Mitigating topological freezing with out-of-equilibrium simulations

Alessandro Nada

Università degli Studi di Torino

DESY/HU Lattice seminar, Berlin, 12th February 2024

based on:

arXiv:2402.06561

in collaboration with:

Claudio Bonanno (IFT Madrid) and Davide Vadacchino (University of Plymouth)

Topological freezing

In Lattice QCD sectors characterized by different values of the topological charge Q emerge in the continuum limit

For $a \to 0$ the transition between these sectors becomes more and more strongly suppressed

$\to$ very **long autocorrelation times** characterize topological observables when standard MCMC algorithms are used

Topological freezing

In Lattice QCD sectors characterized by different values of the topological charge Q emerge in the continuum limit

For $a \to 0$ the transition between these sectors becomes more and more strongly suppressed

$\to$ very **long autocorrelation times** characterize topological observables when standard MCMC algorithms are used

Use of **open boundary conditions [Lüscher and Schaefer; 2011]** in time essentially solves the problem by removing the sectors

Drawback: measurements possible only away from the open boundaries

Methods such as parallel tempering **[Hasenbusch; 2017]** approach the problem in a similar manner

Out-of-equilibrium evolutions for a MCMC

Consider a "guided" MCMC evolution

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

- $c(n)$ is a parameter of the action $S_{c(n)}$ of the model
- the evolution starts from a distribution $q_0 = e^{-S_{c(0)}}/Z_0$, the **prior**, from which we sample ϕ_0 **at equilibrium**
- it goes over n_{step} intermediate steps
- at each step the system evolves using some (e.g. one) regular MC updates with a transition probability $P_{c(n)}(\phi_n \to \phi_{n+1})$
- the transition probability changes along the evolution according to the **protocol** $c(n)$
- the evolution ends at the **target** probability distribution $p = e^{-S_{c(n_{\mathrm{step}})}}/Z_{n_{\mathrm{step}}} \equiv e^{-S}/Z$

The probability distribution (in general not at equilibrium) is

$$q(\phi) = \int [\mathrm{d}\phi_0 \dots \mathrm{d}\phi_{n_{\mathrm{step}}-1}] q_0[\phi_0] \mathcal{P}_{\mathrm{f}}[\phi_0, \dots, \phi]$$

with

$$\mathcal{P}_{\mathrm{f}}[\phi_0, \dots, \phi] = \prod_{n=1}^{n_{\mathrm{step}}} P_{c(n)}(\phi_{n-1} \to \phi_n)$$

Crooks' theorem

One can look at the ratio of the forward and reverse evolutions going through the same intermediate configurations

$$\frac{q_0(\phi_0)\mathcal{P}_{\rm f}[\phi_0,\dots,\phi_{n_{\rm step}}]}{p(\phi)\mathcal{P}_{\rm r}[\phi_{n_{\rm step}},\dots,\phi_0]} = \frac{q_0(\phi_0)\prod_{n=1}^{n_{\rm step}} P_{c(n)}(\phi_{n-1}\to\phi_n)}{p(\phi_{n_{\rm step}})\prod_{n=1}^{n_{\rm step}} P_{c(n)}(\phi_n\to\phi_{n-1})}$$

It is easy to derive **Crooks' theorem** for MCMC **[Crooks; 1999]** if the update algorithm satisfies detailed balance

$$\frac{q_0(\phi_0)\mathcal{P}_{\rm f}[\phi_0,\dots,\phi_{n_{\rm step}}]}{p(\phi)\mathcal{P}_{\rm r}[\phi_{n_{\rm step}},\dots,\phi_0]} = \exp(W-\Delta F)$$

with the generalized **work**

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

and the **free energy** difference

$$\exp(-\Delta F) = \frac{Z}{Z_0}$$

Jarzynski's equality for MCMC

Integrating over the whole trajectory one gets

$$\int [\mathrm{d}\phi_0 \dots \mathrm{d}\phi_{n_{\mathrm{step}}}] q_0(\phi_0) \mathcal{P}_{\mathrm{f}}[\phi_0, \dots, \phi_{n_{\mathrm{step}}}] \exp(-(W - \Delta F)) = 1$$

This is the formal derivation of **Jarzynski's equality [Jarzynski; 1997]** for MCMC

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

The ratio of the two partition functions is computed directly with an average over "forward" non-equilibrium evolutions defined rigorously as

$$\langle \mathcal{A} \rangle_{\mathrm{f}} = \int [\mathrm{d}\phi_0 \dots \mathrm{d}\phi] q_0(\phi_0)\, \mathcal{P}_{\mathrm{f}}[\phi_0, \dots, \phi]\, \mathcal{A}[\phi_0, \dots, \phi]$$

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

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

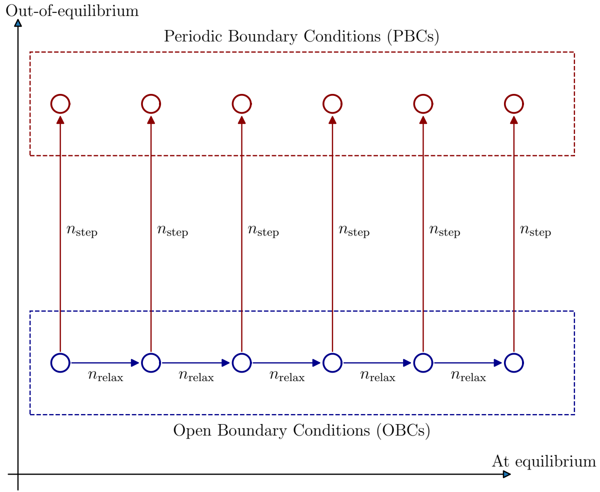
Out-of-equilibrium stochastic evolutions

- the *actual* probability distribution at each step is NOT the equilibrium distribution $\sim \exp(-S_{c(n)})$: it's a non-equilibrium process!
- valid process also without letting the system relax, or far from equilibrium (e.g. n_{step} is "small")
- the $\langle \mathcal{A} \rangle_{\mathrm{f}}$ average is taken over all possible evolutions, so in principle infinite statistics might be needed (more on this later)
- the idea 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

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

- this work: rigorously sample PBC by starting from OBC

A new paradigm to perform MCMC



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!

