Stochastic Normalizing Flows for lattice gauge theory

Alessandro Nada

Università degli Studi di Torino and INFN Torino

Machine-Learning-Based Sampling in Lattice Field Theory and Quantum Chemistry workshop

Bonn, 21-25 October 2024

in collaboration with

Andrea Bulgarelli (UniTo) and Elia Cellini (UniTo)

- For a Yang-Mills theory regularized on the lattice we want to compute **expectation values**

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \prod \mathrm{d}U \, \underbrace{\mathcal{O}(U)}_{\text{measure}} \, \underbrace{\exp(-S_{\mathrm{YM}}(U))}_{\text{sample}}$$

with the probability distribution for $\mathrm{SU}(3)$ YM theory at a given β

$$p(U) = \exp(-S_{\mathrm{YM}}(U))/Z \qquad S_{\mathrm{YM}}(U) = \beta \sum_{x,\mu<\nu} 1 - \frac{1}{3} \mathrm{Re} \mathrm{Tr} \, U_{\mu\nu}(x)$$

then continuum extrapolation $a \to 0$ (roughly $\beta \to \infty$)

- For a Yang-Mills theory regularized on the lattice we want to compute **expectation values**

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \prod \mathrm{d}U \underbrace{\mathcal{O}(U)}_{\text{measure}} \underbrace{\exp(-S_{\mathrm{YM}}(U))}_{\text{sample}}$$

with the probability distribution for $\mathrm{SU}(3)$ YM theory at a given β

$$p(U) = \exp(-S_{\mathrm{YM}}(U))/Z \qquad S_{\mathrm{YM}}(U) = \beta \sum_{x,\mu<\nu} 1 - \frac{1}{3} \mathrm{Re}\mathrm{Tr} U_{\mu\nu}(x)$$

then continuum extrapolation $a \to 0$ (roughly $\beta \to \infty$)

- **(Thermalized) Markov Chain**: elegant and *scalable* numerical solution

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

measure $\mathcal{O}$ on equilibrium configurations

- MCMC algorithms (Metropolis, HMC...) define the transition probability P_p

Critical slowing down

The configurations sampled sequentially in a Markov Chain are **autocorrelated**

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

Measure of this autocorrelation for $\mathcal{O}$:

$$\tau_{\mathrm{int}}(\mathcal{O})$$

$\to$ true independent configurations $= n/2\tau_{\mathrm{int}}(\mathcal{O})$

Critical slowing down

The configurations sampled sequentially in a Markov Chain are **autocorrelated**

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

Measure of this autocorrelation for $\mathcal{O}$:

$$\tau_{\rm int}(\mathcal{O})$$

$\to$ true independent configurations $= n/2\tau_{\rm int}(\mathcal{O})$

Critical slowing down

When a critical point is approached $\tau_{\rm int}$ diverges

The continuum limit $a \to 0$ is a critical point, so

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

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

Topological freezing in lattice gauge theory

- in the continuum: field configurations classified by integer topological charge Q
- on the lattice: topological sectors emerge for $a \to 0$
- Using standard local MCMC algorithms the transition between these sectors is strongly suppressed

Topological freezing in lattice gauge theory

- in the continuum: field configurations classified by integer topological charge Q
- on the lattice: topological sectors emerge for $a \to 0$
- Using standard local MCMC algorithms the transition between these sectors is strongly suppressed

- 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 boundaries mitigate the issue. But: more expensive simulation, more complicated analysis, only for topology
- General solution for critical slowing down?

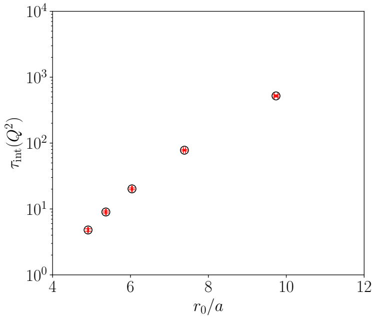


Image courtesy of C. Bonanno

Re-framing critical slowing down: flowing from one distribution to the other

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

Flow-based approach

mapping between the target $p(U)$ and some tractable distribution $q_0(z)$

$\rightarrow$ novel approach to fight critical slowing down

$\rightarrow$ successfully applied in LFTs in 2d: ϕ^4 scalar field theory **[Albergo et al.; 2019]**, **[Kanwar et al.; 2020]**, **[Nicoli et al.; 2020]**, **[Del Debbio et al.; 2021]**, $\mathrm{U}(1)$ **[Singha et al.; 2023]**, $\mathrm{SU}(N)$ **[Boyda et al.; 2020]**

$\rightarrow$ including fermions **[Albergo et al.; 2021]** in $\mathrm{U}(1)$ and $\mathrm{SU}(N)$ **[Abbott et al.; 2022]**, Schwinger model **[Finkenrath et al.; 2022]**, **[Albergo et al.; 2022]**, and QCD **[Abbott et al.; 2022]**

$\rightarrow$ first attempts in 4d **[Abbott et al.; 2023]** with interesting applications **[Abbott et al.; 2024]**

$\rightarrow$ new architectures such as Continuous Normalizing Flows **[Gerdes et al.; 2022]**, **[Caselle et al.; 2023]**, **[Gerdes et al.; 2024]**

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

$\rightarrow$ ...

Re-framing critical slowing down: flowing from one distribution to the other

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

Flow-based approach

mapping between the target $p(U)$ and some tractable distribution $q_0(z)$

$\rightarrow$ novel approach to fight critical slowing down

Normalizing flows do not appear to scale well with the volume (i.e. with the degrees of freedom)

However: same approach is possible stochastically! $\rightarrow$ NE-MCMC

Can we obtain a **clear scaling**?

