ESTIMATION OF THE REDUCED DENSITY MATRIX AND ENTANGLEMENT ENTROPIES USING AUTOREGRESSIVE NETWORKS

Piotr Białas [a], Piotr Korcyl [b], Tomasz Stebel [b]
Dawid Zapolski [c,a]

ᵃInstitute of Applied Computer Science, Jagiellonian University
Łojasiewicza 11, 30-348 Kraków, Poland
ᵇInstitute of Theoretical Physics, Jagiellonian University
Łojasiewicza 11, 30-348 Kraków, Poland
ᶜDoctoral School of Exact and Natural Sciences, Jagiellonian University
Łojasiewicza 11, 30-348 Kraków, Poland

Matrix elements of the density matrix of small quantum systems have recently become experimentally accessible within the cold atoms simulations. Numerical estimation of these quantities is known to be computationally challenging. Tracing out a part of the studied system and hence accessing the reduced density matrices offers similar difficulties, but it is more interesting, as the reduced density matrix is the basis of entanglement entropy measures. In this work, we present a method of estimating the individual elements of the reduced density matrices that uses autoregressive neural networks and the path integral quantization approach. We use a hierarchy of neural networks capable of estimating conditional probabilities of consecutive spins to evaluate elements of reduced density matrices directly. Using the Ising chain as an example, we calculate the continuum limit of the ground state's von Neumann and Rényi bipartite entanglement entropies of an interval built of up to 5 spins. We demonstrate that our architecture is able to estimate all the needed matrix elements with just a single training for a fixed temperature discretization and lattice volume. Our method can be applied to other types of spin chains, possibly with defects, as well as to estimating entanglement entropies of thermal states at non-zero temperature. Hence, it may offer an attractive computational approach as a counterpart to the experimental effort.

1. Introduction

The entanglement spectrum or the Schmidt coefficients of a bipartite quantum many-body system allow us to get a handle on quantum correla-

tions in a pure or mixed state through observables such as entanglement entropy, entanglement negativity [1], or Schmidt gap [2]. The reduced density matrix, for a pure state $|\psi\rangle$ of a system divided into subsystems A and B, can be obtained from the full density matrix

$$\rho = |\psi\rangle\langle\psi|, \tag{1}$$

after a partial trace Tr_B over all degrees of freedom contained in part B

$$\rho_A = \mathrm{Tr}_B\,\rho. \tag{2}$$

Recently, through the development of sophisticated protocols going beyond what is known as quantum state tomography, individual elements of the reduced density matrix have become available experimentally in photon experiments [3, 4] or cold atom simulators [5]. Therefore, the computational access to these matrix elements to make theoretical predictions for systems of medium size, which are being studied experimentally, would be welcome. Multiple Monte Carlo algorithms have been proposed for this task with various limitations [6]. In a frequent approach, where the histogram method is employed, a large span of magnitude of the elements of the reduced density matrix leads to poor precision of the smallest elements.

Knowing the reduced density matrix, the entanglement spectrum can be accessed, as well as all the measures of quantum correlations between two subsystems. This includes, in particular, von Neumann entanglement entropy of A

$$S(A) = -\,\mathrm{Tr}\,\rho_A \log \rho_A, \tag{3}$$

and Rényi entanglement entropy of order n

$$S_n(A) = \frac{1}{1-n}\log \mathrm{Tr}\,\rho_A^n. \tag{4}$$

Over the years, various methods of quantifying entanglement entropy were developed. Their efficiencies depend on the nature of the system and its size. For example, exact diagonalization of the Hamiltonian is limited to small system sizes. The density-matrix renormalization group method (DMRG) uses matrix product states to estimate the ground state [7] and is most effective in one dimension. The replica trick [8] employs the path integral formulation and often finds applications in Quantum Field Theory. In that case, entanglement entropy can be expressed as a ratio of partition functions that can be estimated using Quantum Monte Carlo methods [9, 10]. Such a ratio can also be estimated using Jarzynski's theorem [11–13]. Finally, machine learning techniques are currently being applied to this problem as well [14, 15].

In our previous work [14], we described a method of calculating Rényi entanglement entropies using generative neural networks. It was based on the replica trick and required calculating the partition function of the replica system. Neural networks with explicit probability estimation (*e.g.*, autoregressive networks or normalizing flows) proved to be useful in the context of estimating partition functions [16, 17]. However, the reformulation of the problem using the replica trick only allows for estimating Rényi entanglement entropies and does not give direct access to the reduced density matrix or von Neumann entropy.

