Out-of-equilibrium simulations to fight topological freezing

Alessandro Nada

Università degli Studi di Torino

40th International Symposium on Lattice Field Theory

Fermilab (Batavia, IL), 31st July - 4th August 2023

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 mitigates the problem by removing the sectors

Drawback: strong finite-size effects have to be taken into account

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

Jarzynski's equality in MCMC

Consider two distributions q_0 and p

$$q_0 = \exp(-S_0)/Z_0 \to \cdots \to p = \exp(-S)/Z$$

Ratio of the two partition functions computed directly with an average over **non-equilibrium processes [Jarzynski; 1997]**

$$\frac{Z}{Z_0} = \langle \exp(-W) \rangle_f$$

Jarzynski's equality in MCMC

Consider two distributions q_0 and p

$$q_0 = \exp(-S_0)/Z_0 \to \cdots \to p = \exp(-S)/Z$$

Ratio of the two partition functions computed directly with an average over **non-equilibrium processes [Jarzynski; 1997]**

$$\frac{Z}{Z_0} = \langle \exp\left(-W\right) \rangle_f$$

"Guided" MCMC evolution:

- over n_{step} intermediate steps
- the system evolves using regular MC updates with a transition probability $P_{c(n)}(\phi_n \to \phi_{n+1})$ that changes along the evolution according to a **protocol** $c(n)$
- $c(n)$ interpolates between q_0 and p

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

Jarzynski's equality in MCMC

Consider two distributions q_0 and p

$$q_0 = \exp(-S_0)/Z_0 \to \cdots \to p = \exp(-S)/Z$$

Ratio of the two partition functions computed directly with an average over **non-equilibrium processes [Jarzynski; 1997]**

$$\frac{Z}{Z_0} = \langle \exp\left(-W\right) \rangle_f$$

"Guided" MCMC evolution:

- over n_{step} intermediate steps
- the system evolves using regular MC updates with a transition probability $P_{c(n)}(\phi_n \to \phi_{n+1})$ that changes along the evolution according to a **protocol** $c(n)$
- $c(n)$ interpolates between q_0 and p

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

Along the process we compute the **work**

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

The average on the processes can be written rigorously

$$\frac{Z}{Z_0} = \langle \exp(-W) \rangle_f = \int \mathrm{d}\phi_0\, \mathrm{d}\phi_1 \ldots \mathrm{d}\phi\; q_0(\phi_0)\, P_{\mathrm{f}}[\phi_0, \phi_1, \ldots, \phi]\, \exp(-W)$$

with $P_{\mathrm{f}}[\phi_0, \phi_1, \ldots, \phi] = \prod_{n=0}^{N-1} P_{c(n)}(\phi_n \to \phi_{n+1})$

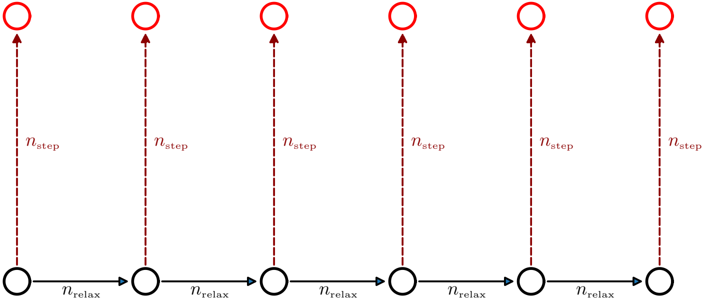
- the *actual* probability distribution at each step is NOT the equilibrium distribution $\sim \exp(-S_{c(n)})$: it's a non-equilibrium process!
- several applications already! interface free-energy**[Caselle et al.; 2016]**, $\mathrm{SU}(3)$ equation of state in 4d**[Caselle et al.; 2018]**, running coupling **[Francesconi et al.; 2020]**, entanglement entropy **[Bulgarelli and Panero; 2023]**

  talk by Andrea Bulgarelli $\to$ QCQI session **Fri 9:20**
- much more general idea! compute v.e.v. for p with