Can we combine it with NFs **in a new architecture**?

Non-equilibrium Monte Carlo

Out-of-equilibrium evolutions

sampling each consecutive step from a sequence of distributions

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

Out-of-equilibrium evolutions

sampling each consecutive step from a sequence of distributions

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

- $c(n)$ is a parameter of the action $S_{c(n)}$ of the model
- start **at equilibrium** from a distribution $q_0 = e^{-S_{c(0)}}/Z_0$, the **prior**
- n_{step} intermediate steps
- at each step: MC update with transition probability $P_{c(n)}(U_n \rightarrow U_{n+1})$
- $P_{c(n)}$ changes along the evolution according to the **protocol** $c(n)$
- end at the **target** probability distribution $p = e^{-S_{c(n_{\mathrm{step}})}}/Z_{n_{\mathrm{step}}} \equiv e^{-S}/Z$

Out-of-equilibrium evolutions

sampling each consecutive step from a sequence of distributions

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

- $c(n)$ is a parameter of the action $S_{c(n)}$ of the model
- start **at equilibrium** from a distribution $q_0 = e^{-S_{c(0)}}/Z_0$, the **prior**
- n_{step} intermediate steps
- at each step: MC update with transition probability $P_{c(n)}(U_n \to U_{n+1})$
- $P_{c(n)}$ changes along the evolution according to the **protocol** $c(n)$
- end at the **target** probability distribution $p = e^{-S_{c(n_{\mathrm{step}})}}/Z_{n_{\mathrm{step}}} \equiv e^{-S}/Z$

"forward" transition probability

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

Crooks' theorem

Look at the ratio of the forward evolution and its reverse

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

Crooks' theorem

Look at the ratio of the forward evolution and its reverse

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

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

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

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

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

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

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

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

A ratio of partition functions is computed directly with an average over "forward" non-equilibrium evolutions

$$\langle \mathcal{A} \rangle_{\mathrm{f}} = \int [\mathrm{d}U_0 \dots \mathrm{d}U] q_0(U_0) \mathcal{P}_{\mathrm{f}}[U_0, \dots, U] \mathcal{A}[U_0, \dots, U]$$

Jarzynski's equality for MCMC

Integrating over all paths gives

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

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

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

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

A ratio of partition functions is computed directly with an average over "forward" non-equilibrium evolutions

$$\langle \mathcal{A} \rangle_{\mathrm{f}} = \int [\mathrm{d}U_0 \dots \mathrm{d}U] q_0(U_0) \mathcal{P}_{\mathrm{f}}[U_0, \dots, U] \, \mathcal{A}[U_0, \dots, U]$$

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

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

Out-of-equilibrium stochastic evolutions

- it's a **non-equilibrium** process!

$$q_n(U_n) \neq \exp(-S_{c(n)}(U_n))/Z_n$$

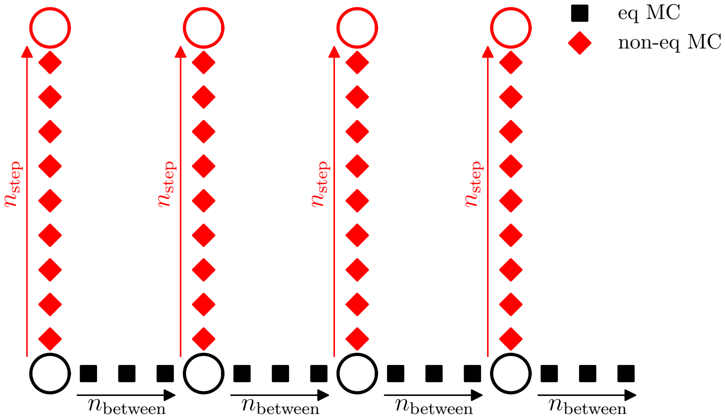
- valid process also far from equilibrium (e.g. n_{step} is "small": $n_{\mathrm{step}} = 1$ is standard reweighting)

- the $\langle \mathcal{A} \rangle_{\mathrm{f}}$ average is taken over all possible evolutions (always true for infinite statistics)

NE-MCMC

This goes beyond computing free energy differences! The same derivation holds if you want to compute v.e.v. of an observable for the target distribution p

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


eq MC
non-eq MC
n_{step}
n_{step}
n_{step}
n_{step}
n_{between}
n_{between}
n_{between}
n_{between}

Applications of Jarzynski's equality in Lattice Field Theory

Several applications in the last 8 years!

- Calculation of the interface free-energy in the Z_2 gauge theory **[Caselle et al.; 2016]**
- $\mathrm{SU}(3)$ pure gauge equation of state in 4d from the pressure **[Caselle et al.; 2018]**
- Renormalized coupling for $\mathrm{SU}(N)$ YM theories **[Francesconi et al.; 2020]**
- Connection with Stochastic Normalizing Flows: first test for ϕ^4 scalar field theory **[Caselle et al.; 2022]**
- Entanglement entropy **[Bulgarelli and Panero; 2023]**, also with (S)NFs **[Bulgarelli et al.; 2024]** ← see Andrea's poster
- Topological unfreezing for $\mathrm{CP}(N-1)$ model **[Bonanno et al.; 2024]**
- Numerical simulations of Effective String Theory **[Caselle et al.; 2024]** ← see Elia's poster

