Flow-Based Sampling for Entanglement Entropy and the Machine Learning of Defects

Andrea Bulgarelli,[1,*] Elia Cellini,[1] Karl Jansen,[2,3] Stefan Kühn,[3] Alessandro Nada,[1] Shinichi Nakajima,[4,5,6] Kim A. Nicoli,[7,8] and Marco Panero[1,9]

[1]Department of Physics, University of Turin and INFN, Turin unit, Via Pietro Giuria 1, I-10125 Turin, Italy
[2]Computation-Based Science and Technology Research Center, The Cyprus Institute, Nicosia, Cyprus
[3]Deutsches Elektronen-Synchrotron DESY, Zeuthen, Germany
[4]Berlin Institute for the Foundations of Learning and Data (BIFOLD), Berlin, Germany
[5]Machine Learning Group, Technische Universität Berlin, Berlin, Germany
[6]RIKEN Center for AIP, Tokyo, Japan
[7]Transdisciplinary Research Area (TRA) Matter, University of Bonn, Germany
[8]Helmholtz Institute for Radiation and Nuclear Physics (HISKP), Bonn, Germany
[9]Department of Physics and Helsinki Institute of Physics, PL 64, FIN-00014 University of Helsinki, Finland

(Received 25 October 2024; revised 16 January 2025; accepted 27 March 2025; published 15 April 2025)

We introduce a novel technique to numerically calculate Rényi entanglement entropies in lattice quantum field theory using generative models. We describe how flow-based approaches can be combined with the replica trick using a custom neural-network architecture around a lattice defect connecting two replicas. Numerical tests for the ϕ^4 scalar field theory in two and three dimensions demonstrate that our technique outperforms state-of-the-art Monte Carlo calculations, and exhibit a promising scaling with the defect size.

DOI: 10.1103/PhysRevLett.134.151601

Introduction—Quantum entanglement is a key property of quantum systems, and has profound implications ranging from high-energy physics [1] to condensed-matter theory [2], holding an important role in probing quantum phases of matter and quantum phase transitions. Also, as quantum simulations are emerging as a new tool to study quantum phenomena [3], characterizing entanglement in quantum many-body systems is increasingly important. For a system with a factorizable Hilbert space and a density matrix ρ, the bipartite entanglement between a subsystem A and its complement B can be quantified by the entanglement entropy

$$S_A = -\mathrm{Tr}(\rho_A \ln \rho_A), \qquad \rho_A = \mathrm{Tr}_B \rho, \tag{1}$$

namely, the von Neumann entropy of the reduced density matrix ρ_A. While the leading term in the entanglement entropy is proportional to the area of the boundary ∂A between A and B [4], and is ultraviolet divergent in a quantum field theory, the derivative of the entanglement entropy with respect to the linear size l of the subsystem A is finite, and is called the entropic c function [5,6]

$$C = \frac{l^{D-1}}{|\partial A|} \frac{\partial S_A}{\partial l}, \tag{2}$$

where D is the number of spacetime dimensions and $|\partial A|$ the area of ∂A. As its name suggests, this quantity provides a measure of the effective number of degrees of freedom of a theory [7–10].

Computing the entanglement entropy directly from Eq. (1) is not possible when ρ_A is not known; tensor networks allow for directly accessing S_A, but scaling these methods beyond $1+1$ dimensions is often challenging. However, introducing the Rényi entropies [11]

$$S_n = \frac{1}{1-n} \ln \mathrm{Tr} \rho_A^n \tag{3}$$

(and the associated entropic c functions C_n), one can obtain S_A as the $n \to 1$ limit of S_n. The advantage of Rényi entropies is that they can be computed through the replica trick [12], expressing the trace in Eq. (3) as the partition function of n copies of the original system, joined together in correspondence of the subsystem A (but not of B). The new geometry, analogous to a Riemann surface, is characterized by the presence of a *defect*, namely the boundary between A and B, breaking some of the spacetime symmetries of the original theory [13]. S_n and C_n can then be expressed in terms of ratios of partition functions in this replica geometry, which can be computed either with tensor networks [14], or with classical [15,16] or quantum Monte Carlo algorithms [17–19]. Classical Monte Carlo calculations, however, are often inefficient in the estimate of ratios of partition functions, since the latter cannot be computed directly as primary observables.

