times (n_{\rm step} + n_{\rm between})$$



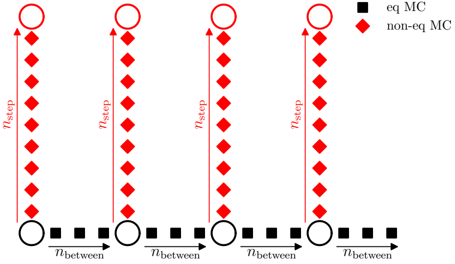
Rescale $n_{\rm step}$ and $n_{\rm between}$ to keep $\hat{\rm ESS}$ and $\tau_{\rm int}$ fixed for $a \to 0$

$$n_{\rm step} \sim \left(\frac{L_d}{a}\right)^3 \sim a^{-3} \qquad n_{\rm between} \sim a^{-2}$$

$$n_{\rm ev} \overbrace{\frac{2\tau_{\rm int}}{\hat{\rm ESS}}}^{\rm fixed} \times (n_{\rm step} + n_{\rm between}) \sim a^{-3}$$

Metric for scaling

$$n_{\text{ev}}^{(\text{eff})} \times (n_{\text{step}} + n_{\text{between}}) \simeq n_{\text{ev}} \frac{2\tau_{\text{int}}}{\hat{\text{ESS}}} \times (n_{\text{step}} + n_{\text{between}})$$



Rescale n_{step} and n_{between} to keep $\hat{\text{ESS}}$ and τ_{int} fixed for $a \to 0$

$$n_{\text{step}} \sim \left(\frac{L_d}{a}\right)^3 \sim a^{-3} \qquad n_{\text{between}} \sim a^{-2}$$

$$n_{\text{ev}} \overbrace{\frac{2\tau_{\text{int}}}{\hat{\text{ESS}}}}^{\text{fixed}} \times (n_{\text{step}} + n_{\text{between}}) \sim a^{-3}$$

Goal of SNF: systematically improve the coefficients

$\rightarrow$ ongoing work on defect coupling layers

in collaboration with C. Bonanno, A. Bulgarelli, E. Cellini, D. Panfalone, D. Vadacchino, L. Verzichelli

Conclusions

General strategy

- Start from NE-MCMC → clear scaling with the degrees of freedom (no training)

$$n_{\rm step} \sim \#\text{d.o.f.} \to \text{constant } \tilde{D}_{\rm KL} \text{ or } \hat{\rm ESS}$$

- Implement lightweight SNF on top of NE-MCMC → improve coefficients of scaling (moderate training)
- Implement more complicated architectures → systematic improvement program (harder training?)

Conclusions

General strategy

- Start from NE-MCMC → clear scaling with the degrees of freedom (no training)

$$n_{\rm step} \sim \#{\rm d.o.f.} \to \text{constant } \tilde{D}_{\rm KL} \text{ or } \hat{\rm ESS}$$

- Implement lightweight SNF on top of NE-MCMC → improve coefficients of scaling (moderate training)
- Implement more complicated architectures → systematic improvement program (harder training?)

Open questions

What happens with less MC and more ML?

Optimal protocols: non-trivial problem, more efficient solutions likely

Same approach, but in continuous time: NETS **[Albergo; 2410.02711]**

Conclusions

General strategy

- Start from NE-MCMC → clear scaling with the degrees of freedom (no training)

$$n_{\rm step} \sim \#{\rm d.o.f.} \rightarrow \text{constant } \tilde{D}_{\rm KL} \text{ or } {\rm E\hat{S}S}$$

- Implement lightweight SNF on top of NE-MCMC → improve coefficients of scaling (moderate training)
- Implement more complicated architectures → systematic improvement program (harder training?)

Open questions

What happens with less MC and more ML?