Computation of free energies and/or sampling problematic distributions

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

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

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

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

Clear "thermodynamic" interpretation

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

$\rightarrow$ 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

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

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

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

which leads to

$$0 < \mathrm{E\hat{S}S} \leq 1$$

Numerical results for $\mathrm{SU}(3)$ in 4 dimensions

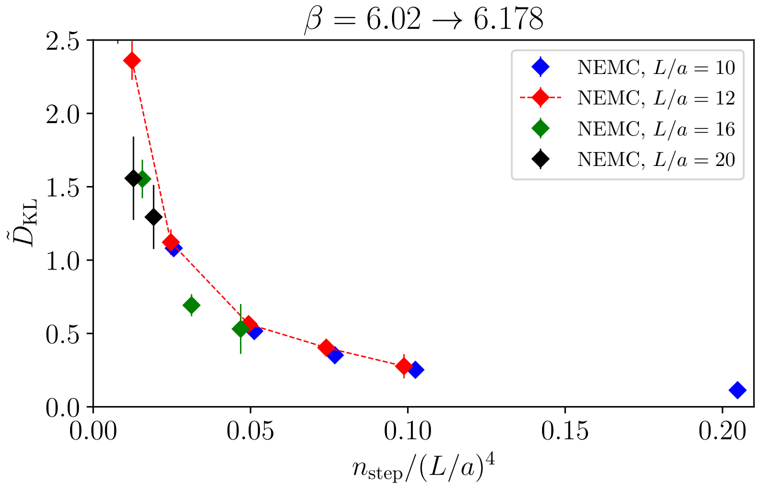
Non-equilibrium strategies for critical slowing down in $\mathrm{SU}(3)$

How to sample frozen topological observables at β_{target} on a L^4 lattice?

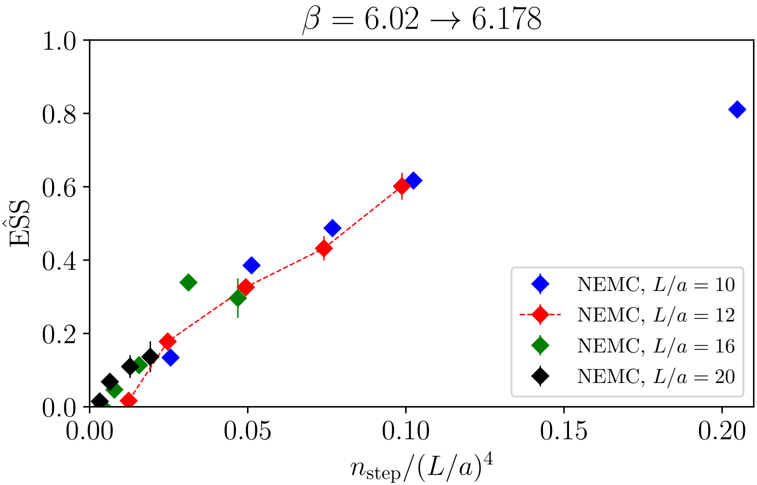
	Evolution in the boundary conditions	**Evolution in β (THIS TALK)**
Prior	thermalized Markov Chain at β_{target} with OBC on a L_d^3 defect	thermalized Markov Chain at $\beta_0 < \beta_{\mathrm{target}}$ ($a_0 > a_{\mathrm{target}}$)
Protocol	Gradually switch on PBC	Gradually increase β (compress the volume)
d.o.f.	$\sim (L_d/a)^3$	$\sim (L/a)^4$
Intermediate sampling	—	possible at any intermediate β
Papers	JHEP 04 (2024) 126 – 2402.06561 with C. Bonanno and D.Vadacchino	2411.XXXX with A. Bulgarelli and E. Cellini

Evolutions in β: volume scaling

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



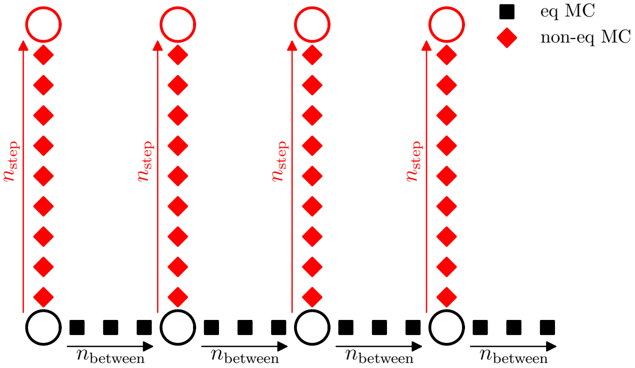
$(1.8\text{fm})^4 \to (1.4\text{fm})^4$ for $L/a = 20$



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?

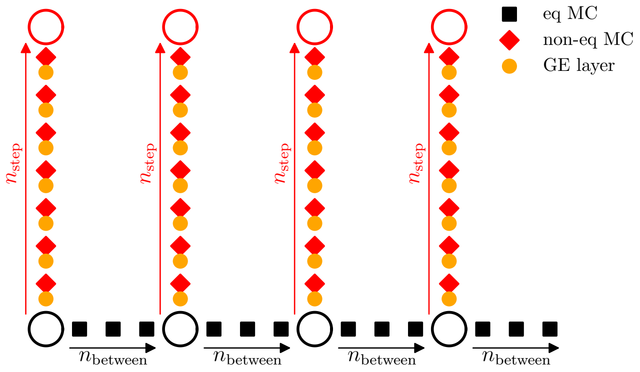


SNFs as systematic improvement of non-equilibrium evolutions

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.; 2020]**)

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



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.; 2020]**)

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

The (generalized) work now is

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

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

Gauge-equivariant layers

Implementation of the coupling layers introduced in **[Nagai and Tomiya; 2021]** and the link-level flow used in **[Abbott et al.; 2023]**

Essentially a stout-smearing transformation **[Morningstar and Peardon; 2003]** with masks to make them invertible (and compute $\log J$)

