Sampling $\mathrm{SU}(3)$ pure gauge theory with out-of-equilibrium evolutions and Stochastic Normalizing Flows

Alessandro Nada

Università degli Studi di Torino and INFN Torino

ML-LFT Workshop

Swansea, 24th-26th July 2024

in collaboration with

Claudio Bonanno (IFT Madrid), Andrea Bulgarelli (UniTo), Elia Cellini (UniTo) and Davide Vadacchino (University of Plymouth)

Critical slowing down in lattice gauge theory

Long autocorrelation times characterize several observables when $a \to 0$

Typical example are **topological observables**: for $a \to 0$ sectors characterized by different values of the topological charge Q emerge

Using standard MCMC algorithms the transition between these sectors is strongly suppressed

Long autocorrelation times characterize several observables when $a \to 0$

Typical example are **topological observables**: for $a \to 0$ sectors characterized by different values of the topological charge Q emerge

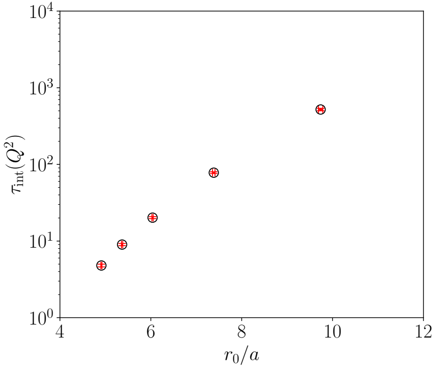
Using standard MCMC algorithms the transition between these sectors is strongly suppressed

This talk: focus on $\mathrm{SU}(3)$ in 4 dimensions

Update algorithm of choice: 1 heat-bath step + 4 over-relaxation steps

Objective: mitigate freezing at $\beta = 6.5$ ($r_0/a \sim 11$)

$\tau_{\mathrm{int}}(Q^2) \sim 10^3$ and $L/a = 36$



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 $p(\phi)$ 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]**

$\rightarrow$ 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 $p(\phi)$ 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! (this talk) $\rightarrow$ better scaling?

Out-of-equilibrium Monte Carlo evolutions

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

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

- $c(n)$ 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 $p = e^{-S_{c(n_{\mathrm{step}})}}/Z_{n_{\mathrm{step}}} \equiv e^{-S}/Z$

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 $p = e^{-S_{c(n_{\mathrm{step}})}}/Z_{n_{\mathrm{step}}} \equiv e^{-S}/Z$

"forward" transition probability

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

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

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

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

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

with the generalized **work**

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

and the **free energy** difference

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

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

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

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

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

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

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 $\langle \exp x \rangle \geq \exp \langle x \rangle$

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

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!

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

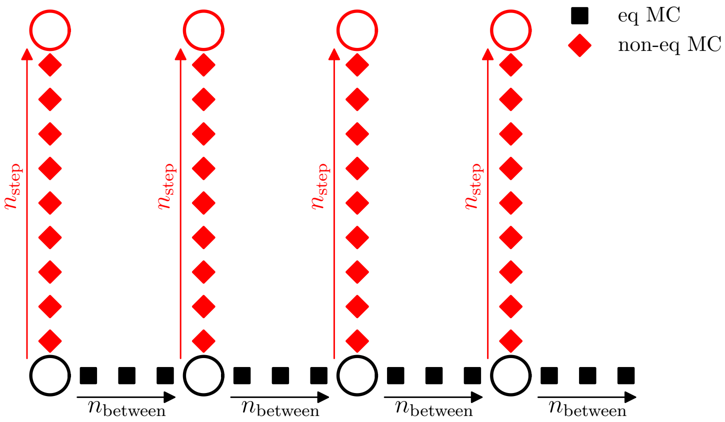
- valid process also far from equilibrium (e.g. n_{step} is "small": $n_{\mathrm{step}} = 1$ is standard reweighting)

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

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

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

A non-equilibrium paradigm to perform MCMC



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]**
- entanglement entropy **[Bulgarelli and Panero; 2023]**
- connection with Stochastic Normalizing Flows: ϕ^4 scalar field theory **[Caselle et al.; 2022]** and Nambu-Goto effective string model **[Caselle et al.; 2023]**
- Topological unfreezing for $\mathrm{CP}(N-1)$ model **[Bonanno et al.; 2023]**

How far are we from equilibrium?

With Normalizing Flows we minimize

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

but $q(U)$ is not tractable in this case

How far are we from equilibrium?

With Normalizing Flows we minimize

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

but $q(U)$ is not tractable in this case

However we can measure the similarity of forward and reverse processes

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

How far are we from equilibrium?

With Normalizing Flows we minimize

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

but $q(U)$ is not tractable in this case

However 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

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

→ measure of how reversible the process is!

