

## Stochastic normalizing flows for lattice field theory

---

**M. Caselle,<sup>a,b</sup> E. Cellini,<sup>a,b,\*</sup> A. Nada<sup>a</sup> and M. Panero<sup>a,b</sup>**

<sup>a</sup>*Department of Physics, University of Turin,  
Via Pietro Giuria 1, I-10125 Turin, Italy*

<sup>b</sup>*INFN, Turin,  
Via Pietro Giuria 1, I-10125 Turin, Italy*

*E-mail: [elia.cellini@unito.it](mailto:elia.cellini@unito.it)*

Stochastic normalizing flows are a class of deep generative models that combine normalizing flows with Monte Carlo updates and can be used in lattice field theory to sample from Boltzmann distributions. In this proceeding, we outline the construction of these hybrid algorithms, pointing out that the theoretical background can be related to Jarzynski's equality, a non-equilibrium statistical mechanics theorem that has been successfully used to compute free energy in lattice field theory. We conclude with examples of applications to the two-dimensional  $\phi^4$  field theory.

*The 39th International Symposium on Lattice Field Theory,  
8th-13th August, 2022,  
Rheinische Friedrich-Wilhelms-Universität Bonn, Bonn, Germany*

---

\*Speaker

## 1. Introduction

Following ref. [1], in this proceeding, we describe how to combine normalizing flows [2] with Monte Carlo updates in a new class of generative models called stochastic normalizing flows [3] which can be easily described in a framework inspired by non-equilibrium statistical mechanics. Normalizing flows are a class of deep generative models used to efficiently evaluate approximations of statistical distributions by mapping them to suitable distributions. In lattice field theory, normalizing flows can sample uncorrelated configurations from Boltzmann distributions [4] and a direct application is to use them to compute physically-interesting thermodynamic observables [5]. Thus, this class of models provides a new, promising route to studying quantum field theories on lattice [4–15]. Furthermore, stochastic normalizing flows are created by combining stochastic updates and normalizing flow layers and share the same theoretical background that underlies Monte Carlo simulations based on Jarzynski's equality [16], a method that found successful application in lattice field theory [17–19].

## 2. Stochastic normalizing flows

In this section, before introducing stochastic normalizing flows [3], we briefly summarize the building blocks of these hybrid algorithms: normalizing flows [2] and out-of-equilibrium Monte Carlo simulations based on Jarzynski's equality [16]. A normalizing flow is a function  $f : \mathbb{R}^d \rightarrow \mathbb{R}^d$ , that expresses a sequence of invertible and differentiable transformations interpolating between a prior distribution  $q_0(\phi_0)$ ,  $\phi_0 \in \mathbb{R}^d$  and the target distribution  $p(\phi) = \exp(-S[\phi])/Z$ ,  $\phi \in \mathbb{R}^d$ . Normalizing flows can be implemented using neural networks by composing  $N$  layers labeled by a natural number  $0 \leq n \leq N$ . Using the change of variable formula it is possible to compute the density of the generated samples:  $q(g(\phi_0)) = q_0(\phi_0) |\det J_g(\phi_0)|^{-1}$ , where  $J_g$  is the Jacobian matrix associated with the map  $g = f^{-1}$ . Therefore, for practical implementation,  $g$  must have a tractable determinant of the Jacobian. Normalizing flows can be trained so that the "learned" distribution  $q$  well approximates the target  $p$ . The training is done by minimizing the Kullback-Leibler divergence:  $D_{KL}(q||p) = \int d\phi [\ln q(\phi) - \ln p(\phi)]$  which is a measure of similarity between the  $q$  and  $p$ . After training, the partition function of the target  $p$  can be computed using a reweighting procedure, also called importance sampling in the machine learning field [5]:

$$\frac{Z}{Z_0} = \int d\phi q(\phi) \tilde{w}(\phi) = \langle \tilde{w}(\phi) \rangle_{\phi \sim q} \quad (1)$$

where we introduce the weight:

$$\tilde{w}(\phi_0) = \exp(-\{S[g(\phi_0)] - S_0[\phi_0] - Q_g\}) \quad (2)$$

and  $q_0(\phi_0) = \exp(-S_0[\phi_0])/Z_0$ . The form of  $Q_g = \sum_{n=0}^{N-1} \ln |\det J_n(\phi_n)|$  depends on the network architectures chosen to implement the layers  $g_n$ . Normalizing flow layers can be combined with Monte Carlo methods to obtain hybrids frameworks, following ref. [1], we compare normalizing flows with the theoretical background of Jarzynski's equality[16]: a non-equilibrium statistical mechanics theorem successfully applied for free energy calculations in Monte Carlo simulations of lattice gauge theories [17–19]. Jarzynski's equality states that equilibrium partition functions ratios

can be calculated as an exponential average over non-equilibrium processes of the dimensionless work  $w$  done on the system:

$$\frac{Z_{\eta_{fin}}}{Z_{\eta_{in}}} = \langle \exp(-w(\phi_0, \phi_1, \dots, \phi_N)) \rangle_f \quad (3)$$

where we introduced a protocol  $\eta(t)$  that drives out of equilibrium the states  $\eta_{in} = \eta(t_{in})$  to  $\eta_{fin} = \eta(t_{fin})$  and can be a set of couplings appearing in the action  $S$ . The average of the equation (3) is taken over all possible trajectories connecting, in the phase space,  $\eta_{in}$  to  $\eta_{fin}$ . Note that only the initial configurations must be in the thermodynamic equilibrium; the system is forced out of equilibrium and is never allowed to relax. These *stochastic evolutions*<sup>1</sup> can be implemented using Monte Carlo updates by discretizing the time  $t$  and computing the work as:

$$w(\phi_0, \phi_1, \dots, \phi_N) = S_{\eta_N}[\phi_N] - S_{\eta_0}[\phi_0] - Q_h(\phi_0, \phi_1, \dots, \phi_N) \quad (4)$$

where  $\phi$  are the degrees of freedom of the system and  $\eta_n = \eta(t_n)$  and  $Q_h$  is the heat exchanged with the environment during the processes. Given a discrete protocol  $\eta$ , it is possible to define a sequence of Boltzmann distributions, at the  $n$ -th step:  $\pi_{\eta_n}[\phi] = \exp(-S_{\eta_n}[\phi])/Z_{\eta_n}$ , the "prior"  $\phi_0$  is sampled from the distribution  $\pi_{\eta_0} = e^{-S_{\eta_0}[\phi_0]}/Z_{\eta_0}$  and the heat can be computed as:  $Q_h(\phi_0, \phi_1, \dots, \phi_N) = \sum_{n=0}^{N-1} \{S_{\eta_{n+1}}[\phi_{n+1}] - S_{\eta_{n+1}}[\phi_n]\}$ . Note that eq. (3) doesn't depend on protocol  $\eta(t)$ . However, there is a limited number of trajectories in the Monte Carlo implementation; hence, the protocol choices impact the overall efficiency of the method. Finally, stochastic normalizing flows can be constructed interleaving normalizing flow layers with stochastic evolution updates: in this hybrid framework partition functions can be calculated using Jarzynski's equality (3) with generalized work  $w(\phi_0, \phi_1, \dots, \phi_N) = S_{\eta_N}[\phi_N] - S_{\eta_0}[\phi_0] - Q(\phi_0, \phi_1, \dots, \phi_N)$  where the "heat"  $Q$  is computed as the sum of  $Q_g$  and  $Q_h$ . Moreover, generic observables  $O$  at  $\eta = \eta_{fin}$  can be computed using Jarzynski's equality:

$$\langle O \rangle_{\eta=\eta_{fin}} = \frac{\langle O(\phi_N) \exp(-w(\phi_0, \phi_1, \dots, \phi_N)) \rangle_f}{\langle \exp(-w(\phi_0, \phi_1, \dots, \phi_N)) \rangle_f} \quad (5)$$