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

with the algebra-valued

$$Q^{(l)}_\mu(x) = 2\left[\Omega^{(l)}_\mu(x)\right]_{\mathrm{TA}}$$

$$\Omega^{(l)}_\mu(x) = \underbrace{C^{(l)}_\mu(x)}_{\text{frozen}}\underbrace{U^\dagger_\mu(x)}_{\text{active}}$$

Sum of frozen staples

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

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

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.; 2022]**

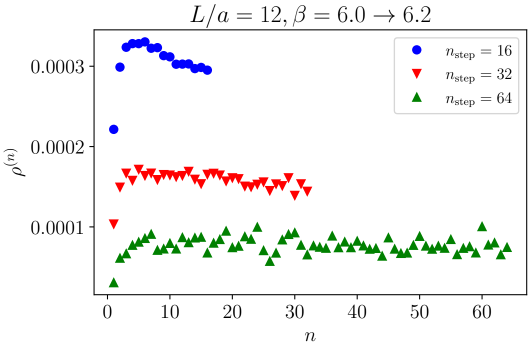
→ memory usage independent of n_{step}!

→ bias in the gradient (no visible effect)

Transferring ρ

Short trainings: 200-1000 epochs enough to saturate

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

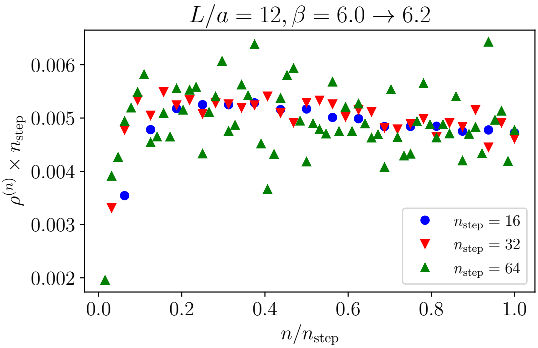


Transferring ρ

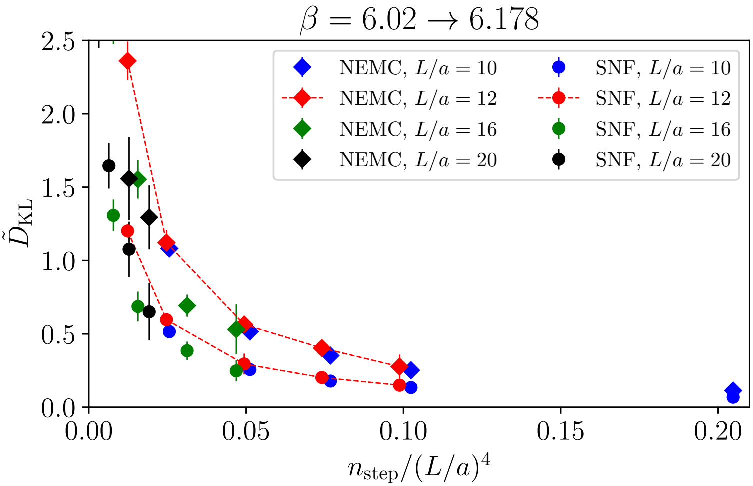
Short trainings: 200-1000 epochs enough to saturate

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

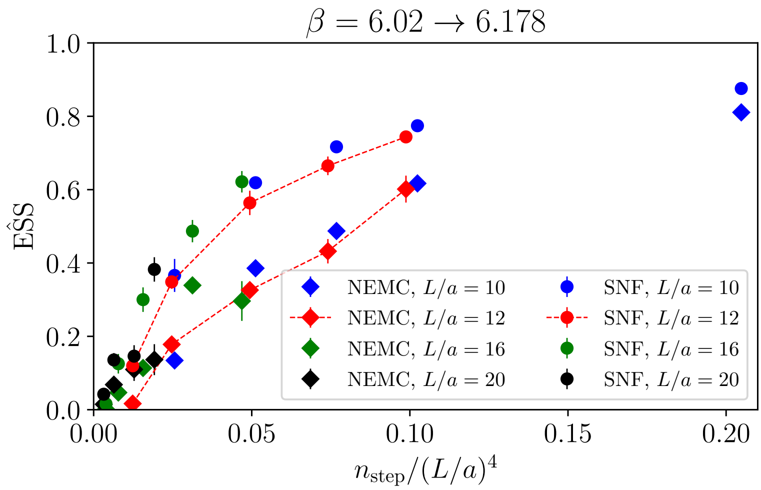
Training only with small n_{step}: clear pattern emerges



SNF volume scaling



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



Main results

- Stochastic approach guarantees a clear scaling with the degrees of freedom

$$n_{\mathrm{step}} \sim \mathrm{d.o.f.} \rightarrow \text{fixed } \tilde{D}_{\mathrm{KL}} \text{ or } \mathrm{ESS}$$

- SNFs improve on NE-MCMC with very cheap training
- thermodynamic understanding of the flow → interpretability

Main results

- Stochastic approach guarantees a clear scaling with the degrees of freedom

$$n_{\mathrm{step}} \sim \mathrm{d.o.f.} \rightarrow \text{fixed } \tilde{D}_{\mathrm{KL}} \text{ or } \mathrm{ESS}$$

- SNFs improve on NE-MCMC with very cheap training
- thermodynamic understanding of the flow $\rightarrow$ interpretability

Future strategy

Systematic improvement over NE-MCMC and current SNFs while retaining the scaling

Better protocols

Huge literature from non-equilibrium SM on optimal protocols

No reason to think linear protocol are particularly efficient

Understand the scaling in β

Better and deeper layers

Include larger loops beyond the plaquette in the smearing

$\rho^{(l)}$ as a neural network as in residual flows **[Abbott et al.; 2023]**

Future prospects

Topological freezing

Implement **SNF for evolutions in the BC**

Base distribution: open BC on a defect L_d $\rightarrow$ target distribution: periodic BC

d.o.f. scales like $(L_d/a)^3$

Work in collaboration with C. Bonanno and D. Vadacchino **[Bonanno et al.; 2024]**

