

Rényi entanglement entropy of spin chain with Generative Neural Networks

Piotr Białas*

Institute of Applied Computer Science, Jagiellonian University, ul. Łojasiewicza 11, 30-348 Kraków, Poland

Piotr Korcyl† and Tomasz Stebel‡

Institute of Theoretical Physics, Jagiellonian University, ul. Łojasiewicza 11, 30-348 Kraków, Poland

Dawid Zapolski§

Faculty of Mathematics and Computer Science, Jagiellonian University, ul. Łojasiewicza 6, 30-348 Kraków, Poland

(Dated: June 11, 2024)

We describe a method to estimate Rényi entanglement entropy of a spin system, which is based on the replica trick and generative neural networks with explicit probability estimation. It can be extended to any spin system or lattice field theory. We demonstrate our method on a one-dimensional quantum Ising spin chain. As the generative model, we use a hierarchy of autoregressive networks, allowing us to simulate up to 32 spins. We calculate the second Rényi entropy and its derivative and cross-check our results with the numerical evaluation of entropy and results available in the literature.

Keywords: Entanglement Entropy, Hierarchical Autoregressive Neural Networks, Monte Carlo simulations, Ising model

I. INTRODUCTION

Quantum entanglement is a phenomenon that underlies many existing and potential applications like quantum cryptography, quantum communication, and quantum computing [1]. The degree of the entanglement between two parts (A and B) of a quantum system can be quantified using the von Neumann *entanglement entropy*

$$S(A) = -\text{Tr} \rho_A \log \rho_A, \quad (1)$$

where ρ_A is the reduced density matrix, i.e. given the full density matrix

$$\rho_{ij} = \frac{\langle i | e^{-\beta H} | j \rangle}{\sum_i \langle i | e^{-\beta H} | i \rangle} \quad (2)$$

of a bipartite system divided into part A and B , the reduced density matrix ρ_A is obtained by tracing out the B part:

$$\rho_A = \text{Tr}_B \rho. \quad (3)$$

The estimation of the von Neumann entanglement entropy requires the full eigenspectrum of the matrix ρ_A . Because the size of the Hilbert space grows exponentially with the size of the system, such calculations are notoriously difficult. Some simplification can be obtained by quantifying quantum entanglement by the quantum Rényi entropy of order n ,

$$S_n(A) = \frac{1}{1-n} \log \text{Tr} \rho_A^n. \quad (4)$$

The von Neumann entanglement entropy can be recovered in the limit of $n \rightarrow 1$. The advantage of Rényi entropy comes from the fact that one can employ the replica trick together with the path integral formalism to rewrite the trace of the power of the reduced density matrix as the ratio of partition functions of systems multiplied in the time direction with appropriate boundary conditions [2]. From this point on, several methods can be applied to access such partition functions. For instance, one can implement a numerical Monte Carlo sampling procedure [3, 4]. Taking as an example the second Rényi entropy, i.e. $S_2(A)$, one has to estimate the ratio of the partition functions of two systems, with a twice time extent differing in boundary conditions. The main difficulty hides behind the fact that the sampling algorithm has to easily transfer configurations with one topology of boundary conditions to the other and back. Failure may result in large autocorrelation time and subsequently large statistical uncertainty on the estimated ratio of partition functions. The problem increases very quickly with the system size and with increasing entanglement entropy. Several techniques have been devised to remedy this problem and ensure the correctness of the outcomes [5, 6]. A separate class of Monte Carlo algorithms is exploiting the Jarzynski equality in non-equilibrium statistics to relate the change in free energy to the ratio of partition functions [6–8]. Among the other approaches aiming at estimating the rate of entanglement, very popular are methods that use tensor networks to approximate the ground state of the system and then find the maximal overlap of that approximated ground state with the most general separable state [9, 10].

In this contribution, we investigate an alternative approach. We do start with the replica trick and replicate the system in the time direction according to the prescription. However, we treat the numerical simulation of the original system and its replicated counterpart sepa-

* piotr.bialas@uj.edu.pl

† piotr.korcyl@uj.edu.pl

‡ tomasz.stebel@uj.edu.pl

§ dawid.zapolski@student.uj.edu.pl

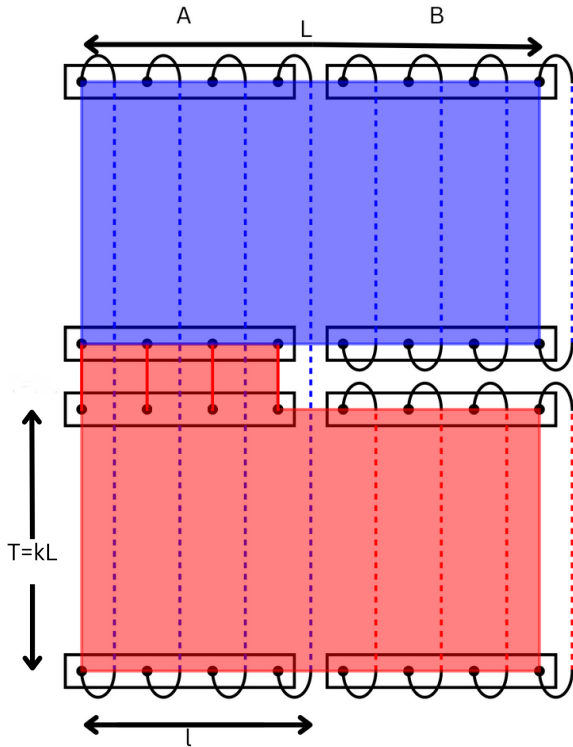


Figure 1. Sketch of spin division in replica trick. The two replicas are denoted by distinct colours: blue and red. A and B denote the two subsystems. L is the physical length of the Ising chain, and l the length of the subsystem A . T is the length in the imaginary time direction of the single replica. The dashed lines show which spins have to be identified to form correct boundary conditions.

rately. We then access the partition functions of each of them independently by employing the recently proposed generative neural network architectures. Once correctly trained, the latter offer direct access to the partition function and other thermodynamic quantities of investigated statistical systems. Our aim is to provide quantitative estimates of the efficiency of such an approach and the possible precision of the results with the currently available neural network architectures. We use the quantum Ising spin chain in the transverse magnetic field as the test bench and study system sizes up to 32 spins, estimating the ground state bipartite second Rényi entanglement entropy as a function of $x = l/L$, where l and L are the lengths of part A and the whole system $A \cup B$, respectively.