Interestingly

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

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

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

Strategies 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**
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

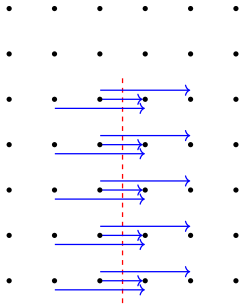
Non-equilibrium evolutions in the boundary conditions

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

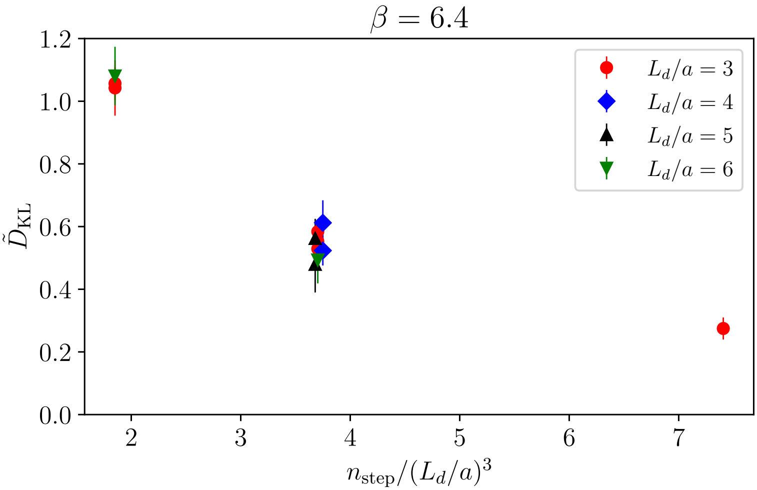
- $\mathrm{CP}(N-1)$ model in 2d **[JHEP 04 (2024) 126, 2402.06561]**

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

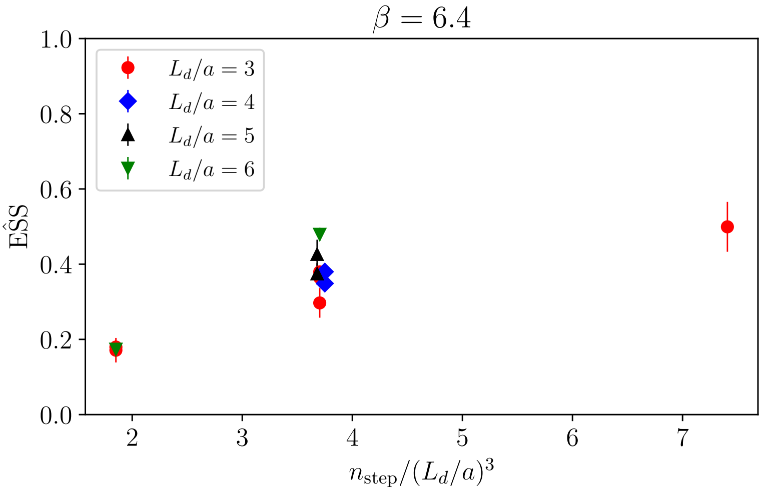
- $\mathrm{SU}(3)$ in 4d: this talk + **poster at Lattice2024**



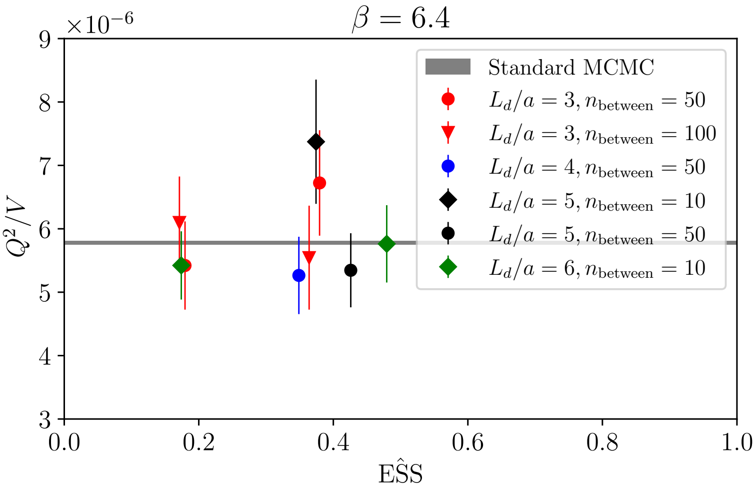
- Parameter controlling the BC is switched linearly until PBC
- Test in 4d $\mathrm{SU}(3)$ at $\beta = 6.4$: scaling with defect and calibration of algorithm for larger βs
- no ML (yet)
- 30^4 lattices at $\beta = 6.4$ ($L = 1.4$fm) with a L_d^3 defect


$\beta = 6.4$
1.2
1.0
0.8
0.6
0.4
0.2
0.0
$\tilde{D}_{\mathrm{KL}}$
2
3
4
5
6
7
$n_{\mathrm{step}}/(L_d/a)^3$
$L_d/a = 3$
$L_d/a = 4$
$L_d/a = 5$
$L_d/a = 6$

