

Estimation of Thermodynamic Observables in Lattice Field Theories with Deep Generative Models

Kim A. Nicoli^{1,*}, Christopher J. Anders,¹ Lena Funcke², Tobias Hartung³, Karl Jansen,⁴ Pan Kessel,^{1,†} Shinichi Nakajima^{1,5} and Paolo Stornati^{4,6}

¹Machine Learning Group, Technische Universität Berlin, Marchstr. 23 10587 Berlin, Germany

²Perimeter Institute for Theoretical Physics, 31 Caroline St N, Waterloo, Ontario N2L 2Y5, Canada

³Department of Mathematics, Kings College London, 80 Kennington Rd, Bishop's, London SE11 6NJ, United Kingdom

⁴NIC, DESY, Zeuthen, Platanenalle 6, 15738 Zeuthen, Germany

⁵RIKEN Center for AIP, 103-0027 Tokyo, Chuo City, Japan

⁶Institut für Physik, Humboldt-Universität, Newtonstraße 15, 12489 Berlin, Germany



(Received 17 July 2020; revised 14 October 2020; accepted 14 December 2020; published 19 January 2021; corrected 21 January 2021)

In this Letter, we demonstrate that applying deep generative machine learning models for lattice field theory is a promising route for solving problems where Markov chain Monte Carlo (MCMC) methods are problematic. More specifically, we show that generative models can be used to estimate the absolute value of the free energy, which is in contrast to existing MCMC-based methods, which are limited to only estimate free energy differences. We demonstrate the effectiveness of the proposed method for two-dimensional ϕ^4 theory and compare it to MCMC-based methods in detailed numerical experiments.

DOI: [10.1103/PhysRevLett.126.032001](https://doi.org/10.1103/PhysRevLett.126.032001)

Introduction.—The free energy of a physical system is of great importance since it can be related to several thermodynamical observables. In particular, at nonzero temperature, it allows one to compute the entropy, the pressure, or, more generally, the equation of state of the considered physical system. For example, QCD at high temperature—as a generic strongly interacting field theory—plays an essential role in the physics of the early Universe and is now extensively probed in large-scale heavy ion experiments [1]. Hence, knowing such thermodynamic quantities from QCD alone is of very high relevance.

The main tool to study strongly coupled field theories, such as QCD, is to discretize them on a spacetime lattice and use Monte Carlo Markov chain (MCMC) methods to numerically calculate the relevant physical quantities. Unfortunately, these thermodynamical quantities are challenging to compute using existing MCMC methods. The fundamental difficulty is that the MCMC is not able to directly estimate the partition function of the lattice field theory. Therefore, the absolute value of the free energy cannot be estimated straightforwardly.

Instead, there are a number of MCMC methods to estimate differences of free energies. One typically chooses a free energy difference $\Delta F = F_b - F_a$ such that F_a is

known either exactly or approximately. One can then deduce the value of the free energy $F_b = \Delta F + F_a$ at the desired point in parameter space. If the free energy F_a is not known exactly, this induces an unwanted approximation error. Most of the methods to estimate ΔF rely on integrating a derivative of the partition function over a trajectory in the parameter space of the lattice field theory [2]. Alternatively, one can use a reweighting procedure to calculate free energy differences between neighboring points of the discretized trajectory and then sum them up [2,3]. These approaches require simulations at each parameter point of the discretized trajectory, which is numerically costly and leads to accumulation of errors. This effect is often the dominant contribution to the error—especially if the trajectory passes a phase transition. Such situations arise, for example, in the context of studying the deconfined phase of SU(3) Yang-Mills theory [4,5]. We stress that the accumulation of the statistical error along the trajectory and the approximation error of its starting point are not independent. The former could be reduced if a better starting point was available. There are also nonequilibrium methods based on Jarzynski's identity to estimate free energy differences without the need for integration [5–7]. However, also these methods require expensive repeated simulations corresponding to an ensemble of nonequilibrium trajectories through phase space. It is therefore desirable to develop methods that allow the direct estimation of the free energy at a given point in parameter space.