The rest of the paper has the following structure. In Section II we briefly describe the details of the replica trick and provide the Hamiltonian of the Ising spin chain, introducing the required definitions of the partition functions. We also describe how generative neural networks based on autoregressive architectures can be used to sample configurations of the two-dimensional statistical spin system. In Section III we provide the overview of our results. In particular, we describe our results for

$S_2(x)$ and its derivative with respect to x , $C_n(x)$, the so-called entropic c -function. We compare our results with those obtained through the Jarzynski inequality and off-equilibrium numerical simulations of Ref.[8]. We conclude and provide some outlook in Section IV.

II. METHOD

A. Rényi entropy and replica trick

Within the path integral formalism, the replica trick [2] allows to express the trace of an arbitrary power of the reduced density matrix ρ_A in terms of a partition function of the system multiplied in the imaginary time direction. More specifically, the quantization of a one-dimensional lattice system with L spins gives rise to a discretized two-dimensional statistical system with dimensions $L \times T$. The division into parts A and B subsists for all imaginary times. We recover the quantized one-dimensional system in the limit $T \rightarrow \infty$ and $\epsilon \rightarrow 0$, where ϵ is the spacing in the imaginary time direction. Keeping T finite corresponds to studying the quantum system at a non-zero temperature. Hence, for sufficiently large T we expect that the dynamics of the quantum system is dominated by its ground state. The extracted Rényi entanglement entropy thus corresponds to the ground state entanglement entropy. In the implementation, we shall always set, $T = kL$ where k is integer ≥ 1 , and investigate how the results change as we increase k .

In Fig. 1 we show schematically the topology of the duplicated system for $n = 2$. Blue and red colours distinguish the two replicas, which are "glued" together at one imaginary time slice only for the subsystem A . The dashed lines show which spins have to be identified to form correct boundary conditions.

For a given subdivision into parts A and B we denote as $Z_n(A)$ the partition function of the full system of replicated n times and the standard partition function (of one replica) by Z . Then, the n -th Rényi entropy is given by [2]

$$S_n(A) = \frac{1}{1-n} \log \frac{Z_n(A)}{Z^n}. \quad (5)$$

B. Ising model

In this manuscript, we consider the quantum Ising model in $D=1+1$ dimensions with the Hamiltonian:

$$\hat{H} = -J \sum_{\langle i,j \rangle} \hat{\sigma}_i^z \hat{\sigma}_j^z - h \sum_i \hat{\sigma}_i^x, \quad (6)$$

where $\hat{\sigma}_i$ are spin operators and their standard representation in the $\hat{\sigma}^z$ basis is given by Pauli matrices. The second term is an interaction with an external (transverse) field. Using the path integral formalism, it can be translated to the two-dimensional classical Ising model

without an external field [11]. For simplicity, we choose J and h such that in the corresponding classical model the couplings between the spins are equal in the imaginary time and spatial directions (together denoted as β). Therefore, the energy of the classical Ising model is given by:

$$E(\mathbf{s}) = -\beta \sum_{\langle i,j \rangle} s_i s_j, \quad (7)$$

where $s = \pm 1$ and the periodic boundary conditions are used. Accordingly, the partition function Z is given as:

$$Z = \sum_{\mathbf{s}} e^{-E(\mathbf{s})}. \quad (8)$$

In order to calculate the partition function of the replicated system $Z_n(A)$, one needs to consider configurations of spins $\mathbf{s}_{(n)}$ of the multiplicated system with interactions dictated by the replica trick. We therefore define the modified energy $E_{(n)}$, for which standard periodic boundary conditions are modified by the cut, as indicated in Fig. 1. The partition function $Z_n(A)$ is then given by:

$$Z_n(A) = \sum_{\mathbf{s}_{(n)}} e^{-E_{(n)}(\mathbf{s}_{(n)})}. \quad (9)$$

In order to eliminate the less interesting constant part and concentrate on the universal coefficient of the l dependent part, it is customary to consider the derivative of S_n w.r.t. l , so-called entropic c-function, which is given by [8]:

$$C_n(l) = \left[\frac{L}{\pi} \sin \left(\frac{\pi l}{L} \right) \right]^{D-1} \frac{1}{|\partial A|} \frac{1}{1-n} \times \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \log \frac{Z_n(l)}{Z_n(l+\epsilon)}. \quad (10)$$

The term $|\partial A|$ corresponds to the boundary of the segment A , which for the one-dimensional system consists of two end-points of the segment, and hence is equal to 2. In practical calculations, we use the approximation of the derivative employing the second-order central finite difference formula:

$$C_n(l) \approx \frac{L}{2\pi} \sin \left(\frac{\pi l}{L} \right) \frac{1}{1-n} \log \frac{Z_n(l - \frac{1}{2})}{Z_n(l + \frac{1}{2})}, \quad (11)$$

where we also have set $D = 2$. In what follows, we shall calculate $C_n(l)$ for fractional values of $l = 3/2, 5/2, 7/2, \dots$. Although, in this formula the standard partition function Z has cancelled out, $C_n(l)$ still requires the estimation of the ratio of two distinct partition functions.

C. Neural Importance Sampling for partition functions

In this contribution, we propose to estimate Rényi entanglement entropy by directly computing the partition

functions (8) and (9) using generative neural networks with explicit probability estimation. We denote the probability distribution, modelled by the neural network, by q_θ , where θ collects all the parameters of the neural network. It is important to distinguish it from the Boltzmann probability distribution, which is the target for training:

$$p(\mathbf{s}) = \frac{1}{Z} e^{-E(\mathbf{s})}. \quad (12)$$

In what follows, we shall focus on autoregressive neural networks (ANNs) [12], because they are well adapted to discrete spin systems. ANNs have proven to be effective samplers of Ising model [13, 14] and can be in principle applied to any model with discrete degrees of freedom [15]. However, it should be noted that our method is not restricted to any particular architecture.

