Flow-based approaches for Monte Carlo acceleration and variance reduction

Alessandro Nada

Università degli Studi di Torino and INFN, Sezione di Torino

Seminar @ Università di Milano-Bicocca, 6th July 2026

Flow-based approaches

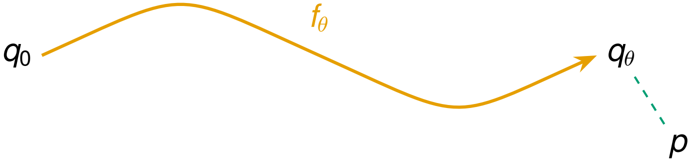
[Albergo, Kanwar, Shanahan; 1904.12072]

Problem: sample from a distribution p is extremely difficult or expensive

Solution: find a *flow* f_θ between a base distribution q_0 and the target p

it transforms the field variables

$$U = f_\theta(U_0) \qquad q_\theta(U) = q_0(U_0)|\det J_{f_\theta}|^{-1} \qquad U_0 \sim q_0$$



Flow-based approaches

[Albergo, Kanwar, Shanahan; 1904.12072]

Problem: sample from a distribution p is extremely difficult or expensive

Solution: find a *flow* f_θ between a base distribution q_0 and the target p

it transforms the field variables

$$U = f_\theta(U_0) \qquad q_\theta(U) = q_0(U_0)|\det J_{f_\theta}|^{-1} \qquad U_0 \sim q_0$$

We eventually sample from $q_\theta \approx p$: it is exact via a reweighting procedure

$$\langle \mathcal{O} \rangle_p = \int \mathcal{D}U \underbrace{q_\theta(U)}_{\text{sample}} \underbrace{\overbrace{\frac{p(U)}{q_\theta(U)}}^{w} \mathcal{O}(U)}_{\text{measure}} = \langle \mathcal{O}(U) w(U) \rangle_q$$

New problem: how to find a flow with *reasonable* weights?

A single framework for different issues?

Flow-based approaches as crucial pieces of future algorithmic development!

1 – Critical slowing down

Flow from a "well-behaved" q_0 towards a p with diverging autocorrelations

Tool: non-equilibrium MC and Stochastic Normalizing Flows

A single framework for different issues?

Flow-based approaches as crucial pieces of future algorithmic development!

1 – Critical slowing down

Flow from a "well-behaved" q_0 towards a p with diverging autocorrelations

Tool: non-equilibrium MC and Stochastic Normalizing Flows

2 – Signal-to-noise ratio problem

Reframe issue using a one-point function evaluated with a source-deformed distribution p

Tool: infinitesimal flows

A non-equilibrium framework to mitigate critical slowing down

In collaboration with C. Bonanno, A. Bulgarelli, E. Cellini, D. Panfalone,

D. Vadacchino, L. Verzichelli

Autocorrelations in Markov Chains

Markov Chain Monte Carlo: a transition probability P_p can generate a sequence of U sampled according to p

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

used to estimate the path integral

$$\langle \mathcal{O} \rangle_p \simeq \frac{1}{N} \sum_{n=0}^{N} \mathcal{O}(U^{(t+n)})$$

Autocorrelations in Markov Chains

Markov Chain Monte Carlo: a transition probability P_p can generate a sequence of U sampled according to p

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

used to estimate the path integral

$$\langle \mathcal{O} \rangle_p \simeq \frac{1}{N} \sum_{n=0}^{N} \mathcal{O}(U^{(t+n)})$$

Subsequent configurations are **autocorrelated**

Autocorrelation measured in the error by $\tau_{\mathrm{int}}(\mathcal{O})$ $\rightarrow$ effectively independent configurations = $N/2\tau_{\mathrm{int}}(\mathcal{O})$

Critical slowing down

[Schaefer, Sommer, Virotta; 1009.5228]

- When a critical point is approached τ_{int} diverges [Wolff; 1990]
- E.g. in the continuum limit $a \to 0$

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

z depends on algorithm/observable

Critical slowing down

[Schaefer, Sommer, Virotta; 1009.5228]

- When a critical point is approached τ_{int} diverges **[Wolff; 1990]**
- E.g. in the continuum limit $a \to 0$

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

z depends on algorithm/observable

Particularly for topological observables Q, Q^2

- On the lattice: emergence of topological sectors (with integer Q) for $a \to 0$
- Transition between sectors strongly suppressed using standard MCMC
- Effective loss of ergodicity

Critical slowing down

[Schaefer, Sommer, Virotta; 1009.5228]

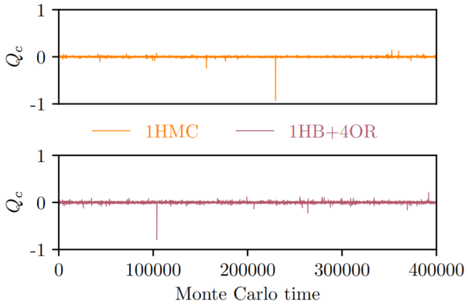
- When a critical point is approached $\tau_{\rm int}$ diverges [Wolff; 1990]
- E.g. in the continuum limit $a \to 0$

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

z depends on algorithm/observable

Particularly for topological observables Q, Q^2

- On the lattice: emergence of topological sectors (with integer Q) for $a \to 0$
- Transition between sectors strongly suppressed using standard MCMC
- Effective loss of ergodicity



[Eichhorn et al.; 2307.04742]

Topological freezing in $\mathrm{SU}(3)$ Yang–Mills

- Strong freezing of topology at $\beta \geq 6.5$ ($r_0/a > 11 \to a < 0.04$fm)
- $\tau_{\mathrm{int}}(Q^2) > 10^3$ with 1 heat-bath step + 4 over-relaxation steps ($z \sim 5.5$)

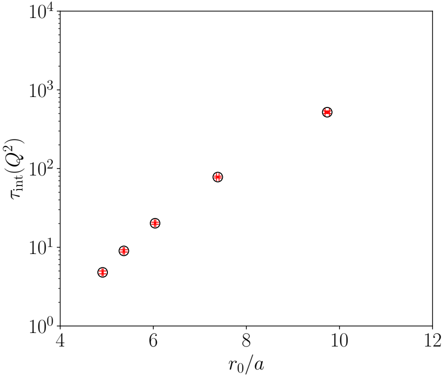
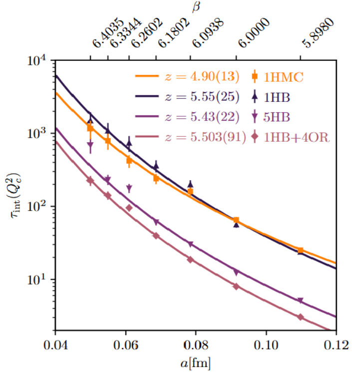


Image courtesy of C. Bonanno

Topological freezing in $\mathrm{SU}(3)$ Yang–Mills

- Strong freezing of topology at $\beta \geq 6.5$ ($r_0/a > 11 \rightarrow a < 0.04$fm)
- $\tau_{\mathrm{int}}(Q^2) > 10^3$ with 1 heat-bath step + 4 over-relaxation steps ($z \sim 5.5$)

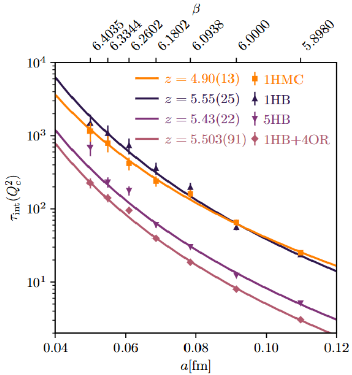


[Eichhorn et al.; 2307.04742]

Topological freezing in $\mathrm{SU}(3)$ Yang–Mills

- Strong freezing of topology at $\beta \geq 6.5$ ($r_0/a > 11 \to a < 0.04$fm)
- $\tau_{\rm int}(Q^2) > 10^3$ with 1 heat-bath step + 4 over-relaxation steps ($z \sim 5.5$)
- **Open Boundary Conditions** remove the sectors and mitigate the issue **[Lüscher and Schaefer; 1105.4749]**

  $\to$ but strong boundary effects
- more general and effective solution needed



[Eichhorn et al.; 2307.04742]

Non-equilibrium Monte Carlo

Out-of-equilibrium evolutions

annealing procedure: sampling each 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

annealing procedure: sampling each 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 **base** distribution $q_0 = e^{-S_{c(0)}}/Z_0$

Out-of-equilibrium evolutions

annealing procedure: sampling each step from a sequence of distributions

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

start **at equilibrium** from the **base** 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** $p = e^{-S}/Z$

Out-of-equilibrium evolutions

annealing procedure: sampling each 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 **base** 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** $p = e^{-S}/Z$

"forward" transition probability

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

Crooks' theorem

if the update algorithm satisfies detailed balance $\rightarrow$

Crooks' theorem for MCMC **[Crooks; 1999]**

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

with the generalized **work** and **free energy** difference

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

Jarzynski equality for MCMC

Define new average over all paths

$$\langle \dots \rangle_{\mathrm{f}} = \int [\mathcal{D}U_0 \dots \mathcal{D}U_{n_{\mathrm{step}}}] q_0(U_0) \mathcal{P}_{\mathrm{f}}[U_0, \dots, U_{n_{\mathrm{step}}}] \dots$$

Integrate Crooks and obtain...