In recent years, this problem has been tackled by means of nonequilibrium Monte Carlo calculations based on

*Contact author: andrea.bulgarelli@unito.it

Jarzynski's equality [20], successfully applied in calculations of various physical quantities [21–23], including Rényi entropies [24] and entropic c functions [25,26].

In parallel with these developments, during the past decade deep generative models have emerged as a new tool to sample Boltzmann-like distributions [27–30]. Of particular interest are autoregressive neural networks [31] and *normalizing flows* (NFs) [32] which provide access to the exact probability distributions of statistical systems and to unbiased estimators of partition functions from the learned variational distributions [33,34].

The main question we address in this Letter is whether deep generative models can also be used to efficiently study entanglement for generic quantum field theories regularized on a lattice. As will be shown below, we give an affirmative answer to this question, and we also present a novel framework in which normalizing flows can be used to study defects in large-scale lattice simulations.

Related work—In recent years, significant efforts were devoted to develop efficient algorithms to estimate ratios of partition functions. In Refs. [25,26], entanglement-related quantities were accurately evaluated through nonequilibrium Markov-chain Monte Carlo (NE-MCMC), leading to a thermodynamic and continuum extrapolation of C_2 in the confining $\mathbb{Z}_2$ gauge theory in $2+1$ dimensions. At the same time, the idea of using NFs to sample Boltzmann distributions rapidly gained momentum: as reviewed in Ref. [29], initial proof-of-concept studies in quantum chemistry [27] and lattice quantum field theory [35–37] were soon followed by works addressing computationally more challenging physical systems, such as lattice quantum chromodynamics in $3+1$ dimensions [38].

Following another line of work, Nicoli *et al.* [33,34,39] demonstrated that asymptotically unbiased estimators of partition functions could be obtained from a trained generative model. While NFs are efficient for low-dimensional systems, they generally exhibit poor scalability with the number of degrees of freedom (d.o.f.). Conversely, NE-MCMC has good scaling properties [40] but often requires a large number of Monte Carlo updates to be effective. To try and combine the advantages of the two methods, *stochastic normalizing flows* (SNFs) were introduced [41,42], and recent studies [43,44] have indeed shown their improved scalability and effectiveness.

In the context of entanglement and replica trick, Białas *et al.* [45] employed autoregressive neural networks to estimate Rényi entropies, acknowledging the poor scaling of this approach for larger systems. In contrast, our proposed method introduces a novel type of coupling layer for normalizing flows that focuses solely on a region near the defect rather than resampling the entire lattice. This straightforward yet effective modification enables our approach to substantially reduce the relevant number of d.o.f. and to overcome the well-known scaling limitations of NFs. For the first time, we demonstrate that NFs can be used to estimate thermodynamic observables in large lattices for $(1+1)$- and $(2+1)$-dimensional scalar field theories.

Lattice field theory and entanglement entropy—We test our proposed methods for a ϕ^4 real scalar field on a lattice Λ; the Euclidean action is

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

a being the lattice spacing, κ the hopping parameter, λ the bare quartic coupling, and $\hat{\mu}$ a positively oriented unit vector along the direction μ. The action is invariant under global $\mathbb{Z}_2$ transformations, $\phi \to \pm\phi$, and the phase diagram of the model is characterized by a critical line of second-order phase transitions in the Ising universality class; therefore, in $1+1$ dimensions, our results can be benchmarked against exact predictions from conformal field theory [12].

The calculation of Rényi entropies through Monte Carlo simulations is based on the replica trick [12]: the system is replicated in n independent copies, joined together in the Euclidean-time direction in the subsystem A, effectively opening a "cut" in the replicated lattice: see Fig. 1 for a sketch for $n=2$. The trace in Eq. (3) is then rewritten as $\mathrm{Tr}\rho_A^n = Z_n/Z^n$, where Z is the partition function of the original model and Z_n the partition function of the replicated system. Similarly, the entropic c function becomes

$$C_n = \frac{l^{D-1}}{|\partial A|}\frac{1}{n-1}\lim_{a\to 0}\frac{1}{a}\ln\frac{Z_n(l)}{Z_n(l+a)}. \tag{5}$$

Nonequilibrium MCMC—Partition-function ratios like the one in Eq. (5) can be computed using Jarzynski's

