

Sampling SU(3) pure gauge theory with Stochastic Normalizing Flows

Andrea Bulgarelli,^a Elia Cellini^a and Alessandro Nada^{a,*}

*^aDipartimento di Fisica, Università degli Studi di Torino and INFN, Sezione di Torino,
Via Pietro Giuria 1, I-10125 Turin, Italy*

E-mail: alessandro.nada@unito.it

Non-equilibrium Monte Carlo simulations based on Jarzynski's equality are a well-understood method to compute differences in free energy and also to sample from a target probability distribution without the need to thermalize the system under study. In each evolution, the system starts from a given base distribution at equilibrium and it is gradually driven out-of-equilibrium while evolving towards the target parameters. If the target distribution suffers from long autocorrelation times, this approach represents a promising candidate to mitigate critical slowing down. Out-of-equilibrium evolutions are conceptually similar to Normalizing Flows and they can be combined into a recently-developed architecture called Stochastic Normalizing Flows (SNFs). In this contribution we first focus on the promising scaling with the volume guaranteed by the purely stochastic approach in the SU(3) lattice gauge theory in 4 dimensions; then, we define an SNF by introducing gauge-equivariant layers between the out-of-equilibrium Monte Carlo updates, and we analyse the improvement obtained as well as the inherited scaling with the volume. Finally, we discuss how this approach can be systematically improved and how simulations of lattice gauge theories in four dimensions for large volumes and close to criticality can be realistically achieved.

*The 41st International Symposium on Lattice Field Theory (LATTICE2024)
28 July - 3 August 2024
Liverpool, UK*

*Speaker

1. Introduction

Flow-based architectures provide an innovative approach to sample from a probability distribution for which standard local Monte Carlo algorithms are highly inefficient, e.g. they suffer from critical slowing down. The idea behind these architectures is straightforward and consists in building a transformation that maps a prior distribution (which is easy to sample from) to the desired target distribution. While this approach goes back to the concept of trivializing maps [1], recent developments in the deep learning community provided a powerful set of tools to build these mappings in the form of Normalizing Flows (NFs) [2], yielding impressive results in low-dimensional lattice field theories (see ref. [3] for a review).

A particularly challenging form of critical slowing down, for example in the case of the SU(3) lattice gauge theory in four dimensions, is the so-called topological freezing [4]: when approaching the continuum limit, the autocorrelation times of topological observables increase exponentially with the inverse of the lattice spacing. Normalizing Flows appear to be a perfect candidate to mitigate this issue: in a flow-based approach, for example, one could map the theory at coarser lattice spacing (corresponding to an inverse bare coupling β_0) to a finer lattice spacing (with $\beta > \beta_0$). However, the performances of NFs appear to scale poorly when the number of degrees of freedom (and/or the dimensionality of the model) increases [5, 6], making their application to realistic volumes in four dimensions a challenge.

Non-equilibrium Markov Chain Monte Carlo (NE-MCMC) simulations based on Jarzynski's equality, which have seen widespread applications in several contexts in lattice field theory in the last years [7–10], also provide a way to mitigate critical slowing down. In this case, the transformation from the prior to the target distribution happens through a non-equilibrium process, which is implemented by updating the system with a specific protocol and driving it out of equilibrium in a controlled way. Indeed, in ref. [11] a severe topological freezing was mitigated by sampling the target distribution with this approach and a scaling of the length of these evolutions with the relevant degrees of freedom was found. Interestingly, these non-equilibrium simulations share a very similar framework with Normalizing Flows, and in refs. [12] they were naturally combined into a new architecture called Stochastic Normalizing Flows (SNFs) [12, 13], that has been recently used for numerical simulations of effective string theory [14]. In particular, the updates of (NE-MCMC) are interleaved with the building blocks of Normalizing Flows and in this way a hybrid architecture that mixes deterministic and stochastic transformations is created.

In this contribution we introduce for the first time a SNF for the SU(3) Yang-Mills theory in four spacetime dimensions and we test it on a mapping from coarser to finer lattice spacings. We analyse the features characterizing the training of the parameters of this architecture and we compare its performances with NE-MCMC. Most importantly, we confirm the findings of ref. [11] concerning the scaling of NE-MCMC and we find that the same applies to SNFs as well.