### 3. Application to scalar field theory

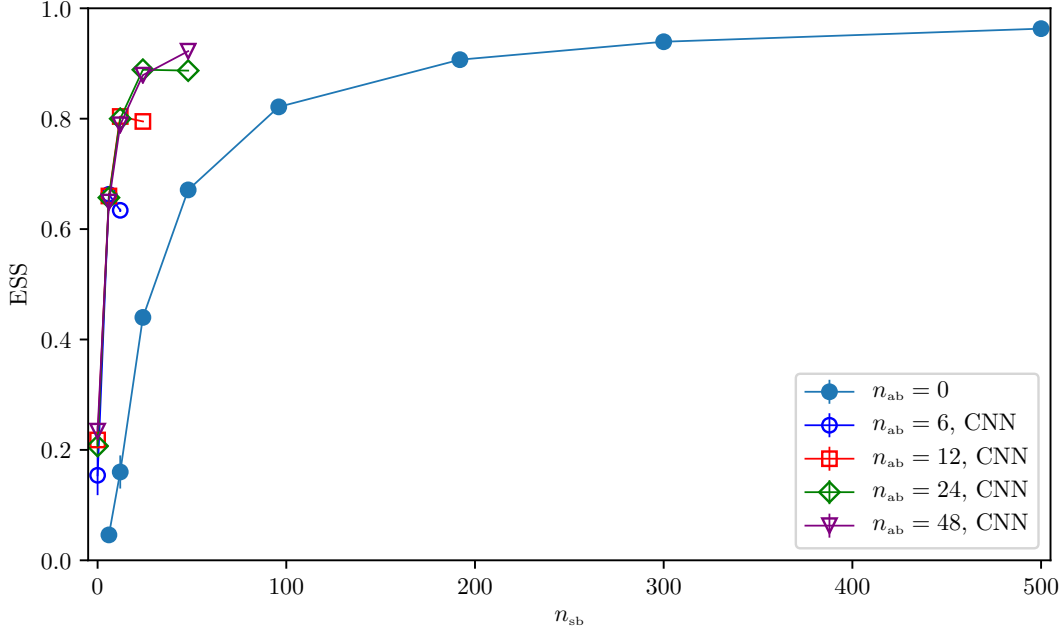
In this section, we show some results in the two-dimensional  $\phi^4$  interacting field theory, more extended studies can be found in [1]. The theory is regularized on a lattice  $\Lambda$  of size  $L_t \times L_s$ , with lattice spacing  $a$  and periodic boundary conditions along both dimensions. We defined  $N_t = L_t/a$  and  $N_s = L_s/a$  as the number of sites in the temporal and spatial directions. The Euclidean action is:

$$S(\phi) = \sum_{x \in \Lambda} -2\kappa \sum_{\mu=0,1} \phi(x)\phi(x+\hat{\mu}) + (1-2\lambda)\phi(x)^2 + \lambda\phi(x)^4. \quad (6)$$

For each algorithm: normalizing flows (NFs), stochastic evolutions (SEs) and stochastic normalizing flows (SNFs), we sample the prior from a normal distribution with  $\mu = 0$  and  $\sigma = 0.5$ , thus we recover (6) with  $\kappa = 0$  and  $\lambda = 0$ . This choice simplifies the protocol needed for SEs and SNFs. To compare the performances, we use the effective sample sizes (ESS):

$$ESS = \frac{\langle \tilde{w} \rangle_f^2}{\langle \tilde{w}^2 \rangle_f} \quad (7)$$

<sup>1</sup>An equivalent algorithm is the *annealed importance sampling* [20]