ANNs utilize the probability product rule, where q_θ is factorized into a product of conditional probabilities

$$q_\theta(\mathbf{s}) = \prod_{i=1}^{N_{spin}} q_\theta(s^i | s^1, s^2, \dots, s^{i-1}). \quad (13)$$

One samples spin configurations from q_θ by ancestral sampling, i.e. by fixing the spins one after another based on the conditional probabilities and the values of previously fixed spins.

The aim of the training is to tune the parameters θ so that q_θ is close to p with respect to the Kullback-Leibler divergence:

$$D_{KL}(q_\theta | p) = \sum_{\mathbf{s}} q_\theta(\mathbf{s}) \log \left(\frac{q_\theta(\mathbf{s})}{p(\mathbf{s})} \right). \quad (14)$$

Once the network is trained, one can use it to sample configurations and calculate partition functions. We first note that Eq.(8) can be written as:

$$Z = \sum_{\mathbf{s}} q_\theta(\mathbf{s}) \frac{e^{-\beta E(\mathbf{s})}}{q_\theta(\mathbf{s})} \equiv \langle \hat{w}(\mathbf{s}) \rangle_{q_\theta}, \quad (15)$$

where we defined the weights $\hat{w}(\mathbf{s}) = e^{-\beta E(\mathbf{s})}/q_\theta(\mathbf{s})$ and the average of the r.h.s. is performed over the probability distribution q_θ . The latter can be approximated by the standard arithmetic mean from $N \gg 1$ configurations, distributed according to the distribution q_θ :

$$Z \approx \frac{1}{N} \sum_{i=1}^N \hat{w}(\mathbf{s}_i) \equiv \hat{Z}_N, \quad \mathbf{s}_i \sim q_\theta, \quad (16)$$

where by $\mathbf{s}_i \sim q_\theta$ we mean that configurations $\mathbf{s}_i$ are drawn from the distribution q_θ .

Note that although the above formula formally does not require q_θ to be close to p , if this is not the case, the uncertainty of $\hat{Z}_N$ will be large. The procedure of sampling configurations from the approximate probability distribution provided by the neural network together

with the estimation of the partition function was proposed in Ref. [16] and named Neural Importance Sampling (NIS).

In this work, NIS is applied to quantum entanglement entropy for the first time. Since we evaluate each partition function separately, we avoid the problem of large ratios of partition functions, which in principle can render typical Monte Carlo estimators ineffective due to exponentially large autocorrelation times.

D. Numerical setup

In order to improve the scaling of the numerical cost of the training of neural networks, we use the hierarchical algorithm proposed in Ref.[14]. Compared to the original autoregressive networks approach [12], the single autoregressive neural network is replaced by a recursive hierarchy of much smaller neural networks. We have adapted the geometry of these networks to the particularities of the replica topology. The details are described in Appendix A.

The estimation of entanglement entropy Eq.(5) and its derivative Eq.(11) requires two different partition functions for each value of k , l and L . For the calculations presented below, we have independently trained neural networks for different values of $k = 1, \dots, 8$, system size $L = 8, 16, 32$, and each value of $l = 1, \dots, L - 1$. We profited from transfer learning by training the neural networks at $l = L/2$ and using the trained networks as the initial state for the training at $l = 1, l = 2$, and proceeding in this way until $l = L - 1$.

III. RESULTS

In what follows we concentrate on the second Rényi entropy, $n = 2$, which is the easiest to calculate numerically as the replica trick requires only a duplicated system. In principle, our method applies to any $n > 1$, however, in the numerical simulations we are limited by the total number of spins in the system for which the networks can be trained. We first consider the entropic c-functions C_2 (11) leaving entropy S_2 itself to the end of this section.

We focus on a single inverse temperature,

$$\beta = \frac{1}{2} \log(1 + \sqrt{2}) \approx 0.4406868,$$

which would correspond to the critical temperature in the classical Ising model at $L \rightarrow \infty$. This allows us to compare our results with analytic predictions from conformal field theory [2] as well as the numerical results of the same quantity presented in Ref.[8]. The approach works for any value of β , though.

To test our method, we start with a small system size of $L = 8$, where the numerically exact results can be obtained using the transfer matrix method. For this purpose we express the partition functions Z and Z_2 as a trace of powers of the transfer matrix. The transfer matrix in this case has size $2^L \times 2^L$ and for small L can be diagonalized exactly. The details of this method can be found in Appendix B. In Fig. 2 we plot the results of entropic c-function Eq.(11) obtained for $L = 8$ and $x = l/L = 1.5/8$. The horizontal axis represents the time extent of the system in units of L . One can clearly see that when k increases, the entropic c-function converges to the ground state's result (we show this $k = \infty$ result as a red horizontal line, it was obtained using transfer matrix). One can show (for example by investigating the expressions obtained using the transfer matrix approach) that the contributions from excited states decay exponentially with k . We note that the difference between the results obtained using NIS approach (green dots) and the ones obtained using the transfer matrix method (purple dots) cannot be resolved at the scale of the plot. Therefore, we provide an inset where we investigate the ratio of the results obtained using two methods. One notices that NIS gives correct results within errors with per-mile accuracy. The uncertainties of the NIS result are obtained using the bootstrap method.

Having tested our method against the transfer matrix approach, we can now move to larger system sizes where the diagonalization of the transfer matrix cannot be efficiently performed. In order to estimate the Rényi entanglement entropy of the ground state, we need to ensure that the path integrals are dominated by that state. From the data in Fig. 2 one can conclude that the time extent of $k = 8$ is enough for $L = 8$ (see also Ref. [8], where authors claim that $k = 8$ is sufficiently large). For larger systems, we propose to perform an extrapolation with $k \rightarrow \infty$ to enforce the ground state dominance. Before we describe the details of the extrapolation, let us show in Fig. 3 the results of the entropic c-function for $L = 32$ as a function of the subsystem size $x = l/L$. The purple dots indicate the results obtained using $k = 8$ and we have checked that, within uncertainties, the results for $k = 6, 7, 8$ agree with each other. We first observe that the points are distributed such that the anti-symmetry of the entropic c-function, $C_2(1 - x) = -C_2(x)$, is satisfied, as it should be for $k \rightarrow \infty$. However, the uncertainties are significant, and the expected dependence, called scaling function, $\frac{1}{16} \cos(\pi x)$, (calculation performed in [2] for $L \rightarrow \infty$) is not visible. In the case of NIS, the size of the uncertainties is directly related to the quality of the training of the neural networks, i.e. how well q_θ approximates the probability distribution p . For large neural networks (with large L and/or k), the approximation is worse, resulting in bigger uncertainties.