2. Out-of-equilibrium evolutions

In a non-equilibrium Markov Chain Monte Carlo (NE-MCMC) a sequence of configurations is generated using a transition probability $P_{c(n)}$ that changes along the transformation, i.e.

$$U_0 \xrightarrow{P_{c(1)}} U_1 \xrightarrow{P_{c(2)}} U_2 \xrightarrow{P_{c(3)}} \dots \xrightarrow{P_{c(n_{\text{step}})}} U_{n_{\text{step}}} \quad (1)$$

The first configuration U_0 is sampled from the prior distribution $q_0 = \exp(-S_0)/Z_0$, while the last transition probability $P_{c(n_{\text{step}})}$ is the one corresponding to the target distribution $p = \exp(-S)/Z$. The transition probabilities $P_{c(n)}$ are defined by a set of changing parameters $c(n)$, called the *protocol*, that controls the intermediate actions used to update the system. Formally, the overall out-of-equilibrium evolution is defined by a forward transition probability

$$\mathcal{P}_f[U_0, \dots, U] = \prod_{n=1}^{n_{\text{step}}} P_{c(n)}(U_{n-1} \rightarrow U_n). \quad (2)$$

Using Crooks' theorem [15] one can re-express the ratio of \mathcal{P}_f with the transition probability of the reverse transformation \mathcal{P}_r

$$\frac{q_0 \mathcal{P}_f[U_0, \dots, U]}{p \mathcal{P}_r[U, \dots, U_0]} = \exp(W - \Delta F) \quad (3)$$

with $\Delta F = -\log Z/Z_0$ and W being the (dimensionless) work performed on the system in a given transformation while changing the parameters controlling the protocol, i.e.

$$W = \sum_{n=0}^{n_{\text{step}}-1} \{S_{c(n+1)}[U_n] - S_{c(n)}[U_n]\}. \quad (4)$$

Integrating eq. (3) over all possible transformation yields Jarzynski's equality [16]

$$\langle \exp(-W) \rangle_f = \frac{Z}{Z_0} = \exp(-\Delta F). \quad (5)$$

The average $\langle \dots \rangle_f$ is formally defined as

$$\langle \mathcal{A} \rangle_f = \int dU_0 \dots dU q_0(U_0) \mathcal{P}_f[U_0, \dots, U] \mathcal{A}[U_0, \dots, U]. \quad (6)$$

While Jarzynski's equality provides a powerful tool to compute directly (ratios of) partition functions, the scope of non-equilibrium MCMC is much wider. Indeed, it is possible to sample any vacuum expectation value on the target distribution p with a reweighting-like formula:

$$\langle \mathcal{O} \rangle_p = \frac{\langle \mathcal{O} \exp(-W) \rangle_f}{\langle \exp(-W) \rangle_f}; \quad (7)$$

on the right-hand side the work W is calculated along the whole non-equilibrium evolution while the observable \mathcal{O} is computed only on the last configuration $U_{n_{\text{step}}}$.

It is clear that any hope of tackling critical slowing down on the target distribution p requires a careful evaluation of the effectiveness of NE-MCMC in sampling p : in particular, for relatively small values of n_{step} we expect an overlap problem to appear between p and the generated distribution at the end of the non-equilibrium transformation, i.e.

$$q(U) = \int dU_0 \dots dU_{n_{\text{step}}-1} q_0(U_0) \mathcal{P}_f[U_0, \dots, U], \quad (8)$$

which is generally intractable. One possible figure of merit is the Kullback-Leibler (KL) divergence between the forward and reverse transition probabilities, that we can write as

$$D_{\text{KL}}(q_0 \mathcal{P}_f \| p \mathcal{P}_r) = \langle W \rangle_f + \log \frac{Z}{Z_0} = \langle W \rangle_f - \Delta F. \quad (9)$$

This quantity measures the similarity between the same evolution running forwards and backwards for a given protocol $c(n)$: as expected, if it is zero, then the transformation is reversible and $\langle W \rangle_f = \Delta F$ ¹. Furthermore, it is easy to derive that $D_{\text{KL}}(q||p) \leq D_{\text{KL}}(q_0 \mathcal{P}_f || p \mathcal{P}_f)$, thus putting an upper bound on the similarity of $q(U)$ and $p(U)$.