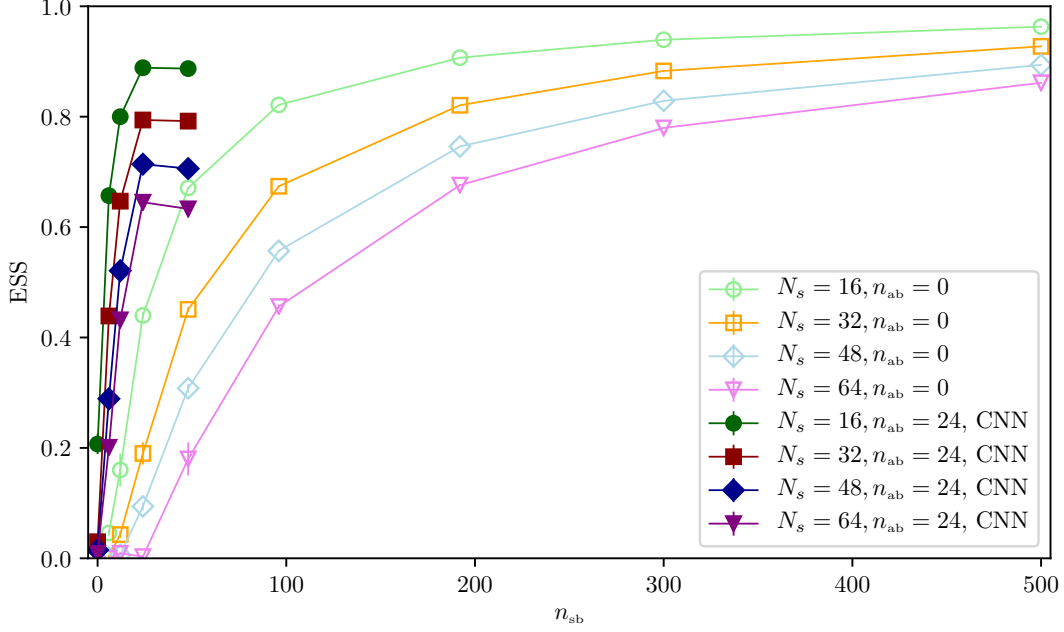


**Figure 1:** ESS as a function of  $n_{sb}$  for different flows.  $16 \times 8$  lattices at  $\kappa = 0.2$ ,  $\lambda = 0.022$ . The filled  $n_{ab} = 0$  point represents SEs, while the empty points represent SNFs. In the limit  $n_{sb} = 0$  and  $n_{ab} \neq 0$  NFs are recovered.

which is defined in the interval  $[0, 1]$  and is equal to 1 when the generated distribution  $q$  is equal to the target.

For normalizing flow layers, we implement the RealNVP [21] affine coupling layers and enforce  $\mathbb{Z}_2$  equivariance as [10]. The networks used are minimal convolutional neural networks with just one layer, two feature maps, and  $3 \times 3$  kernel. Since each coupling layer transforms only half of the lattice sites (we use a "checkerboard" even-odd, partitioning), we define one affine block as the sum of two affine layers and introduce  $n_{ab}$  as the number of deterministic, affine blocks. For SEs, we fix a linear protocol  $\eta$  interpolating between the initial and final action parameters, the  $n$ -th "layer" is defined by the protocol parameters  $\eta_n = \eta(t_n)$  that are used to update the configurations with the action  $S_{\eta_n}$ . We implemented local heatbath updates and defined the number of stochastic blocks as  $n_{sb}$ . SNFs are built by alternating affine blocks with stochastic updates. The models (NFs and SNFs) are trained by minimizing the loss function  $-\langle \ln \tilde{w} \rangle_f$  which is the Kullback-Leibler divergence minus  $\ln Z/Z_0$ . To update the model parameters, we use ADAM [22] with  $10^4$  steps, set the initial learning rate to 0.0005, and use ReduceLRonPlateau scheduler with patience of 500 steps. The software used is written using Pytorch [23]. All the numerical experiments have been performed on an NVIDIA Volta V100 GPU with 16 GB.

In fig. (1), we fixed the volume of the lattices to  $16 \times 8$  and set the couplings to  $\kappa = 0.2$  and  $\lambda = 0.022$ , which lie in the symmetric phase of the theory. In fig. (2) we fixed  $n_{ab} = 24$  and  $N_t = 8$  and we vary  $n_{sb}$ . Both fig. (1) and (2) show that SNFs can reach high ESS using a small number of  $n_{sb}$ ; this is highly relevant because the number of stochastic updates is drastically reduced. We also observed a performance peak when  $n_{sb} = n_{ab}$ , which shows that the layer produced by one