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]**

How far are we from equilibrium?

Ideally we would like

$$\tilde{D}_{\mathrm{KL}}(q\|p) = \int \mathrm{d}\phi \; q(\phi) \log \left(\frac{q(\phi)}{p(\phi)} \right) \qquad\qquad q(\phi) = \int [\mathrm{d}\phi_0 \dots \mathrm{d}\phi_{n_{\mathrm{step}}-1}] q_0(\phi_0) \mathcal{P}_{\mathrm{f}}[\phi_0, \dots, \phi]$$

but the generated distribution $q(\phi)$ is not tractable!

However we can measure the "quality" of the out-of-equilibrium evolutions by comparing forward and reverse processes!

$$\tilde{D}_{\mathrm{KL}}(q_0\mathcal{P}_{\mathrm{f}}\|p\mathcal{P}_{\mathrm{r}}) = \int [\mathrm{d}\phi_0 \dots \mathrm{d}\phi] \, q_0(\phi_0)\mathcal{P}_{\mathrm{f}}[\phi_0, \dots, \phi] \log \frac{q_0(\phi_0)\mathcal{P}_{\mathrm{f}}[\phi_0, \dots, \phi]}{p(\phi)\mathcal{P}_{\mathrm{r}}[\phi, \phi_{n_{\mathrm{step}}-1}, \dots, \phi_0]}$$

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

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

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 approximate estimator

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

$\rightarrow$ very common metric to evaluate generative models in the deep-learning community

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

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

which leads to the constraint

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

Non-equilibrium evolutions for topological observables

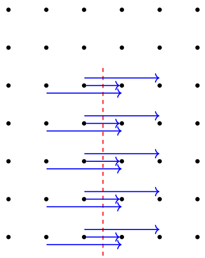
The CP^{N-1} model with a defect

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

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

where the changing value $c(n)$ will be clear in the following

Geometric definition of the topological charge Q

$$Q_{\mathrm{geo}}[U] = \frac{1}{2\pi} \sum_x \Im \{\log [\Pi_{01}(x)]\} \in \mathbb{Z},$$

with $\Pi_{\mu\nu}(x) \equiv U_\mu(x) U_\nu(x+\hat{\mu}) \bar{U}_\mu(x+\hat{\nu} \bar{U}_\nu(x)$

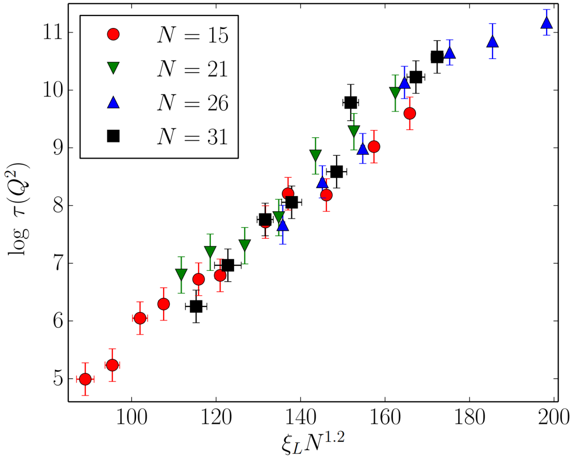
We look at the topological susceptibility