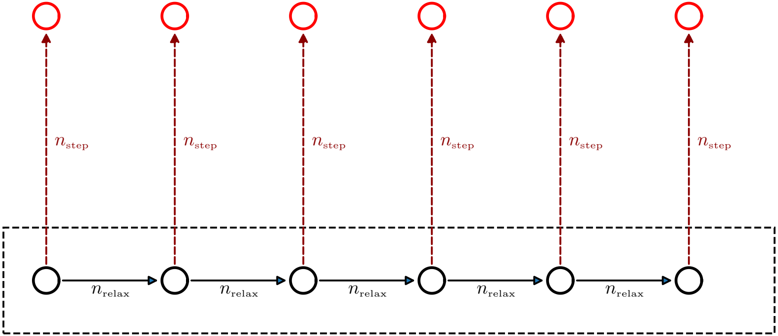
$$\langle \mathcal{O} \rangle = \frac{\langle \mathcal{O}(\phi) \exp(-W(\phi_0 \to \phi)) \rangle_{\mathrm{f}}}{\langle \exp(-W(\phi_0 \to \phi)) \rangle_{\mathrm{f}}}$$

- this work: rigorously sample PBC by starting from OBC!

A new paradigm to perform MCMC

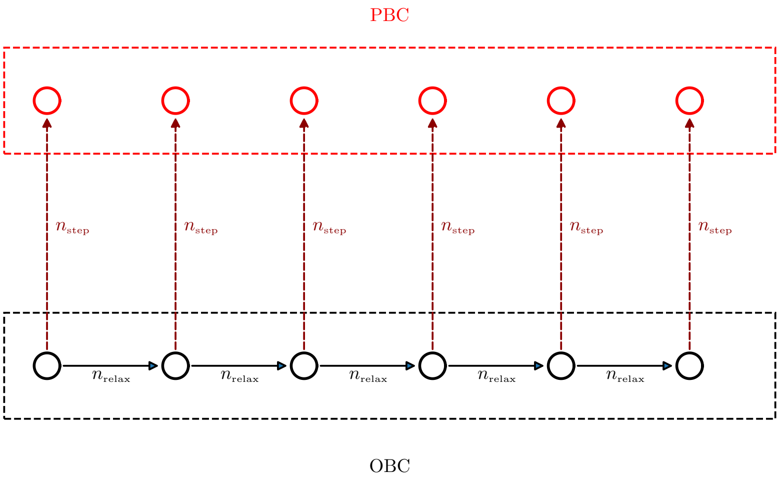


A new paradigm to perform MCMC



OBC

A new paradigm to perform MCMC



A common framework with Normalizing Flows

We measure the "quality" of the out-of-equilibrium evolutions with

$$\tilde{D}_{\mathrm{KL}}(q_0 P_{\mathrm{f}} \| p P_{\mathrm{r}}) = \int \prod_{i=0}^{N} \mathrm{d}\phi_i \, q_0(\phi_0) P_{\mathrm{f}}[\phi_0, \phi_1, \dots, \phi_N] \log \frac{q_0(\phi_0) P_{\mathrm{f}}[\phi_0, \phi_1, \dots, \phi_N]}{p(\phi_N) P_{\mathrm{r}}[\phi_N, \phi_{N-1}, \dots, \phi_0]}$$

$\rightarrow$ measure of how reversible the process is!