Jarzynski equality for MCMC

Define new average over all paths

$$\langle \dots \rangle_{\mathrm{f}} = \int [\mathcal{D}U_0 \dots \mathcal{D}U_{n_{\mathrm{step}}}] q_0(U_0) \mathcal{P}_{\mathrm{f}}[U_0, \dots, U_{n_{\mathrm{step}}}] \dots$$

Integrate Crooks and obtain...

Jarzynski equality for MCMC [Jarzynski; 1997]

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

Jarzynski equality for MCMC

Define new average over all paths

$$\langle \ldots \rangle_{\mathrm{f}} = \int [\mathcal{D}U_0 \ldots \mathcal{D}U_{n_{\mathrm{step}}}] q_0(U_0) \mathcal{P}_{\mathrm{f}}[U_0, \ldots, U_{n_{\mathrm{step}}}] \ldots$$

Integrate Crooks and obtain...

Jarzynski equality for MCMC [Jarzynski; 1997]

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

Using Jensen's inequality $\langle \exp x \rangle \geq \exp \langle x \rangle$ we get the Second Law of Thermodynamics

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

A very similar formula holds for v.e.v.

NEMC estimator

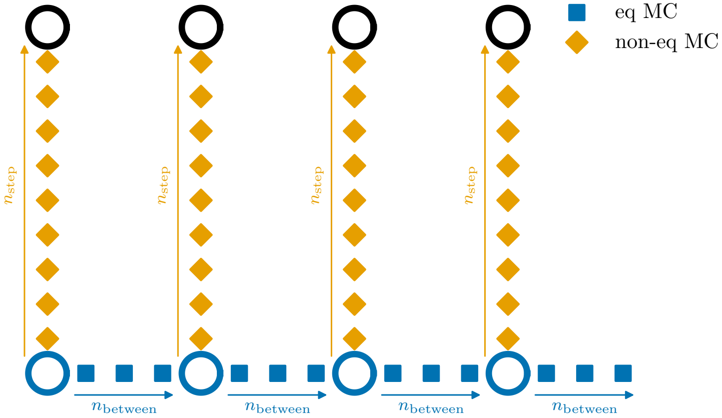
$$\boxed{\langle \mathcal{O} \rangle_p = \frac{\langle \mathcal{O} \ \exp(-W) \rangle_{\mathrm{f}}}{\langle \exp(-W) \rangle_{\mathrm{f}}} = \langle \mathcal{O} \ \exp(-W_{\mathrm{d}}) \rangle_{\mathrm{f}}}$$

with the <u>dissipated work</u> $W_{\mathrm{d}} = W - \Delta F$

$$\langle \exp\left(-W_{\mathrm{d}}\right) \rangle_{\mathrm{f}} = 1$$

Important for later: $\langle W_{\mathrm{d}} \rangle_{\mathrm{f}} \geq 0$ is a measure of dissipation and characterizes the evolution

A non-equilibrium paradigm to perform Monte Carlo



Jarzynski equality and lattice field theory

Computation of free energies and/or sampling problematic distributions

Jarzynski equality and lattice field theory

Computation of free energies and/or sampling problematic distributions

- Interface free-energy in the Z_2 gauge theory **[Caselle et al.; 1604.05544]**
- Equation of state (pressure) in 4d $\mathrm{SU}(3)$**[Caselle et al.; 1801.03110]**
- Renormalized coupling in $\mathrm{SU}(N)$ YM theories **[Francesconi et al.; 2003.13734]**

- Entanglement entropy **[Bulgarelli and Panero; 2304.03311]**, also with (S)NFs **[Bulgarelli et al.; 2410.14466]**
- Numerical simulations of Effective String Theory **[Caselle et al.; 2409.15937]**

- Casimir effect in $O(N)$ models **[Bulgarelli et al.; 2505.20403]**

Jarzynski equality and lattice field theory

Computation of free energies and/or sampling problematic distributions

- Interface free-energy in the Z_2 gauge theory **[Caselle et al.; 1604.05544]**
- Equation of state (pressure) in 4d $\mathrm{SU}(3)$**[Caselle et al.; 1801.03110]**
- Renormalized coupling in $\mathrm{SU}(N)$ YM theories **[Francesconi et al.; 2003.13734]**
- Connection with Stochastic Normalizing Flows in ϕ^4 **[Caselle et al.; 2201.08862]** and $\mathrm{SU}(3)$ **[Bulgarelli et al.; 2412.00200]**
- Entanglement entropy **[Bulgarelli and Panero; 2304.03311]**, also with (S)NFs **[Bulgarelli et al.; 2410.14466]**
- Numerical simulations of Effective String Theory **[Caselle et al.; 2409.15937]**
- Topological unfreezing for $\mathrm{CP}^{(N-1)}$ model **[Bonanno et al.; 2402.06561]** and 4d $\mathrm{SU}(3)$ **[Bonanno et al.; 2510.25704]**
- Casimir effect in $O(N)$ models **[Bulgarelli et al.; 2505.20403]**

How far are we from equilibrium?

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

$\rightarrow$ measure the similarity of forward and reverse processes

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

How far are we from equilibrium?

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

$\rightarrow$ measure the similarity of forward and reverse processes

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

Clear “thermodynamic” interpretation

$$\tilde{D}_{\mathrm{KL}}(q_0\mathcal{P}_{\mathrm{f}} \| p\mathcal{P}_{\mathrm{r}}) = \langle W \rangle_{\mathrm{f}} - \Delta F \equiv \underbrace{\langle W_{\mathrm{d}} \rangle_{\mathrm{f}} \geq 0}_{\text{Second Law}}$$

$\rightarrow$ measure of **reversibility**!

Note that

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

The effective sample size

We want our estimator to have small variance

$$\langle \mathcal{O} \rangle = \langle \mathcal{O} \ \exp(-W_{\mathrm{d}}) \rangle_{\mathrm{f}}$$

Reasonable approximation: minimize

$$\mathrm{Var}(\exp(-W_{\mathrm{d}})) = \langle \exp(-2W_{\mathrm{d}}) \rangle_{\mathrm{f}} - 1$$

In generative models often one uses the **Effective Sample Size**

$$\hat{\mathrm{ESS}} = \frac{\langle \exp(-W) \rangle_{\mathrm{f}}^2}{\langle \exp(-2W) \rangle_{\mathrm{f}}} = \frac{1}{\langle \exp(-2W_{\mathrm{d}}) \rangle_{\mathrm{f}}} = \frac{1}{1 + \mathrm{Var}(\exp(-W_{\mathrm{d}}))}$$

for which

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

NEMC in BC – $\mathrm{SU}(3)$ Yang–Mills

Can we use NEMC to improve current methods for topological unfreezing?

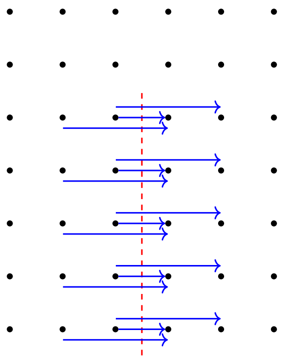
NEMC in BC – $\mathrm{SU}(3)$ Yang–Mills

Can we use NEMC to improve current methods for topological unfreezing?

Proposal: **NEMC in boundary conditions**

- Base dist: Open BC on a L_d^3 defect
- Protocol: switch $0 \to \beta$ only on boundary defect links
- # of relevant dof $\sim (L_d/a)^3$
- “Dual” to Parallel Tempering
- Goal: full mitigation in $\mathrm{SU}(3)$ pure gauge

[Bonanno, AN, Vadacchino; 2402.06561]
[Bonanno et al.; 2510.25704]

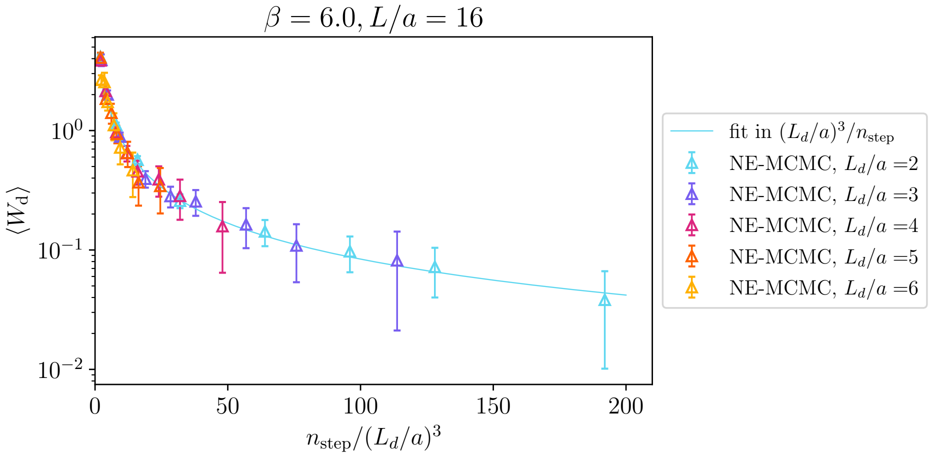


NEMC in BC: defect size scaling

[Bonanno et al.; 2510.25704]

How much doest it cost changing β on a $(L_d/a)^3$ defect?