$$\chi = \frac{1}{V} \langle Q^2_{\mathrm{geo}} \rangle$$

after applying a few steps of cooling

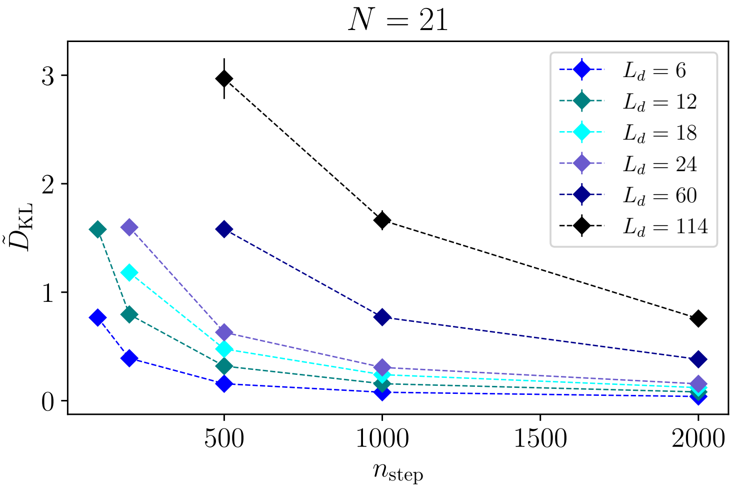
Topology freezing in the CP^{N-1} model

From **[Bonanno et al.; 2018]**



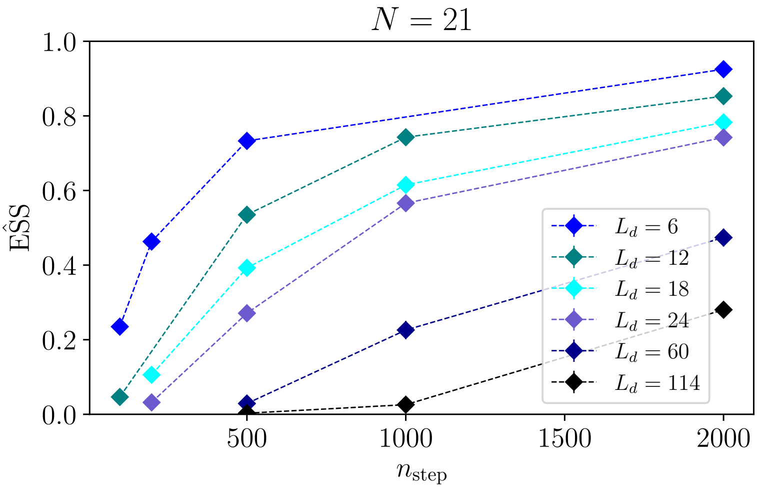
We will study $(N = 21, \beta = 0.7)$ with $\tau(Q^2) \sim 6 \times 10^4$ and $(N = 41, \beta = 0.65)$ with $\tau(Q^2) \sim 5 \times 10^5$

"Slower" evolutions allow for better (but more expensive) sampling



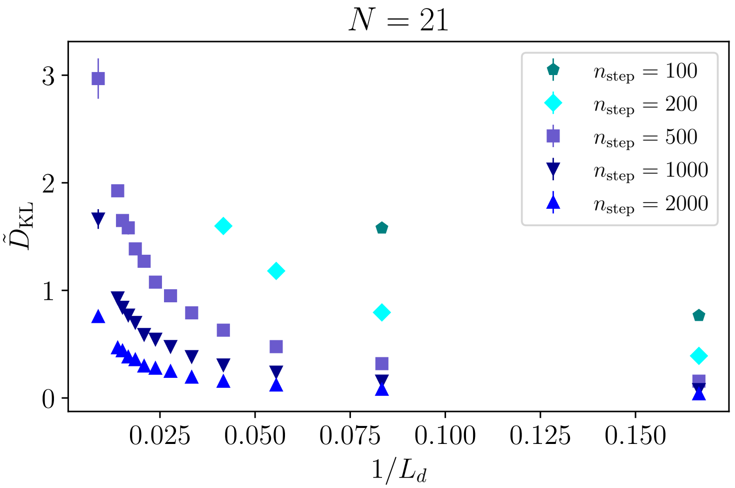
Results for $N = 21$, $\beta = 0.7$, $V = 114^2$

"Slower" evolutions allow for better (but more expensive) sampling



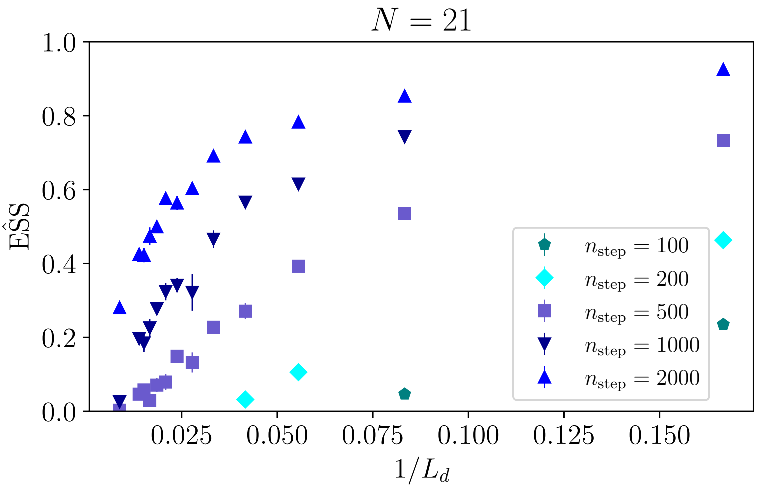
Results for $N = 21$, $\beta = 0.7$, $V = 114^2$

Larger defects require larger n_{step}



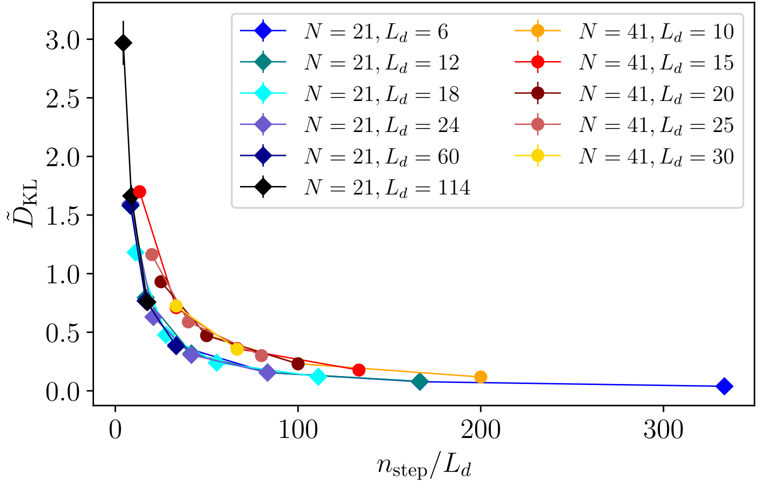
Results for $N = 21,\ \beta = 0.7,\ V = 114^2$