Clear "thermodynamic" interpretation

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

Metric

Effective Sample Size as metric to evaluate architectures

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

ESS $= 1 \rightarrow$ equilibrium

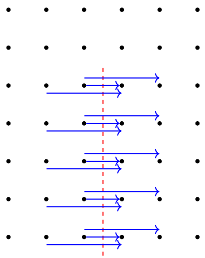
The CP^{N-1} model with a defect

Improved action

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

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

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

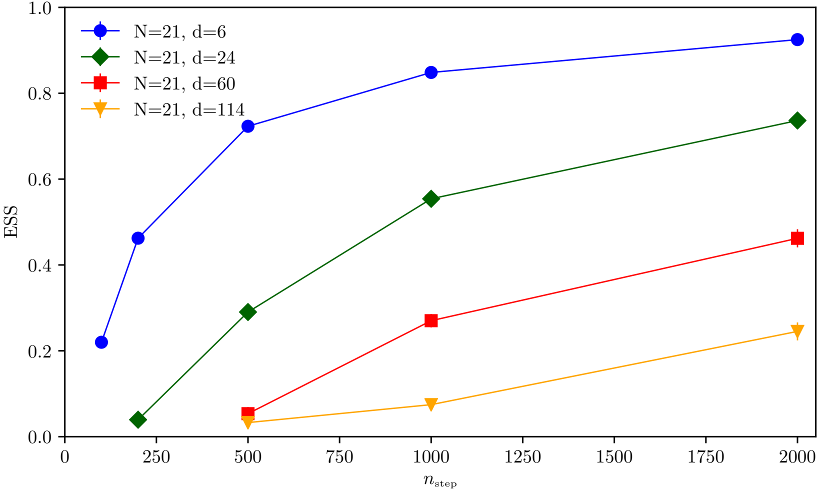


The $k_\mu^{(n)}(x)$ regulate the boundary conditions along a given defect D

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

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

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



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

Topological observables and benchmarks

Geometric definition of the topological charge Q

$$Q = \frac{1}{2\pi} \sum_x \mathrm{Im}\, \log \Pi_{12}(x)$$

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 topological susceptibility

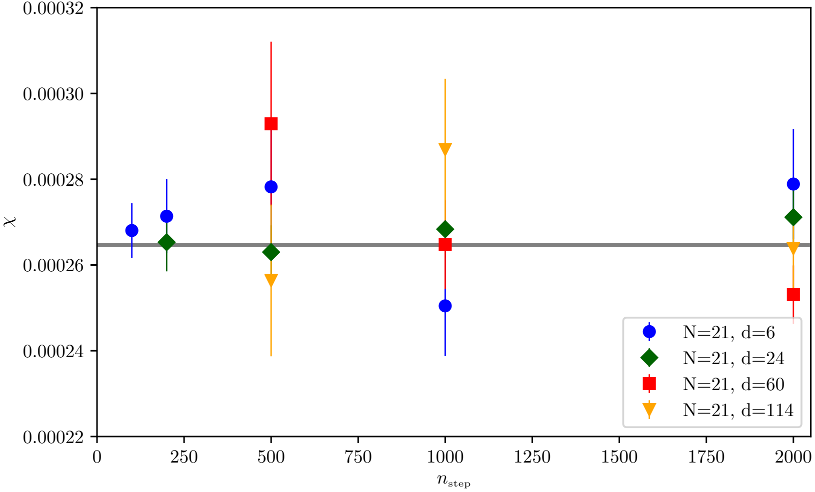
$$\chi = \frac{1}{V} \langle Q^2 \rangle$$

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
- decorrelation of topological charge improved thanks to OBC replica
- observable computed on PBC replica

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

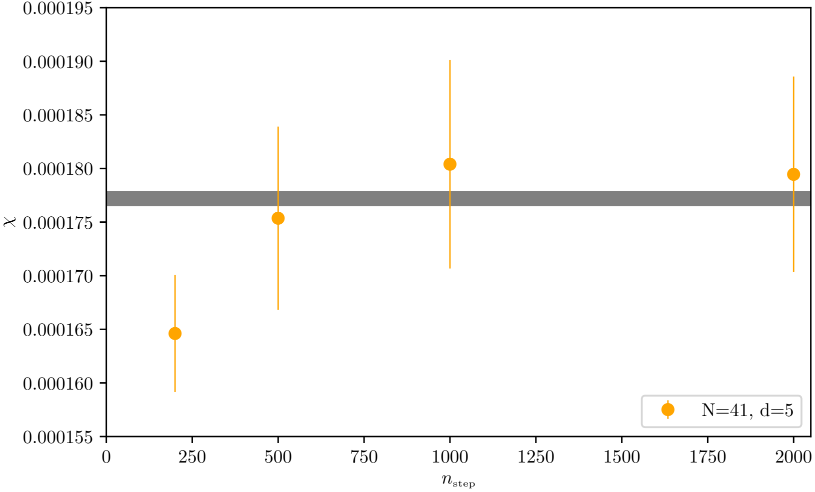
Note that with OBC $\rightarrow \tau_{int}(\chi) \sim 50$ **(preliminary)**