In the following, we will propose such a method based on deep generative machine learning models. As we will discuss, our method comes with rigorous error estimators

Published by the American Physical Society under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/). Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI. Funded by SCOAP³.

and asymptotic guarantees. Over the last years, deep generative models have been applied with great success to generate, for example, high-resolution images, natural speech, and text (see [8] for an overview). In [9], a machine-learning-based regression algorithm for determining action parameters from an ensemble of field configurations is proposed and [10] uses a neural network to predict the structure of phase transitions from field configurations. References [11–15] conjecture a relation between restricted Boltzmann machines and quantum fields in the context of the holographic duality. In recent works [16–18], deep generative models have also been used in the context of lattice quantum field theories (see also [19,20]). The main objective of these works was to reduce the integrated autocorrelation of the simulations. In contrast, this Letter demonstrates that deep generative models can be used to estimate quantities that are not (directly) obtainable by MCMC approaches.

We also note that generative models have been used in [21] to estimate free energy differences in the context of statistical mechanics by combining these models with the Zwanzig free energy perturbation method [22]. Contrary to this approach, our method estimates the absolute value of the free energy. We furthermore note that the free energy can also be directly computed using the tensor renormalization group method; see [23] for an application to ϕ^4 theory. For other novel approaches to obtain thermodynamic quantities and, in particular, the equation of state, see [24,25].

In the following, we will give a brief overview of relevant aspects of lattice field theories and generative models. We will then discuss how generative models can be used to estimate the free energy and compare this approach to MCMC-based methods in numerical experiments.

Lattice field theory.—A lattice field theory can be described by an action $S(\phi)$. In the following, we will consider (Euclidean) real scalar field theory for concreteness, i.e., $\phi(x) \in \mathbb{R}$ for each lattice site $x \in \Lambda$ of the lattice Λ . The path integral then reduces to an ordinary high-dimensional integral. Therefore, expectation values of operators $\mathcal{O}(\phi)$ can be calculated by

$$\langle \mathcal{O} \rangle = \frac{1}{Z} \int \mathcal{D}[\phi] \mathcal{O}(\phi) \exp[-S(\phi)],$$

where we defined $\mathcal{D}[\phi] = \prod_{x \in \Lambda} d[\phi(x)]$ and the partition function Z is given by

$$Z = \int \mathcal{D}[\phi] \exp[-S(\phi)].$$

If we impose periodic boundary conditions in time for a lattice with temporal extent N_T , the theory is at finite temperature $T = (1/\beta) = (1/N_T a)$, where a denotes the lattice spacing. The free energy is then defined by

$$F = -T \ln(Z) \quad (1)$$

and can be related to the pressure $p = -(F/V)$, where V denotes the spatial volume of the lattice Λ whose number of lattice sites we denote by $|\Lambda|$. Similarly, the entropy H can be obtained from the free energy by $F = U - TH$, where U is the internal energy.

Deep generative models.—We focus on a particular subclass of generative models called normalizing flows (see [26] for a recent review). These flows are distributions q_θ with learnable parameters θ . They also have the appealing property that they allow for efficient sampling and calculation of the probability of the samples.

In more detail, these flows are constructed by defining an invertible neural network g_θ . For a brief overview of neural networks, we refer the reader to the Supplemental Material [27]. The samples $\phi \in \mathbb{R}^{|\Lambda|}$ are obtained by applying this network to samples $z \in \mathbb{R}^{|\Lambda|}$ drawn from a simple prior distribution q_Z such as a standard normal $\mathcal{N}(0, 1)$,

$$\phi = g_\theta(z), \quad z \sim q_Z. \quad (2)$$

Since the network g_θ is invertible by assumption, it then follows by the change of variable theorem that $\phi \sim q_\theta$ with

$$q_\theta(\phi) = q_Z[g_\theta^{-1}(\phi)] \left| \frac{dg_\theta}{dz} \right|^{-1}. \quad (3)$$