Larger defects require larger n_{step}



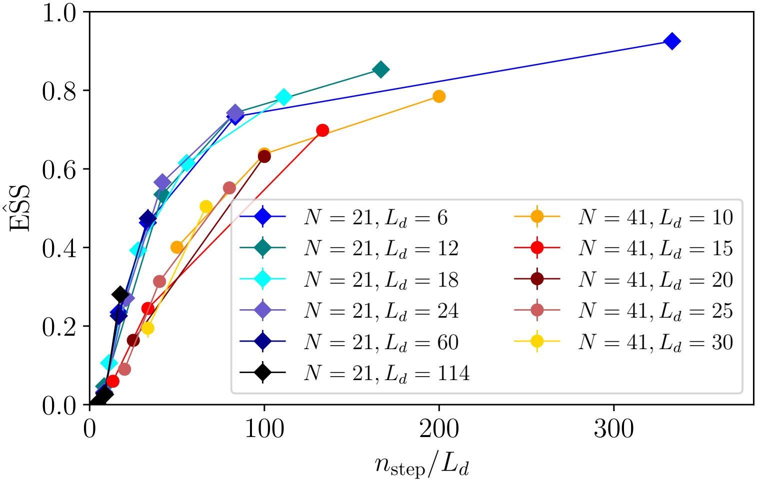
Results for $N = 21$, $\beta = 0.7$, $V = 114^2$

Simple description: the efficiency of the reweighting depends uniquely on $n_{\rm step}/L_d$



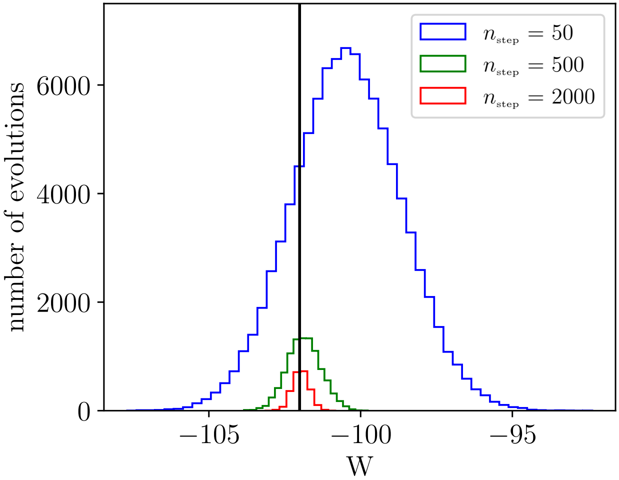
Results for $N = 21,\ \beta = 0.7,\ V = 114^2$ and $N = 41,\ \beta = 0.65,\ V = 132^2$

Simple description: the efficiency of the reweighting depends uniquely on n_{step}/L_d



Results for $N = 21$, $\beta = 0.7$, $V = 114^2$ and $N = 41$, $\beta = 0.65$, $V = 132^2$

Far- and not-so-far from equilibrium: the distribution of the work W and its variance

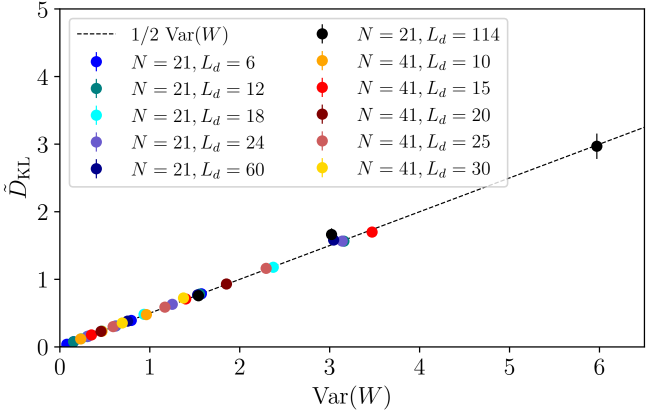


$N = 21, \ \beta = 0.7, \ V = 114^2$

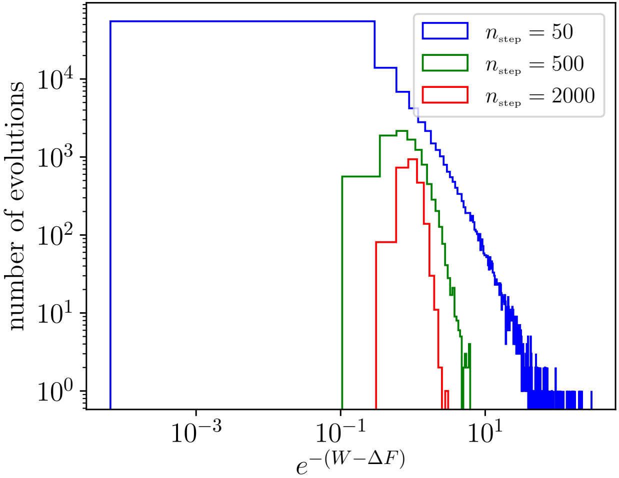
Variance of the work distribution and the $\tilde{D}_{\mathrm{KL}}$ divergence are tightly related!