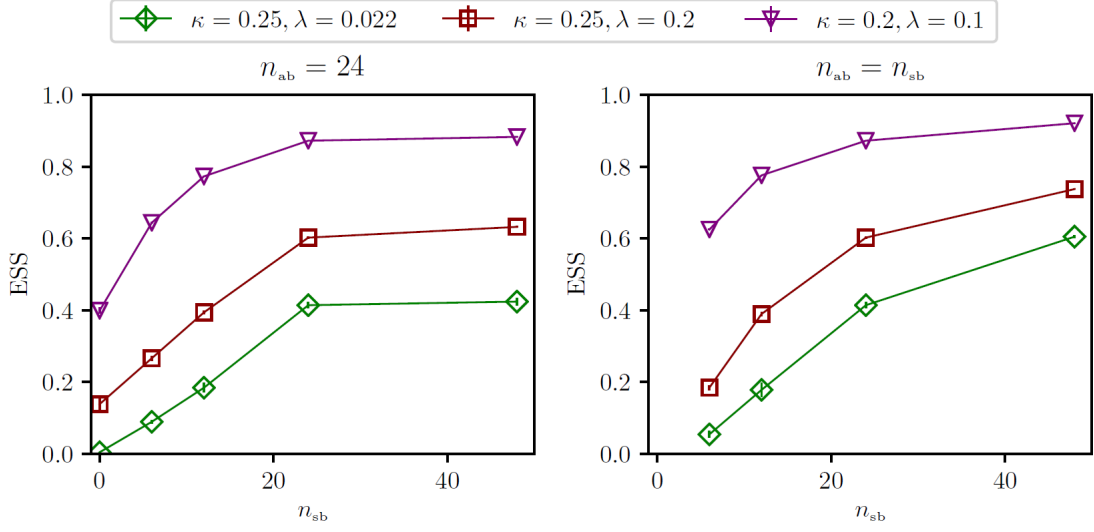


**Figure 2:** ESS as function of  $n_{sb}$  and  $N_s$  for SEs and SNFs with  $n_{ab} = 24$ . Lattices of size  $N_t = 8$  at  $\kappa = 0.2$ ,  $\lambda = 0.022$ .

deterministic and one stochastic block is the most expressive ingredient of SNFs. In fig. (3), we target different action parameters finding the same trends as the other study, namely the saturation of the ESS when  $n_{ab} > n_{sb}$ . We remark that all the tests are performed in the symmetric phase; a different type of SNFs may be required to reproduce our results in the unbroken symmetry phase. In fig. (4), we compare Gaussian and identity initializations for flow parameters; for all the measures in this proceeding, we initialized the parameters of flows with random Gaussian values. Furthermore, following ref. [24], we found that identity initialization for NF layers can speed training; this choice reduces the untrained SNFs to SEs, offering suitable starting points for training procedures.

#### 4. Conclusion

In this proceeding, based on [1], we have discussed the relationship between normalizing flows and non-equilibrium Monte Carlo simulations related to Jarzynski's equality (stochastic evolutions): both, providing deterministic or stochastic maps, give an effective way to compute partition functions and sample target distributions. Moreover, stochastic normalizing flows can be constructed by combining normalizing flows and stochastic evolutions; in this hybrid framework, the non-equilibrium protocol helps deterministic blocks to find suitable paths, while normalizing flows provide highly expressive maps. Since Monte Carlo updates are ergodic, stochasticity ameliorates mode-collapsing of normalizing flows. Despite the potential of these novel algorithms, the accuracy of the measures depends on a wide range of largely undiscovered possibilities; further research is required, and out-of-equilibrium thermodynamics can be exploited as a theoretical tool to drive the investigation of novel algorithms.



**Figure 3:** ESS as a function of  $n_{sb}$  SNFs with  $n_{sb} = 24$  (left panel) and  $n_{sb} = n_{ab}$  (right panel). Lattices of size  $16 \times 8$  and different values of the target parameters. To the best of our knowledge, the green target parameters should lie very close to the critical point.