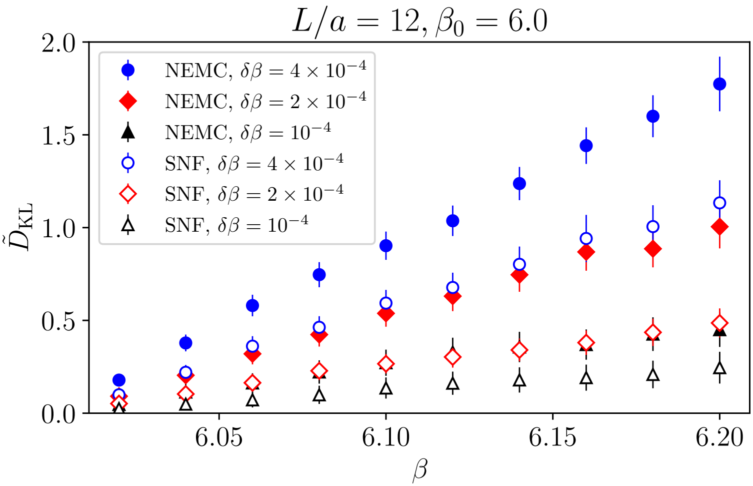
Optimal protocols: non-trivial problem, more efficient solutions likely

Same approach, but in continuous time: NETS **[Albergo; 2410.02711]**

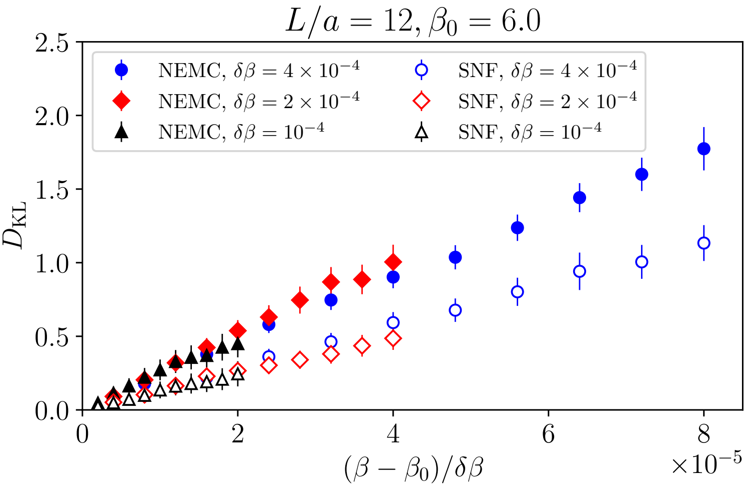
Characterize thermalization processes?

Completely different applications: signal-to-noise → **Guilherme's talk on Wednesday**

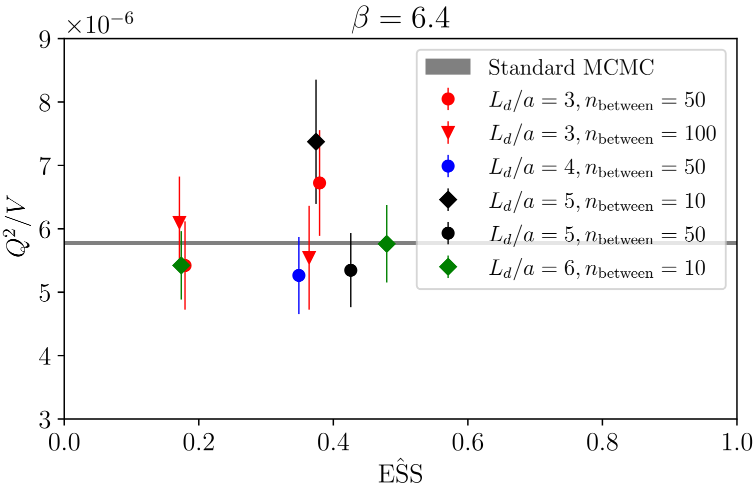
Thank you for your attention!