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

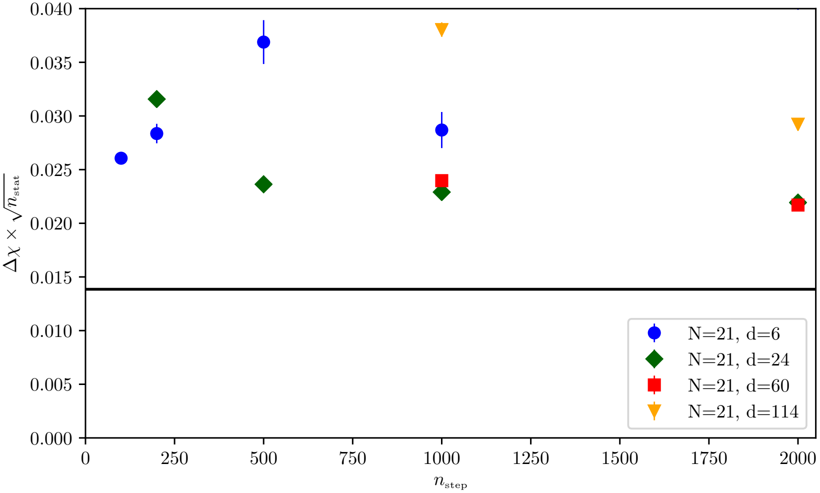
Topological susceptibility for various protocols for $N = 41$, $\beta = 0.65$, $V = 132^2$ (roughly similar numerical effort)

Note that with OBC $\rightarrow \tau_{int}(\chi) \sim 170$ **(preliminary)**

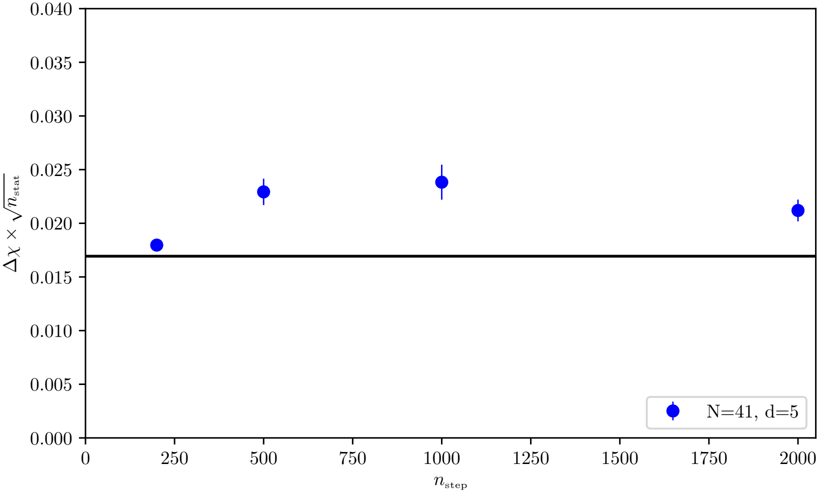


Black band is from parallel tempering **[Bonanno et al.; 2019]** $\rightarrow$ with $\times \sim 40$ numerical cost

Efficiency of various methods as error $\times\sqrt{\text{effort}}$ for $N = 21, \beta_L = 0.7$ **(preliminary)**

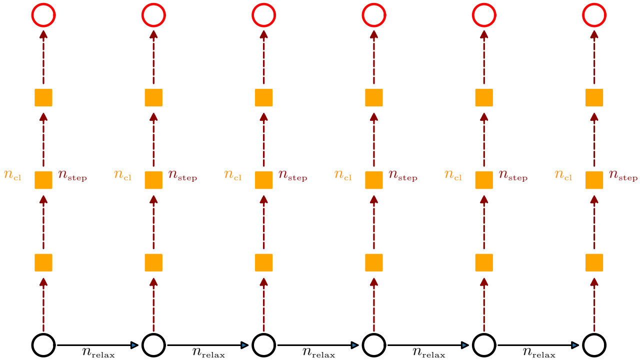


Efficiency of various methods as error $\times\sqrt{\text{effort}}$ for $N = 41, \beta_L = 0.65$ **(preliminary)**