In order to reduce the uncertainties, instead of taking just the results for the largest k available, we propose to extrapolate to infinite k using an exponential Ansatz. This allows us to gain statistical precision by

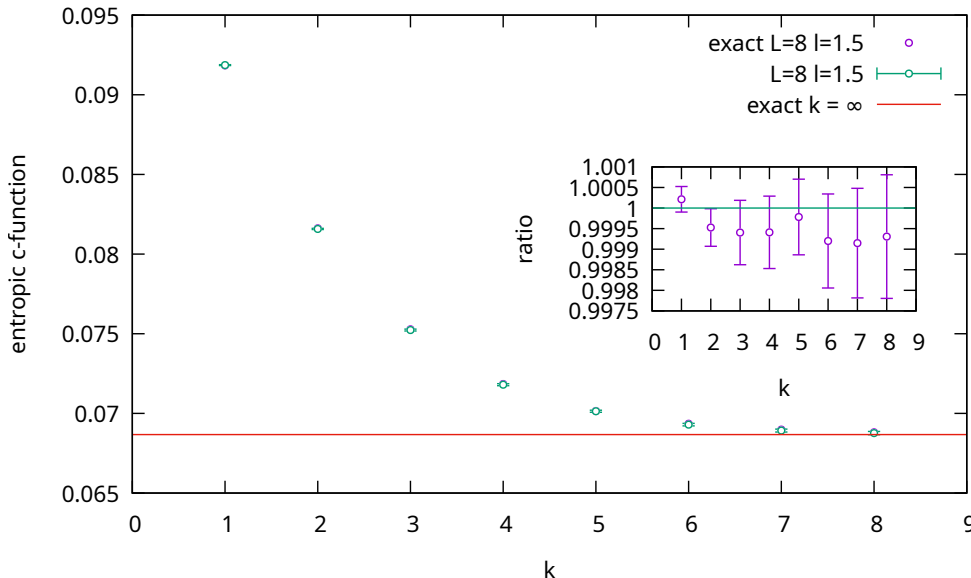


Figure 2. The comparison of the entropic c -function obtained in two independent ways: violet data come from the exact transfer matrix method (see Appendix B) and have no statistical uncertainties; green data points were generated with the NIS approach and are shown with their statistical uncertainties. Red horizontal line denotes the exact ground state result obtained by taking $k \rightarrow \infty$ in the transfer matrix method. The studied system was the Ising spin chain with $L = 8$ spins and the subsystem A had length $l = 1.5$. The inset shows the ratio of the values obtained by the two methods.

using data at several values of k . Using the property of anti-symmetry of $C_2(x)$ when $k \gg 1$, we can simultaneously fit dependence on k for $x = x_0 < 0.5$ and $x = 1 - x_0 > 0.5$. We use the following model for the fit:

$$\begin{aligned} f(k) &= a + be^{-mk} & \text{for } x = x_0, \\ f(k) &= -a + \hat{b}e^{-mk} & \text{for } x = 1 - x_0. \end{aligned} \quad (17)$$

One should note that we also assumed that the data can be described by a single excited state with an energy gap of m , which is the same for $x = x_0$ and $x = 1 - x_0$. This is motivated by transfer matrix method expressions. Therefore, we do not expect that the fit will correctly describe data points with $k \sim 1$. With this procedure, we fit four parameters $a, b, \hat{b}, m$ to 16 points of $C_2(k, x)$ ($k = 2, \dots, 8$ and two values of x : x_0 and $1 - x_0$). The values $C_2(k = \infty, x_0)$ and $C_2(k = \infty, 1 - x_0)$ are given respectively by a and $-a$.

An example of such fit is shown in Fig. 4, where we show data for two values of l : $l = 6.5$ and $l = 25.5$. Data points are obtained by averaging over all the available statistics. Curves correspond to the combined fit of Eq.(17) to all the data points except $k = 1$, which exhibits large excited states contributions and is not described by our simple Ansatz.

The results obtained by the described fitting procedure are shown in Fig. 3 as green dots. One clearly sees that the errors were significantly reduced compared to the result using just $k = 8$ (purple points).

In Fig. 5 we show entropic c -function as function of x for $L = 16$ and $L = 32$. One can see that by increasing L

we approach the theoretical result of $\frac{1}{16} \cos(\pi x)$ (named as scaling function) which is obtained in the $L \rightarrow \infty$ limit. Note that in this figure we plot $C_2(x)$ only for $x > 0.5$ as the points for $x < 0.5$ are not independent due to assumption in the fitting procedure that $C_2(1 - x) = -C_2(x)$.

To further check our results against the literature, we use the parametrization by Bulgarelli and Panero [8]:

$$C_2(x) = \frac{1}{16} \cos(\pi x) + \frac{\kappa}{2L} \cot(\pi x) \quad (18)$$

where the Authors have fitted $\kappa = 0.162$ for $L = 32$. The second term is the first-order correction due to the finite size L . As is demonstrated in Fig. 5, we get a perfect agreement with the results obtained in [8] using the Jarzynski's equality.

As for the entanglement entropy S_2 itself, we plot the result as a function of x in Fig. 6. We compare the results of $L = 16$ and $L = 32$ at $k = 8$. The entanglement entropy is maximal at $x = 0.5$ (where two subsystems are of the same size) and grows with L . We also clearly observe the symmetry property, namely $S_2(x) = S_2(1 - x)$.