Further applications

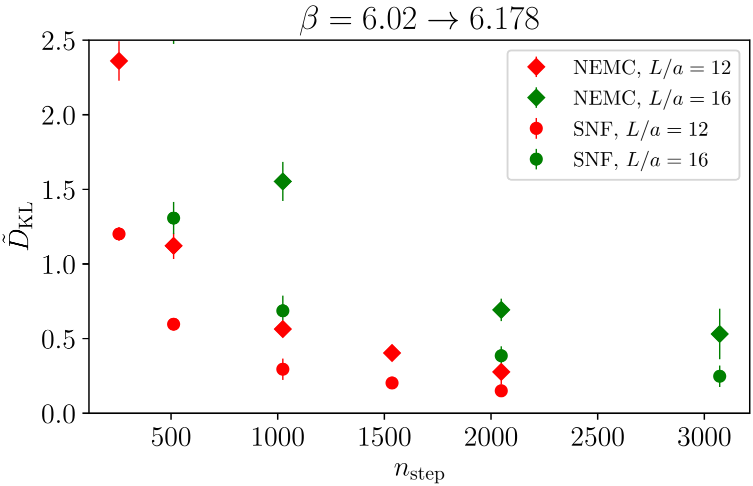
Use SNFs and NE-MCMC to quantitatively study thermalization processes

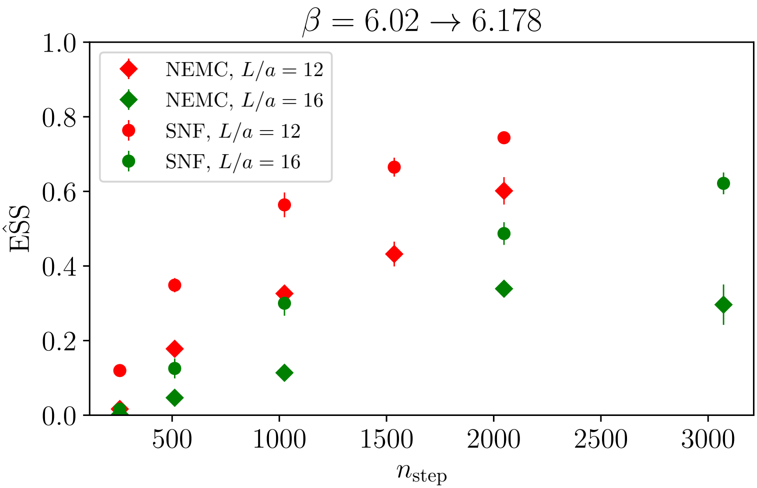
Possible application to master-field simulations

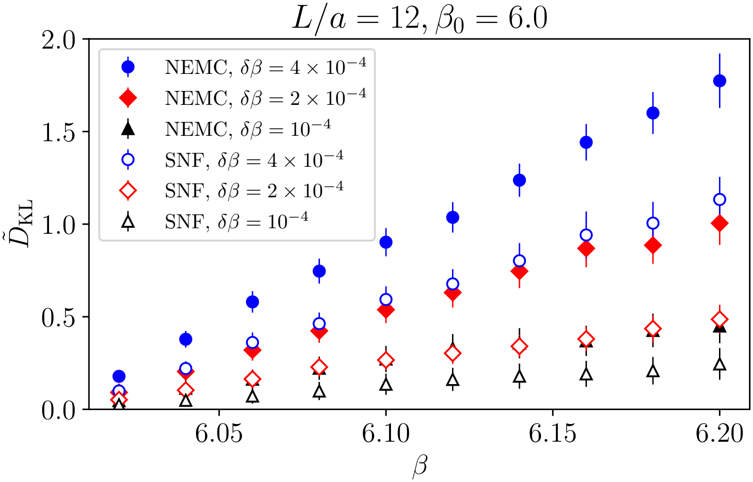
Rethinking thermalization of MCMC

Thank you for your attention!

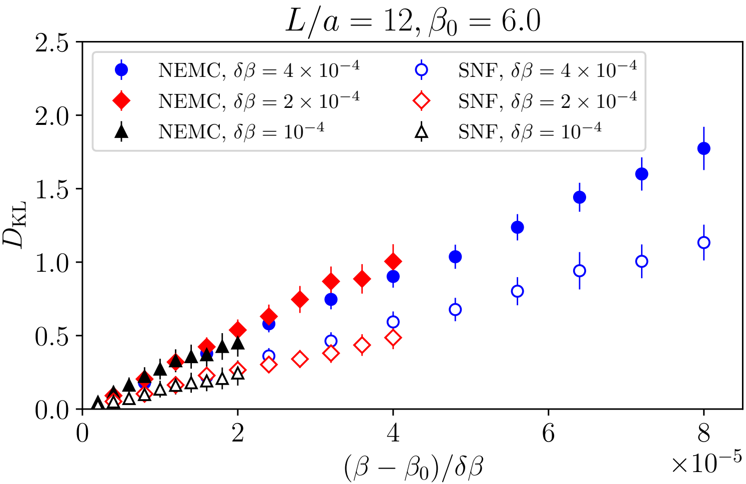
Improvements over purely stochastic approach




$\beta = 6.02 \to 6.178$
1.0
0.8
0.6
0.4
0.2
0.0
$\hat{\mathrm{ESS}}$
500
1000
1500
2000
2500
3000
n_{step}
NEMC, $L/a = 12$
NEMC, $L/a = 16$
SNF, $L/a = 12$
SNF, $L/a = 16$