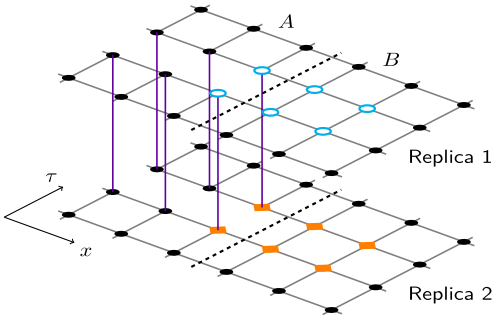

FIG. 1. $(1+1)$-dimensional lattice with two replicas (τ is the Euclidean-time direction). Purple links connect different replicas; dashed lines separate A and B. When defect coupling layers act on the configuration, the lattice is divided in three parts: the environment (black sites), which does not enter the coupling layer; frozen sites (empty cyan circles), that are the neural network input; active sites (orange diamonds), which are transformed by the layer.

equality [20]: an initial equilibrium distribution $q_0 = \exp(-S_i)/Z_i$ is driven toward a final distribution $p = \exp(-S)/Z$ by a sequence of updates, following a protocol $b(j)$, with a varying transition probability $P_{b(j)}$ that satisfies detailed balance

$$\phi_0 \xrightarrow{P_{b(1)}} \phi_1 \xrightarrow{P_{b(2)}} \dots \xrightarrow{P_{b(n_{\text{step}})}} \phi_{n_{\text{step}}}, \tag{6}$$

driving it out of equilibrium. The ratio of partition functions corresponding to the initial and the final distributions is then expressed as

$$\frac{Z}{Z_i} = \langle \exp(-W) \rangle_f, \tag{7}$$

where $\langle \dots \rangle_f$ denotes an average over a set of such evolutions, and W is the (dimensionless) work done on the system:

$$W = \sum_{j=0}^{n_{\text{step}}-1} S_{b(j+1)}(\phi_j) - S_{b(j)}(\phi_j) \tag{8}$$

$$= S(\phi_{n_{\text{step}}}) - S_i(\phi_0) - Q, \tag{9}$$

with $S_{b(n_{\text{step}})} = S$, $S_{b(0)} = S_i$, and Q being the (dimensionless) heat. The deviation from equilibrium of a given evolution can be quantified by the Kullback–Leibler (KL) divergence [46] between the forward and reverse transition probabilities:

$$D_{\text{KL}}(q_0 \mathcal{P}_f \| p \mathcal{P}_r) = \langle W \rangle_f + \ln \frac{Z}{Z_i}, \tag{10}$$

with $\mathcal{P}_f(\phi_0, \dots, \phi) = \prod_{j=0}^{n_{\text{step}}-1} P_{j+1}(\phi_j \to \phi_{j+1})$, and $\mathcal{P}_r$ the same for a reversed evolution. This approach is equivalent to annealed importance sampling [47].

Normalizing flows—Deep generative models can be used to directly estimate partition functions. In particular, NFs [32,48] allow one to sample highly nontrivial probability distributions. A NF is a composition of diffeomorphisms between probability distributions, which can be implemented as coupling layers g_l. Concatenating several coupling layers,

$$g_\theta(\phi_0) = (g_N \circ \dots \circ g_1)(\phi_0), \tag{11}$$

one maps a configuration ϕ_0, sampled from a base distribution q_0, into a configuration $\phi = g_\theta(\phi_0)$ from the learned distribution q_θ. Compared to other generative models, NFs give access to the exact likelihood of the variational density

$$q_\theta(\phi) = q_0(g_\theta^{-1}(\phi)) J_{g_\theta}^{-1}, \tag{12}$$

where J_{g_θ} is the Jacobian determinant of the transformation g_θ. By training the parameters of each layer (collectively denoted by θ) one can approximate the target distribution p with the variational *Ansatz* q_θ. While NFs allow us to compute the partition function of a target system [34], here we evaluate ratios of partition functions, transforming samples from the distribution $q_0 = \exp(-S_i)/Z_i$ to the target distribution $p = \exp(-S)/Z$. The variational density q_θ is constructed by minimizing the KL divergence

$$D_{\text{KL}}(q_\theta \| p) = \langle -\ln \tilde{w} \rangle_{q_\theta} + \ln \frac{Z}{Z_i}, \tag{13}$$

with respect to θ, and with the weight

$$\tilde{w} = \exp\left(-(S(\phi) - S_i(g_\theta^{-1}(\phi)) - \ln J_{g_\theta})\right). \tag{14}$$

Using q_0 as the prior distribution one can show that

$$\frac{Z}{Z_i} = \langle \tilde{w} \rangle_{q_\theta}, \tag{15}$$

where the average $\langle \tilde{w} \rangle_{q_\theta}$ is computed on configurations from the q_θ distribution.