The architecture of the neural network g_θ is chosen such that (i) invertibility of g_θ and (ii) efficient evaluation of the Jacobian determinant $|dg_\theta/dz|$ are ensured. A particular example of such an architecture is nonlinear independent component estimation (NICE) [28], for which the neural network g_θ consists of invertible coupling layers $y^l: \mathbb{R}^{|\Lambda|} \rightarrow \mathbb{R}^{|\Lambda|}$, i.e.,

$$g_\theta(z) = (y^L \circ y^{L-1} \circ \dots \circ y^1)(z). \quad (4)$$

Invertibility and efficient evaluation of Jacobian determinant is then ensured by splitting the components of the layer $y^l = (y_u^l, y_d^l)$ in two parts $y_u^l \in \mathbb{R}^{|\Lambda|-k}$ and $y_d^l \in \mathbb{R}^k$ for given $k \in \{1, |\Lambda| - 1\}$. The layer $y^{l+1} = (y_u^{l+1}, y_d^{l+1})$ is then recursively defined by

$$\begin{aligned} y_u^{l+1} &= y_u^l, \\ y_d^{l+1} &= y_d^l + m(y_u^l), \end{aligned} \quad (5)$$

where m is another neural network (not necessarily satisfying the two requirements from above). Because of the splitting, this can be easily inverted by

$$\begin{aligned} y_u^l &= y_u^{l+1}, \\ y_d^l &= y_d^{l+1} - m(y_u^{l+1}), \end{aligned}$$

and the determinant of the Jacobian is given by

$$\det \frac{\partial y^{l+1}}{\partial y^l} = \begin{vmatrix} \frac{\partial y_u^{l+1}}{\partial y_u^l} & \frac{\partial y_d^{l+1}}{\partial y_d^l} \\ \frac{\partial y_d^{l+1}}{\partial y_u^l} & \frac{\partial y_u^{l+1}}{\partial y_d^l} \end{vmatrix} = \begin{vmatrix} \mathbb{I} & 0 \\ * & \mathbb{I} \end{vmatrix} = 1.$$

The total Jacobian determinant is then $|dg_\theta/dz| = 1$ since it is the product of the Jacobian determinant of each layer.

Training.—We want to train a generative model that samples field configurations $\phi \sim q_\theta$ approximately from the path-integral distribution

$$p(\phi) = \frac{1}{Z} \exp[-S(\phi)]. \quad (6)$$

For this, the Kullback-Leibler (KL) divergence [29] between the normalizing flow q_θ and the target distribution p is minimized, i.e.,

$$\text{KL}(q_\theta||p) = \int \mathcal{D}[\phi] q_\theta(\phi) \ln \left(\frac{q_\theta(\phi)}{p(\phi)} \right) = \beta(F_q - F),$$

where we have defined the variational free energy

$$\beta F_q = \mathbb{E}_{\phi \sim q_\theta} [S(\phi) + \ln q_\theta(\phi)], \quad (7)$$

as well as the expectation value $\mathbb{E}_{\phi \sim q}[\mathcal{O}] = \int \mathcal{D}[\phi] q_\theta(\phi) \mathcal{O}(\phi)$ and the free energy $F = -(1/\beta) \ln(Z)$. This divergence vanishes if and only if the distributions q and p are identical [30].

The KL divergence is minimized by gradient descent with respect to the parameter θ of the flow q_θ . Since the free energy F does not depend on the flow q , the variational free energy F_q can equivalently be minimized. Therefore, the training procedure does not require a target distribution (6) with a tractable partition function Z . Using the explicit expression for the probability of the flow (3), we can rewrite the variational free energy as

$$\beta F_q = \mathbb{E}_{z \sim q_z} \left[S[g_\theta(z)] - \ln \left| \frac{dg_\theta}{dz} \right| (z) + \ln q_z(z) \right].$$

In training, the expectation value is approximated by its Monte Carlo estimate. In machine learning, this approach of learning a model from an unnormalized target distribution is very well established [31–34]. Recently, the same method has been used in the context of lattice field theories [16]. Furthermore, this approach has been applied to quantum chemistry [35] and statistical physics [36–38].

The variational free energy does not allow us to infer the value of the KL divergence since the free energy F is not known. In order to alleviate this shortcoming, we define the random variable $C(\phi) = S(\phi) + \ln q_\theta(\phi)$, which is related to the variational free energy by $\beta F_q = \langle C \rangle_{q_\theta}$. In the Supplemental Material [27], we show that

$$\text{KL}(q_\theta||p) = \frac{1}{2} \text{Var}_{q_\theta}(C) + \mathcal{O}(\mathbb{E}_{q_\theta}[|w-1|^3]),$$

where we have defined the importance weight $w(\phi) = [p(\phi)/q_\theta(\phi)]$. Thus convergence of training will result in a small variance $\text{Var}_{q_\theta}(C)$. In practice, a Monte Carlo estimate of this quantity can be calculated without any significant overhead during training as $C(\phi)$ is also needed for Monte Carlo estimation of the variational free energy F_q , see (7). It is therefore advisable to closely monitor the variance of C during training.