Another relevant metric we use to evaluate out-of-equilibrium evolutions is the Effective Sample Size, defined in this context as

$$\text{ESS} = \frac{\langle \exp(-W) \rangle_f^2}{\langle \exp(-2W) \rangle_f}. \quad (10)$$

It provides an approximation of the ratio of the variance of an observable O computed with the estimator of eq. (7) on independent evolutions and the variance of the same observable computed on independent configurations sampled directly from p .

3. Stochastic Normalizing Flows for SU(3) Yang-Mills theory

Stochastic Normalizing Flows [12, 13] are an extremely promising architecture that can systematically improve the efficiency of out-of-equilibrium evolutions. In particular, in each evolution non-equilibrium MCMC updates are alternated with so-called *coupling layers* (here denoted as g_l), i.e. deterministic transformations that are generally used to build Normalizing Flow architectures. Compared to a NE-MCMC (see eq. (1)), a sequence of configurations in a forward pass in a SNF looks like the following

$$U_0 \xrightarrow{g_1} g_1(U_0) \xrightarrow{P_{c(1)}} U_1 \xrightarrow{g_2} g_2(U_1) \xrightarrow{P_{c(2)}} U_2 \xrightarrow{g_3} \dots \xrightarrow{P_{c(n_{\text{step}})}} U_{n_{\text{step}}} \quad (11)$$

where in each step a configuration is first transformed using the layer g_n and then updated using a Monte Carlo algorithm with transition probability $P_{c(n)}$.

In this work we use gauge-equivariant coupling layers g_l inspired by the ones introduced in refs. [17, 18]. Each link is transformed according to a stout-smearing transformation [19], applied using 8 masks (4 for each space-time direction and 2 for even-odd sites) to ensure invertibility and a manageable computation of the Jacobian. In particular we have that each active link $U_\mu(x)$ is transformed as

$$U'_\mu(x) = g_l(U_\mu(x)) = \exp\left(Q_\mu^{(l)}(x)\right) U_\mu(x) \quad (12)$$

with

$$Q_\mu^{(l)}(x) = 2 \left[\Omega_\mu^{(l)}(x) \right]_{\text{TA}} \quad (13)$$

where TA is the traceless and antihermitian operation. $\Omega_\mu^{(l)}(x)$ is a sum of untraced plaquettes, namely

$$\Omega_\mu^{(l)}(x) = C_\mu^{(l)}(x) U_\mu^\dagger(x). \quad (14)$$

Here $U_\mu(x)$ is the active link, while the $C_\mu^{(l)}(x)$ is a weighted sum of the staples $S_{\mu\nu}(x)$, that are composed exclusively of frozen links, i.e. links which are not transformed in the current mask:

$$C_\mu(x) = \sum_{\nu \neq \mu} \rho_{\mu\nu}^{(l)}(x) S_{\mu\nu}(x) \quad (15)$$

¹The positivity of the KL divergence implies that $\langle W \rangle_f \geq \Delta F$, i.e. the Second Law of thermodynamics for these types of MCMC transformations out of equilibrium

The stout-smearing parameters $\rho_{\mu\nu}^{(l)}(x)$ in eq. (15) have to be determined using a training procedure. In this work we simply set them to be invariant under translations and (discrete) rotations, i.e. $\rho_{\mu\nu}^{(l)}(x) = \rho^{(l)}$; we also set them to be shared by all masks in each layer l . Concerning the training procedure itself, we minimize the same KL divergence of eq. (9), with the only difference being the use of a "generalized" work, defined as

$$W = \sum_{n=0}^{n_{\text{step}}-1} S_{c(n+1)}(g_n(U_n)) - S_{c(n)}(g_n(U_n)) - \log |\det J_n(U_n)| \quad (16)$$

which takes into account the logarithm of the determinant of the Jacobian J_n of the coupling layer g_n . Using this definition of work, the formulae of the previous section still hold: indeed SNFs can be used both for the computation of ratios of partition functions using Jarzynski's equality (5) and for the sampling of generic observables using eq. (7).