$L/a = 12, \beta_0 = 6.0$
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

SNF scaling in β

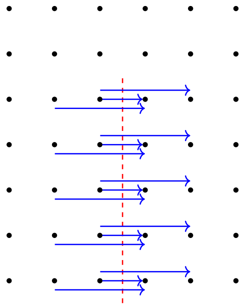


$\rightarrow$ $\delta\beta \sim \beta - \beta_0$ for fixed $\tilde{D}_{\mathrm{KL}}$ (linear protocol)

Based on work with **Claudio Bonanno** and **Davide Vadacchino**

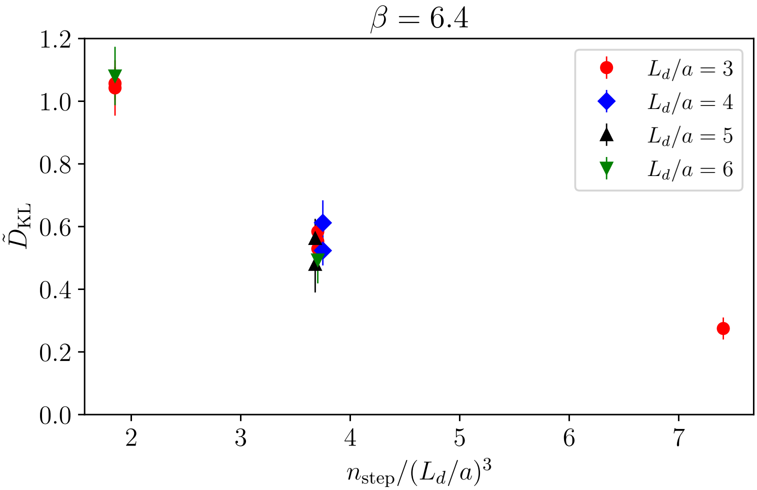
- $\mathrm{CP}(N-1)$ model in 2d **[JHEP 04 (2024) 126, 2402.06561]**

  Promising results: $\tau_{\mathrm{int}} \sim 10^5$ tamed to effectively a few thousands + length of non-equilibrium evolutions scales with defect size
- $\mathrm{SU}(3)$ in 4d: **poster at Lattice2024**

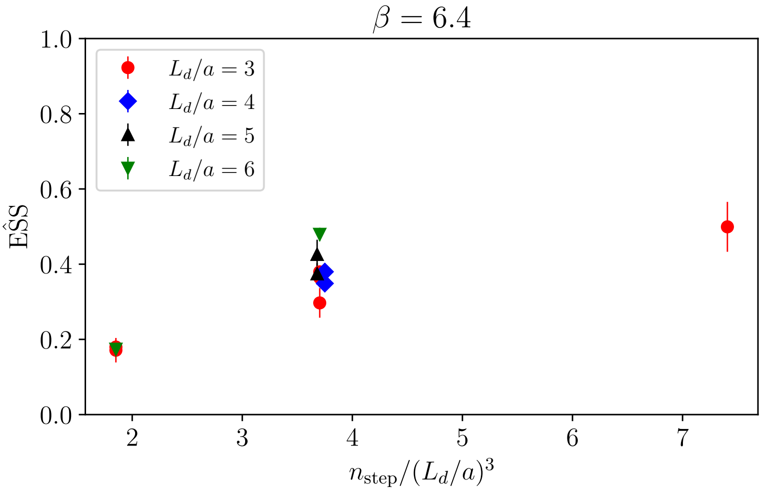


- Parameter controlling the BC is switched linearly until PBC
- Test in 4d $\mathrm{SU}(3)$ at $\beta = 6.4$: scaling with defect and calibration of algorithm for larger βs
- no ML (yet)
- 30^4 lattices at $\beta = 6.4$ ($L = 1.4$fm) with a L_d^3 defect

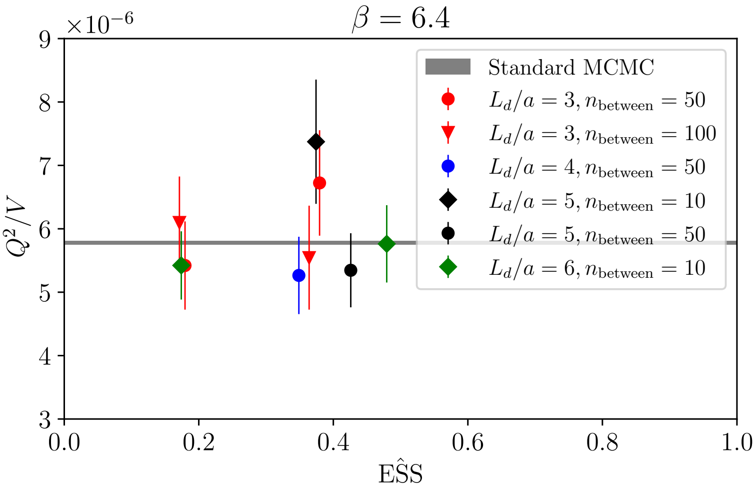
Switching BC in $\mathrm{SU}(3)$: scaling with the defect size