Stochastic normalizing flows—The similarity of Eqs. (7) and (15) makes it natural to combine the two approaches in a more general architecture, called *stochastic normalizing flows* (SNFs) [41,42]. In practice, NE-MCMC updates are interleaved with coupling layers,

$$\phi_0 \xrightarrow{g_1} g_1(\phi_0) \xrightarrow{P_{b(1)}} \phi_1 \xrightarrow{g_2} \dots \xrightarrow{P_{b(n_{\text{step}})}} \phi_{n_{\text{step}}}. \tag{16}$$

Equation (7) is still valid if a generalized work is used,

$$W_{\text{SNF}} = S(\phi_{n_{\text{step}}}) - S_i(\phi_0) - Q - \ln J_{g_\theta}, \tag{17}$$

and if the parameters of the coupling layers g_l are trained minimizing the KL divergence of Eq. (10) using W_{SNF}.

Crucially, compared to NFs, SNFs benefit from a more favorable scaling with the d.o.f., inherited from the underlying NE-MCMC, as demonstrated in [44], while it was shown that standard NFs suffer from poor scaling [49–51]. Furthermore, SNFs are generally cheaper than NE-MCMC, since coupling layers are more computationally efficient than Monte Carlo updates.

Proposed method—The partition functions in Eq. (5) are associated with systems that differ only for a localized set of links on the lattice. This led us to introduce the *defect coupling layer* as a new architecture to study the entropic c function, where the flow acts on a codimension-2 defect, e.g., $(L/a)^{D-2}$. For a pictorial representation see Fig. 1. In the lth coupling layer, while most of the system (the "environment") is not affected by the transformation, a subset of the lattice is updated by means of a RealNVP affine transformation [52],

$$\phi^{l+1}_{\text{active}} = \exp(-|s(\phi^{l}_{\text{frozen}})|)\phi^{l}_{\text{active}} + t(\phi^{l}_{\text{frozen}}), \quad (18)$$

where s and t are the outputs of an odd neural network, enforcing the $\mathbb{Z}_2$ equivariance [34,49]. A related architecture has been discussed in Ref. [53], with a different active-frozen partitioning; here, we exploit an even-odd replica decomposition, transforming one replica at a time while keeping the others frozen. The neural networks we used are fully connected (FCNN) and convolutional (CNN).

Numerical tests—In what follows, we consider NE-MCMC as the state-of-the-art baseline to benchmark our method. Although SNFs feature better scaling for larger volumes, plain NFs are known to be more computationally efficient for smaller systems; therefore, it is worth investigating both approaches. We compare the various methods by studying their effective sample size (ESS),

$$\text{ESS} = \frac{\langle e^{-w}\rangle^2}{\langle e^{-2w}\rangle}, \quad (19)$$

with $w = W$ for NE-MCMC, $w = W_{\text{SNF}}$ for SNF, and $w = -\ln\tilde{w}$ for NF. Note that $0 \leq \text{ESS} \leq 1$, and that the larger the ESS, the smaller the variance of the estimator of Z/Z_{i} [34,40].

Figure 2, top row, displays a comparison of architectures that were trained to compute the ratio (5) with $l/a = 1$, for the scalar theory in $1+1$ dimensions on a lattice of sizes $T/a \times L/a = 128 \times 16$ and for a critical value of the couplings, $(\kappa_c, \lambda_c) = (0.2758297, 0.03)$ [54]. The trained models have been then transferred to other values of the parameters without further retraining.

For all volumes we simulated and for a large set of couplings close to the critical point, both SNFs and NFs significantly outperform the NE-MCMC, with the NFs being the ones with the largest ESS. As the models are transferred to larger volumes, the ESS remains constant, which is not surprising since the patch where the transformation of Eq. (18) acts is independent of the volume of the lattice and effectively encodes all relevant information. This allows one to study C_2 with high precision for large volumes without further retraining (see Fig. 2, top right panel). When κ is varied, the NFs are still the models with the best performances for a significant range of couplings around the critical point.