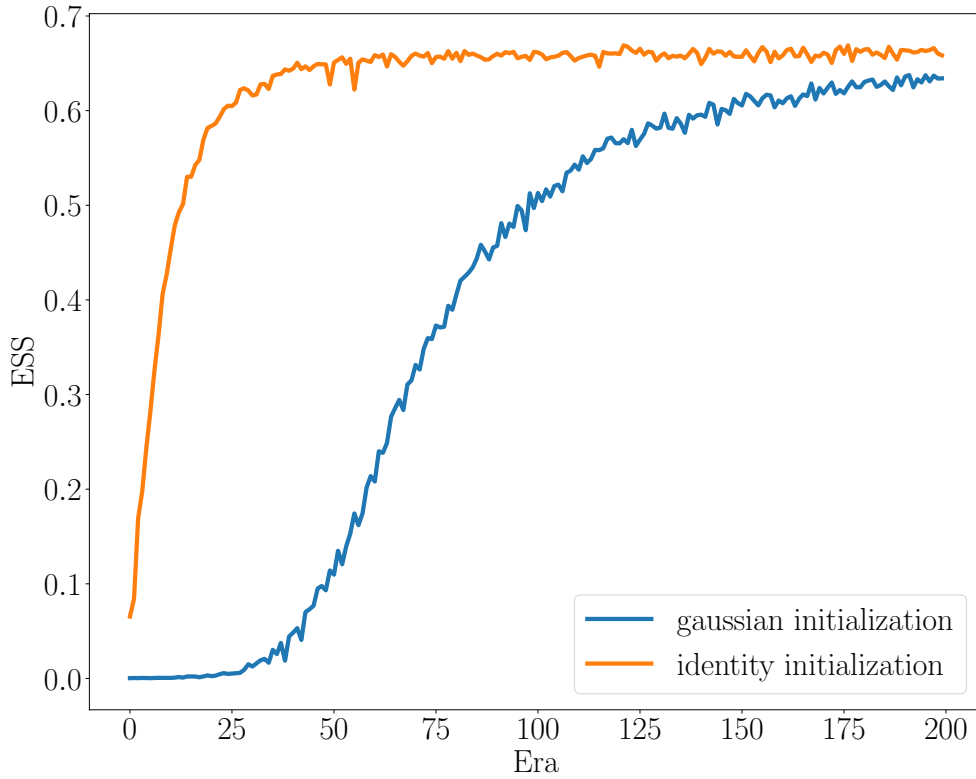
Stochastic evolution has been successfully applied to high-precision lattice gauge theory studies [17–19]. Hence, natural developments of our contribution will examine the extension of these works using stochastic normalizing flows. We showed it is possible to use stochasticity to improve simple normalizing flows. Therefore, other improvements of this work include the study of the stochastic counterparts of state-of-art flow-based samplers like continuous normalizing flows [14, 15] and the application to the lattice field theory of extensions of stochastic normalizing flows [24].

## Acknowledgments

The numerical simulations were run on machines of the Consorzio Interuniversitario per il Calcolo Automatico dell’Italia Nord Orientale (CINECA). We acknowledge support from the SFT Scientific Initiative of INFN. This work was partially supported by the Simons Foundation grant 994300 (Simons Collaboration on Confinement and QCD Strings). Part of the numerical functions used in the present work are based on ref. [25].

## References

- [1] M. Caselle, E. Cellini, A. Nada and M. Panero, *Stochastic normalizing flows as non-equilibrium transformations*, *JHEP* **07** (2022) 015 [2201.08862].
- [2] D. Rezende and S. Mohamed, *Variational inference with normalizing flows*, in *International conference on machine learning*, pp. 1530–1538, PMLR, 2015.
- [3] H. Wu, J. Köhler and F. Noé, *Stochastic normalizing flows*, *Advances in Neural Information Processing Systems* **33** (2020) 5933.



**Figure 4:** Learning curve of ESS for different SNFs initializations. One era corresponds to the average of 10 optimizer updates.  $16 \times 8$  lattices at  $\kappa = 0.2$ ,  $\lambda = 0.022$ .