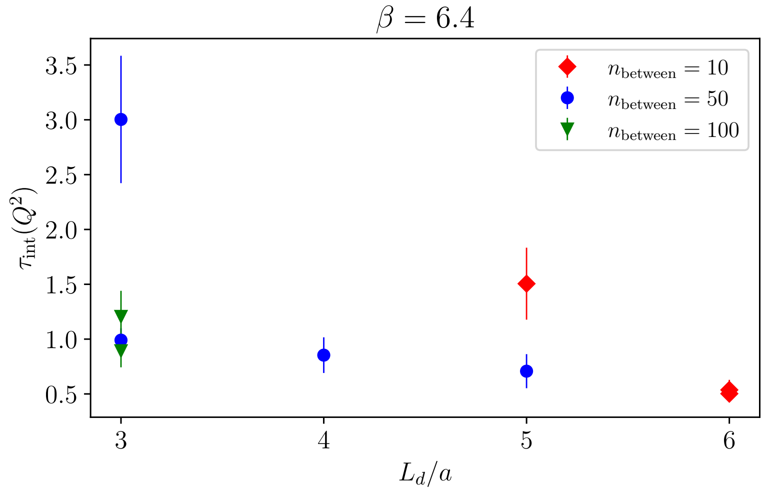
Switching BC in $\mathrm{SU}(3)$: scaling with the defect size



Switching BC in SU(3): topology



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



Non-equilibrium evolutions in β

Implemented for the $\mathrm{SU}(3)$ equation of state [with M. Caselle and M. Panero, PRD 98 (2018) 5, 054513, 1801.03110] on large lattices

Generalization to SNF: this talk + talk at Lattice2024, in collaboration with **Andrea Bulgarelli** and **Elia Cellini**

Strategy

- Inverse coupling β is increased linearly until target value is reached
- Aim: analyze scaling with volume $(L/a)^4$
- Set MCMC standard for flow-based approach
- No topology yet (charge not frozen yet)

Setup:

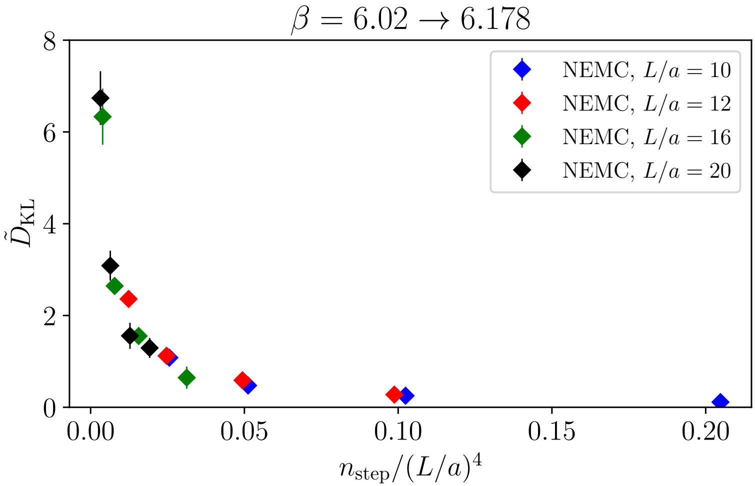
- Increasingly large lattices, from $L/a = 10$ to $L/a = 20$
- "Jump" in β:

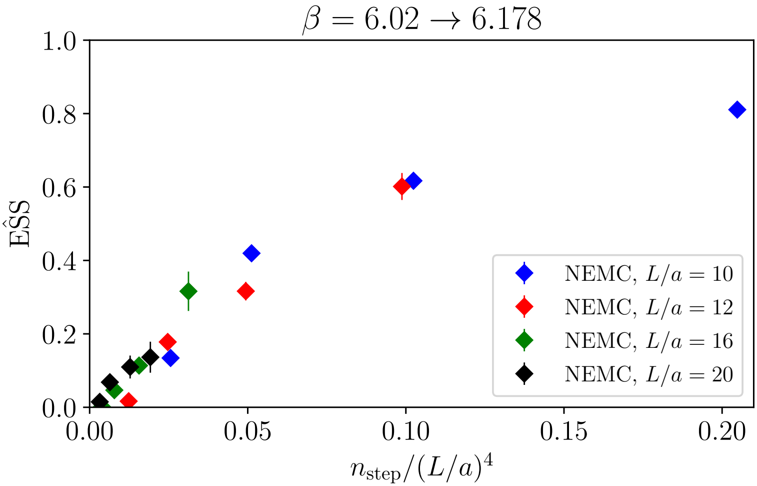
$$6.02 \to 6.178$$

corresponding to

$$(1.8\mathrm{fm})^4 \to (1.4\mathrm{fm})^4$$

for $L/a = 20$

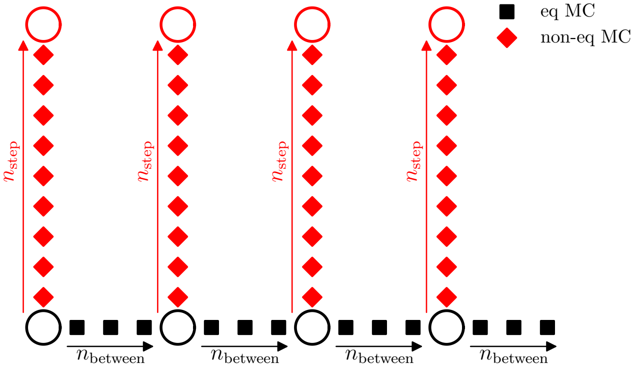

$\beta = 6.02 \to 6.178$
$\tilde{D}_{\mathrm{KL}}$
$n_{\mathrm{step}}/(L/a)^4$
0
2
4
6
8
0.00
0.05
0.10
0.15
0.20
NEMC, $L/a = 10$
NEMC, $L/a = 12$
NEMC, $L/a = 16$
NEMC, $L/a = 20$


$\beta = 6.02 \to 6.178$
1.0
0.8
0.6
0.4
0.2
0.0
$\hat{\mathrm{ESS}}$
0.00
0.05
0.10
0.15
0.20
$n_{\mathrm{step}}/(L/a)^4$
NEMC, $L/a = 10$
NEMC, $L/a = 12$
NEMC, $L/a = 16$
NEMC, $L/a = 20$

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?