The results in $2+1$ dimensions share some similarities with those in $1+1$ dimensions. We trained the model for $l/a = 1$, $(L/a)^2 = 8^2$, $T/a = 32$ and at the critical point $(\kappa_c, \lambda_c) = (0.18670475, 0.1)$ [55], and again we transferred without retraining. The behavior at different values of κ is qualitatively similar to the previous case, with the flow-based algorithm outperforming the NE-MCMC for a large range of couplings; moreover, as the lattice volume is increased, NFs always display the largest ESS; see Fig. 2, bottom row. In contrast to the $(1+1)$-dimensional case, however, the ESS decreases with the size of the lattice: this behavior is not unexpected as in $2+1$ dimensions the number of d.o.f. on which the model acts grows as L/a.

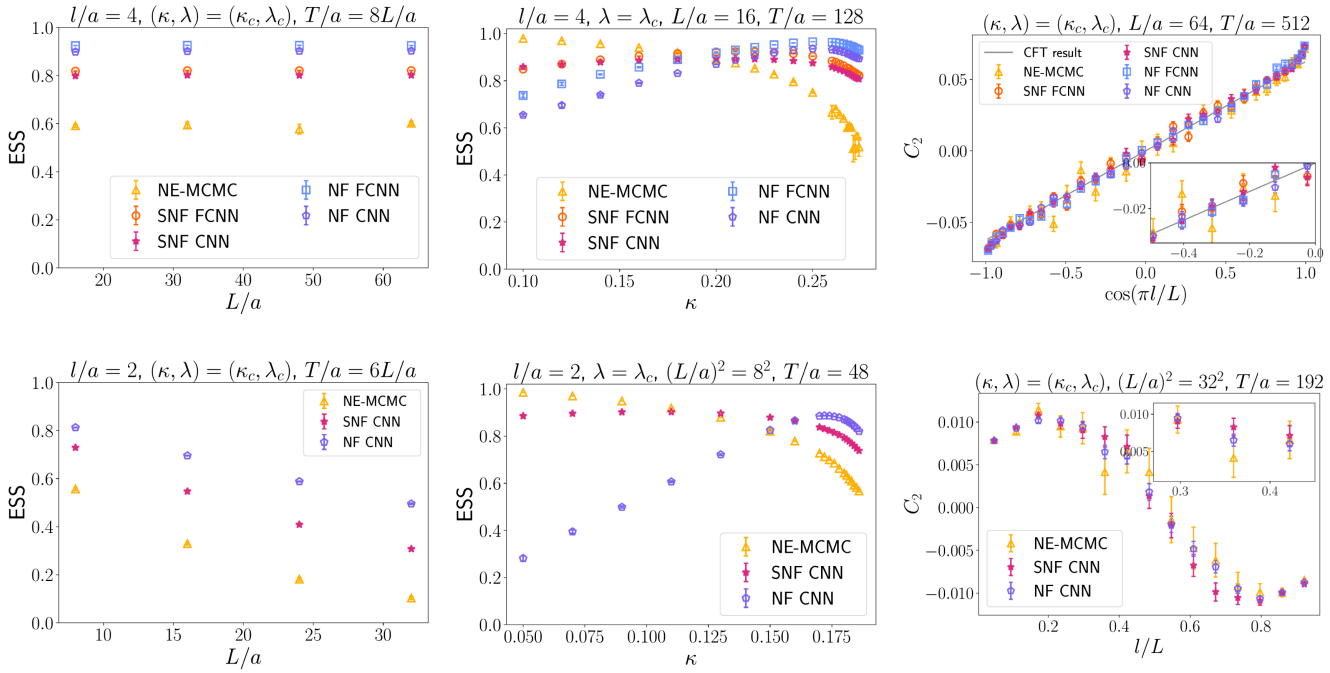

FIG. 2. Top row: results in $1+1$ dimensions. Bottom row: results in $2+1$ dimensions. Left panels: quality of the sampling as the model, trained at the smaller value of the volume, is transferred to larger volumes. Central panel: transfer in the hopping parameter κ of the model trained at (κ_c, λ_c). Right panel: estimate of the critical behavior of C_2; in $1+1$ dimensions it is compared with the analytical solution [12]. In all the plots, the quantities in the titles are fixed.