clear scaling $n_{\mathrm{step}} \sim (L_d/a)^3$

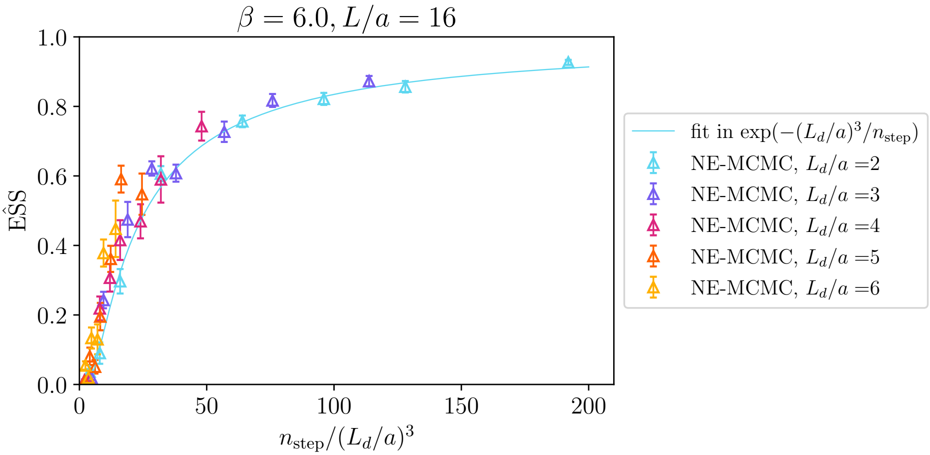


NEMC in BC: defect size scaling

[Bonanno et al.; 2510.25704]

How much doest it cost changing β on a $(L_d/a)^3$ defect?

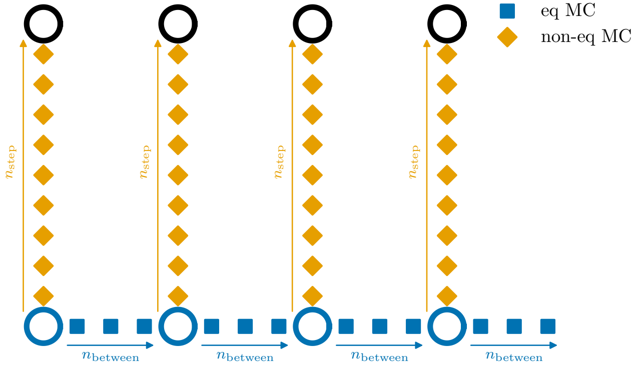
clear scaling $n_{\mathrm{step}} \sim (L_d/a)^3$



Stochastic Normalizing Flows

A systematic improvement of non-equilibrium MC

Alternate non-equilibrium MC updates with Normalizing Flow transformations

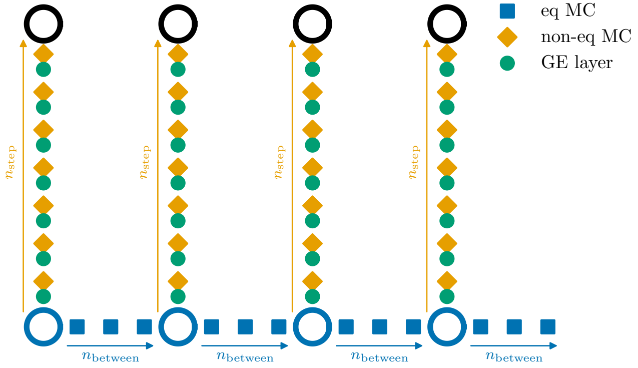


A systematic improvement of non-equilibrium MC

Alternate non-equilibrium MC updates with Normalizing Flow transformations

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

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



A systematic improvement of non-equilibrium MC

Alternate non-equilibrium MC updates with Normalizing Flow transformations

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]**

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

- each term trained separately (no memory issues) **[Matthews at al.; 2201.13117]**
- the trained g_n systematically remove dissipation

Gauge-equivariant layers [Nagai and Tomiya; 2103.11965] [Abbott et al.; 2305.02402]

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

$$\boxed{U'_\mu(x) = g_l(U_\mu(x)) = \exp\left(Q_\mu^{(l)}(x)\right) \underbrace{U_\mu(x)}_{\text{active}}}$$

with the algebra-valued

$$Q_\mu^{(l)}(x) = 2\left[\Omega_\mu^{(l)}(x)\right]_{\mathrm{TA}} \qquad \Omega_\mu^{(l)}(x) = \underbrace{C_\mu^{(l)}(x)}_{\text{frozen}}\underbrace{U_\mu^\dagger(x)}_{\text{active}}$$

- Sum of frozen staples

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

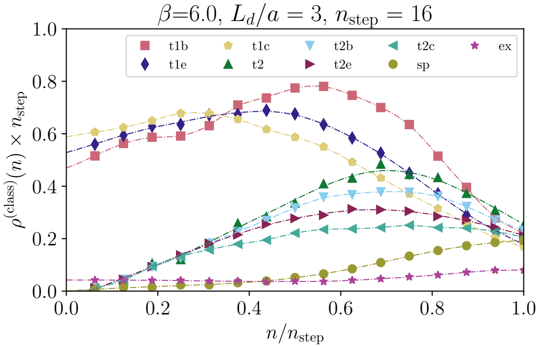
- Invertibility + easy computation of $\log J$ guaranteed by $2\times D = 8$ masks

Learning $\rho_{\mu\nu}(x)$

[Bonanno et al.; 2510.25704]

Stout smearing applied **only** in the vicinity of the defect

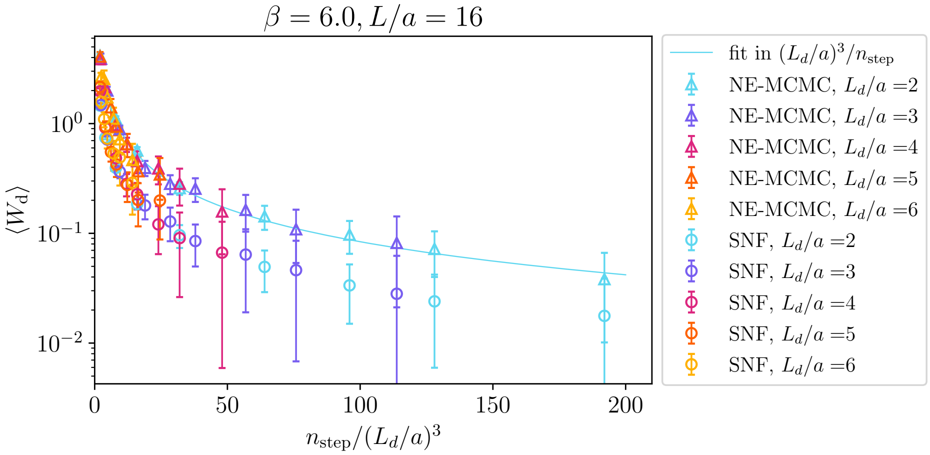
No translational invariance $\rightarrow$ 9 "classes" of parameters given by symmetries of defect links



Parameters perfectly linear in n_{step} $\rightarrow$ training only for $n_{\text{step}} = 16$, then transfer

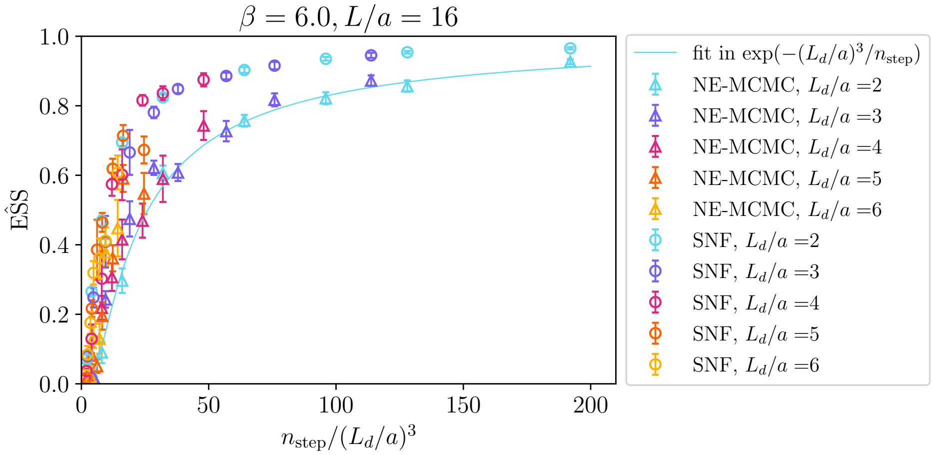
SNF in BC: defect size scaling

[Bonanno et al.; 2510.25704]



SNF in BC: defect size scaling

[Bonanno et al.; 2510.25704]



Mitigating topological freezing

A scalable unfreezing strategy

[Bonanno et al.; 2510.25704]

Effective cost

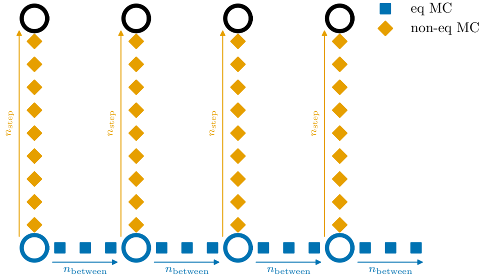
$$n_{\mathrm{ev}}^{(\mathrm{eff})} \times \underbrace{(n_{\mathrm{step}} + n_{\mathrm{between}})}_{\text{cost of 1 sample}} \simeq n_{\mathrm{ev}} \frac{2\tau_{\mathrm{int}}}{\mathrm{E\hat{S}S}} \times (n_{\mathrm{step}} + n_{\mathrm{between}})$$