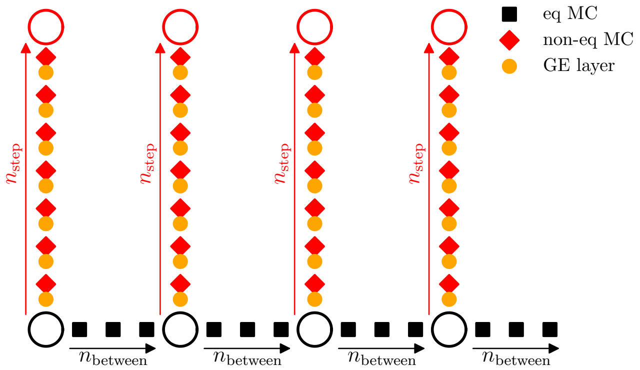


SNFs as systematic improvement of non-equilibrium evolutions

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]**)

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



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.; 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

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

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

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

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

$$\Omega_\mu(x) = \underbrace{C_\mu(x)}_{\text{frozen}}\underbrace{U^\dagger_\mu(x)}_{\text{active}}$$

Sum of frozen staples

$$C_\mu(x) = \sum_{\nu\neq\mu} \rho_{\mu\nu}(x)\, \underbrace{S_{\mu\nu}(x)}_{\text{staple}}$$

in this work: $\rho_{\mu\nu}(x) \longrightarrow \rho$

$\rightarrow$ 1 parameter per mask/8 parameters per layer

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

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

Short trainings: 200-1000 epochs

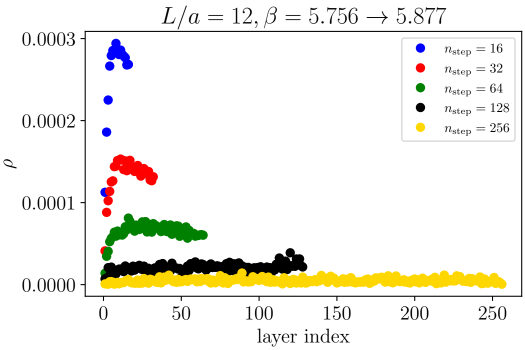
Memory issues for large n_{step} and large volumes

Practical solution: train each layer separately during the non-equilibrium evolution $\rightarrow$ reminiscent of CRAFT **[Matthews at al.; 2022]**

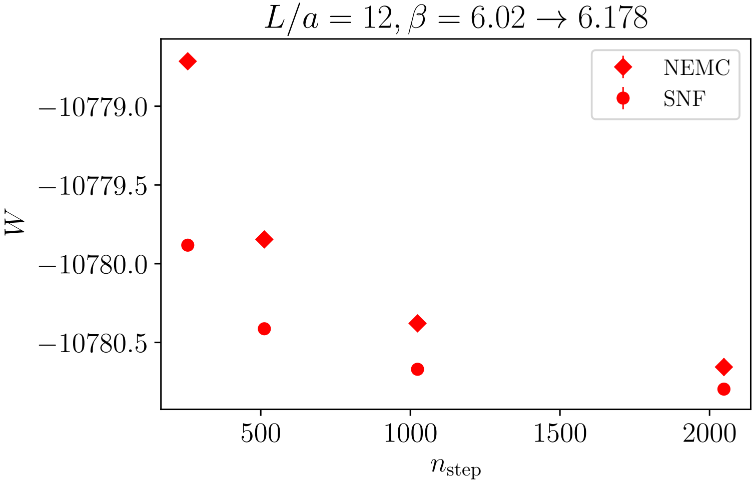
Heavy use of **transfer learning** for each $\beta_0 \rightarrow \beta$ evolution:

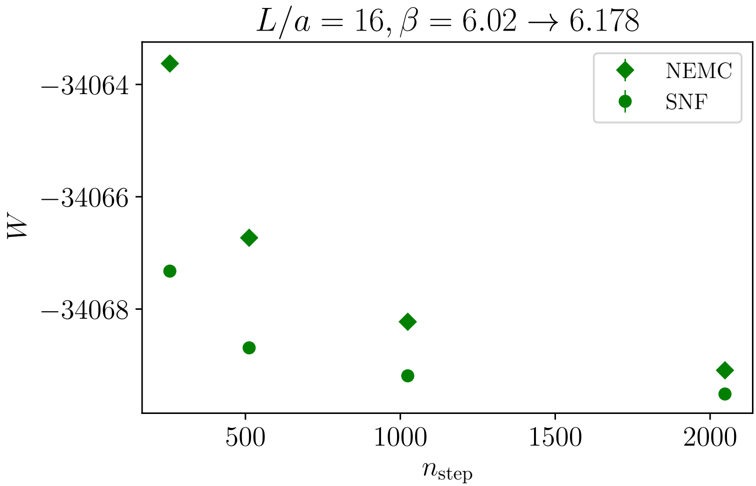
- training only at small volumes
- training only with small n_{step}: global interpolation of ρ