Elegant result from **[Nicoli; 2020]**:

$$\tilde{D}_{\mathrm{KL}} \simeq \frac{1}{2}\mathrm{Var}(W)$$

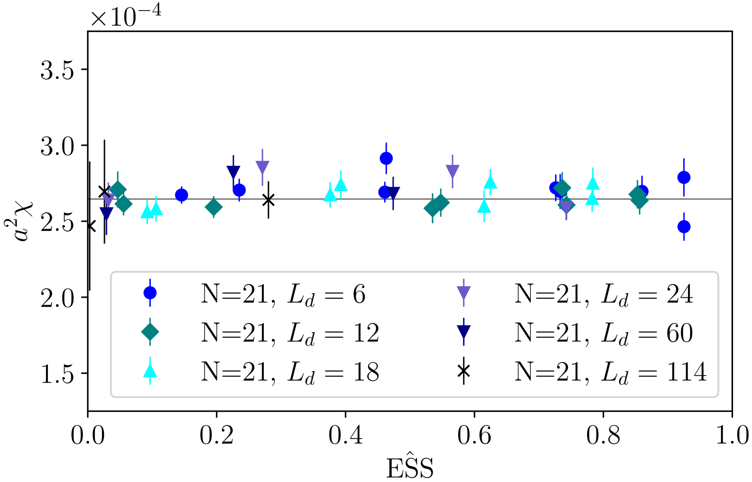


Far- and not-so-far from equilibrium: the distribution of $\exp(-W)$ and rare events



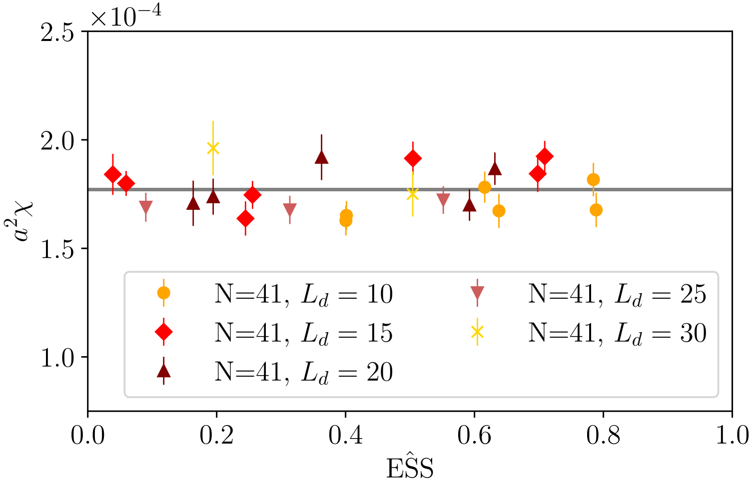
$N = 21, \ \beta = 0.7, \ V = 114^2$

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



Black band is from parallel tempering **[Bonanno et al.; 2019]**

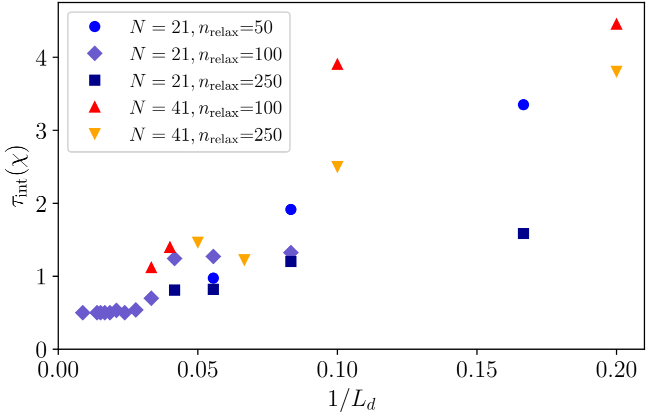
Topological susceptibility for various protocols for $N = 41$, $\beta = 0.65$, $V = 132^2$ (roughly similar numerical effort)



Black band is from parallel tempering **[Bonanno et al.; 2019]**

Autocorrelation times

A quick view of the scaling of autocorrelations times with L_d



Since each non-equilibrium evolution has a sizable cost ($n_{\rm step}$ MC updates), we keep $n_{\rm relax}$ large to avoid large autocorrelations

Periodic Boundary Conditions

For $(N = 21, \beta = 0.7) \to \tau_{\rm int}(\chi) \sim 6 \times 10^4$

For $(N = 41, \beta = 0.65) \to \tau_{\rm int}(\chi) \sim 5 \times 10^5$

Non-equilibrium evolutions

Autocorrelations depend on the choice of $(n_{\rm step}, n_{\rm relax}, L_d)$

$\tau_{\rm int}(\chi) \sim 0.5 - 5$

How to compare these values with the autocorrelations obtained with this method?

Multiply them by the cost of each measurement in terms of Monte Carlo updates, i.e. $n_{\rm step} + n_{\rm relax}$

The "effective" autocorrelation times range between 160 and 5000 both for $N = 21$ and $N = 41$ depending on the protocol

A proper analysis of the error must take into account the fact that we are out-of-equilibrium!