In this work, we propose a method that yields the full eigenspectrum of small reduced density matrices. It relies on a trained hierarchy of autoregressive neural networks. The employed architecture is conditioned on the ket and bra defining the corresponding element of the reduced density matrix. We demonstrate that the architecture is able to learn the required conditional probabilities for all the matrix elements for a given system size and discretization of temperature. Using the Ising chain as an example, we are able to simulate a spin chain of a total length of 32 spins, obtaining the elements of the reduced density matrix for the subsystem A in an interval of 1 to 5 spins. Later, we focus on the von Neumann and Rényi entropies as they are well-studied measures of quantum correlation. We perform a full continuum extrapolation of the von Neumann entanglement entropy of a path-quantized quantum Ising chain, finding very good agreement with Conformal Field Theory results.

In the following, we introduce the quantum Ising chain in its path-quantized form, then move to the detailed description of the method itself and present our results.

2. Quantum Ising chain

The Hamiltonian of the 1D transverse Ising model with L spins is given by

$$H = -J \sum_{i=0}^{L-1} \hat{\sigma}_i^z \otimes \hat{\sigma}_{i+1}^z - h \sum_{i=0}^{L-1} \hat{\sigma}_i^x \,, \tag{5}$$

where $\hat{\sigma}_i^k$ is a kth Pauli matrix associated with site i. We assume periodic boundary conditions.

For a system in thermal equilibrium at an inverse temperature $\beta = 1/k_{\text{B}}T$, the density matrix is given by

$$\rho(\beta) = \frac{1}{Z} \sum_i \mathrm{e}^{-\beta E_i} \left| \Phi^i \right\rangle \left\langle \Phi^i \right| \,, \tag{6}$$

where $\left| \Phi^i \right\rangle$ are eigenstates of the Hamiltonian and

$$Z(\beta) = \operatorname{Tr} \rho(\beta) = \sum_i \langle \Phi_i | \, \mathrm{e}^{-\beta H} | \Phi_i \rangle \tag{7}$$

is the partition function.

In what follows, we are interested in density matrix elements in the product states basis, $|\mu\rangle, |\nu\rangle \in \{ |s_1\rangle \otimes |s_2\rangle \otimes \ldots |s_L\rangle \}$,

$$\rho_{\mu\nu}(\beta) = \frac{1}{Z} \langle \mu | \, \mathrm{e}^{-\beta H} | \nu \rangle \,, \tag{8}$$

where $|s_i\rangle$ is an eigenvector of $\hat{\sigma}_i^z$ operator.

Applying the path integral formalism, we discretize the $[0, \beta]$ interval by dividing it into m small intervals $\Delta\tau$, $\beta = m\Delta\tau$ [18]. Then, the matrix elements of the 1D quantum Ising model can be calculated using the classical 2D Ising model with anisotropic couplings on the $L \times (m+1)$ lattice — see Fig. 1 for case $L = 8$ and $m = 16$. The numbers and marks on Fig. 1 denote classical (with values -1 or 1) spins.

The energy of such a classical Ising model is given by

$$
\begin{aligned}
\tilde{E}(\boldsymbol{s}) \;=\; & -\sum_{j=0}^{L-1} J_\tau \sum_{i=0}^{m-1} s_{i,j} \cdot s_{i+1,j} \\
& -\sum_{j=0}^{L-1} J_s \sum_{i=1}^{m-1} s_{i,j} \cdot s_{i,j+1} \\
& -\frac{J_s}{2} \sum_{j=0}^{L-1} (s_{0,j} \cdot s_{0,j+1} + s_{m,j} \cdot s_{m,j+1}) \,,
\end{aligned}
\tag{9}
$$

where

$$
\begin{aligned}
J_s &\equiv J_s(J, h, \Delta\tau) \;=\; \Delta\tau J \,, \\
J_\tau &\equiv J_\tau(J, h, \Delta\tau) \;=\; \operatorname{arctanh} \mathrm{e}^{-2\Delta\tau h} \,.
\end{aligned}
\tag{10}
$$