4. Numerical results

We test the scaling of both NE-MCMC and SNFs on the SU(3) pure lattice gauge theory in 4 dimensions for transformations in the lattice spacing. We use the plaquette (Wilson) action and the Monte Carlo update of choice during the out-of-equilibrium evolutions is composed of 1 step of heatbath plus 4 steps of over-relaxation. The system is thermalized at a given initial value of the inverse bare coupling β_0 and then it is driven out of equilibrium using a linear protocol in β :

$$\beta(n) = \beta_0 + n \frac{\beta_t - \beta_0}{n_{\text{step}}}, \quad (17)$$

with $\beta_t > \beta_0$ being the target value of the inverse bare coupling.

We studied lattice sizes $L/a = [10, 12, 16, 20]$ and we focused our effort on a protocol in β from $\beta_0 = 6.02$ to $\beta_t = 6.178$. This approximately corresponds to a change in the physical size of the system of 1.8fm \rightarrow 1.4fm for a $L/a = 20$ lattice. We remark that perfectly compatible results have been found also for the $\beta = 5.896 \rightarrow 6.037$ protocol².

In order to find the optimal parameters of the SNF we minimize the KL divergence of eq. (9) using an Adam optimizer. The peculiar structure of the work of eq. (16) is such that we can train along each forward pass each layer separately, in a procedure reminiscent of ref. [20], thus keeping the memory constant for any value of n_{step} . Training of the SNFs was performed only for architectures with $n_{\text{step}} = 16, 32, 64$ for all volumes; the length of the training was 800 epochs for $n_{\text{step}} = 16, 32$ and 400 epochs for $n_{\text{step}} = 64$. Results for the smearing parameters on the largest lattice are presented in fig. 1.

We observed that the values of ρ roughly collapse on the same curve if rescaled with n_{step} : following this intuition we then implemented a global interpolation of these results obtaining a function valid for all n_{step} . Possibly due to the relative simplicity of the layers used in this work, we observed no differences when sampling at larger n_{step} between a direct training procedure for

²We remark that at these values of the inverse coupling, using a heatbath+overrelaxation algorithm on a standard equilibrium simulation, topological observables are not frozen yet. We leave the application of these architectures on finer lattice spacings (and correspondingly larger volumes) to future work.

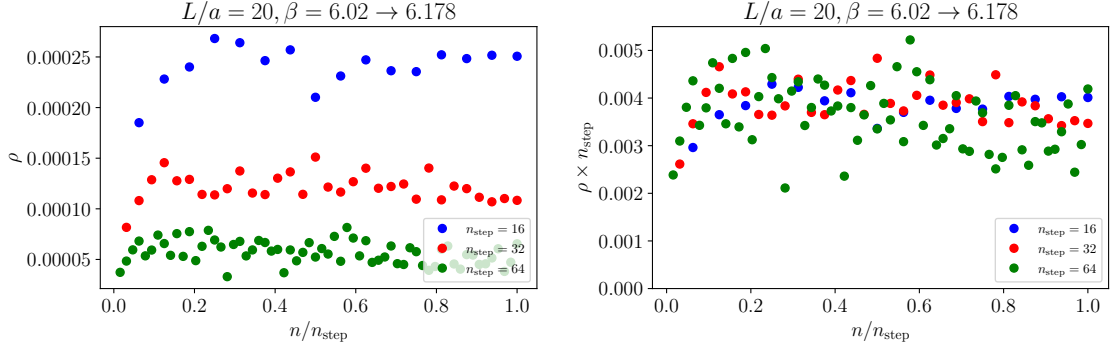


Figure 1: Value of the learned parameter ρ along the SNF for the n -th layer (left panel) and the same value rescaled for n_{step} (right panel), for a training performed on a $L/a = 20$ lattice for three values of n_{step} .

that architecture and the interpolation carried out uniquely with $n_{\text{step}} = 16, 32, 64$. In the following all results have been obtained with the latter. While we performed this training procedure for each volume separately, we noticed that transfer learning between the volumes seems to be possible as well. Namely, we trained ρ on $L/a = 12, 16, 20$ lattices and then applied the results on a SNF with $n_{\text{step}} = 1024$ for the largest lattice $L/a = 20$, noticing no discernible differences in the KL divergence or in the ESS.