Assessing the efficiency of the non-equilibrium evolutions

The effect on the variance of the protocols used in out-of-equilibrium evolutions can be decomposed into **two distinct** contributions

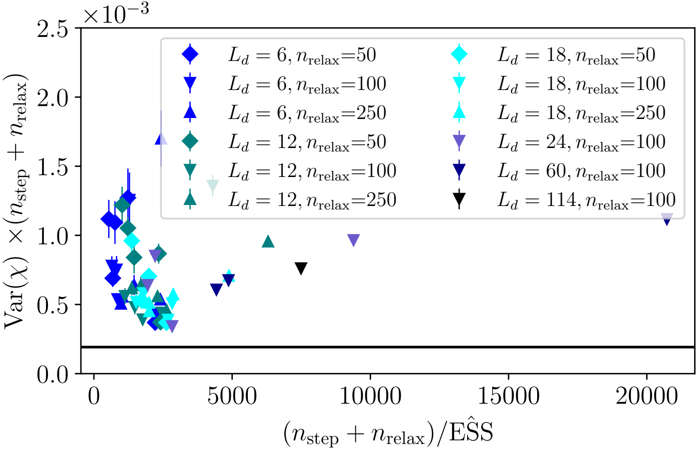
- the $\hat{\mathrm{ESS}}$, that takes into account the effect of the reweighting procedure. It depends **only** on n_{step} and L_d
- τ_{int}, that takes into account the effect of the autocorrelations (as in a normal MC chain). It depends on n_{relax}, L_d and (mildly) on n_{step}

The figure that we will use to assess the efficiency of the method is

$$\mathrm{Var}(\chi)_{\mathrm{NE}} \times (n_{\mathrm{step}} + n_{\mathrm{relax}}) \simeq \mathrm{Var}(\chi)_p \frac{2\tau_{\mathrm{int}}}{\hat{\mathrm{ESS}}} \times (n_{\mathrm{step}} + n_{\mathrm{relax}})$$

Efficiency comparison

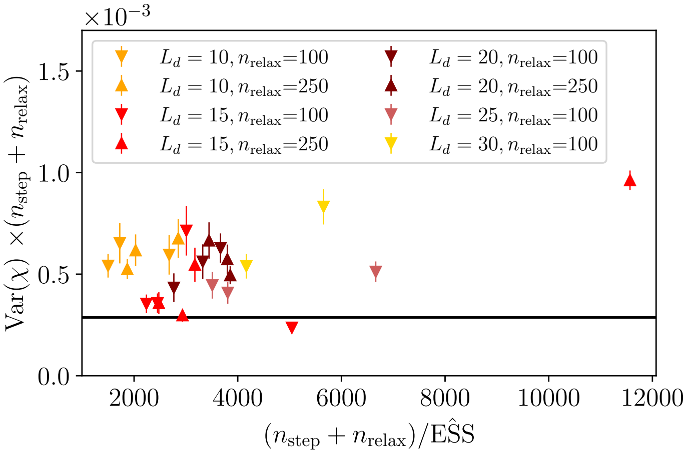
Comparison of variance × cost of one measurement with PTBC



1 measurement with NE costs $n_{\mathrm{step}} + n_{\mathrm{relax}}$ while 1 measurement with PTBC costs n_{replicas}

Efficiency comparison

Comparison of variance $\times$ cost of one measurement with PTBC



1 measurement with NE costs $n_{\mathrm{step}} + n_{\mathrm{relax}}$ while 1 measurement with PTBC costs n_{replicas}

Conclusions and future outlook

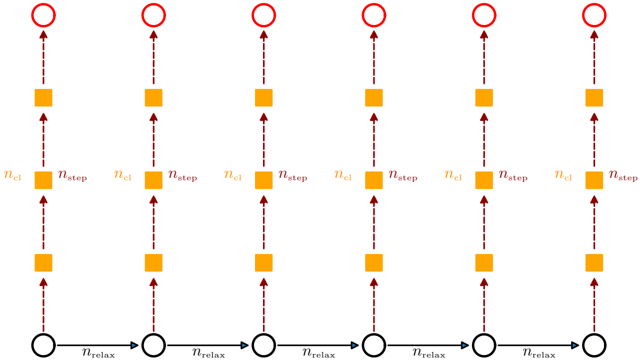
Conclusions

- Out-of-equilibrium simulations are a realistic and effective approach to mitigate critical slowing down
- The features of "well-behaved" probability distributions can be exploited by moving to more complicated target distributions within the same simulation
- Sampling of observables during such evolutions is possible with a particular reweighting
- The effect on the error can be understood quantitatively with the use of specific metrics and its efficiency can be studied rigorously
- New paradigm for MCMC with large τ_{int}: giving up equilibrium in favour of "guided" non-equilibrium simulations
- Generalizations connect with the framework of Normalizing Flows in a non-trivial manner

Natural extension: SNFs

Stochastic Normalizing Flows alternate MC updates with coupling layers **[Wu et al.; 2020],[Caselle et al.; 2022]**

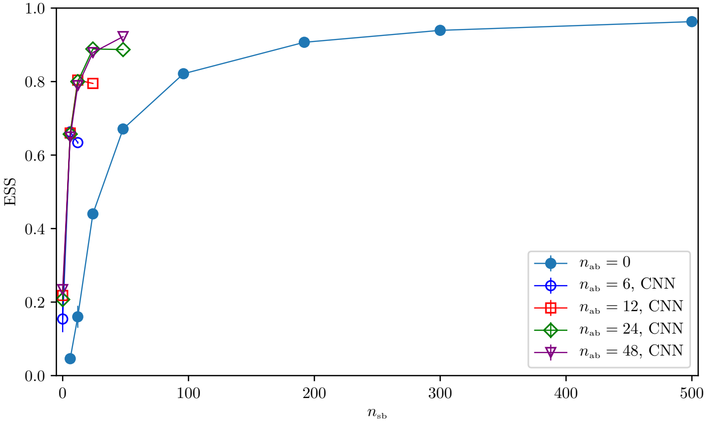
$$\phi_0 \to g_1(\phi_0) \overset{P_{c(1)}}{\to} \phi_1 \to g_2(\phi_1) \overset{P_{c(2)}}{\to} \dots \overset{P_{c(n_{\rm step})}}{\to} \phi$$



essentially share the same loss $\tilde{D}_{\rm KL}$ and same simulation structure