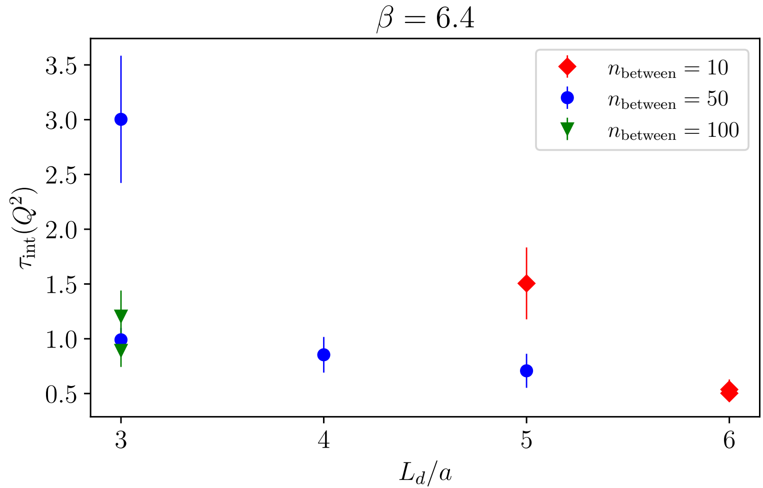
Switching BC in $\mathrm{SU}(3)$: scaling with the defect size



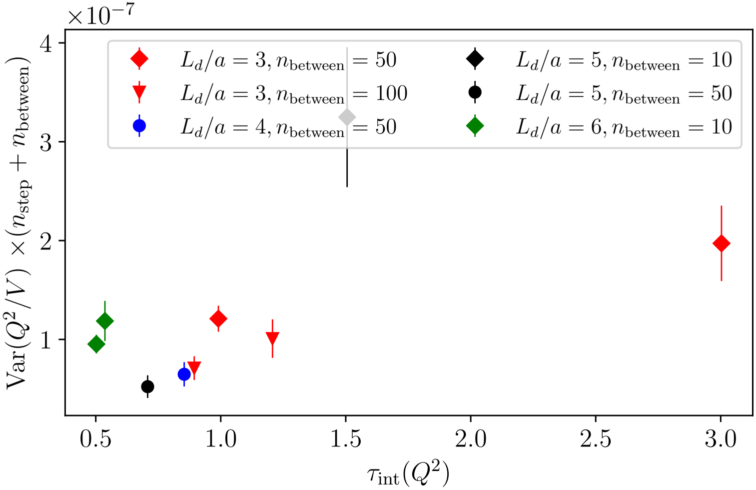
Switching BC in SU(3): topology



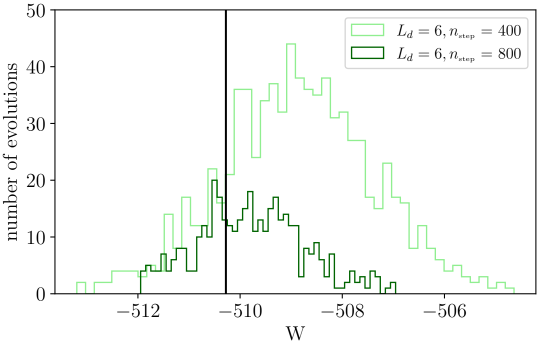
Switching BC in SU(3): autocorrelations

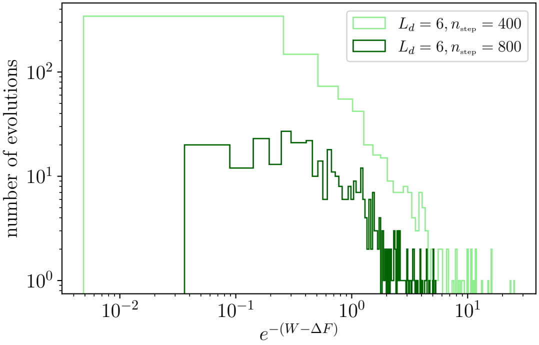


Switching BC in SU(3): efficiency



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





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

A connection to traditional reweighting

A typical reweighting procedure is meant to sample a distribution p using a (close enough) distribution q_0. It can be written as

$$\langle \mathcal{O} \rangle_{\mathrm{RW}} = \frac{\langle \mathcal{O}(\phi) \exp(-\Delta S) \rangle_{q_0}}{\langle \exp(-\Delta S) \rangle_{q_0}}$$

It is just Jarzynski's equality for $n_{\mathrm{step}} = 1$, see the work

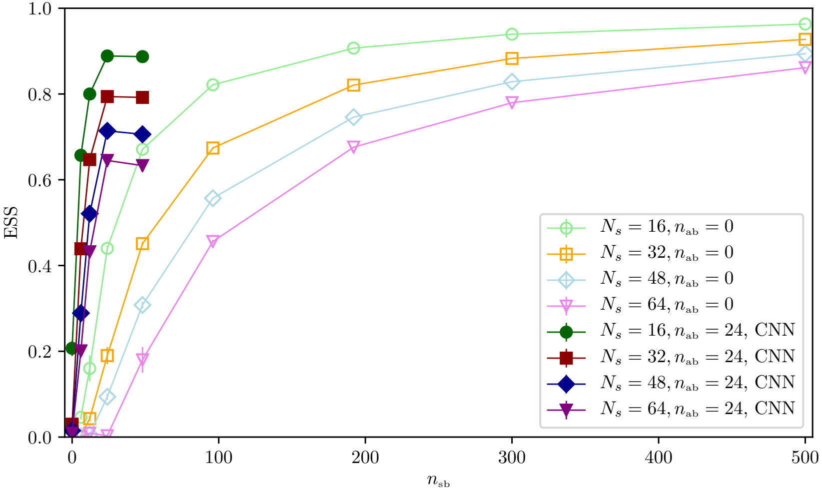
$$W = \sum_{n=0}^{n_{\mathrm{step}}-1} \left\{ S_{c(n+1)}\left[\phi_n\right] - S_{c(n)}\left[\phi_n\right] \right\} = \Delta S(\phi_0)$$

with ϕ_0 sampled from q_0

- It's important to note that there is no issue with the fact that ΔS itself can be large
- The real issue is that the *distribution* of ΔS (and in general of W) can lead to an extremely poor estimate of $\Delta F \rightarrow$ highly inefficient sampling
- The exponential average can be tricky when very far from equilibrium!