IV. CONCLUSIONS

In this contribution, we presented the method of calculating entanglement entropy using generative neural networks. Such neural networks were recently used for

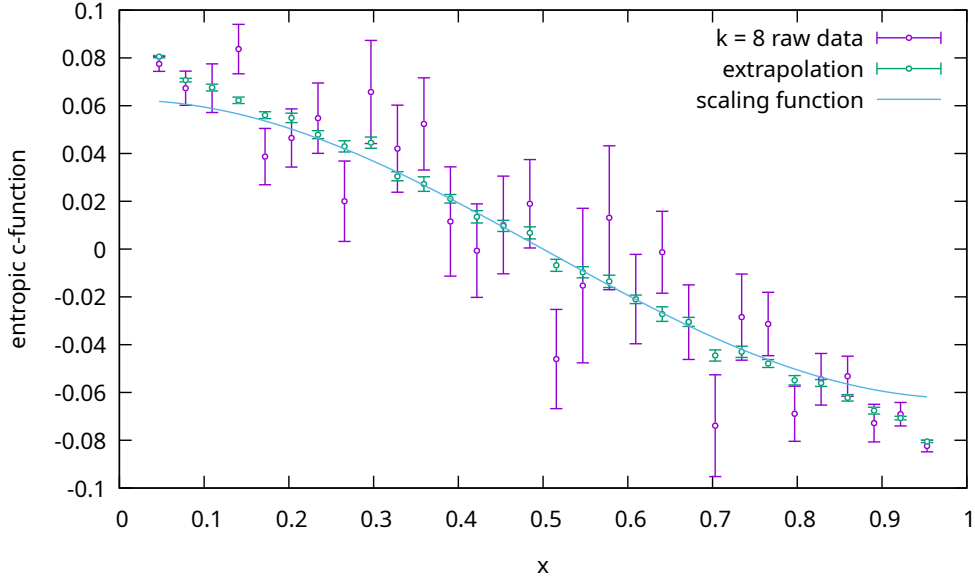


Figure 3. Entropic c -function for $L = 32$ as a function of x . Violet data points correspond to the direct evaluation with $k = 8$, whereas the green data points were obtained as a result of the combined extrapolation fit to data with $k = 2, \dots, 8$ as described in the text. The solid curve shows the theoretical expectation for infinite L , the leading term of Eq. (18).

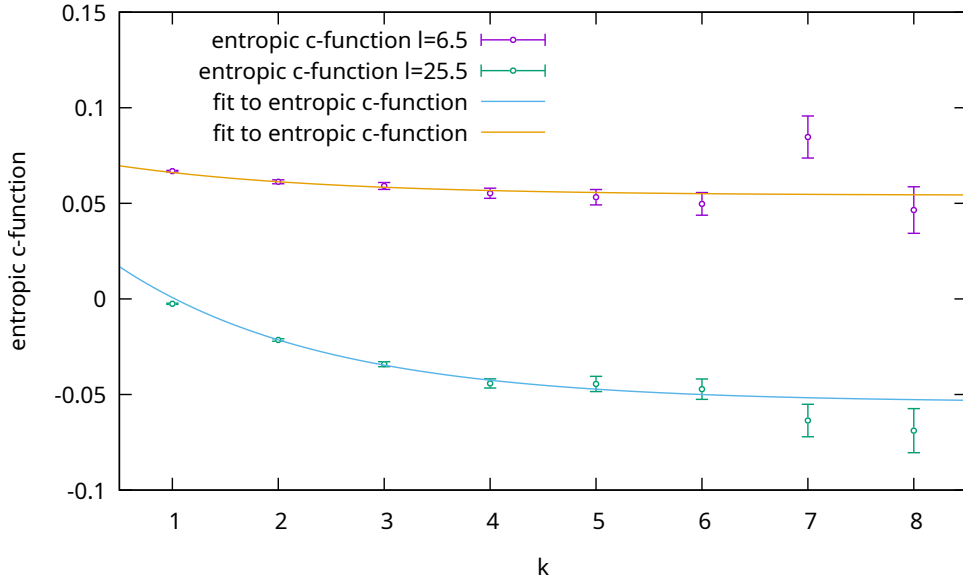


Figure 4. Example of the combined fit of Eq. (17) for $L = 32$. The data points at $k = 1$ were excluded from the fit because of the large contributions of excited states.

simulations of physically interesting quantities in spin systems [16–19] and lattice field theories [20, 21]. The key property of those architectures is that they provide explicit probabilities of the generated samples, hence giving access to thermodynamical observables like free energy or entropy. In this manuscript, we have extended the applicability of neural sampling algorithms towards

quantum information theory and calculated the entanglement entropy in the one-dimensional quantum Ising model. We compared our results with other methods and found agreement within the statistical uncertainties of our method.

The system sizes that we consider here are rather small and there exist other methods that can be used in this

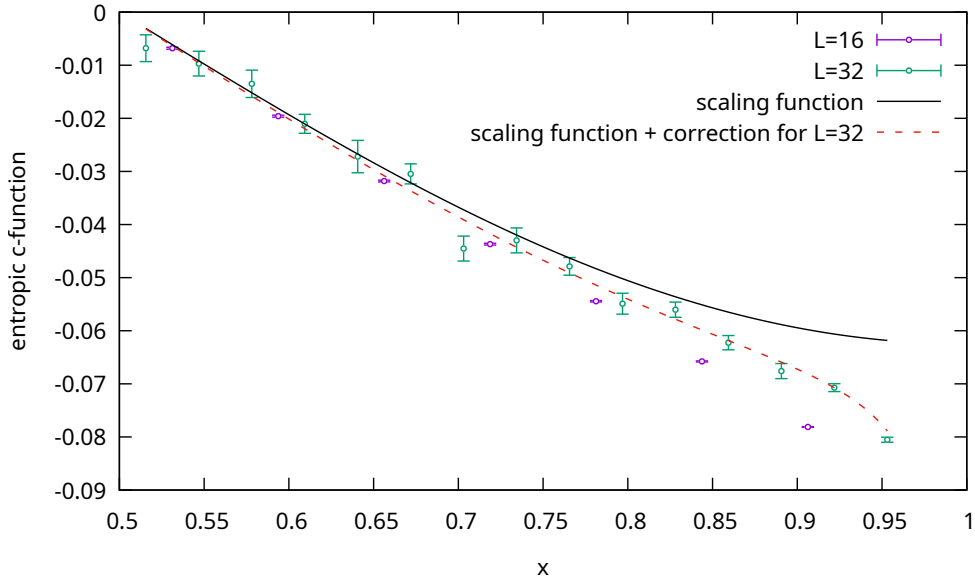


Figure 5. Entropic c-function for $L = 16$ and $L = 32$ obtained using autoregressive networks (violet and green points). The black solid line (scaling function) denotes the leading term in Eq. (18), $\sim \cos(\pi x)$. Red dashed line denotes full expression (18), namely with subleading term $\sim \cot(\pi x)$ included.