Estimation of thermodynamical observables.—The partition function Z can be rewritten as

$$Z = \int \mathcal{D}[\phi] q_\theta(\phi) \tilde{w}(\phi), \quad (8)$$

where we have defined the unnormalized importance weight $\tilde{w}(\phi) = \{(\exp[-S(\phi)])/(q_\theta(\phi))\}$. Therefore, the partition function can be estimated by Monte Carlo sampling as follows:

$$\hat{Z} = \frac{1}{N} \sum_{i=1}^N \tilde{w}(\phi_i) \quad \text{with} \quad \phi_i \sim q_\theta. \quad (9)$$

We emphasize that the sampling procedure does not need to be sequential (as for a Markov chain). As a result, it can very efficiently be parallelized and does not suffer from autocorrelation. From \hat{Z} , one can then easily estimate the free energy by

$$\hat{F} = -T \ln \hat{Z}. \quad (10)$$

From the free energy (10), one can then straightforwardly obtain estimates for the pressure and entropy, as explained above. The estimator (10) has been extensively studied in the context of training an importance weighted variational autoencoder [39–41]. It was shown in [39] that it is a statistically consistent estimator if q has support larger or equal to the target p . In [40], its variance and bias were derived using the delta method (see also [37,41]). For convenience, we summarize the relevant results in the Supplemental Material [27]. Alternatively, one can use the jackknife method to estimate the bias and variance [42].

Numerical experiments.—We apply the proposed method to two-dimensional real scalar field theory with action

$$S = \sum_{x \in \Lambda} -2\kappa \sum_{\hat{\mu}=1}^2 \varphi(x) \varphi(x + \hat{\mu}) + (1 - 2\lambda) \varphi(x)^2 + \lambda \varphi(x)^4,$$

where κ is the hopping parameter and λ denotes the bare coupling constant of the theory. The action is invariant under \mathbb{Z}_2 transformations, i.e., $\phi \rightarrow -\phi$. Figure 1 shows the

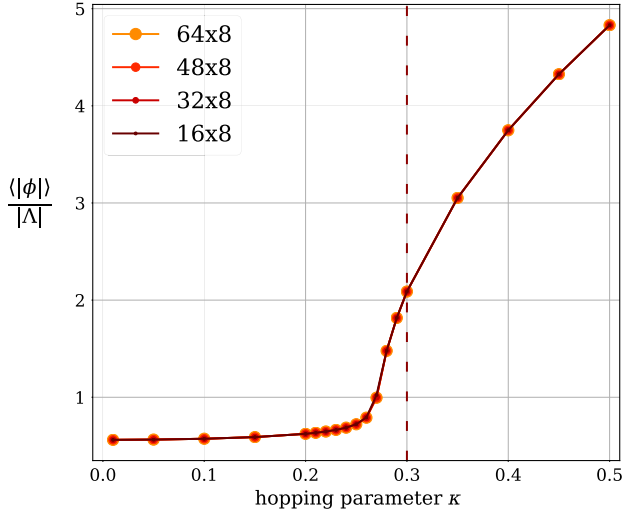


FIG. 1. Absolute magnetization density as a function of hopping parameter κ for bare coupling $\lambda = 0.022$. Results for various lattice sizes overlap. The values were estimated with an overrelaxed Hamiltonian Monte Carlo (HMC) [43–46]. The dashed line denotes the hopping parameter value $\kappa = 0.3$ for the free energy estimation in the numerical experiments.

absolute magnetization $\langle |\phi| \rangle$ as a function of the hopping parameter κ . As the hopping parameter κ increases, spontaneous magnetization is observed.

In the following, we will estimate the free energy F_e at $\lambda_e = 0.022$ and $\kappa_e = 0.3$ for lattice sizes $|\Lambda| = N_L \times N_T$ of 64×8 , 48×8 , 32×8 , and 16×8 with both the flow- and MCMC-based methods.