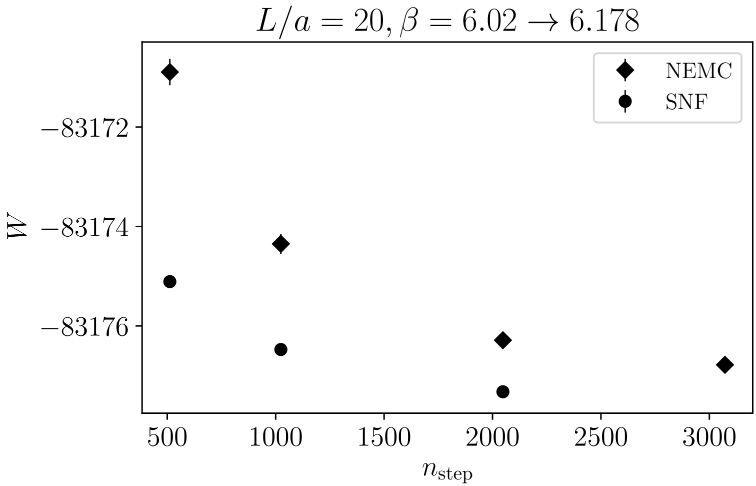
No retraining!

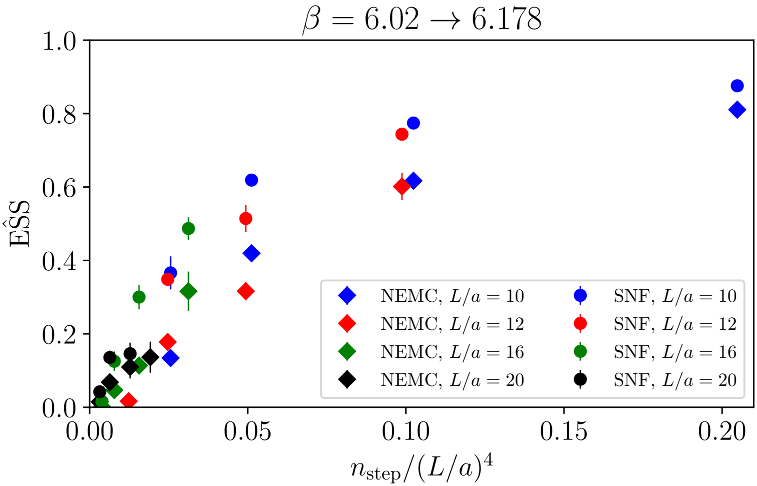


Improvements over purely stochastic approach




$L/a = 16, \beta = 6.02 \to 6.178$
NEMC
SNF
−34064
−34066
−34068
W
500
1000
1500
2000
n_{step}


$L/a = 20, \beta = 6.02 \to 6.178$
NEMC
SNF
W
−83172
−83174
−83176
500
1000
1500
2000
2500
3000
n_{step}


$\beta = 6.02 \rightarrow 6.178$
1.0
0.8
0.6
0.4
0.2
0.0
$\hat{\mathrm{ESS}}$
0.00
0.05
0.10
0.15
0.20
$n_{\mathrm{step}}/(L/a)^4$
NEMC, $L/a = 10$
NEMC, $L/a = 12$
NEMC, $L/a = 16$
NEMC, $L/a = 20$
SNF, $L/a = 10$
SNF, $L/a = 12$
SNF, $L/a = 16$
SNF, $L/a = 20$

Conclusions and future prospects

Stochastic approach guarantees a clear scaling with the degrees of freedom

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

while providing a thermodynamic understanding of the flow

Conclusions and future prospects

Stochastic approach guarantees a clear scaling with the degrees of freedom

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

while providing a thermodynamic understanding of the flow

Overall strategy

systematically improve on stochastic approach by machine-learning deterministic transformations between MC steps

Future improvements

Better protocols (huge literature from non-eq SM): only linear protocols were used in this work!