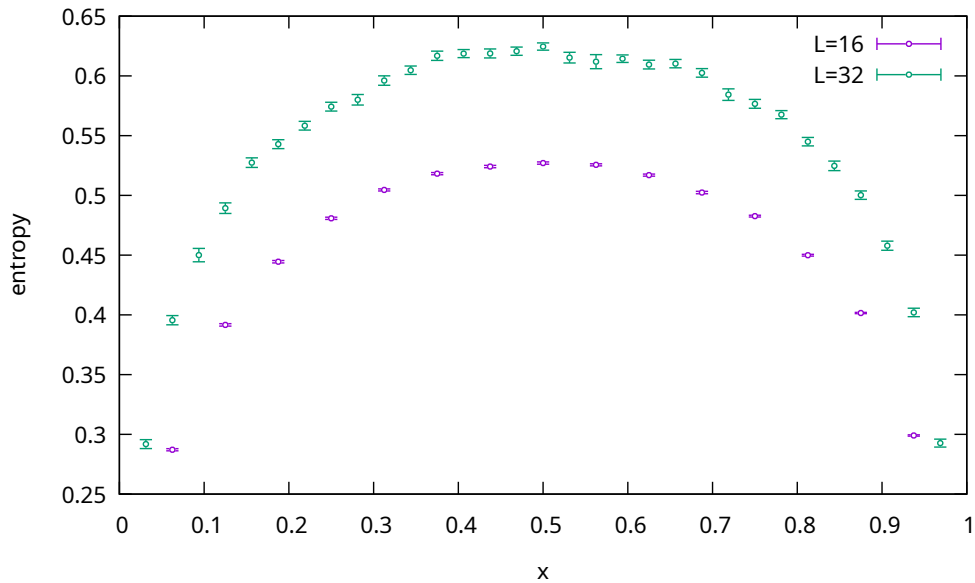


Figure 6. Rényi entropy S_2 for $L = 16$ and $L = 32$ for $k = 8$.

context, resulting in better precision. Nevertheless, our results can be seen as a first step towards developing a new, universal technique for evaluating entanglement entropy. Indeed, the method we used is quite general and can be used in any system where network-based sampling is possible. With the rapid development of deep neural network algorithms, which we are currently witnessing, we believe that our method can be competitive with tra-

ditional methods like Quantum Monte Carlo or tensor networks in the future. In particular, the progress in the normalizing flow architectures [22–26] paves the way toward evaluating entanglement entropy in lattice field theories.

ACKNOWLEDGMENTS

Computer time allocation 'plnglft' on the Athena supercomputer hosted by AGH Cyfronet in Kraków, Poland was used through the Polish PLGRID consortium. T.S. and D.Z. acknowledge the support of the Polish National Science Center (NCN) Grant No. 2021/43/D/ST2/03375. P.K. acknowledges the support of the Polish National Science Center (NCN) grant No. 2022/46/E/ST2/00346. This research was partially funded by the Priority Research Area Digiworld under the program Excellence Initiative – Research University at the Jagiellonian University in Kraków. We also acknowledge very fruitful discussions with Leszek Hadasz on entanglement entropy.

Appendix A: Configurations generation

In this work, we use hierarchical autoregressive neural networks (HAN) [14] which are modifications of Variational Autoregressive Networks (VAN) [12]. The details of the two architectures can be found for example in Appendix B of [17] and we refer the reader to this appendix for more details. In this contribution we used the same architecture of HAN as described in [17], with modifications due to the different topology of the configurations, which we shall describe below.

In Fig. 7 we show a schematic representation of the configuration of two replicas with $k = 1$ and $L = 8$ "glued together". The blue vertical line separates two subsystems of the spin chain. Different shapes of marks denote the spins fixed at a given stage of the hierarchy:

1. We start with fixing the spins denoted by the numbers from 1 to 16 using an autoregressive network. Then, the values of the spins are copied according to the pattern shown in Fig. 7: the repeated numbers denote copies of the spins. This procedure assures the correct topology of the generated configuration: on the left side the replicas are "glued" whereas on the right side each replica is "independent".
2. At the next step of the hierarchy, we generate the spins denoted with "×" mark. For this purpose we use autoregressive networks which depend on the spins that were previously fixed, but only those surrounding the given area. These are the networks introduced by the HAN algorithm [14].
3. At this stage, as the L values are powers of 2, all the remaining spins form squared areas. We proceed with the filling of those areas with the "cross" shapes, which we denote with the "+" mark in Fig. 7. For the systems with $L > 8$, we continue dividing "squared" areas into 4 smaller "squares" using the "cross" shapes, until single spins are left unfixed.