Using the flow method, we can directly estimate these free energies. We modify the NICE architecture to ensure that the flow q_θ is invariant under \mathbb{Z}_2 transformations, i.e., $q_\theta(\phi) = q_\theta(-\phi)$. By the definition (3) of q_θ , an odd function $g_\theta(-z) = -g_\theta(z)$ implies \mathbb{Z}_2 invariance of q_θ . The map g_θ is odd if all its coupling blocks y^l are odd, see (4). The latter condition can be ensured by choosing an odd neural network m for the coupling (5), which we achieve by using tanh nonlinearities and vanishing biases for the network m .

After training has completed, the free energy is then computed using the proposed estimator (10). For error analysis, we use both the jackknife as well as the delta method and check that they lead to consistent error estimates. In many applications, generative models suffer from “mode dropping” [47]; i.e., some modes of the target p are not captured by the model q_θ . For our specific estimation method, however, a simple consistency check can be performed ensuring that mode dropping does not occur. To this end, we estimate $Z = (\mathbb{E}_p[\tilde{w}^{-1}])^{-1}$ by a single Markov chain at the target point in parameter space and ensure that this leads to a compatible estimate, see the Supplemental Material [27].

For MCMC methods, we use a reweighting procedure [2,3], which is significantly more involved and uses the

relation $F_e = \Delta F_{eb} + F_b$. Here, F_b is the free energy at $\kappa_b = 0$ and $\lambda_b = \lambda_e$. The value of F_b can be analytically calculated, since for vanishing Hopping parameter κ ,

$$F(\lambda) = -|\Lambda|T \ln z(\lambda),$$

where $|\Lambda|$ denotes the number of sites of the lattice Λ and

$$z(\lambda) = \sqrt{\frac{1-2\lambda}{4\lambda}} \exp\left(\frac{(1-2\lambda)^2}{8\lambda}\right) K_{\frac{1}{2}}\left(\frac{(1-2\lambda)^2}{8\lambda}\right),$$

with K_n being the Bessel function of the second kind. We prove this relation in the Supplemental Material [27]. The free energy difference $\Delta F_{eb} = F_e - F_b = -T \ln(Z_e/Z_b)$ can be obtained by

$$\mathbb{E}_{p_b} \left[\frac{\exp(-S_e)}{\exp(-S_b)} \right] = \frac{1}{Z_b} \int \mathcal{D}[\phi] e^{-S_b(\phi)} \frac{e^{-S_e(\phi)}}{e^{-S_b(\phi)}} = \frac{Z_e}{Z_b}.$$

We estimate this expectation value with an overrelaxed HMC algorithm [43–46]. In practice, the variance of the estimator will become prohibitively large if the two distributions p_b and p_e do not have sufficient overlap. We therefore choose intermediate distributions p_{i_1}, \dots, p_{i_K} , ensuring that neighboring distributions p_{i_k} and $p_{i_{k+1}}$ have sufficient overlap. The free energy difference can then be obtained by

$$\Delta F_{eb} = \Delta F_{e,i_K} + \Delta F_{i_K i_{K-1}} + \dots + \Delta F_{i_1 b}.$$

In our numerical experiments, we keep $\lambda = 0.022$ fixed and only vary the hopping parameter κ of the intermediate distributions p_i . We choose a difference in hopping parameter of $\delta\kappa = 0.01$ for $\kappa \in [0.2, 0.3]$ and $\delta\kappa = 0.05$ for all other intermediate hopping parameters κ . We therefore use $K = 14$ Markov chains with 400 000 steps each. Thus, a total number of 5.6×10^6 configurations is used for estimation. For a detailed analysis of the dependence of our results on this choice of $\delta\kappa$, we refer the reader to the Supplemental Material [27].

The error analysis is performed with both the uwerr [48] and jackknife methods, which are checked to lead to consistent estimates. We again refer the reader to the Supplemental Material [27] for a more detailed description.