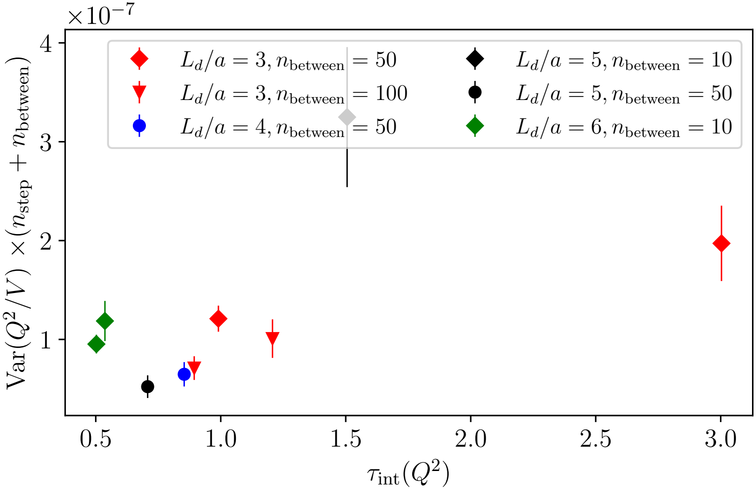
Better and deeper layers: include larger loops beyond the plaquette + ρ as a neural network **[Abbott et al.; 2023]**

Future implementations

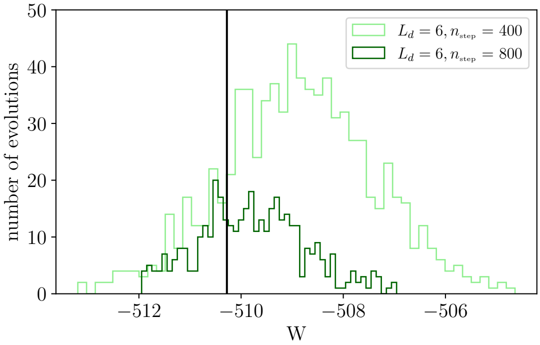
Implement **SNF for evolutions in the BC**

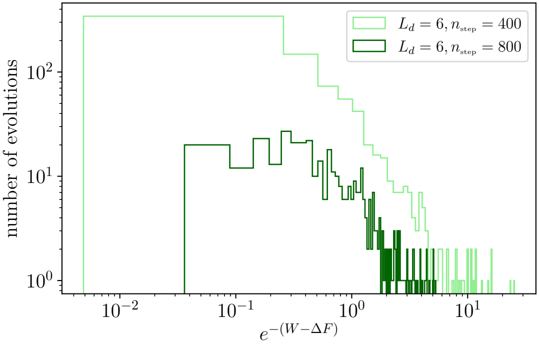
Push **SNFs/evolutions in β** at finer lattice spacings

Thank you for your attention!

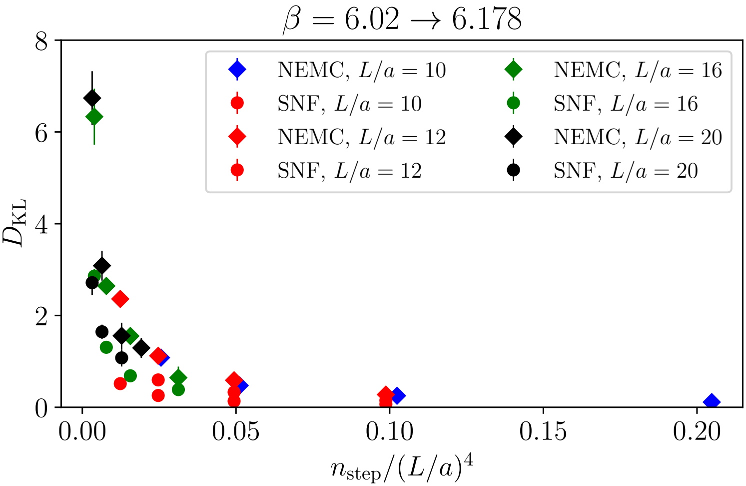

$\times 10^{-7}$
4
3
2
1
$\text{Var}(Q^2/V) \times (n_{\text{step}} + n_{\text{between}})$
0.5
1.0
1.5
2.0
2.5
3.0
$\tau_{\text{int}}(Q^2)$
$L_d/a = 3, n_{\text{between}} = 50$
$L_d/a = 3, n_{\text{between}} = 100$
$L_d/a = 4, n_{\text{between}} = 50$
$L_d/a = 5, n_{\text{between}} = 10$
$L_d/a = 5, n_{\text{between}} = 50$
$L_d/a = 6, n_{\text{between}} = 10$

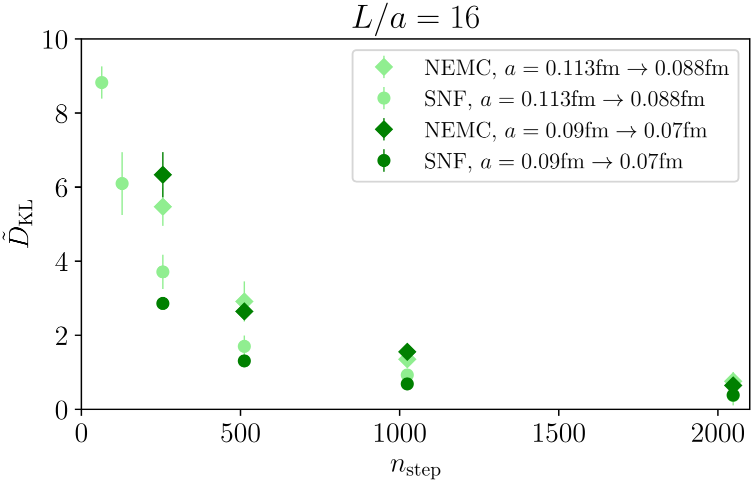
Switching BC in $\mathrm{SU}(3)$: work histograms