A scalable unfreezing strategy

[Bonanno et al.; 2510.25704]

Effective cost

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



Keep $n_{\mathrm{ev}}^{(\mathrm{eff})}$ fixed for $a \to 0$ by rescaling:

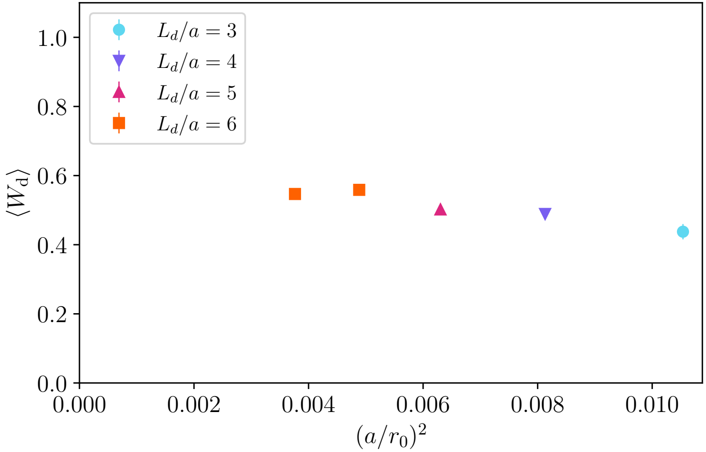
$$n_{\mathrm{step}} \sim \left(\frac{L_d}{a}\right)^3 \sim a^{-3} \qquad n_{\mathrm{between}} \sim a^{-2}$$

- unfreezing guaranteed by OBC defect
- exact removal of OBC finite-size effects guaranteed by NEMC/SNF

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

[Bonanno et al.; in preparation]

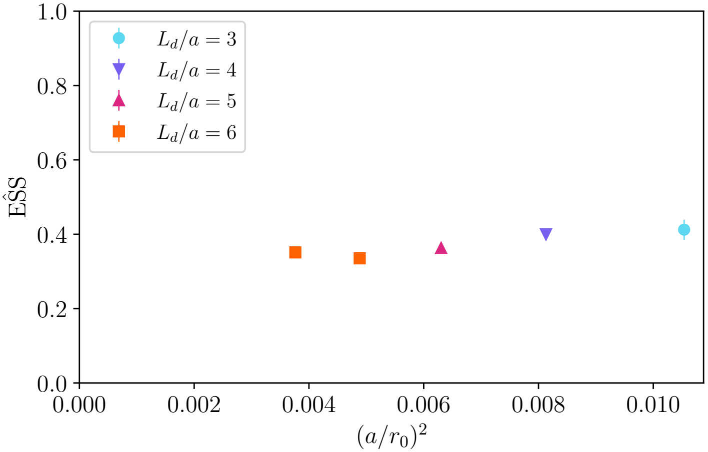
rescaling n_{step} and n_{between} $\to$ keep ESS and $\tau_{\mathrm{int}}(Q^2)$ fixed up to $\beta = 6.8$



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

[Bonanno et al.; in preparation]

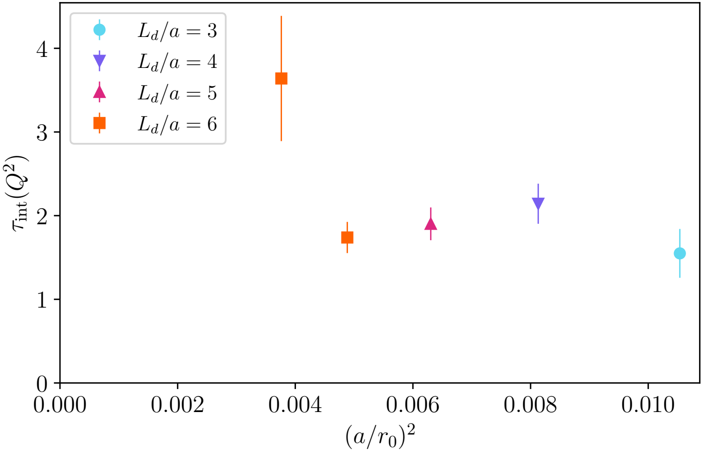
rescaling n_{step} and n_{between} $\rightarrow$ keep ESS and $\tau_{\mathrm{int}}(Q^2)$ fixed up to $\beta = 6.8$



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

[Bonanno et al.; in preparation]

rescaling n_{step} and n_{between} $\rightarrow$ keep ESS and $\tau_{\mathrm{int}}(Q^2)$ fixed up to $\beta = 6.8$



Conclusions – 1

Topological freezing is mitigated

effective cost of simulation $\sim a^{-3}$

Conclusions – 1

Topological freezing is mitigated

effective cost of simulation $\sim a^{-3}$

Much more general strategy!

- non-equilibrium framework $\rightarrow$ clear scaling with d.o.f.

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

- Stochastic Normalizing Flows $\rightarrow$ systematic improvement program to minimize dissipation

More critical slowing down... [Bonanno, AN, Vadacchino; in progress]

Deconfinement transition in $\mathrm{SU}(N)$ Yang–Mills has very interesting physics:

- latent heat
- interface tension
- bubble nucleation
- ...

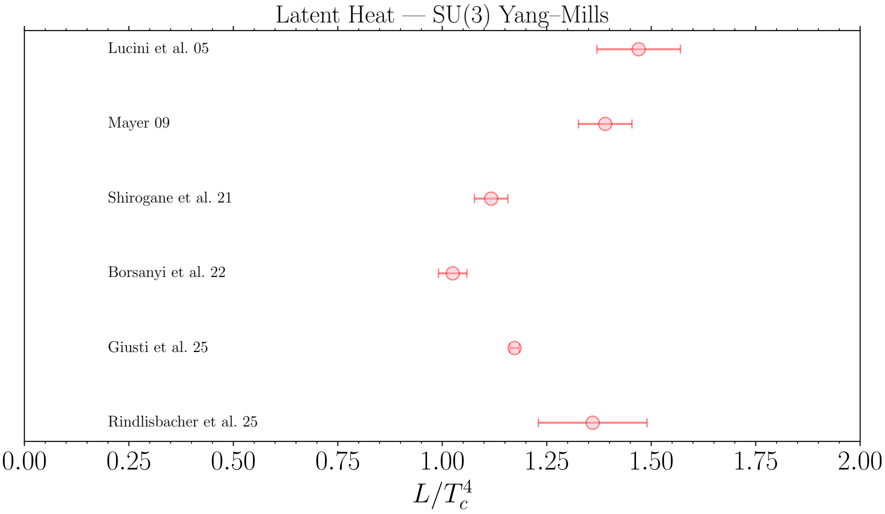
$\rightarrow$ relevant for strongy-coupled dark matter models

But: in $D = 4$ first-order phase transition for $N \geq 3$

$\rightarrow$ very severe critical slowing down at T_c

More critical slowing down...

[Bonanno, AN, Vadacchino; in progress]



Correlation functions with infinitesimal flows

In collaboration with P. Butti, G. Catumba, L. Spatscheck

Signal-to-noise ratio problem

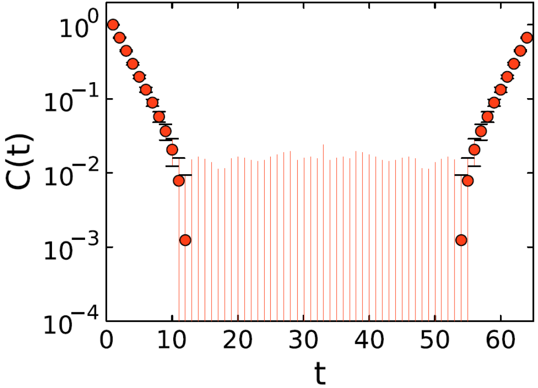
Euclidean correlator $C(t) \to$ extract masses and matrix elements...