Figure 2 shows that the estimates of both the flow and MCMC methods are compatible within error bars. However, the trajectory of the MCMC method has to pass the critical region, which is challenging due to critical slowing down. The flow-based estimate can be directly performed at the desired point in parameter space and therefore does not suffer from this problem. This conceptual difference leads to a significantly more precise estimate by the flow-based method. For regions in parameter space that do not require the crossing of a phase transition, MCMC-based methods have errors of comparable order of magnitude

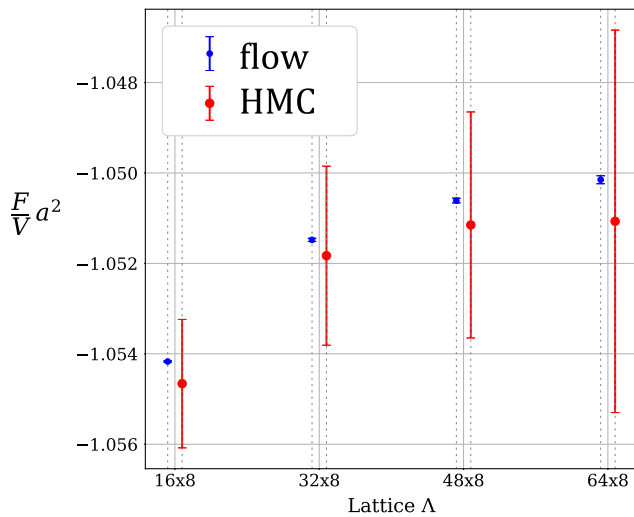


FIG. 2. Estimate of free energy density at $\lambda = 0.022$ and $\kappa = 0.3$ obtained by both the flow- and MCMC-based methods for various lattice sizes. MCMC estimates are obtained from integrating free energy differences. Both methods use the same number of samples (5.6 M) for estimation. Errors are obtained with the delta and uwerr method [48] for flow and HMC, respectively (see the Supplemental Material [27] for jackknife error analysis and Appendix of [48] for theory of the jackknife method).

(see Supplemental Material [27]). The ability of the flow to perform direct estimates is both of practical as well as of conceptual importance. For example, in finite-temperature QCD, one often uses a trajectory whose initial free energy is approximated by the hadron resonance model (see, for example, [2]), leading to an undesirable systematic error. Furthermore, summing up the free energy differences along the trajectory leads to an accumulation of errors. This effect is often the dominant contribution to the error and is particularly pronounced in situations for which the trajectory has to cross a phase transition. Such situations are of great practical relevance, for example, in the deconfined phase of SU(3) gauge theory [4,5]. We stress that both error sources are related since the initial free energy is the starting point of the trajectory.

Conclusion.—In this Letter, we have proposed a method to directly estimate the free energy and hence thermodynamical observables of lattice field theories using deep generative models. This method is of great conceptual appeal as it avoids cumbersome integration through parameter space and does not require an exactly or approximately known integration constant. Future work will focus on scaling this approach to four-dimensional gauge theories. Recent work has successfully constructed flows that are manifestly gauge invariant [17,18]. This recent progress, combined with the enormous ongoing advances in deep learning, makes it very promising that our method can be applied to non-Abelian gauge theories, and ultimately QCD, in the not too distant future.

This work was supported in part by the German Ministry for Education and Research (BMBF) under Grants No. 01IS14013A-E, No. 01GQ1115, No. 01GQ0850, No. 01IS18025A, and No. 01IS18037A. This work is also supported by the Fraunhofer Heinrich-Hertz-Institut (HHI) and by the grant funded by the DFG (EXC 2046/1, Project-ID 390685689). P. S. thanks the Helmholtz Einstein International Berlin Research School in Data Science (HEIBRiDS) for funding. Research at Perimeter Institute is supported by the Government of Canada through the Department of Innovation, Science and Economic Development Canada and by the Province of Ontario through the Ministry of Economic Development, Job Creation and Trade. We want to express our gratitude for valuable feedback by Klaus-Robert Müller, Wojciech Samek, Nils Strodthoff, Alessandro Nada, and Stefan Kühn.