SNFs for ϕ^4 at various volumes

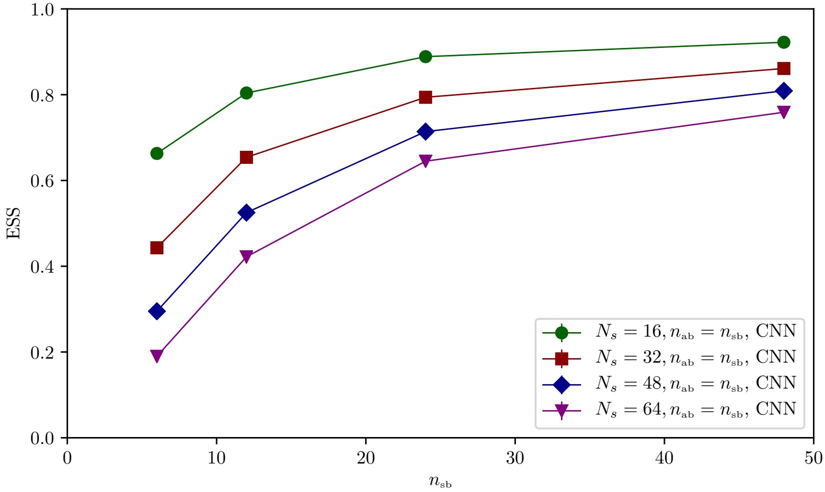
Training length: 10^4 epochs for all volumes. Slowly-improving regime reached fast



Interesting behaviour for all volumes: a peak for

SNFs for ϕ^4 at various volumes

SNFs with $n_{sb} = n_{ab}$ as a possible recipe for efficient scaling



Taking cues from the $\mathrm{SU}(3)$ e.o.s.

Large-scale application: computation of the $\mathrm{SU}(3)$ equation of state **[Caselle et al.; 2018]**

Goal: extract the pressure with Jarzynski's equality

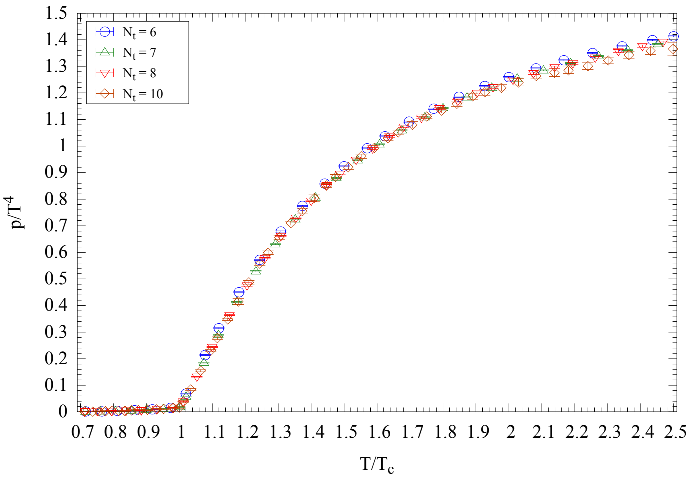
$$\frac{p(T)}{T^4} - \frac{p(T_0)}{T_0^4} = \left(\frac{N_t}{N_s}\right)^3 \log\langle e^{-W_{\mathrm{SU}(N_c)}}\rangle_{\mathrm{f}}$$

evolution in β_g (inverse coupling) $\rightarrow$ changes lattice spacing a $\rightarrow$ changes temperature $T = 1/(aN_t)$

Prior: thermalized Markov chain at a certain $\beta_g^{(0)}$

For systems with many d.o.f. (i.e. large volumes), JE works when N is large, i.e. evolution is slow (and expensive)

SU(3) e.o.s. with Jarzynski's equality



Large volumes (up to $160^3 \times 10$) and very fine lattice spacings $\beta \simeq 7$

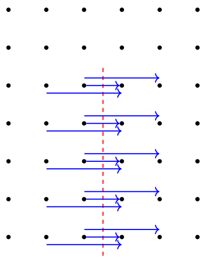
The CP^{N-1} model with a defect

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

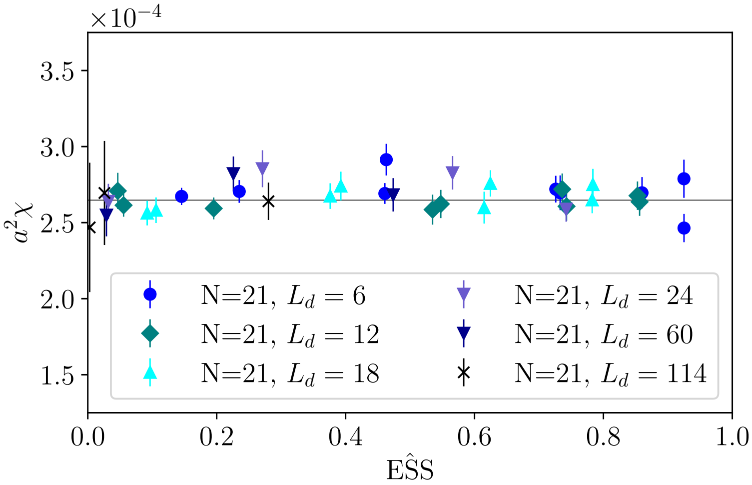


The $k_\mu^{(n)}(x)$ regulate the boundary conditions along a given defect D

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

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