Signal and noise decay with different behaviours: their ratio may decrease exponentially in t

$$\frac{C(t)}{\text{error}[C(t)]} \sim \frac{e^{-\Delta t}}{\sqrt{N_{\text{samples}}}}$$

QCD $\to$ Parisi-Lepage argument

scalar theory $\to$ variance is approx constant



Correlation functions from the generating functional

We want to use a different estimator for

$$\boxed{C(t) = \langle \mathcal{O}_t \mathcal{O}_0 \rangle_{q_0} - \langle \mathcal{O} \rangle_{q_0}^2} \qquad q_0(\phi) = \frac{1}{Z} e^{-S}$$

with the zero-momentum projection

$$\mathcal{O}_t[\phi] = \frac{1}{V_s} \sum_{\vec{x}} \phi(\vec{x}, t)$$

Correlation functions from the generating functional

We want to use a different estimator for

$$\boxed{C(t) = \langle \mathcal{O}_t \mathcal{O}_0 \rangle_{q_0} - \langle \mathcal{O} \rangle_{q_0}^2} \qquad q_0(\phi) = \frac{1}{Z} e^{-S}$$

with the zero-momentum projection

$$\mathcal{O}_t[\phi] = \frac{1}{V_s} \sum_{\vec{x}} \phi(\vec{x}, t)$$

Obtain the same quantity by

$$\boxed{C(t) = \frac{\partial}{\partial J} \langle \mathcal{O}_t \rangle_p \bigg|_{J=0}}$$

with an inclusion of a source term in the action

$$p(\phi) = \frac{1}{Z_J} e^{-S_J} \qquad S_J = S - J\mathcal{O}_0$$

Reframing the StN problem

Evaluate $\langle \dots \rangle_p$ via reweighting:

$$\langle \mathcal{O}_t \rangle_p = \langle w \mathcal{O}_t \rangle_{q_0}$$

with the normalized weights

$$w(\phi) = \frac{p(\phi)}{q_0(\phi)} = \frac{Z}{Z_J} e^{S-(S-J\mathcal{O}_0(\phi))} = \frac{e^{J\mathcal{O}_0(\phi)}}{\langle e^{J\mathcal{O}_0} \rangle_{q_0}}$$

Reframing the StN problem

Evaluate $\langle \dots \rangle_p$ via reweighting:

$$\langle \mathcal{O}_t \rangle_p = \langle w \mathcal{O}_t \rangle_{q_0}$$

with the normalized weights

$$w(\phi) = \frac{p(\phi)}{q_0(\phi)} = \frac{Z}{Z_J} e^{S-(S-J\mathcal{O}_0(\phi))} = \frac{e^{J\mathcal{O}_0(\phi)}}{\langle e^{J\mathcal{O}_0} \rangle_{q_0}}$$

Let us formally fix $J = \varepsilon$ and expand around ε **[Catumba, Ramos; 2502.15570]**

$$w(\phi) = 1 + \varepsilon(\mathcal{O}_0 - \langle \mathcal{O}_0 \rangle_{q_0}) + \dots$$

$$\langle \mathcal{O}_t \rangle_p = \langle \mathcal{O}_t \rangle_{q_0} + \varepsilon C(t) + \dots$$

StN degradation due to fluctuation in the weights! $\rightarrow$ overlap problem

Infinitesimal flows

see also [Abbott et al.; 2603.02984, 2606.15986]

Transform samples with a residual flow:

$$\tilde{\phi} = T(\phi) = \phi + \varepsilon f(\phi)$$

with f being an (invertible) vector field

Infinitesimal flows

see also [Abbott et al.; 2603.02984, 2606.15986]

Transform samples with a residual flow:

$$\tilde{\phi} = T(\phi) = \phi + \varepsilon f(\phi)$$

with f being an (invertible) vector field

New distribution

$$q(\tilde{\phi}) = q_0(\phi) |\det J_T|^{-1}$$

with new weights

$$\tilde{w}(\phi) = \frac{p}{q} = \frac{Z}{Z_J} e^{-S_J(\tilde{\phi}) + S(\phi) + \log\det J_T}$$

Infinitesimal flows

see also [Abbott et al.; 2603.02984, 2606.15986]

Transform samples with a residual flow:

$$\tilde{\phi} = T(\phi) = \phi + \varepsilon f(\phi)$$

with f being an (invertible) vector field

New distribution

$$q(\tilde{\phi}) = q_0(\phi) |\det J_T|^{-1}$$

with new weights

$$\tilde{w}(\phi) = \frac{p}{q} = \frac{Z}{Z_J} e^{-S_J(\tilde{\phi}) + S(\phi) + \log\det J_T}$$

Expanding:

$$\boxed{\tilde{w}(\phi) = 1 + \varepsilon(\underbrace{\mathrm{Tr} J_f - f \cdot \nabla S}_{\text{Stein operator: } \mathcal{T}_r f} + \mathcal{O}_0 - \langle \mathcal{O}_0 \rangle_{q_0}) + \ldots}$$

[Liu, Wang; 1608.04471]

Note that $\langle w \rangle_{q_0} = 1$ as $\langle \mathcal{T}_r f \rangle_{q_0} = 0$ (i.b.p.)

Minimal variance condition

How to find an f that renders the weights constant?

For NFs one usually minimizes the KL divergence...

$$\boxed{\tilde{D}_{\mathrm{KL}}(q||p) = -\langle \log \tilde{w} \rangle_{q_0}}$$

Minimal variance condition

How to find an f that renders the weights constant?

For NFs one usually minimizes the KL divergence...

$$\boxed{\tilde{D}_{\mathrm{KL}}(q||p) = -\langle \log \tilde{w} \rangle_{q_0}}$$

...which is perfectly natural! Approx. same as minimizing $\tilde{w}$ variance

$$\tilde{D}_{\mathrm{KL}}(q||p) = \frac{1}{2}\mathrm{Var}_{q_0}(\tilde{w}) + \mathcal{O}(\langle(\tilde{w}-1)^3\rangle_{q_0})$$

Minimal variance condition

How to find an f that renders the weights constant?

For NFs one usually minimizes the KL divergence...

$$\boxed{\tilde{D}_{\mathrm{KL}}(q||p) = -\langle \log \tilde{w} \rangle_{q_0}}$$

...which is perfectly natural! Approx. same as minimizing $\tilde{w}$ variance

$$\tilde{D}_{\mathrm{KL}}(q||p) = \frac{1}{2}\mathrm{Var}_{q_0}(\tilde{w}) + \mathcal{O}(\langle(\tilde{w}-1)^3\rangle_{q_0})$$

Expanding in ε

$$\boxed{\tilde{D}_{\mathrm{KL}}(q||p) = \frac{\varepsilon^2}{2}\Big(\langle \mathrm{Tr} J_f^2 + f^T H_S f - 2f\cdot\nabla\mathcal{O}_0\rangle_{q_0} + \mathrm{Var}_{q_0}(\mathcal{O}_0)\Big) + \dots}$$

Architecture and pipeline

Several iterations of architectures for $\tilde{\phi} = \phi + \varepsilon f_\theta(\phi)$

$$f_\theta(\phi) = \mathrm{iFFT}\left[\frac{\delta(\mathbf{p})}{\hat{p}^2 + M_\theta(\phi_p)}\right] + C_\theta(\phi)$$

inspired by the exact transformation for the non-interacting theory

$$\tilde{\phi}_p = \phi_p + \frac{1}{2}\frac{\delta(\mathbf{p})}{\hat{p}^2 + \hat{m}^2}\varepsilon$$

Architecture and pipeline

Several iterations of architectures for $\tilde{\phi} = \phi + \varepsilon f_\theta(\phi)$

$$f_\theta(\phi) = \mathrm{iFFT}\left[\frac{\delta(\mathbf{p})}{\hat{p}^2 + M_\theta(\phi_p)}\right] + C_\theta(\phi)$$

inspired by the exact transformation for the non-interacting theory

$$\tilde{\phi}_p = \phi_p + \frac{1}{2}\frac{\delta(\mathbf{p})}{\hat{p}^2 + \hat{m}^2}\varepsilon$$

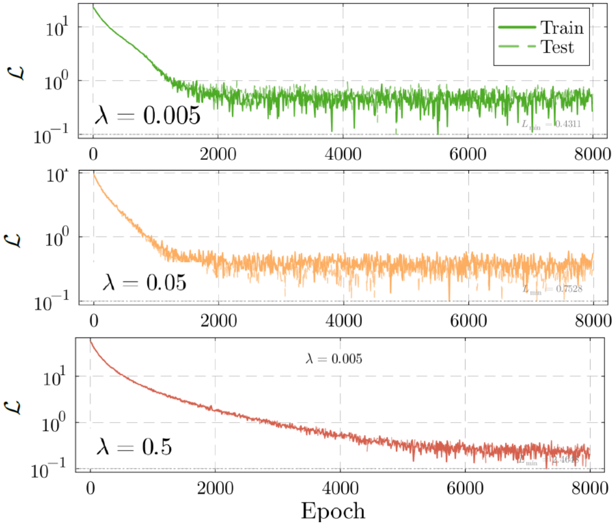
Training

- sample ϕ from $q_0 \sim e^{-S}$
- flow through f_θ
- compute KL divergence at $\mathcal{O}(\varepsilon^2)$
- minimize using standard ML toolbox (and repeat as necessary)

Sampling

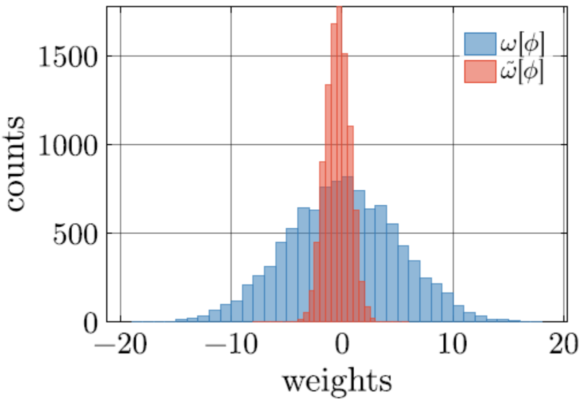
- sample ϕ from $q_0 \sim e^{-S}$
- flow through f_θ
- compute weights $\tilde{w}$ at $\mathcal{O}(\varepsilon)$
- extract $C(t)$ analytically or using **[Catumba+Ramos; 2502.15570]**

Some results in $\lambda\phi^4$ – training



Some results in $\lambda\phi^4$ – training

weights distribution before and after training



Back to the correlator

What do we compute exactly in the end?