*kim.a.nicoli@tu-berlin.de

†pan.kessel@gmail.com

- [1] W. Busza, K. Rajagopal, and W. Van Der Schee, Heavy ion collisions: The big picture and the big questions, *Annu. Rev. Nucl. Part. Sci.* **68**, 339 (2018).
- [2] O. Philipsen, The QCD equation of state from the lattice, *Prog. Part. Nucl. Phys.* **70**, 55 (2013).
- [3] P. de Forcrand, M. D’Elia, and M. Pepe, ‘t Hooft Loop in SU(2) Yang-Mills Theory, *Phys. Rev. Lett.* **86**, 1438 (2001).
- [4] L. Giusti and M. Pepe, Equation of state of the SU(3) Yang-Mills theory: A precise determination from a moving frame, *Phys. Lett. B* **769**, 385 (2017).
- [5] 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**, 054513 (2018).
- [6] C. Jarzynski, Equilibrium free-energy differences from nonequilibrium measurements: A master-equation approach, *Phys. Rev. E* **56**, 5018 (1997).
- [7] C. Jarzynski, Nonequilibrium Equality for Free Energy Differences, *Phys. Rev. Lett.* **78**, 2690 (1997).
- [8] I. Goodfellow, Y. Bengio, and A. Courville, *Deep Learning* (MIT Press, Cambridge, MA, 2016).
- [9] P. E. Shanahan, D. Trewartha, and W. Detmold, Machine learning action parameters in lattice quantum chromodynamics, *Phys. Rev. D* **97**, 094506 (2018).
- [10] D. Bachtis, G. Aarts, and B. Lucini, A mapping of distinct phase transitions to a neural network, *Phys. Rev. E* **102**, 053306 (2020).
- [11] J.-W. Lee, Quantum fields as deep learning, *J. Korean Phys. Soc.* **76**, 684 (2020).
- [12] K. Hashimoto, S. Sugishita, A. Tanaka, and A. Tomiya, Deep learning and the AdS/CFT correspondence, *Phys. Rev. D* **98**, 046019 (2018).
- [13] K. Hashimoto, S. Sugishita, A. Tanaka, and A. Tomiya, Deep learning and holographic QCD, *Phys. Rev. D* **98**, 106014 (2018).
- [14] K. Hashimoto, AdS/CFT correspondence as a deep Boltzmann machine, *Phys. Rev. D* **99**, 106017 (2019).