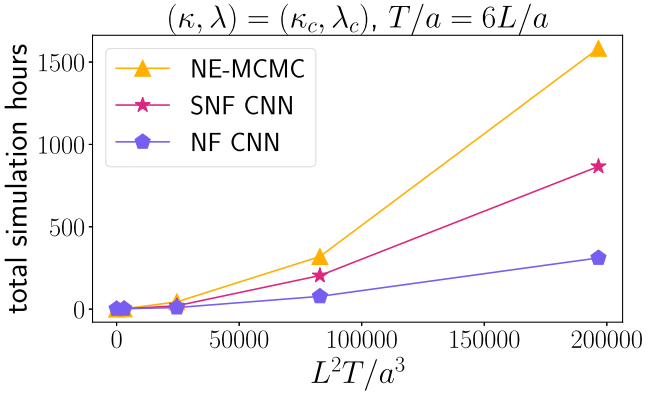

FIG. 3. Total numerical cost in wall-clock hours to compute the entropic c function in $2+1$ dimensions for increasingly large volumes at fixed target accuracy. We observe that the larger the volume of the target system, the more evident the advantage. All models used for sampling were trained at the smallest volume, for fixed $l/a = 1$ at the critical point.

Nonetheless, the ESS of the flow-based samplers is still large enough to perform a high-precision study of C_2 for large lattices ($T/a \times L^2/a^2 = 192 \times 32^2$ in the bottom right panel of Fig. 2).

Finally, we studied the total cost for training, thermalization, and sampling, for each method. In Fig. 3 we plot the cumulative cost to perform the simulations up to a given lattice volume, to compute the entropic c function at the critical point in $2+1$ dimensions, similar to the study in Fig. 2, bottom right panel.

This Letter presents, for the first time in the lattice literature, evidence that flow-based methods *outperform* MCMC-based simulations across a physically relevant range of volumes, establishing a new state of the art. As it is known that SNFs exhibit a more favorable scaling with the degrees of freedom, we expect them to hold promise of being more effective than NFs for larger volumes. We refer the reader to Supplemental Material [56] for additional results on the scaling for both NFs and SNFs.

Conclusions—In this Letter, we introduced a novel method to compute physical observables relevant for studying entanglement in quantum systems. Our approach represents a new state-of-the-art method for the numerical evaluation of Rényi entropies in the Lagrangian approach. Specifically, we proposed a new type of coupling layer for normalizing flows, which acts on a reduced number of degrees of freedom, and is particularly efficient to study lattice defects.

This Letter is the first example of flow-based sampling outperforming state-of-the-art baselines for large lattices. The scaling of our method in $2+1$ dimensions is especially promising when employing SNFs. The implementation of *direct transfer learning* enables efficient sampling across different system volumes, coupling constant values and defect lengths without the need for retraining. This key advantage allows for a single, inexpensive training on small lattices, facilitating sampling from different configurations of the theory.

Our approach has a broad range of potential applications, including the study of interfaces in spin models and the sampling of different topological sectors in lattice gauge theories. We leave the study of these subjects for future work.

Acknowledgments—The authors thank M. Caselle and L. Funcke for insightful comments and discussions. This work is funded by the European Union's Horizon Europe Framework Programme (HORIZON) under the ERA Chair scheme with Grant Agreement No. 101087126. This work is supported with funds from the Ministry of Science, Research and Culture of the State of Brandenburg within the Center for Quantum Technologies and Applications (CQTA). A. N. and E. C. acknowledge support and A. B. acknowledges partial support by the Simons Foundation Grant No. 994300 (Simons Collaboration on Confinement and QCD Strings). A. N. 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. 2022ZTPK4E. A. B., E. C., A. N., and M. P. acknowledge support from the SFT Scientific Initiative of INFN. K. A. N. is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) as part of the CRC 1639 NuMeriQS—Project No. 511713970. S. N. is supported by the German Ministry for Education and Research (BMBF) as BIFOLD—Berlin Institute for the Foundations of Learning and Data (BIFOLD24B), and by the European Union—HORIZON MSCA Doctoral Networks programme project AQTIVATE (101072344). The numerical simulations were run on machines of the Consorzio Interuniversitario per il Calcolo Automatico dell'Italia Nord Orientale (CINECA).