- [4] M.S. Albergo, G. Kanwar and P.E. Shanahan, *Flow-based generative models for markov chain monte carlo in lattice field theory*, *Phys. Rev. D* **100** (2019) 034515.
- [5] K.A. Nicoli, C.J. Anders, L. Funcke, T. Hartung, K. Jansen, P. Kessel et al., *Estimation of thermodynamic observables in lattice field theories with deep generative models*, *Phys. Rev. Lett.* **126** (2021) 032001.
- [6] G. Kanwar, M.S. Albergo, D. Boyda, K. Cranmer, D.C. Hackett, S. Racaniere et al., *Equivariant flow-based sampling for lattice gauge theory*, *Physical Review Letters* **125** (2020) 121601.
- [7] D. Boyda, G. Kanwar, S. Racanière, D.J. Rezende, M.S. Albergo, K. Cranmer et al., *Sampling using  $SU(n)$  gauge equivariant flows*, *Phys. Rev. D* **103** (2021) 074504.
- [8] D.C. Hackett, C.-C. Hsieh, M.S. Albergo, D. Boyda, J.-W. Chen, K.-F. Chen et al., *Flow-based sampling for multimodal distributions in lattice field theory*, [2107.00734](#).

- [9] R. Abbott et al., *Gauge-equivariant flow models for sampling in lattice field theories with pseudofermions*, [2207.08945](#).
- [10] L. Del Debbio, J. Marsh Rossney and M. Wilson, *Efficient modeling of trivializing maps for lattice  $\phi^4$  theory using normalizing flows: A first look at scalability*, *Phys. Rev. D* **104** (2021) 094507.
- [11] J.M. Pawłowski and J.M. Urban, *Flow-based density of states for complex actions*, [2203.01243](#).
- [12] J.-L. Wynen, E. Berkowitz, S. Krieg, T. Luu and J. Ostmeyer, *Machine learning to alleviate Hubbard-model sign problems*, *Phys. Rev. B* **103** (2021) 125153 [[2006.11221](#)].
- [13] J. Finkenrath, *Tackling critical slowing down using global correction steps with equivariant flows: the case of the Schwinger model*, [2201.02216](#).
- [14] P. de Haan, C. Rainone, M.C.N. Cheng and R. Bondesan, *Scaling Up Machine Learning For Quantum Field Theory with Equivariant Continuous Flows*, [2110.02673](#).
- [15] M. Gerdes, P. de Haan, C. Rainone, R. Bondesan and M.C.N. Cheng, *Learning Lattice Quantum Field Theories with Equivariant Continuous Flows*, [2207.00283](#).
- [16] C. Jarzynski, *Nonequilibrium equality for free energy differences*, *Phys. Rev. Lett.* **78** (1997) 2690.
- [17] M. Caselle, G. Costagliola, A. Nada, M. Panero and A. Toniato, *Jarzynski's theorem for lattice gauge theory*, *Phys. Rev. D* **94** (2016) 034503.
- [18] O. Francesconi, M. Panero and D. Preti, *Strong coupling from non-equilibrium Monte Carlo simulations*, *JHEP* **07** (2020) 233 [[2003.13734](#)].
- [19] 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.
- [20] R.M. Neal, *Annealed importance sampling*, *Statistics and computing* **11** (2001) 125.
- [21] L. Dinh, J. Sohl-Dickstein and S. Bengio, *Density estimation using Real NVP*, [1605.08803](#).
- [22] D.P. Kingma and J. Ba, *Adam: A Method for Stochastic Optimization*, [1412.6980](#).
- [23] A. Paszke, S. Gross, F. Massa, A. Lerer, J. Bradbury, G. Chanan et al., *Pytorch: An imperative style, high-performance deep learning library*, in *Advances in Neural Information Processing Systems*, H. Wallach, H. Larochelle, A. Beygelzimer, F. d'Alché-Buc, E. Fox and R. Garnett, eds., vol. 32, Curran Associates, Inc., 2019.
- [24] A.G.D.G. Matthews, M. Arbel, D.J. Rezende and A. Doucet, *Continual Repeated Annealed Flow Transport Monte Carlo*, [2201.13117](#).
- [25] M.S. Albergo, D. Boyda, D.C. Hackett, G. Kanwar, K. Cranmer, S. Racanière et al., *Introduction to Normalizing Flows for Lattice Field Theory*, [2101.08176](#).