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

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

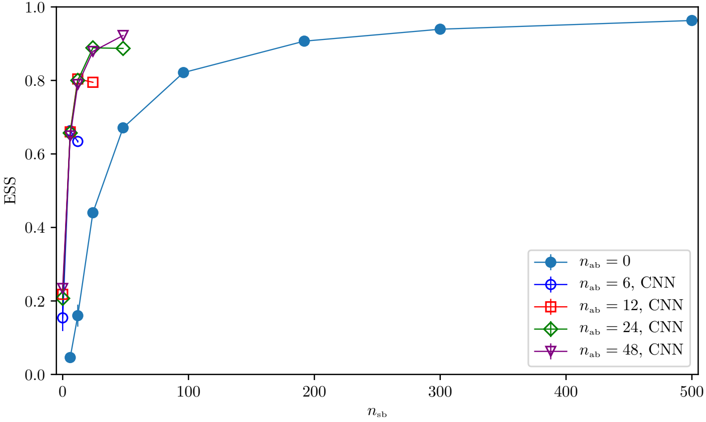


essentially share the same loss $\tilde{D}_{\mathrm{KL}}$ and same simulation structure

Encouraging results from SNFs in a toy model

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

see **poster** by **Joe Marsh Rossney** on equivariant NFs for CP^{N-1}!

Thank you for your attention!

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

Jarzynski's equality [Jarzynski; 1997] is a beautiful result from non-equilibrium statistical mechanics

$$\left\langle \exp\left(-\frac{W}{T}\right)\right\rangle_f = \exp\left(-\frac{\Delta F}{T}\right)$$

valid for a given process f

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

$$\exp\left(-\frac{\Delta F}{T}\right) = \left\langle \exp\left(-\frac{W}{T}\right)\right\rangle_f \geq \exp\left(-\frac{\langle W\rangle_f}{T}\right)$$

and we get the Second Law of thermodynamics

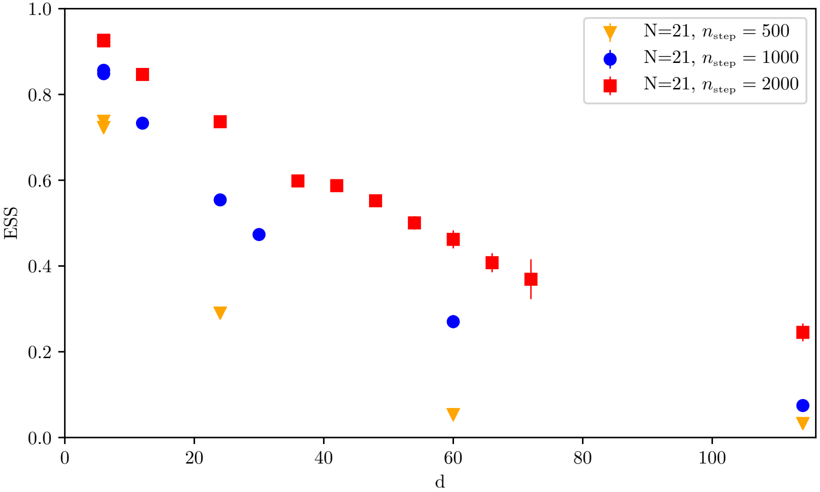
$$\langle W\rangle_f \geq \Delta F$$

Apart from the real world, it can be proved for several processes

$\rightarrow$ most relevantly for us: **Markov Chain Monte Carlo** for lattice field theory

Numerical experiments with various defect sizes (up to full OBC)