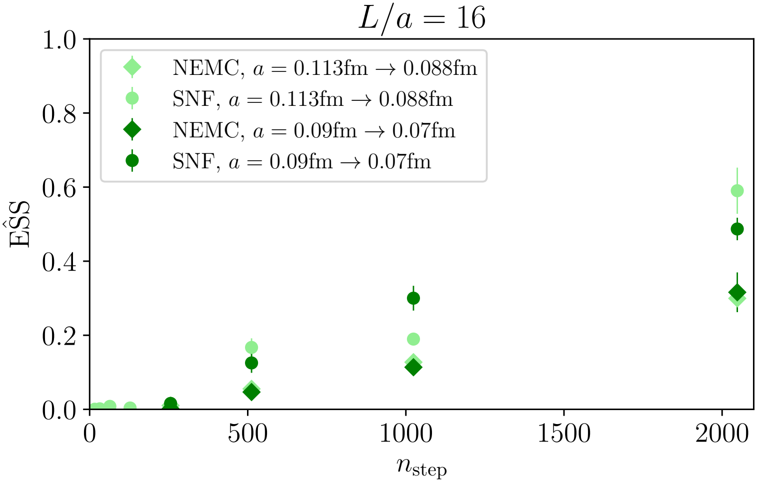




SNF evolutions in β: volume scaling




$L/a = 16$
10
8
6
4
2
0
0
500
1000
1500
2000
$\tilde{D}_{\mathrm{KL}}$
n_{step}
NEMC, $a = 0.113\mathrm{fm} \to 0.088\mathrm{fm}$
SNF, $a = 0.113\mathrm{fm} \to 0.088\mathrm{fm}$
NEMC, $a = 0.09\mathrm{fm} \to 0.07\mathrm{fm}$
SNF, $a = 0.09\mathrm{fm} \to 0.07\mathrm{fm}$


$L/a = 16$
NEMC, $a = 0.113\text{fm} \to 0.088\text{fm}$
SNF, $a = 0.113\text{fm} \to 0.088\text{fm}$
NEMC, $a = 0.09\text{fm} \to 0.07\text{fm}$
SNF, $a = 0.09\text{fm} \to 0.07\text{fm}$
$\hat{\text{ESS}}$
1.0
0.8
0.6
0.4
0.2
0.0
0
500
1000
1500
2000
n_{step}

The Second Law of Thermodynamics

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

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

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 $\Delta F \to$ highly inefficient sampling
- The exponential average can be tricky when very far from equilibrium!

A common framework: Stochastic Normalizing Flows

Jarzynski's equality is the same formula used to extract Z in NFs

$$\frac{Z}{Z_0} = \langle \tilde{w}(\phi) \rangle_{\phi \sim q_N} = \langle \exp(-W) \rangle_{\mathrm{f}}$$

The exponent of the weight is always of the form (note that for NFs $\langle \dots \rangle_{\phi \sim q_N} = \langle \dots \rangle_{\mathrm{f}}$)

$$W(\phi_0, \dots, \phi_N) = S(\phi_N) - S_0(\phi_0) - Q(\phi_1, \dots, \phi_N)$$

Normalizing Flows

$$\phi_0 \to \phi_1 = g_1(\phi_0) \to \dots \to \phi_N$$

$$\text{"} Q \text{"} = \log J = \sum_{n=0}^{N-1} \log |\det J_n(\phi_n)|$$

stochastic non-equilibrium evolutions

$$\phi_0 \overset{P_{c(1)}}{\to} \phi_1 \overset{P_{c(2)}}{\to} \dots \overset{P_{c(N)}}{\to} \phi_N$$

$$Q = \sum_{n=0}^{N-1} S_{c(n+1)}(\phi_{n+1}) - S_{c(n+1)}(\phi_n)$$

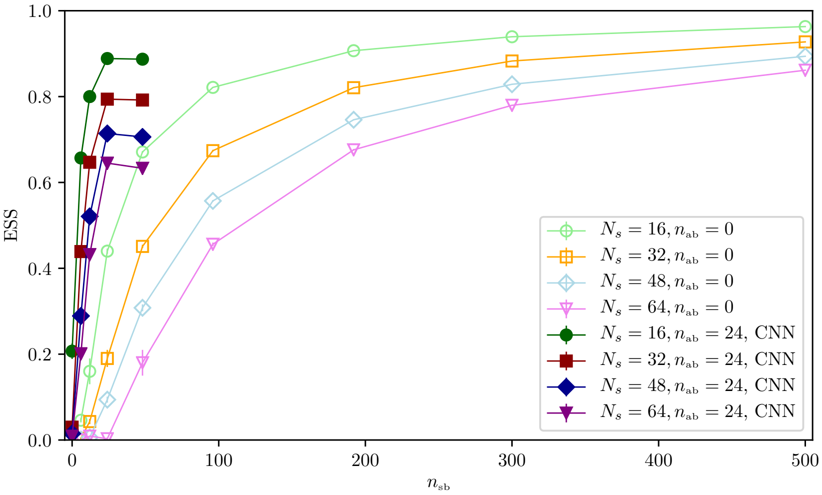
Stochastic Normalizing Flows (introduced in **[Wu et al.; 2020]**)

$$\phi_0 \to g_1(\phi_0) \overset{P_{c(1)}}{\to} \phi_1 \to g_2(\phi_1) \overset{P_{c(2)}}{\to} \dots \overset{P_{c(N)}}{\to} \phi_N$$

$$Q = \sum_{n=0}^{N-1} S_{c(n+1)}(\phi_{n+1}) - S_{c(n+1)}(g_n(\phi_n)) + \log |\det J_n(\phi_n)|$$