$L/a = 12, \beta_0 = 6.0$
2.0
1.5
1.0
0.5
0.0
$\tilde{D}_{\mathrm{KL}}$
NEMC, $\delta\beta = 4 \times 10^{-4}$
NEMC, $\delta\beta = 2 \times 10^{-4}$
NEMC, $\delta\beta = 10^{-4}$
SNF, $\delta\beta = 4 \times 10^{-4}$
SNF, $\delta\beta = 2 \times 10^{-4}$
SNF, $\delta\beta = 10^{-4}$
6.05
6.10
6.15
6.20
β

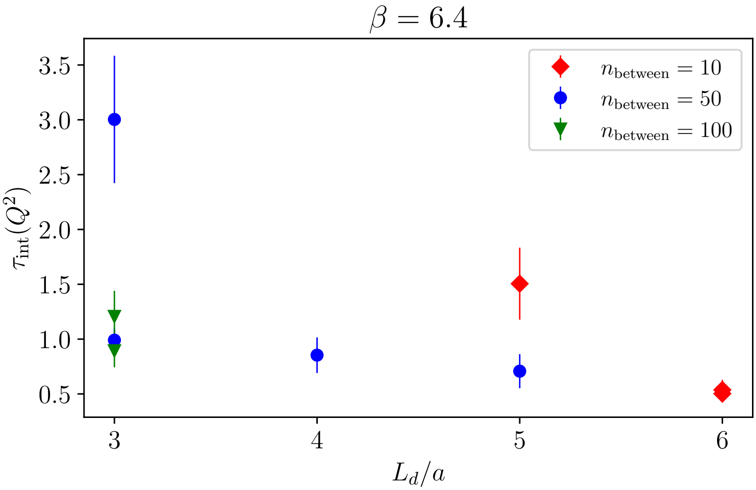
Switching β in SU(3): scaling in $\beta - \beta_0$



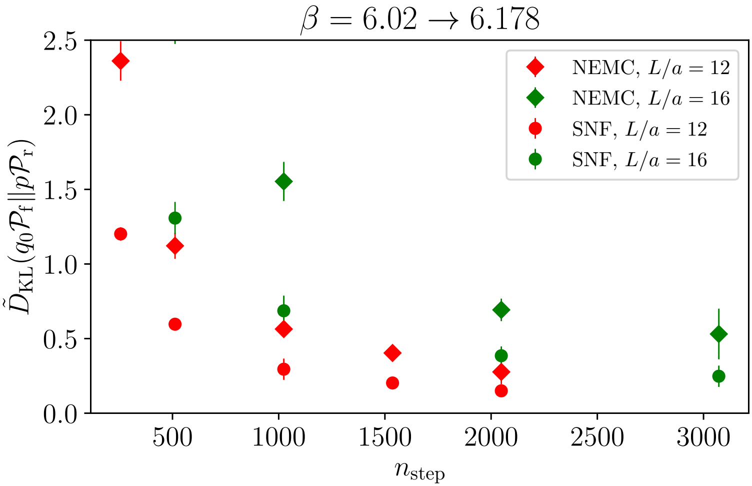
$\rightarrow \delta\beta \sim \beta - \beta_0$ for fixed $\tilde{D}_{\mathrm{KL}}$ (linear protocol)


$\beta = 6.4$
$\times 10^{-6}$
9
8
7
6
5
4
3
Q^2/V
0.0
0.2
0.4
0.6
0.8
1.0
$\hat{\mathrm{ESS}}$
Standard MCMC
$L_d/a = 3, n_{\mathrm{between}} = 50$
$L_d/a = 3, n_{\mathrm{between}} = 100$
$L_d/a = 4, n_{\mathrm{between}} = 50$
$L_d/a = 5, n_{\mathrm{between}} = 10$
$L_d/a = 5, n_{\mathrm{between}} = 50$
$L_d/a = 6, n_{\mathrm{between}} = 10$

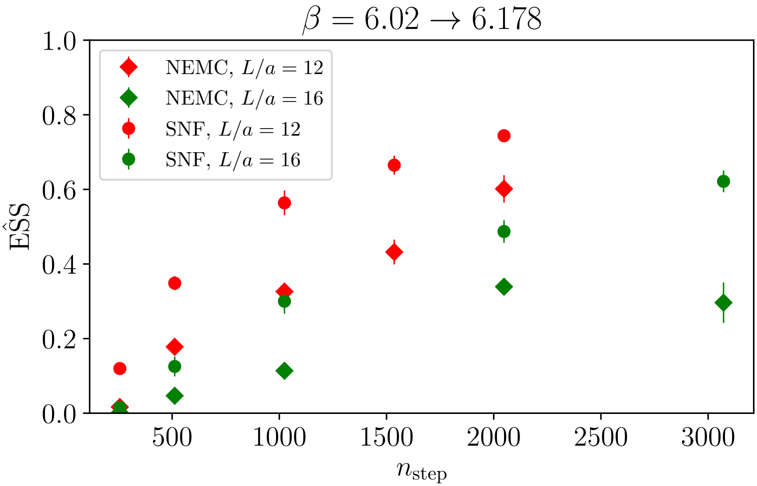
Switching BC in $\mathrm{SU}(3)$: autocorrelations