Expand the improved reweigthing

$$\langle \mathcal{O}_t \rangle_p = \langle \tilde{w} \mathcal{O}_t \rangle_{q_0} = \langle \mathcal{O}_t \rangle_{q_0} + \varepsilon \left(C(t) + \underbrace{\langle \mathcal{T}_r f \cdot \mathcal{O}_t + f \cdot \nabla \mathcal{O}_t \rangle_{q_0}}_{=0 \text{ (i.b.p.)}} \right) + \ldots$$

The extra piece disappears $\rightarrow$ at order ε we still get $C(t)$!

Back to the correlator

What do we compute exactly in the end?

Expand the improved reweigthing

$$\langle \mathcal{O}_t \rangle_p = \langle \tilde{w} \mathcal{O}_t \rangle_{q_0} = \langle \mathcal{O}_t \rangle_{q_0} + \varepsilon \big(C(t) + \underbrace{\langle \mathcal{T}_r f \cdot \mathcal{O}_t + f \cdot \nabla \mathcal{O}_t \rangle_{q_0}}_{=0 \text{ (i.b.p.)}} \big) + \dots$$

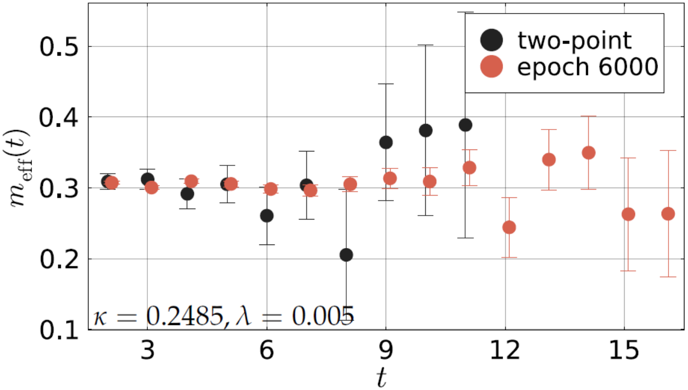
The extra piece disappears $\rightarrow$ at order ε we still get $C(t)$!

But actually we are designing a control variate

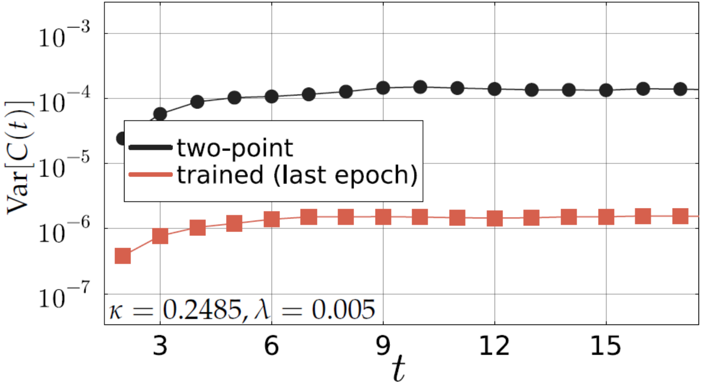
$$\mathcal{O}_t (\mathcal{O}_0 - \langle \mathcal{O} \rangle_{q_0}) + \mathcal{T}_r f \cdot \mathcal{O}_t + f \cdot \nabla \mathcal{O}_t$$

same expectation value, different variance

Some results in $\lambda\phi^4$ – sampling



Some results in $\lambda\phi^4$ – sampling



Conclusions – 2

- Stochastic Automatic Differentiation **[Catumba, Ramos; 2502.15570]** can provide **automatically** the ε and ε^2 terms!

- Architecture still in development $\rightarrow$ stout smearing-like transformations being tested

- A solution of the minimization condition

$$\mathcal{T}_r f^* + \mathcal{O}_0 - \langle \mathcal{O}_0 \rangle = 0$$

can be found using a Feynman-Kac approach **[Albergo, Kanwar; 2603.00252]**

Thank you for your attention!

A strategy for thermodynamics

Deconfinement transition $T \to T_c$

- (thermal) critical slowing down
- $\tau_{\text{int}}(U_{\mu\nu}), \tau_{\text{int}}(P), \dots$ diverge
- latent heat + transition features + equation of state difficult to determine

NEMC in β (temperature)

[Bonanno, AN, Vadacchino; in progress]

- Base dist: $T_0 \neq T_c$ at given β_0, N_t
- Protocol: switch $\beta_0 \to \beta_c(N_t)$
- # dof $\sim L_t \times L_s^3$

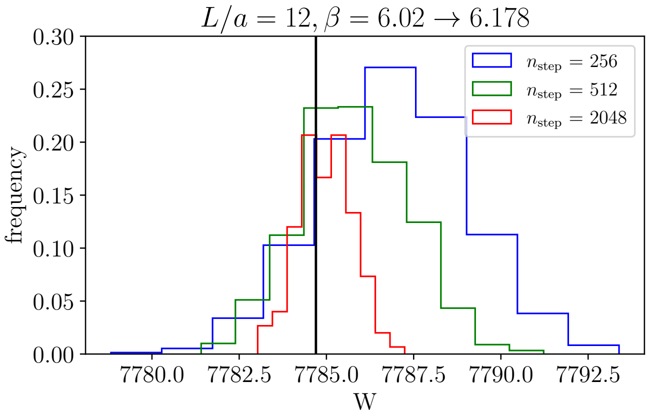
[Caselle, AN, Panero; 1801.03110]
[Bulgarelli, Cellini, AN; 2412.00200]

NEMC in β in $\mathrm{SU}(3)$

[Bulgarelli, Cellini, AN; 2412.00200]

Test: changing β on a $(L/a)^4$ system

Far from critical point!



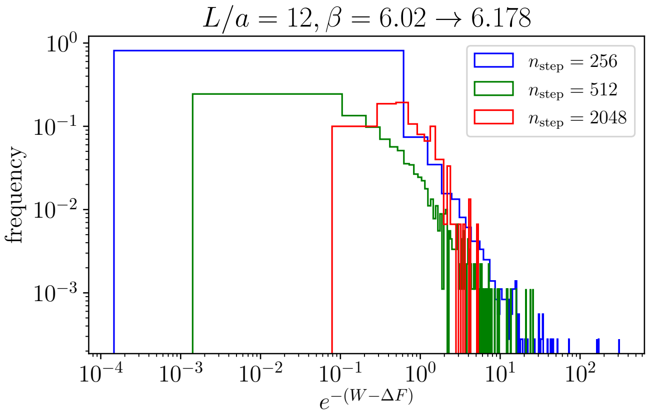
$\tilde{D}_{\mathrm{KL}} = \langle W \rangle_{\mathrm{f}} - \Delta F$ decreases with increasing n_{step}

NEMC in β in $\mathrm{SU}(3)$

[Bulgarelli, Cellini, AN; 2412.00200]

Test: changing β on a $(L/a)^4$ system

Far from critical point!

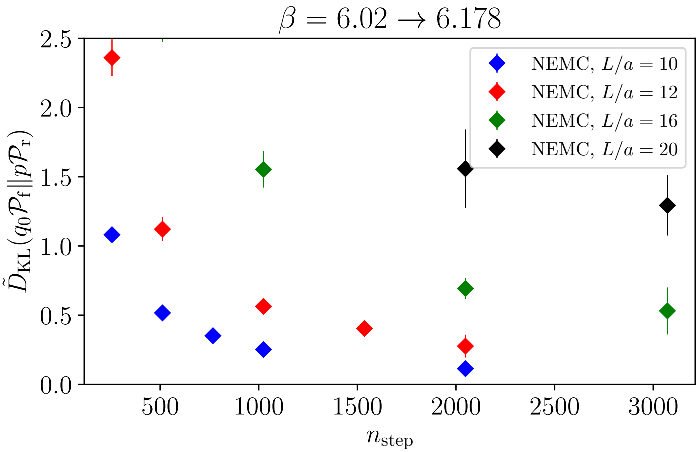


$\mathrm{Var}(\exp(-W_{\mathrm{d}})) = \frac{1}{\hat{\mathrm{ESS}}} - 1$ decreases with increasing n_{step}

NEMC in β: volume scaling

[Bulgarelli, Cellini, AN; 2412.00200]

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

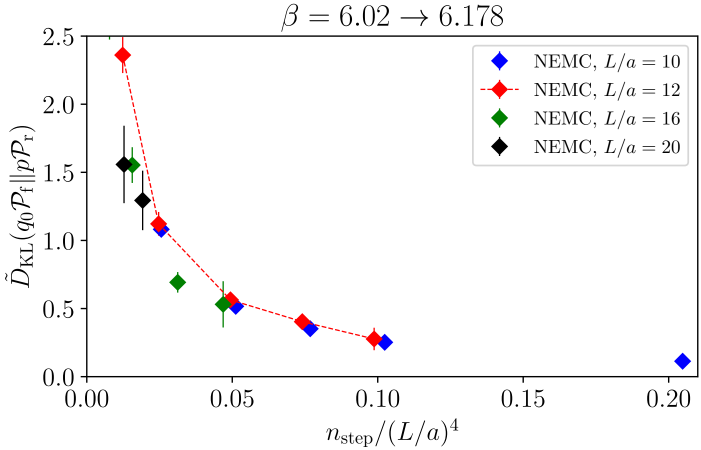


NEMC in β: volume scaling

[Bulgarelli, Cellini, AN; 2412.00200]