- [15] T. Akutagawa, K. Hashimoto, and T. Sumimoto, Deep learning and AdS/QCD, *Phys. Rev. D* **102**, 026020 (2020).
- [16] 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**, 034515 (2019).
- [17] G. Kanwar, M. S. Albergo, D. Boyda, K. Cranmer, D. C. Hackett, S. Racanière, D. J. Rezende, and P. E. Shanahan, Equivariant Flow-Based Sampling for Lattice Gauge Theory, *Phys. Rev. Lett.* **125**, 121601 (2020).
- [18] D. Boyda, G. Kanwar, S. Racanière, D. J. Rezende, M. S. Albergo, K. Cranmer, D. C. Hackett, and P. E. Shanahan, Sampling using $SU(n)$ gauge equivariant flows, [arXiv:2008.05456](https://arxiv.org/abs/2008.05456).
- [19] J. M. Pawłowski and J. M. Urban, Reducing autocorrelation times in lattice simulations with generative adversarial networks, *Mach. Learn. Sci. Technol.* **1**, 045011 (2020).
- [20] A. Tanaka and A. Tomiya, Towards reduction of autocorrelation in HMC by machine learning, [arXiv:1712.03893](https://arxiv.org/abs/1712.03893).
- [21] P. Wirnsberger, A. J. Ballard, G. Papamakarios, S. Abercrombie, S. Racanière, A. Pritzel, D. J. Rezende, and C. Blundell, Targeted free energy estimation via learned mappings, [arXiv:2002.04913](https://arxiv.org/abs/2002.04913).
- [22] R. W. Zwanzig, High-temperature equation of state by a perturbation method. II. Polar gases, *J. Chem. Phys.* **23**, 1915 (1955).
- [23] S. Akiyama, D. Kadoh, Y. Kuramashi, T. Yamashita, and Y. Yoshimura, Tensor renormalization group approach to four-dimensional complex ϕ^4 theory at finite density, *J. High Energy Phys.* **09** (2020) 177.
- [24] M. Asakawa, T. Hatsuda, E. Itou, M. Kitazawa, H. Suzuki (F. Collaboration), Thermodynamics of $SU(3)$ gauge theory from gradient flow on the lattice, *Phys. Rev. D* **90**, 011501 (R) (2014).
- [25] L. Giusti and M. Pepe, Equation of State of a Relativistic theory from a Moving Frame, *Phys. Rev. Lett.* **113**, 031601 (2014).
- [26] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, Normalizing flows for probabilistic modeling and inference, [arXiv:1912.02762](https://arxiv.org/abs/1912.02762).
- [27] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/PhysRevLett.126.032001> for detailed proofs of theorems, detailed calculations, further insights into the experiments and a deeper overview on neural networks as well as the type of algorithm we used.
- [28] L. Dinh, D. Krueger, and Y. Bengio, Nice: Non-linear independent components estimation, [arXiv:1410.8516](https://arxiv.org/abs/1410.8516).
- [29] David J. C. Mac Kay, *Information Theory, Inference and Learning Algorithms* (Cambridge University Press, Cambridge, England, 2003).
- [30] More precisely, the distributions p and q have to be identically only almost everywhere, i.e., up to a set of measure zero.
- [31] D. P. Kingma and M. Welling, Auto-encoding variational Bayes, [arXiv:1312.6114](https://arxiv.org/abs/1312.6114).
- [32] D. J. Rezende, S. Mohamed, and D. Wierstra, Stochastic backpropagation and approximate inference in deep generative models, [arXiv:1401.4082](https://arxiv.org/abs/1401.4082).
- [33] D. J. Rezende and S. Mohamed, Variational inference with normalizing flows, [arXiv:1505.05770](https://arxiv.org/abs/1505.05770).
- [34] T. Müller, B. McWilliams, F. Rousselle, M. Gross, and J. Novák, Neural importance sampling, *ACM Transactions on Graphics (TOG)* **38**, 1 (2019).
- [35] F. Noé, S. Olsson, J. Köhler, and H. Wu, Boltzmann generators: Sampling equilibrium states of many-body systems with deep learning, *Science* **365**, eaaw1147 (2019).
- [36] D. Wu, L. Wang, and P. Zhang, Solving Statistical Mechanics Using Variational Autoregressive Networks, *Phys. Rev. Lett.* **122**, 080602 (2019).
- [37] K. A. Nicoli, S. Nakajima, N. Strodthoff, W. Samek, K.-R. Müller, and P. Kessel, Asymptotically unbiased estimation of physical observables with neural samplers, *Phys. Rev. E* **101**, 023304 (2020).
- [38] K. Nicoli, P. Kessel, N. Strodthoff, W. Samek, K.-R. Müller, and S. Nakajima, Comment on “solving statistical mechanics using vns”: Introducing savant-vns enhanced by importance and MCMC sampling, [arXiv:1903.11048](https://arxiv.org/abs/1903.11048).
- [39] Y. Burda, R. Grosse, and R. Salakhutdinov, Importance weighted autoencoders, [arXiv:1509.00519](https://arxiv.org/abs/1509.00519).
- [40] S. Nowozin, Debiasing evidence approximations: On importance-weighted autoencoders and jackknife variational inference, ICLR 2018 (2018), <https://openreview.net/forum?id=HyZoi-WRb>.
- [41] Y. W. Teh, D. Newman, and M. Welling, A collapsed variational Bayesian inference algorithm for latent Dirichlet allocation, in *Proceedings of the 19th International Conference on Neural Information Processing Systems* (2007), pp. 1353–1360.
- [42] C. Gattringer and C. Lang, *Quantum Chromodynamics on the Lattice: An Introductory Presentation*, Vol. 788 (Springer Science & Business Media, New York, 2009).
- [43] S. L. Adler, Over-relaxation method for the Monte Carlo evaluation of the partition function for multiquadratic actions, *Phys. Rev. D* **23**, 2901 (1981).
- [44] C. Whitmer, Over-relaxation methods for Monte Carlo simulations of quadratic and multiquadratic actions, *Phys. Rev. D* **29**, 306 (1984).
- [45] D. J. E. Callaway and A. Rahman, Lattice gauge theory in the microcanonical ensemble, *Phys. Rev. D* **28**, 1506 (1983).
- [46] Z. Fodor and K. Jansen, Overrelaxation algorithm for coupled gauge-Higgs systems, *Phys. Lett. B* **331**, 119 (1994).
- [47] C. M. Bishop, *Pattern Recognition and Machine Learning* (Springer, New York, 2006), pp. 468–469.
- [48] U. Wolff *et al.* (A. Collaboration), Monte Carlo errors with less errors, *Comput. Phys. Commun.* **156**, 143 (2004).

Correction: A proof change request for the fourth affiliation was misinterpreted and has been set right.