SNFs for ϕ^4 at various volumes

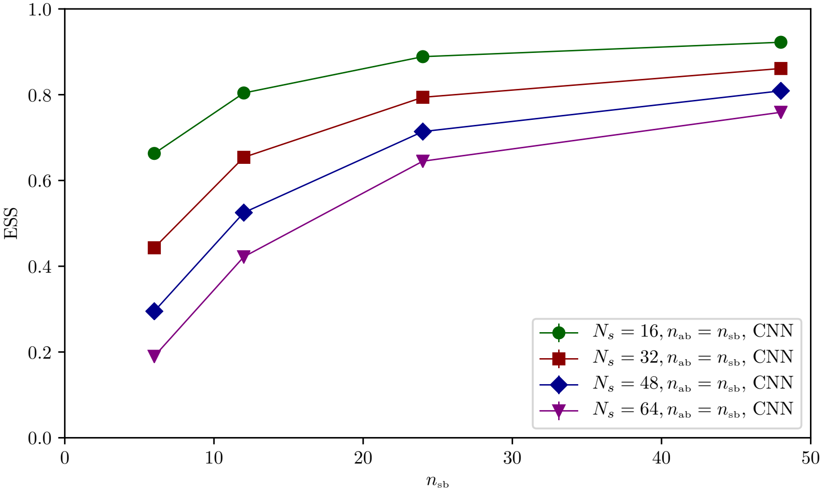
Training length: 10^4 epochs for all volumes. Slowly-improving regime reached fast



Interesting behaviour for all volumes: a peak for $n_{\mathrm{sb}} = n_{\mathrm{ab}}$?

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]**

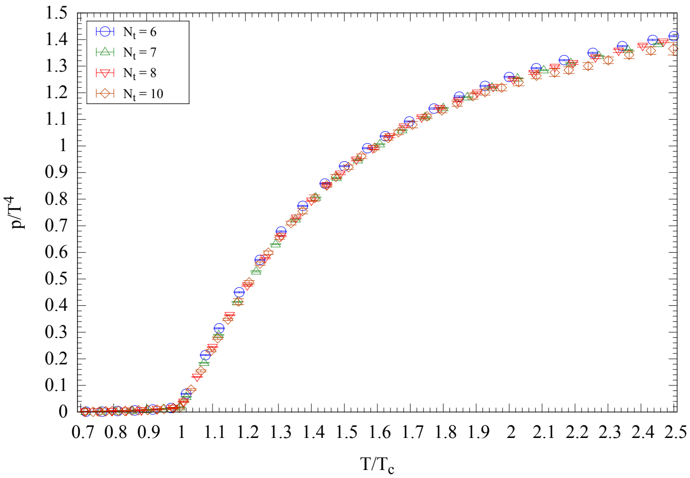
Goal: extract the pressure with Jarzynski's equality

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

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)



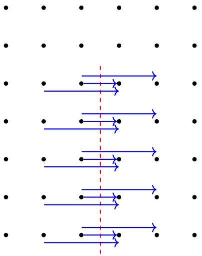
Large volumes (up to $160^3 \times 10$) and very fine lattice spacings $\beta \simeq 7$

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



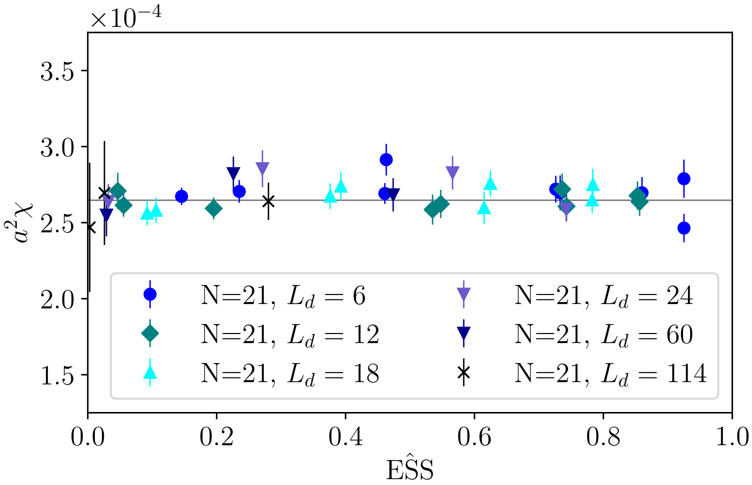
The $k_\mu^{(n)}(x)$ regulate the boundary conditions along a given defect D

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

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