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

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

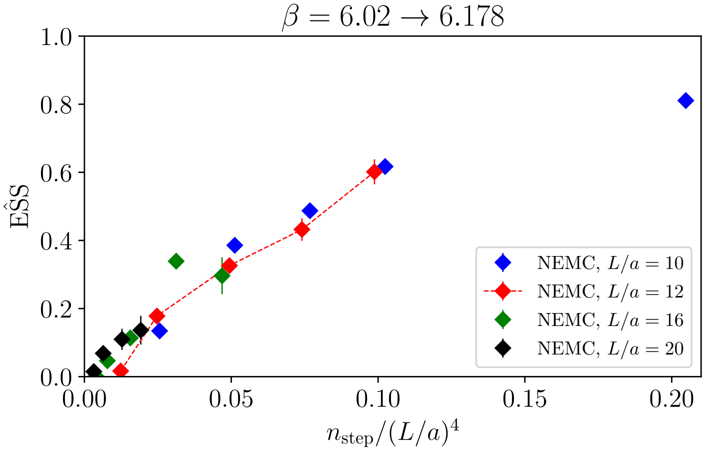


NEMC in β: volume scaling

[Bulgarelli, Cellini, AN; 2412.00200]

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

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

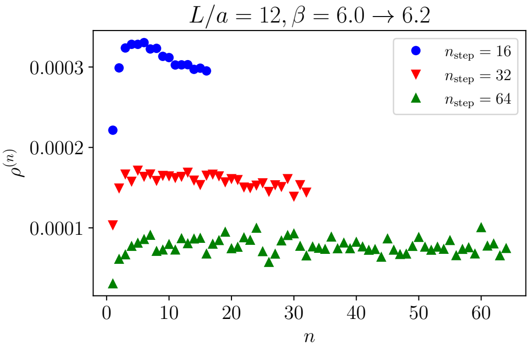


Transferring ρ in SNFs

[Bulgarelli, Cellini, AN; 2412.00200]

Short trainings: 200-1000 epochs enough to saturate

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



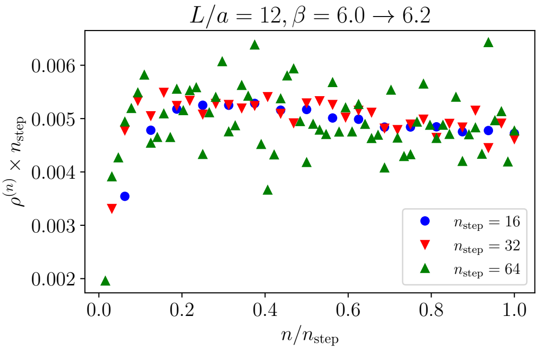
- global interpolation of ρ from trainings at $n_{\text{step}} = 16, 32, 64$
- $\rho^{(l)}$ extrapolated to large $n_{\text{step}} \to$ no retraining!
- Heavy use of **transfer learning** for each $\beta_0 \to \beta$ evolution
- Transfer learning also possible between different volumes

Transferring ρ in SNFs

[Bulgarelli, Cellini, AN; 2412.00200]

Short trainings: 200-1000 epochs enough to saturate

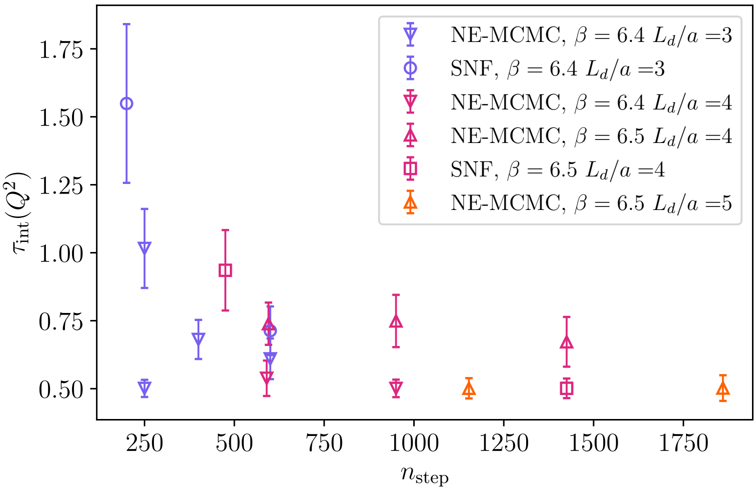
Training only with small $n_{\rm step}$: clear pattern emerges



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

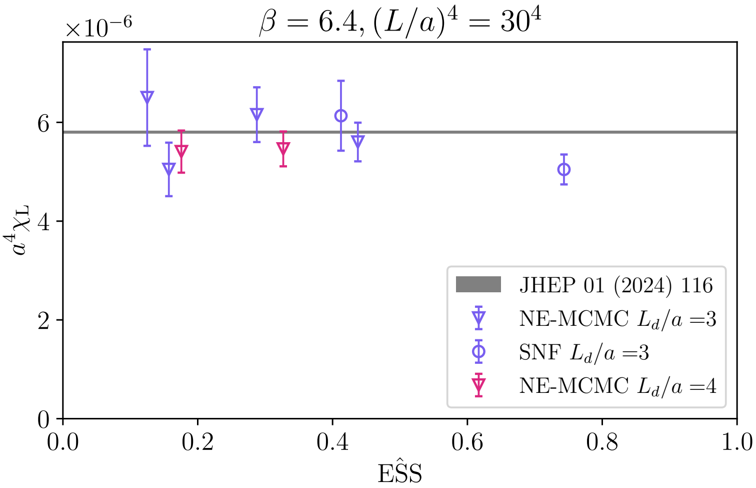
Switching BC in $\mathrm{SU}(3)$: τ_{int}

[Bonanno et al.; 2510.25704]



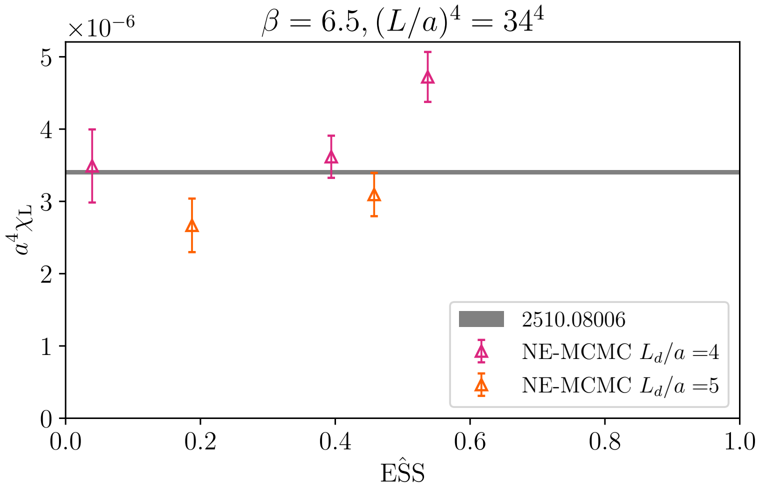
Switching BC in $\mathrm{SU}(3)$: sanity check

[Bonanno et al.; 2510.25704]



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

[Bonanno et al.; 2510.25704]



Discrete Normalizing Flows

First, most common implementation:

- build NFs as a combination of discrete transformations (coupling layers)

$$f_\theta = g_{n_l} \circ \cdots \circ g_2 \circ g_1$$

- implement partitioning of d.o.f. at each step (masking)

$$U^{(l)} \to \left(U^{(l)}_{\mathrm{active}}, U^{(l)}_{\mathrm{frozen}} \right)$$

$\to$ invertibility and triangular Jacobian guaranteed

$$g_n : \begin{cases} U^{(l)}_{\mathrm{frozen}} = U^{(l)}_{\mathrm{frozen}} \\ U^{(l)}_{\mathrm{active}} = W_{\theta_l}(U^{(l)}_{\mathrm{frozen}}) \cdot U^{(l)}_{\mathrm{active}} \end{cases}$$

the function W_{θ_l} can be parametrized with neural networks

Continuous Normalizing Flows

No discrete coupling layers: learn an ODE instead

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

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

with Z_t given by the force of

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

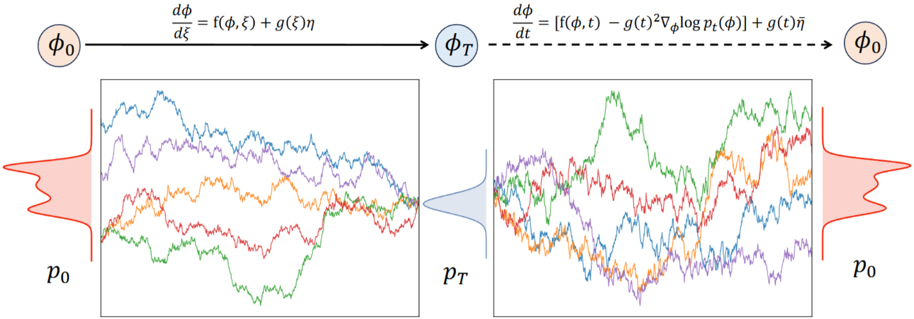
Machine-learn the c_i coefficients!