Excellent results in ϕ^4 theory in 2d **[Caselle et al.; 2022]**

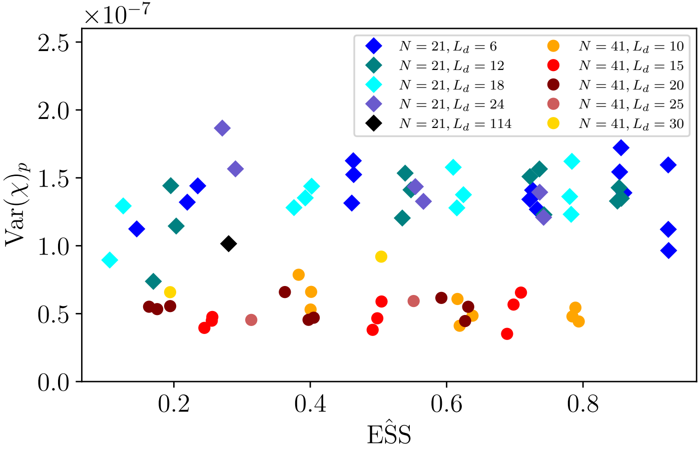
With a proper NN+MC architecture same efficiency as non-equilibrium evolutions with $\sim 1/10$ of MC updates



Idea: **systematically** improve out-of-equilibrium evolutions using SNFs

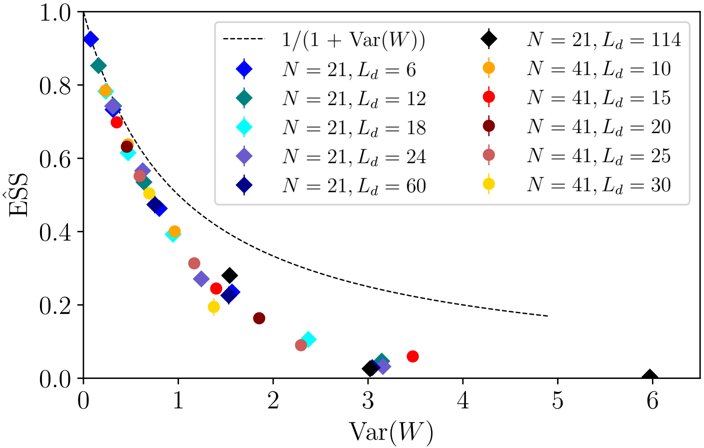
Thank you for your attention!

$$\mathrm{Var}(\chi)_p = \langle (a^2\chi)^2 \rangle - \langle a^2\chi \rangle^2 \simeq \mathrm{Var}(\chi)_{\mathrm{NE}} \frac{\hat{\mathrm{ESS}}}{2\tau_{\mathrm{int}}}$$



Checking the estimator for the ESS

$$\frac{1}{\hat{\mathrm{ESS}}} \simeq 1 + \mathrm{Var}(W), \qquad \mathrm{Var}(W) \ll 1$$



Efficiency-wise **Parallel Tempering** is our benchmark (mainly results from **[Bonanno et al.; 2019]**)

- proposed for $2d$ CP^{N-1} **[Hasenbusch; 2017]**, recently implemented for $4d$ $\mathrm{SU}(N)$ pure-gauge **[Bonanno et al.; 2021, 2022]**
- consider a collection of N_r lattice replicas that differ for the value of $c(r)$, each updated with standard methods
- after updates, propose swaps among configurations via Metropolis test
- other ingredients: hierarchic updates + translation of periodic replica
- decorrelation of topological charge improved thanks to OBC replica
- observable computed on PBC replica

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

Normalizing flows: structure and training

NFs are a deterministic mapping

$$g_\theta(\phi_0) = (g_N \circ \cdots \circ g_1)(\phi_0) \qquad\qquad \phi_0 \sim q_0$$

composed of N invertible transformations $\rightarrow$ the **coupling layers** g_n

The generated distribution for ϕ_N is

$$q_N(\phi_N) = q_0(g_\theta^{-1}(\phi_N)) \prod_n |\det J_n(\phi_n)|^{-1}$$

g_n chosen to be invertible and with an easy-to-compute Jacobian

Training procedure minimizes the **Kullback-Leibler** divergence: measure of the "similarity" between two distributions

$$\tilde{D}_{\mathrm{KL}}(q_N \| p) = \int \mathrm{d}\phi \, q_N(\phi) \log \frac{q_N(\phi)}{p(\phi)}$$

Sampling (not the only possibility: see independent MH)

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \mathrm{d}\phi \, \mathcal{O}(\phi) q_N(\phi) \frac{p(\phi)}{q_N(\phi)} = \frac{Z_0}{Z} \int \mathrm{d}\phi \, \underbrace{q_N(\phi)}_{\text{sample}} \underbrace{\mathcal{O}(\phi)\tilde{w}(\phi)}_{\text{measure}} = \frac{\langle \mathcal{O}(\phi)\tilde{w}(\phi)\rangle_{\phi\sim q_N}}{\langle \tilde{w}(\phi)\rangle_{\phi\sim q_N}}$$

with a weight

$$\tilde{w}(\phi) = \exp\left(-\left\{S[\phi] - S_0[g_\theta^{-1}(\phi)] - \log J\right\}\right)$$