[1] I. R. Klebanov, D. Kutasov, and A. Murugan, Nucl. Phys. **B796**, 274 (2008).
[2] G. Vidal, J. I. Latorre, E. Rico, and A. Kitaev, Phys. Rev. Lett. **90**, 227902 (2003).
[3] M. C. Bañuls *et al.*, Eur. Phys. J. D **74**, 165 (2020).
[4] J. Eisert, M. Cramer, and M. B. Plenio, Rev. Mod. Phys. **82**, 277 (2010).
[5] H. Casini and M. Huerta, J. Phys. A **40**, 7031 (2007).
[6] T. Nishioka and T. Takayanagi, J. High Energy Phys. 01 (2007) 090.
[7] A. B. Zamolodchikov, Pis'ma Zh. Eksp. Teor. Fiz. **43**, 565 (1986) [JETP Lett. **43**, 730 (1986)].
[8] H. Casini and M. Huerta, Phys. Lett. B **600**, 142 (2004).
[9] H. Casini and M. Huerta, Phys. Rev. D **85**, 125016 (2012).
[10] H. Casini, E. Testé, and G. Torroba, Phys. Rev. Lett. **118**, 261602 (2017).
[11] G. Vidal, J. Mod. Opt. **47**, 355 (2000).
[12] P. Calabrese and J. L. Cardy, J. Stat. Mech. (2004) P06002.