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



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 with Normalizing flows

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

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

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

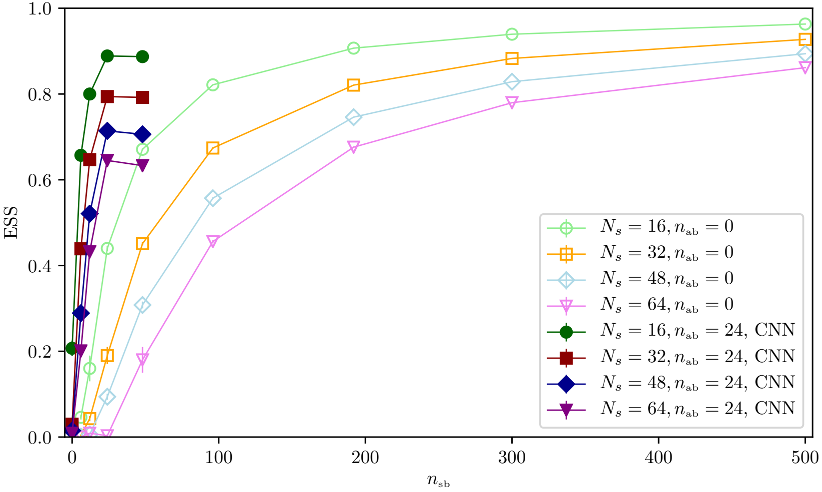
Stochastic Normalizing Flows (introduced in **[Wu et al.; 2020]**)

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

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

SNFs for ϕ^4 at various volumes

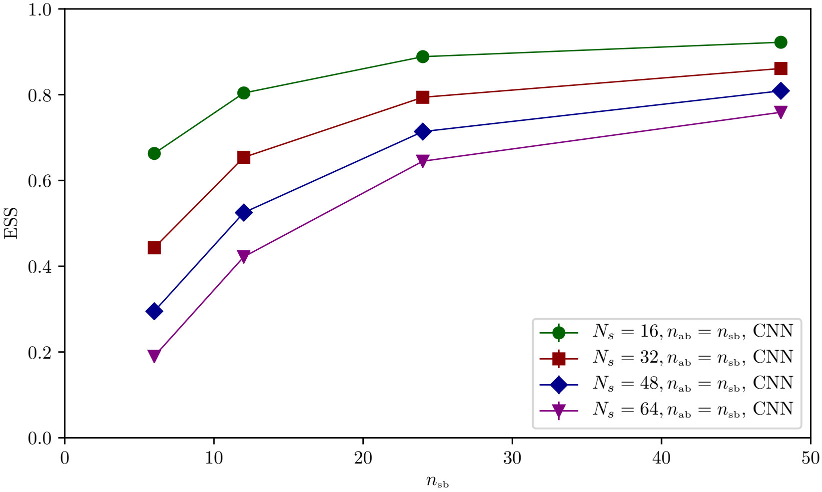
Training length: 10^4 epochs for all volumes. Slowly-improving regime reached fast



Interesting behaviour for all volumes: a peak for

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]**

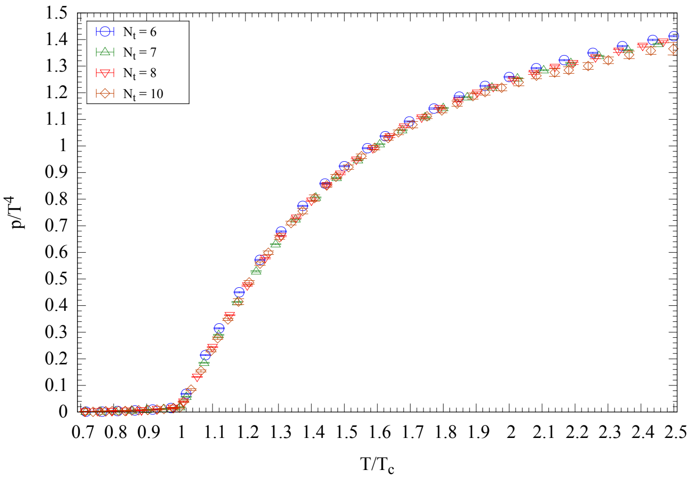
Goal: extract the pressure with Jarzynski's equality

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

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 $\beta \simeq 7$