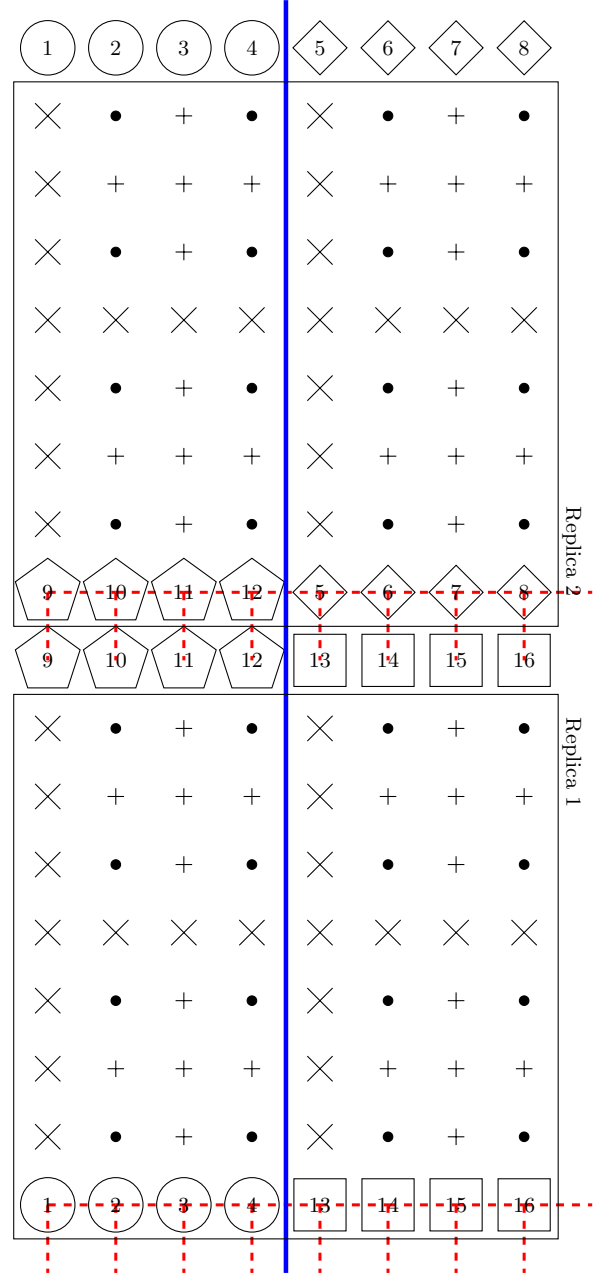


Figure 7. Sketch of spin division in replica trick. Numbers denote spins fixed first, using the first stage of hierarchy. Repeated number means that the spin is not generated but copied. Spins fixed at other stages of hierarchy are denoted with ×, +, • marks. Red dashed lines denote connections (interactions) between spins, which were removed when evaluating energy of the configuration.

4. The hierarchy ends with fixing separated single spins (denoted by "•" in Fig. 7) - single spins can be drawn using heat bath algorithm, where its probability depends only on the values of the four neighbours (see for example (C1) from Ref. [17] for the formula).

The above algorithm, which we described for $k = 1$,

can be straightforwardly generalized for $k > 1$ case. For this purpose, we change the pattern of "x" spins in the second stage of the algorithm - they form a ladder with multiple rungs.

One should notice that the number of independent spins generated in total is kL^2 for each replica (in Fig. 7 for $k = 1$, we denote them by frames).

From the technical point of view, the definition of the energy of a configuration has to be adapted to the particular topology imposed by the replica trick and its implementation as shown schematically in Fig. 7. In particular, some pairwise interactions between spins, which define the energy of the configuration, need to be adjusted. In Fig. 7 we denote connections which need to be *removed* using red dashed lines. All the other connections (not denoted with red dashed lines), which would appear in the periodic boundary conditions with nearest neighbour interactions, are assumed to be present. The connections in the horizontal direction (for example between spin 1 and 2 in Replica 1) are removed to prevent double counting (the connection between 1 and 2 is already included in the first row). Some of the connections in the vertical direction need to be removed for the same reason (these are connections that join spins with the same number, e.g. 9 and 9). Finally, some vertical connections are removed due to the shape of the replica system (for example, the connection between 5 and 13).

Appendix B: Transfer matrix method

For small sizes ($L \leq 10$) we can obtain exact results using the transfer matrix approach. We define the matrix between two rows of spins s and s' as

$$\mathcal{A}_{s';s} = \exp \left(\beta \sum_{i=0}^{L-1} s'^i \cdot s^{i+1} + \beta \sum_{i=0}^{L-1} s^i \cdot s'^i \right). \quad (\text{B1})$$

This is a $2^L \times 2^L$ matrix. The partition function of the $L \times M$ spin system is equal to

$$Z_{L,M} = \text{Tr } \mathcal{A}^M. \quad (\text{B2})$$

To calculate the partition function $Z_2(l)$ of the two replicas, we split the indices of the matrix $\mathcal{A}$ into two parts: A consisting of l spins and B containing the remaining $L - l$ spins.

$$s = s_A, s_B$$

With such split, the partition function of the two $L \times M$ replicas connected as depicted in Figure 1 can be written

as

$$\begin{aligned} Z_2 &= \sum_{s_A, s_B, s'_A, s'_B} \mathcal{A}_{s'_A, s_B; s'_A, s'_B}^M \mathcal{A}_{s_A, s'_B; s_A, s_B}^M \\ &= \sum_{s'_B} \mathcal{A}_{B:s_B; s'_B}^M \mathcal{A}_{B:s'_B; s_B}^M = \text{Tr } \mathcal{A}_B^{2M} \end{aligned} \quad (\text{B3})$$

where

$$\mathcal{A}_{B:s'_B; s_B}^M \equiv \sum_{s_A} \mathcal{A}_{s_A, s'_B; s_A, s_B}^M \quad (\text{B4})$$

Providing that we can calculate $\mathcal{A}^M$, such sums can be easily performed numerically for small L . The power $\mathcal{A}^M$ can be calculated by noting that

$$\mathcal{A}_{s';s} = p_{s'} \mathcal{T}_{s';s} p_s^{-1}, \quad (\text{B5})$$

where

$$\mathcal{T}_{s';s} = \exp \left(\beta \frac{1}{2} \sum_{i=0}^{L-1} (s^i \cdot s^{i+1} + s'^i \cdot s'^{i+1}) + \beta \sum_{i=0}^{L-1} s_i \cdot s'_i \right) \quad (\text{B6})$$

and

$$p_{s'} = \exp \left(\beta \frac{1}{2} \sum_{i=0}^{L-1} s'^i \cdot s'^{i+1} \right).$$

Please note that there is no implied summation in (B5). Formula (B5) also entails

$$\mathcal{A}_{s';s}^M = p_{s'} \mathcal{T}_{s';s}^M p_s^{-1} \quad (\text{B7})$$

Contrary to $\mathcal{A}$ matrix $\mathcal{T}$ is symmetric and can be written in the form

$$\mathcal{T}^M = \mathcal{P} \cdot \mathcal{D}^M \cdot \mathcal{P}^T$$

where $\mathcal{D}$ is the diagonal eigenvalues matrix, and $\mathcal{P}$ is the orthogonal matrix of eigenvectors. From (B7) we finally obtain

$$\mathcal{A}_{s',s}^{kL} = p_{s'} (\mathcal{P} \cdot \mathcal{D}^{kL} \cdot \mathcal{P}^T)_{s',s} p_s^{-1}.$$

All these calculations can be easily performed using NumPy library. The implementation can be found in the accompanying notebook.