Get Z directly by sampling from q_N **[Nicoli et al.; 2020]**

$$Z = \int \mathrm{d}\phi \, \exp(-S[\phi]) = Z_0 \int \mathrm{d}\phi \, q_N(\phi)\tilde{w}(\phi) = Z_0 \langle \tilde{w}(\phi)\rangle_{\phi\sim q_N}$$

Train minimizing

$$\tilde{D}_{\mathrm{KL}}(q_N \| p) = -\langle \log \tilde{w}(\phi)\rangle_{\phi\sim q_N} + \log \frac{Z}{Z_0}$$

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

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

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

stochastic non-equilibrium evolutions

$$\phi_0 \overset{P_{\eta_1}}{\to} \phi_1 \overset{P_{\eta_2}}{\to} \dots \overset{P_{\eta_N}}{\to} \phi_N$$

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

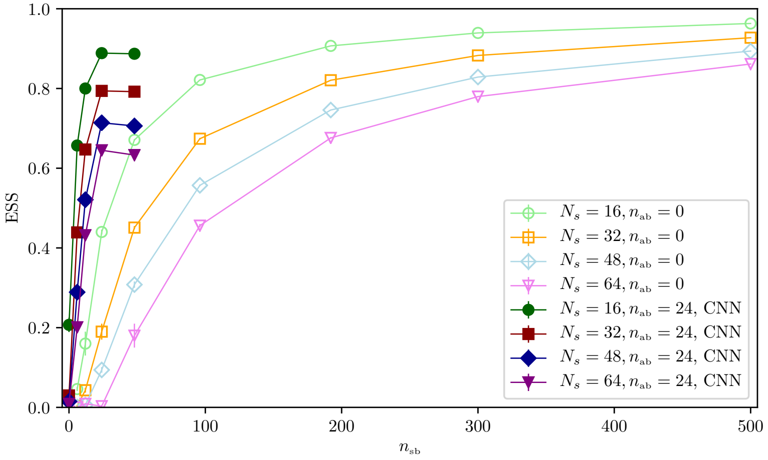
Stochastic Normalizing Flows (introduced in **[Wu et al.; 2020]**)

$$\phi_0 \to g_1(\phi_0) \overset{P_{\eta_1}}{\to} \phi_1 \to g_2(\phi_1) \overset{P_{\eta_2}}{\to} \dots \overset{P_{\eta_N}}{\to} \phi_N$$

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

SNFs for ϕ^4 at various volumes

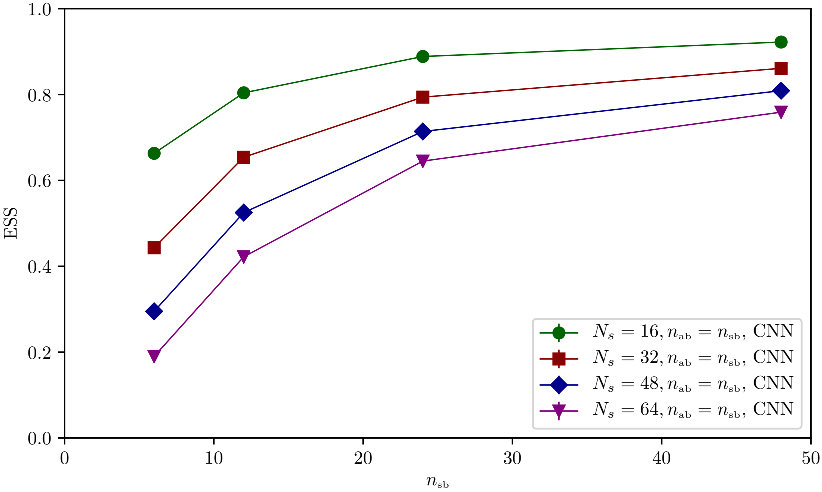
Training length: 10^4 epochs for all volumes. Slowly-improving regime reached fast



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

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

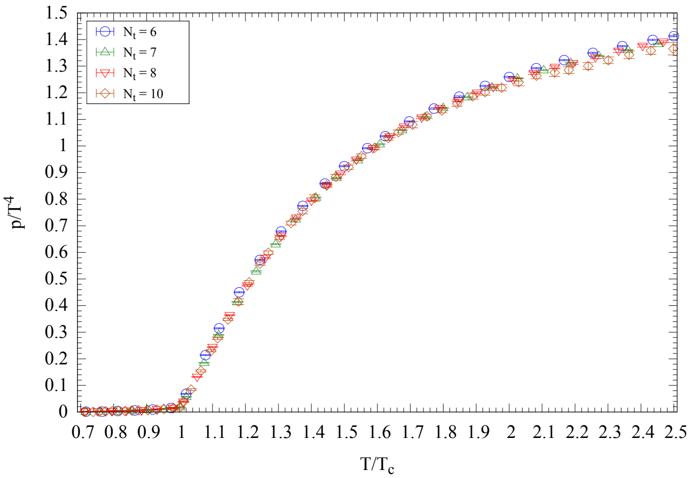
$$\frac{p(T)}{T^4} - \frac{p(T_0)}{T_0^4} = \left(\frac{N_t}{N_s}\right)^3 \log\langle e^{-W_{\mathrm{SU}(N_c)}}\rangle_{\mathrm{f}}$$

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)

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



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