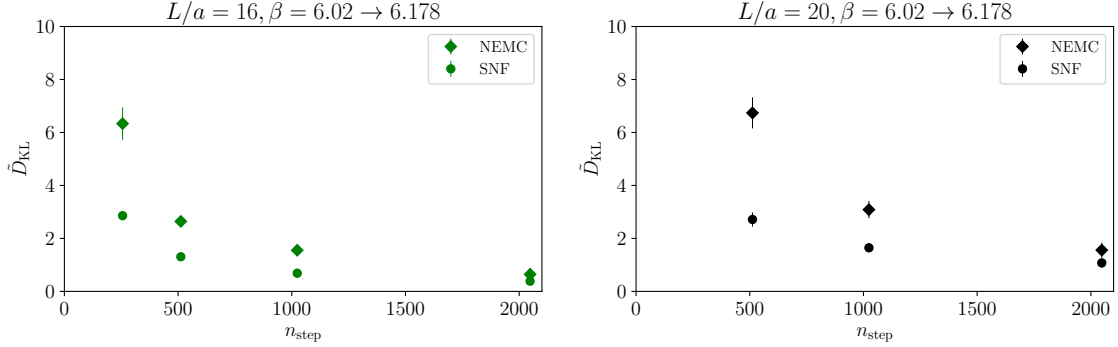


Figure 2: Results for the KL divergence for $L/a = 16$ (left panel) and $L/a = 20$ (right panel).

In fig. 2 results for the KL divergence of eq. (9) are presented for the two largest volumes, both for the NE-MCMC and for the trained SNFs, for several values of n_{step} . For increasing n_{step} , i.e. for slower evolutions closer to equilibrium, D_{KL} decreases rather quickly for all architectures and as expected the results indicate a convergence to 0, i.e. to a reversible transformation. Crucially, SNFs are roughly a factor 2 more efficient than the standard NE-MCMC, as the same value of D_{KL} is reached for about half the value of n_{step} . We stress the fact that this improvement came at a very modest cost: the values of ρ were trained on values of n_{step} which are much smaller than the ones tested here. Thus, the computational cost of the training is much smaller than the one used for sampling.

In fig. 3 we show results for the performance of both for NE-MCMC and for SNF samplers for multiple volumes. As already observed in ref. [11], if we keep the KL divergence or the ESS

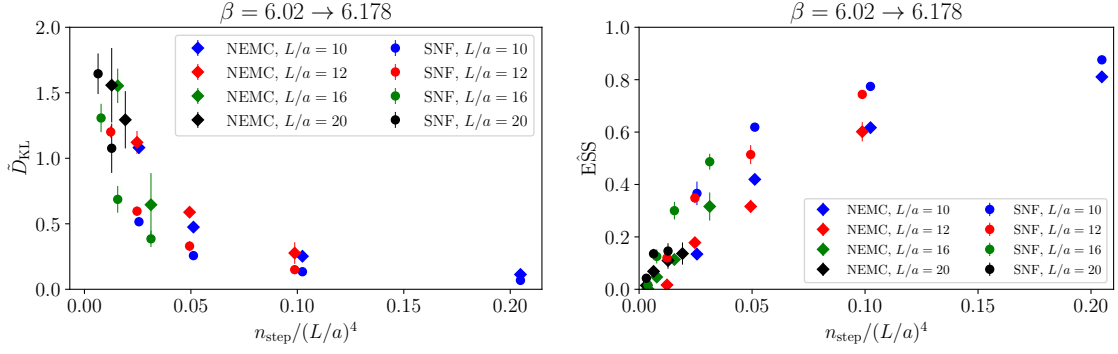


Figure 3: Results for the KL divergence (left panel) and for the ESS (right panel) for all the lattice sizes analysed in this study.

fixed the number of steps n_{step} for NE-MCMC scales with the degrees of freedom that are changed throughout a non-equilibrium evolution. In this case β is changed over all links, so the d.o.f. of interest go with the volume of the lattice. This is confirmed by the collapse of the points on essentially the same curve for all volumes of interest. The same scaling appears to hold also for trained SNFs, and this represents one of the main results of this work. All volumes present the same factor of improvement over the NE-MCMC, which is roughly 2 both for the D_{KL} and the ESS: this suggests that the scaling with the volume (or, more precisely, with the degrees of freedom of the system interested by the evolution) of SNFs is inherited fully from the NE-MCMC. In this view, the coupling layers simply enhance the non-equilibrium evolution making it more efficient for a limited cost in training.