[1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, "Quantum entanglement," *Reviews of*

Modern Physics, vol. 81, pp. 865–942, 6 2009.

[2] P. Calabrese and J. Cardy, "Entanglement entropy and

- conformal field theory,” *Journal of Physics A: Mathematical and Theoretical*, vol. 42, p. 504005, dec 2009.
- [3] M. B. Hastings, I. González, A. B. Kallin, and R. G. Melko, “Measuring renyi entanglement entropy in quantum monte carlo simulations,” *Phys. Rev. Lett.*, vol. 104, p. 157201, Apr 2010.
 - [4] S. Humeniuk and T. Roscilde, “Quantum Monte Carlo calculation of entanglement Renyi entropies for generic quantum systems,” *Phys. Rev. B*, vol. 86, p. 235116, 2012.
 - [5] D. J. Luitz, X. Plat, N. Laflorencie, and F. Alet, “Improving entanglement and thermodynamic rényi entropy measurements in quantum monte carlo,” *Phys. Rev. B*, vol. 90, p. 125105, Sep 2014.
 - [6] J. Zhao, B.-B. Chen, Y.-C. Wang, Z. Yan, M. Cheng, and Z. Y. Meng, “Measuring rényi entanglement entropy with high efficiency and precision in quantum monte carlo simulations,” *npj Quantum Materials*, vol. 7, June 2022.
 - [7] J. D’Emidio, “Entanglement entropy from nonequilibrium work,” *Phys. Rev. Lett.*, vol. 124, p. 110602, Mar 2020.
 - [8] A. Bulgarelli and M. Panero, “Entanglement entropy from non-equilibrium Monte Carlo simulations,” *JHEP*, vol. 06, p. 030, 2023.
 - [9] J. Eisert, “Entanglement and tensor network states,” *Modeling and Simulation*, vol. 3, p. 520, 2013.
 - [10] K. Okunishi, T. Nishino, and H. Ueda, “Developments in the tensor network — from statistical mechanics to quantum entanglement,” *Journal of the Physical Society of Japan*, vol. 91, jun 2022.
 - [11] J. B. Kogut, “An introduction to lattice gauge theory and spin systems,” *Reviews of Modern Physics*, vol. 51, p. 659, 10 1979.
 - [12] D. Wu, L. Wang, and P. Zhang, “Solving Statistical Mechanics Using Variational Autoregressive Networks,” *Phys. Rev. Lett.*, vol. 122, p. 080602, Mar. 2019.
 - [13] P. Białas, P. Korcyl, and T. Stebel, “Analysis of autocorrelation times in neural Markov chain Monte Carlo simulations,” *Phys. Rev. E*, vol. 107, no. 1, p. 015303, 2023.
 - [14] P. Białas, P. Korcyl, and T. Stebel, “Hierarchical autoregressive neural networks for statistical systems,” *Comput. Phys. Commun.*, vol. 281, p. 108502, 2022.
 - [15] P. Białas, P. Czarnota, P. Korcyl, and T. Stebel, “Simulating first-order phase transition with hierarchical autoregressive networks,” *Phys. Rev. E*, vol. 107, no. 5, p. 054127, 2023.
 - [16] 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*, vol. 101, p. 023304, Feb. 2020.
 - [17] P. Białas, P. Korcyl, and T. Stebel, “Mutual information of spin systems from autoregressive neural networks,” *Phys. Rev. E*, vol. 108, no. 4, p. 044140, 2023.
 - [18] I. Biazzo, “The autoregressive neural network architecture of the Boltzmann distribution of pairwise interacting spins systems,” *Communications Physics*, vol. 6, p. 296, Dec. 2023.
 - [19] I. Biazzo, D. Wu, and G. Carleo, “Sparse Autoregressive Neural Networks for Classical Spin Systems,” *arXiv e-prints*, p. arXiv:2402.16579, Feb. 2024.
 - [20] K. A. Nicoli, C. J. Anders, L. Funcke, T. Hartung, K. Jansen, P. Kessel, S. Nakajima, and P. Stornati, “Estimation of thermodynamic observables in lattice field theories with deep generative models,” *Phys. Rev. Lett.*, vol. 126, p. 032001, Jan 2021.
 - [21] 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.*, vol. 5, no. 9, pp. 526–535, 2023.
 - [22] R. Abbott *et al.*, “Normalizing flows for lattice gauge theory in arbitrary space-time dimension,” 5 2023.
 - [23] M. Caselle, E. Cellini, and A. Nada, “Sampling the lattice Nambu-Goto string using Continuous Normalizing Flows,” *JHEP*, vol. 02, p. 048, 2024.
 - [24] R. Abbott, M. S. Albergo, D. Boyda, K. Cranmer, D. C. Hackett, G. Kanwar, S. Racanière, D. J. Rezende, F. Romero-López, P. E. Shanahan, B. Tian, and J. M. Urban, “Gauge-equivariant flow models for sampling in lattice field theories with pseudofermions,” *Phys. Rev. D*, vol. 106, p. 074506, Oct 2022.
 - [25] K. A. Nicoli, C. J. Anders, L. Funcke, K. Jansen, S. Nakajima, and P. Kessel, “NeuLat: a toolbox for neural samplers in lattice field theories,” *PoS*, vol. LATTICE2023, p. 286, 2024.
 - [26] 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, “Applications of flow models to the generation of correlated lattice QCD ensembles,” *Phys. Rev. D*, vol. 109, no. 9, p. 094514, 2024.