Improvements over purely stochastic approach



Improvements over purely stochastic approach



Clausius inequality for an (isothermal) transformation from state A to state B

$$\frac{Q}{T} \leq \Delta S$$

If we use

$$\begin{cases} Q = & \Delta E - W \quad \text{(First Law)} \\ F \stackrel{\text{def}}{=} & E - ST \end{cases}$$

the Second Law becomes

$$W \geq \Delta F$$

where the equality holds for reversible processes.

Moving from thermodynamics to statistical mechanics, we know that actually

$$\langle W \rangle_f \geq \Delta F = F_B - F_A$$

for a given "forward" process f from A to B

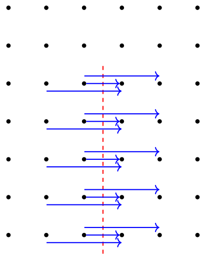
The CP^{N-1} model with a defect

Improved action

$$S_L^{(r)} = -2N\beta_L \sum_{x,\mu} \left\{ k_\mu^{(n)}(x) c_1 \Re\left[\bar{U}_\mu(x)\bar{z}(x+\hat{\mu})z(x)\right] + k_\mu^{(n)}(x+\hat{\mu})k_\mu^{(n)}(x) c_2 \Re\left[\bar{U}_\mu(x+\hat{\mu})\bar{U}_\mu(x)\bar{z}(x+2\hat{\mu})z(x)\right] \right\}$$

with $z(x)$ a vector of N complex numbers with $\bar{z}(x)z(x) = 1$ and $U_\mu(x) \in U(1)$

$c_1 = 4/3$ and $c_2 = -1/12$ are Symanzik-improvement coefficients



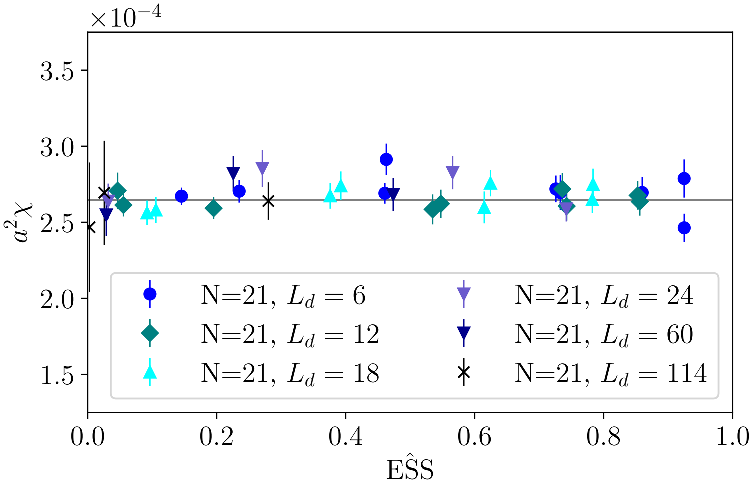
The $k_\mu^{(n)}(x)$ regulate the boundary conditions along a given defect D

$$k_\mu^{(n)}(x) \equiv \begin{cases} c(n) & x \in D \wedge \mu = 0\,; \\ 1 & \text{otherwise.} \end{cases}$$

at a given step n of the out-of-equilibrium evolution protocol $c(n)$

Topological susceptibility for various protocols for $N = 21$, $\beta_L = 0.7$, $V = 114^2$ (roughly similar numerical effort)

Note that with OBC $\rightarrow \tau_{int}(\chi) \sim 50$



Black band is from parallel tempering **[Bonanno et al.; 2019]** $\rightarrow$ with $\times \sim 100$ numerical cost