5. Conclusions and future work

In this contribution we have presented the first implementation of Stochastic Normalizing Flows for SU(3) Yang-Mills theory simulations on the lattice. We tested this approach for a mapping between two probability distributions at different values of β for various volumes and we compared it with its purely stochastic counterpart, i.e. non-equilibrium Monte Carlo simulations.

We provided evidence for several promising features of this architecture: first, the training procedure is simple and computationally cheap. The parameters of the coupling layers, through an interpolation, can be rather effortlessly transferred to architectures with larger values of n_{step} ; furthermore, training length is constant for all lattice sizes and transfer of parameters to larger volumes is feasible with no retraining. Second, SNFs outperform NE-MCMC by roughly a factor 2 and, even more importantly, appear to inherit the same scaling with the volume of non-equilibrium simulations. Indeed, our results strongly support the idea that larger volumes can be efficiently sampled by scaling n_{step} with the number of degrees of freedom interested by the transformation: in this case $n_{\text{step}} \sim (L/a)^4$ for fixed ESS or D_{KL} .

The numerical evidence provided in this work represents a crucial step forward in the construction of a flow-based approach that can tackle critical slowing down in lattice gauge theories in four dimensions, with a particular focus towards an efficient mitigation of topological freezing.

However, the strategy followed in this work to change β over all the links of the lattice becomes very expensive at finer lattice spacings. Indeed, a realistic volume of $L \sim 1.4$ fm at $\beta > 6.4$ would require lattices with $L/a > 30$ and the corresponding n_{step} would be in the order of tens of thousands, depending on the starting prior distribution. However, the SNF presented in this work can be repurposed for a different approach, i.e. for out-of-equilibrium simulations that change the boundary condition of the system, see refs. [11, 21]: in this mapping n_{step} scales only with the size of the boundary changed along the evolution and not with the lattice volume, while exploiting open boundary conditions for the unfreezing of the topological charge.

We conclude by stressing the fact that the current SNF architecture can be systematically improved in several directions. For example, in line with the features of the link-level flow of ref. [18] the stout smearing parameter ρ can be essentially modelled as the output of a simple neural network: the increase in expressivity of the gauge-equivariant coupling layers and its effect on the performances of SNFs is a direction we intend to pursue in future works. Moreover, the linear protocol used in this work for both NE-MCMC and SNFs was chosen uniquely for its simplicity. However, in the last few years a large effort has been focused on optimal protocols in non-equilibrium simulations, see ref. [22] for a review: thus, the development of more efficient protocols in lattice field theory will be a key factor for cost-effective non-equilibrium simulations in the future.

Acknowledgments

We thank S. Bacchio, C. Bonanno, M. Caselle, L. Giusti, P. Kessel, S. Nakajima, K. Nicoli, M. Panero, D. Panfalone, A. Patella, S. Schaefer, A. Tomiya, D. VDACCHINO, L. Vaitl and L. Verzhicelli for insightful comments and discussions. The numerical simulations were run on machines of the Consorzio Interuniversitario per il Calcolo Automatico dell’Italia Nord Orientale (CINECA). A. Nada acknowledges support and A. Bulgarelli and E. Cellini acknowledge partial support by the Simons Foundation grant 994300 (Simons Collaboration on Confinement and QCD Strings). A. Nada acknowledges support from the European Union - Next Generation EU, Mission 4 Component 1, CUP D53D23002970006, under the Italian PRIN “Progetti di Ricerca di Rilevante Interesse Nazionale – Bando 2022” prot. 2022TJFCYB. All the authors acknowledge support from the SFT Scientific Initiative of INFN.