[Gerdes et al.; 2207.00283], **[Caselle et al.; 2307.01107]**, **[Gerdes et al.; 2410.13161]**

Diffusion models

[Wang, Aarts, Zhou; 2309.17082]

First introduce noise in the data (forward Langevin process), then **denoise** it (backward)



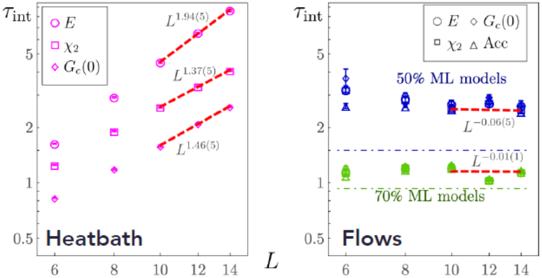
Data from target is required!

Connection to stochastic quantization!

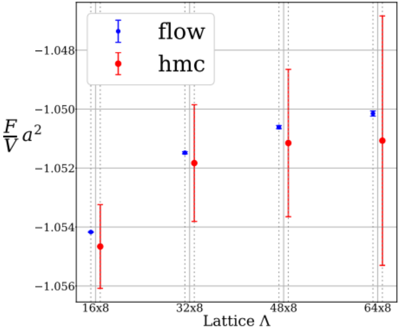
The path until now - scalar field theory

Natural toy model to start with: ϕ^4 scalar field theory in 2d

[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]** + . . .

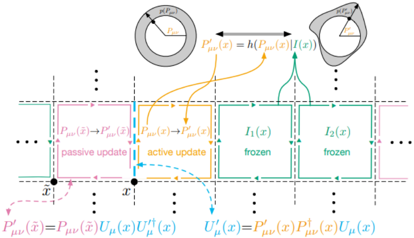
The path until now - gauge theories

What about gauge theories? $\rightarrow$ gauge-equivariant architectures

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

$$P'_{\mu\nu}(x) = h(P_{\mu\nu}(x)|I(x))$$

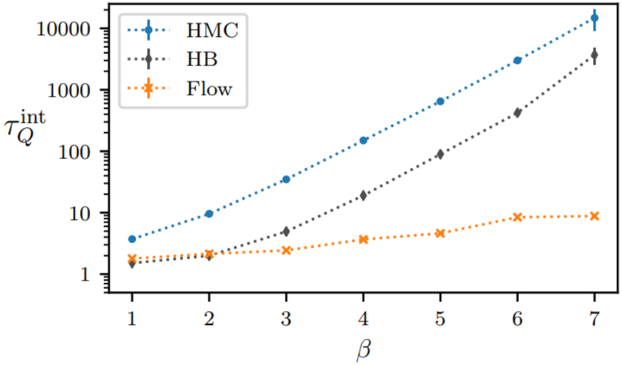
$$U'_{\mu}(x) = P'_{\mu\nu}(x)P^{\dagger}_{\mu\nu}(x)U_{\mu}(x)$$



- **link-level flow** (more later!)
- **continuous NFs** are naturally equivariant

The path until now - gauge theories

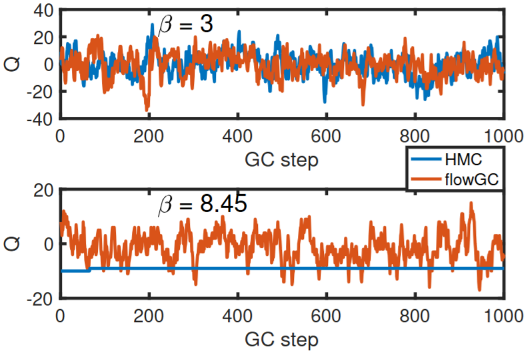
$\mathrm{U}(1)$ gauge theory in 2d **[Kanwar et al.; 2003.06413]**



but lots more work in other gauge theories (see next slide)

The path until now - gauge theories

Schwinger model **[Finkerath; 2201.02216]**



but lots more work in other gauge theories (see next slide)

The path until now

ϕ^4 scalar field theory in 2d $\rightarrow$ **[Albergo, Kanwar, Shanahan; 1904.12072] [Nicoli et al.; 2007.07115] [Del Debbio et al.; 2105.12481] [Gerdes et al.; 2207.00283] [Albandea et al.; 2302.08408] [Wang, Aarts, Zhou; 2309.17082]**

$\mathrm{U}(1)$ and $\mathrm{SU}(N)$ in 2d $\rightarrow$ **[Kanwar et al.; 2003.06413] [Boyda et al.; 2008.05456] [Bacchio et al.; 2212.08469] [Gerdes et al.; 2410.13161] [Zhu et al.; 2502.05504] [Vega et al.; 2510.26081] [Aarts et al.; 2601.19552]**

fermionic theories **[Albergo et al.; 2106.05934]** $\rightarrow$ Schwinger model **[Finkenrath et al.; 2201.02216] [Albergo et al.; 2202.11712]** and $\mathrm{SU}(N)$ in 2d **[Abbott et al.; 2207.08945]**

approaching $\mathrm{SU}(3)$ in 4d **[Abbott et al.; 2305.02402] [Abbott et al.; 2401.10874]**

different models: Nambu-Goto string **[Caselle et al.; 2307.01107]** and Hubbard model **[Schuh et al.; 2506.17015]**

The path until now

ϕ^4 scalar field theory in 2d $\to$ **[Albergo, Kanwar, Shanahan; 1904.12072] [Nicoli et al.; 2007.07115] [Del Debbio et al.; 2105.12481] [Gerdes et al.; 2207.00283] [Albandea et al.; 2302.08408] [Wang, Aarts, Zhou; 2309.17082]**

$\mathrm{U}(1)$ and $\mathrm{SU}(N)$ in 2d $\to$ **[Kanwar et al.; 2003.06413] [Boyda et al.; 2008.05456] [Bacchio et al.; 2212.08469] [Gerdes et al.; 2410.13161] [Zhu et al.; 2502.05504] [Vega et al.; 2510.26081] [Aarts et al.; 2601.19552]**

fermionic theories **[Albergo et al.; 2106.05934]** $\to$ Schwinger model **[Finkenrath et al.; 2201.02216] [Albergo et al.; 2202.11712]** and $\mathrm{SU}(N)$ in 2d **[Abbott et al.; 2207.08945]**

approaching $\mathrm{SU}(3)$ in 4d **[Abbott et al.; 2305.02402] [Abbott et al.; 2401.10874]**

different models: Nambu-Goto string **[Caselle et al.; 2307.01107]** and Hubbard model **[Schuh et al.; 2506.17015]**

discrete NFs // continuous NFs // diffusion models

The path until now

ϕ^4 scalar field theory in 2d $\rightarrow$ **[Albergo, Kanwar, Shanahan; 1904.12072] [Nicoli et al.; 2007.07115] [Del Debbio et al.; 2105.12481] [Gerdes et al.; 2207.00283] [Albandea et al.; 2302.08408] [Wang, Aarts, Zhou; 2309.17082]**

$\mathrm{U}(1)$ and $\mathrm{SU}(N)$ in 2d $\rightarrow$ **[Kanwar et al.; 2003.06413] [Boyda et al.; 2008.05456] [Bacchio et al.; 2212.08469] [Gerdes et al.; 2410.13161] [Zhu et al.; 2502.05504] [Vega et al.; 2510.26081] [Aarts et al.; 2601.19552]**

fermionic theories **[Albergo et al.; 2106.05934]** $\rightarrow$ Schwinger model **[Finkenrath et al.; 2201.02216] [Albergo et al.; 2202.11712]** and $\mathrm{SU}(N)$ in 2d **[Abbott et al.; 2207.08945]**

approaching $\mathrm{SU}(3)$ in 4d **[Abbott et al.; 2305.02402] [Abbott et al.; 2401.10874]**

different models: Nambu-Goto string **[Caselle et al.; 2307.01107]** and Hubbard model **[Schuh et al.; 2506.17015]**

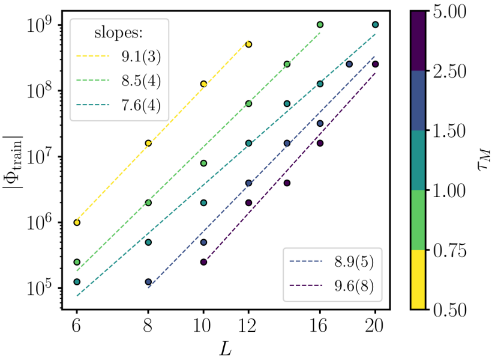
Very effective in critical slowing down mitigation in lower-dimensional theories!

How do flow-based samplers scale?

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

We require two features:

- better inference $\to$ improved scaling of the sampling with respect to MCMC
- manageable training costs $\to$ learning f_θ scales reasonably
- for a fixed trained model: $\mathrm{ESS}(V) = \mathrm{ESS}(V_0)^{V/V_0}$
- difficult to determine in general **[Abbott et al.; 2211.07541]**



worrying scaling of training costs for fixed models
[Del Debbio et al.; 2105.12481]