[13] L. Bianchi, M. Meineri, R. C. Myers, and M. Smolkin, J. High Energy Phys. 07 (2016) 076.
[14] A. Coser, L. Tagliacozzo, and E. Tonni, J. Stat. Mech. (2014) P01008.
[15] P. V. Buividovich and M. I. Polikarpov, Nucl. Phys. **B802**, 458 (2008).
[16] M. Caraglio and F. Gliozzi, J. High Energy Phys. 11 (2008) 076.
[17] M. B. Hastings, I. González, A. B. Kallin, and R. G. Melko, Phys. Rev. Lett. **104**, 157201 (2010).
[18] J. D'Emidio, Phys. Rev. Lett. **124**, 110602 (2020).
[19] J. Zhao, B.-B. Chen, Y.-C. Wang, Z. Yan, M. Cheng, and Z. Y. Meng, Materials **7**, 69 (2022).
[20] C. Jarzynski, Phys. Rev. Lett. **78**, 2690 (1997).
[21] M. Caselle, G. Costagliola, A. Nada, M. Panero, and A. Toniato, Phys. Rev. D **94**, 034503 (2016).
[22] M. Caselle, A. Nada, and M. Panero, Phys. Rev. D **98**, 054513 (2018).
[23] O. Francesconi, M. Panero, and D. Preti, J. High Energy Phys. 07 (2020) 233.
[24] V. Alba, Phys. Rev. E **95**, 062132 (2017).
[25] A. Bulgarelli and M. Panero, J. High Energy Phys. 06 (2023) 030.
[26] A. Bulgarelli and M. Panero, J. High Energy Phys. 06 (2024) 041.
[27] F. Noé, S. Olsson, J. Köhler, and H. Wu, Science **365**, eaaw1147 (2019).
[28] D. Wu, L. Wang, and P. Zhang, Phys. Rev. Lett. **122**, 080602 (2019).
[29] K. Cranmer, G. Kanwar, S. Racanière, D. J. Rezende, and P. E. Shanahan, Nat. Rev. Phys. **5**, 526 (2023).
[30] Y. Tang, J. Liu, J. Zhang, and P. Zhang, Nat. Commun. **15**, 1117 (2024).
[31] A. van den Oord, N. Kalchbrenner, L. Espeholt, K. Kavukcuoglu, O. Vinyals, and A. Graves, in Advances in Neural Information Processing Systems (2016), Vol. **29**, pp. 4797–4805, arXiv:1606.05328.
[32] G. Papamakarios, E. Nalisnick, D. J. Rezende, S. Mohamed, and B. Lakshminarayanan, J. Mach. Learn. Res. **22**, 2617 (2021).
[33] K. A. Nicoli, S. Nakajima, N. Strodthoff, W. Samek, K.-R. Müller, and P. Kessel, Phys. Rev. E **101**, 023304 (2020).
[34] K. A. Nicoli, C. J. Anders, L. Funcke, T. Hartung, K. Jansen, P. Kessel, S. Nakajima, and P. Stornati, Phys. Rev. Lett. **126**, 032001 (2021).
[35] M. S. Albergo, G. Kanwar, and P. E. Shanahan, Phys. Rev. D **100**, 034515 (2019).
[36] G. Kanwar, M. S. Albergo, D. Boyda, K. Cranmer, D. C. Hackett, S. Racanière, D. J. Rezende, and P. E. Shanahan, Phys. Rev. Lett. **125**, 121601 (2020).
[37] M. Caselle, E. Cellini, and A. Nada, J. High Energy Phys. 02 (2024) 048.
[38] R. Abbott, A. Botev, D. Boyda, D. C. Hackett, G. Kanwar, S. Racanière, D. J. Rezende, F. Romero-López, P. E. Shanahan, and J. M. Urban, Phys. Rev. D **109**, 094514 (2024).
[39] K. A. Nicoli, C. J. Anders, T. Hartung, K. Jansen, P. Kessel, and S. Nakajima, Phys. Rev. D **108**, 114501 (2023).
[40] C. Bonanno, A. Nada, and D. Vadacchino, J. High Energy Phys. 04 (2024) 126.
[41] H. Wu, J. Köhler, and F. Noe, in Advances in Neural Information Processing Systems (2020), Vol. **33**, pp. 5933–5944, arXiv:2002.06707.
[42] M. Caselle, E. Cellini, A. Nada, and M. Panero, J. High Energy Phys. 07 (2022) 015.
[43] M. Caselle, E. Cellini, and A. Nada, J. High Energy Phys. 02 (2025) 090.
[44] A. Bulgarelli, E. Cellini, and A. Nada, arXiv:2412.00200.
[45] P. Białas, P. Korcyl, T. Stebel, and D. Zapolski, Phys. Rev. E **110**, 044116 (2024).
[46] S. Kullback and R. A. Leibler, Ann. Math. Stat. **22**, 79 (1951).
[47] R. M. Neal, Stat. Comput. **11**, 125 (2001).
[48] D. Rezende and S. Mohamed, in *Proceedings of the 32nd International Conference on Machine Learning* (PMLR, 2015), Vol. 37, pp. 1530–1538.
[49] L. Del Debbio, J. Marsh Rossney, and M. Wilson, Phys. Rev. D **104**, 094507 (2021).
[50] R. Abbott *et al.*, arXiv:2305.02402.
[51] R. Abbott *et al.*, Eur. Phys. J. A **59**, 257 (2023).
[52] L. Dinh, J. Sohl-Dickstein, and S. Bengio, in Proceedings of the International Conference on Learning Representations (2017), arXiv:1605.08803.
[53] R. Abbott, D. Boyda, D. C. Hackett, G. Kanwar, F. Romero-López, P. E. Shanahan, J. M. Urban, and M. S. Albergo, Proc. Sci., LATTICE2023 (**2024**) 011 [arXiv:2404.11674].
[54] P. Bosetti, B. De Palma, and M. Guagnelli, Phys. Rev. D **92**, 034509 (2015).
[55] M. Hasenbusch, J. Phys. A **32**, 4851 (1999).
[56] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevLett.134.151601 for details on simulations, algorithm performance across theory parameters, and a preliminary scaling analysis. Supplemental Material also includes Refs. [57–61].
[57] F. Joswig, S. Kuberski, J. T. Kuhlmann, and J. Neuendorf, Comput. Phys. Commun. **288**, 108750 (2023).
[58] A. G. D. G. Matthews, M. Arbel, D. J. Rezende, and A. Doucet, in *Proceedings of the International Conference on Machine Learning* (PMLR, 2022), pp. 15196–15219, arXiv:2201.13117.
[59] A. Bulgarelli, E. Cellini, and A. Nada, Proc. Sci., LATTICE2024 (**2025**) 040 [arXiv:2409.18861].
[60] D. P. Kingma and J. Ba, arXiv:1412.6980.
[61] D. Schuh, J. Kreit, E. Berkowitz, L. Funcke, T. Luu, K. A. Nicoli, and M. Rodekamp, Proc. Sci., LATTICE2024 (**2025**) 069 [arXiv:2501.07371].