References

- [1] M. Luscher, *Trivializing maps, the Wilson flow and the HMC algorithm*, *Commun. Math. Phys.* **293** (2010) 899 [0907.5491].
- [2] D. Rezende and S. Mohamed, *Variational inference with normalizing flows*, in *International conference on machine learning*, pp. 1530–1538, PMLR, 2015.
- [3] K. Cranmer, G. Kanwar, S. Racanière, D. J. Rezende and P. E. Shanahan, *Advances in machine-learning-based sampling motivated by lattice quantum chromodynamics*, *Nature Rev. Phys.* **5** (2023) 526 [2309.01156].
- [4] ALPHA collaboration, S. Schaefer, R. Sommer and F. Virota, *Critical slowing down and error analysis in lattice QCD simulations*, *Nucl. Phys. B* **845** (2011) 93 [1009.5228].

- [5] R. Abbott et al., *Aspects of scaling and scalability for flow-based sampling of lattice QCD*, *Eur. Phys. J. A* **59** (2023) 257 [2211.07541].
- [6] D. Albandea, L. Del Debbio, P. Hernández, R. Kenway, J. Marsh Rossney and A. Ramos, *Learning trivializing flows*, *Eur. Phys. J. C* **83** (2023) 676 [2302.08408].
- [7] M. Caselle, G. Costagliola, A. Nada, M. Panero and A. Toniato, *Jarzynski's theorem for lattice gauge theory*, *Phys. Rev. D* **94** (2016) 034503.
- [8] M. Caselle, A. Nada and M. Panero, *QCD thermodynamics from lattice calculations with nonequilibrium methods: The $SU(3)$ equation of state*, *Phys. Rev. D* **98** (2018) 054513 [1801.03110].
- [9] O. Francesconi, M. Panero and D. Preti, *Strong coupling from non-equilibrium Monte Carlo simulations*, *JHEP* **07** (2020) 233 [2003.13734].
- [10] A. Bulgarelli and M. Panero, *Entanglement entropy from non-equilibrium Monte Carlo simulations*, *JHEP* **06** (2023) 030 [2304.03311].
- [11] C. Bonanno, A. Nada and D. Vadacchino, *Mitigating topological freezing using out-of-equilibrium simulations*, *JHEP* **04** (2024) 126 [2402.06561].
- [12] M. Caselle, E. Cellini, A. Nada and M. Panero, *Stochastic normalizing flows as non-equilibrium transformations*, *JHEP* **07** (2022) 015 [2201.08862].
- [13] H. Wu, J. Köhler and F. Noé, *Stochastic normalizing flows*, *Advances in Neural Information Processing Systems* **33** (2020) 5933.
- [14] M. Caselle, E. Cellini and A. Nada, *Numerical determination of the width and shape of the effective string using Stochastic Normalizing Flows*, 2409.15937.
- [15] G. E. Crooks, *Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences*, *Phys. Rev. E* **60** (1999) 2721.
- [16] C. Jarzynski, *Nonequilibrium equality for free energy differences*, *Phys. Rev. Lett.* **78** (1997) 2690.
- [17] A. Tomiya and Y. Nagai, *Gauge covariant neural network for 4 dimensional non-abelian gauge theory*, 2103.11965.
- [18] R. Abbott et al., *Normalizing flows for lattice gauge theory in arbitrary space-time dimension*, 2305.02402.
- [19] C. Morningstar and M. J. Peardon, *Analytic smearing of $SU(3)$ link variables in lattice QCD*, *Phys. Rev. D* **69** (2004) 054501 [hep-lat/0311018].
- [20] A. G. D. G. Matthews, M. Arbel, D. J. Rezende and A. Doucet, *Continual Repeated Annealed Flow Transport Monte Carlo*, in *International Conference on Machine Learning*, pp. 15196–15219, PMLR, 2022, 2201.13117.

- [21] C. Bonanno, A. Nada and D. VDACCHINO, *Topological susceptibility of $SU(3)$ pure-gauge theory from out-of-equilibrium simulations*, *PoS LATTICE2024* (2024) 415.
- [22] S. Blaber and D. A. Sivak, *Optimal control in stochastic thermodynamics*, *J. Phys. Commun.* **7** (2023) 033001 [[2212.00706](#)].