Note that the last term of (9) is an artifact of the discretization we have chosen — the horizontal couplings between spins in the first and last row are divided by $1/2$ to avoid double counting. In the horizontal direction of the 2D classical system, we have periodic boundary conditions since we are considering a periodic quantum spin chain. In the vertical direction, open boundary conditions are imposed (the first and last rows do not interact with each other).
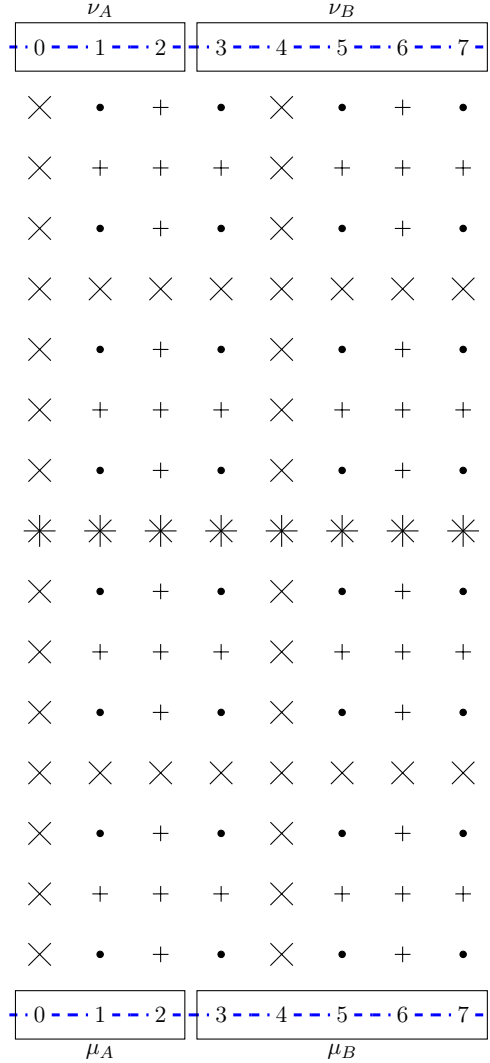
Fig. 1. The structure of the spin configurations generated during reduced density matrix estimation, plotted for $L = 8$, $m = 16$, and $l = 3$. Each node (number or mark) represents one classical spin (value -1 or 1). The first and last rows are divided into two parts (ν_A, ν_B) and (μ_A, μ_B), those rows are sampled first, the same numbers indicate identical spins. The rest of the configuration is filled in order: first: $*$, then: $\times$, then: $+$, and finally $\bullet$. The blue dashed lines mark connections with halved couplings.

To evaluate the density matrix (8), we fix the spins in the first ($i = 0$) and last ($i = m$) row of the spin configuration (see Fig. 1). We perform the sum

$$\left[\rho^{\text{cl}}(m, \Delta\tau)\right]_{\mu\nu} = \frac{1}{Z^{\text{cl}}(m, \Delta\tau)} \sum_{\boldsymbol{s}\backslash\{\mu,\nu\}} e^{-\tilde{E}(\boldsymbol{s})}, \tag{11}$$

where $\nu, \mu \in \{1, -1\}^L$ denote the first and last row of the configuration, respectively. The sum is performed over all remaining spins in the 2D system. The partition function in (11) is given as

$$Z^{\text{cl}}(m, \Delta\tau) = \sum_{\boldsymbol{s}|_{\mu=\nu}} e^{-\tilde{E}(\boldsymbol{s})}, \tag{12}$$

where the first and last rows are the same, and the summation is performed over all spins in the configuration. Note that, due to the normalization, we have $\text{Tr}\rho^{\text{cl}}(m, \Delta\tau) = 1$, as required.

The quantum density matrix and partition function are obtained in the continuum limit

$$\left[\rho^{\text{cl}}(m, \Delta\tau)\right]_{\mu\nu} \rightarrow \rho_{\mu\nu}(\beta), \tag{13}$$

$$Z^{\text{cl}}(m, \Delta\tau) \rightarrow Z(\beta), \tag{14}$$

when $m \rightarrow \infty$, $\Delta\tau \rightarrow 0$, and $m\Delta\tau = \beta = \text{const}$. Note that in this limit, anisotropy increases, $J_s \rightarrow 0$, $J_\tau \rightarrow \infty$.

We now divide the full spin chain of length L into two segments, A and B of lengths l and $L - l$, respectively. In Fig. 1, the first and last rows of the 2D configuration were divided into parts A and B, $\mu = (\mu_A, \mu_B)$, $\nu = (\nu_A, \nu_B)$. To calculate ρ_A, we keep spins fixed in μ_A and ν_A, and perform the summation over all possible configurations of spins in part B, where spins in μ_B are the same as in ν_B (in Fig. 1, the same numbers indicate identical spins)

$$\left[\rho_A^{\text{cl}}(m, \Delta\tau)\right]_{\mu_A\nu_A} = \sum_{\mu_B, \mu_B=\nu_B} \left[\rho^{\text{cl}}(m, \Delta\tau)\right]_{\mu\nu}. \tag{15}$$

The last equation can also be written using (11)

$$\left[\rho_A^{\text{cl}}(m, \Delta\tau)\right]_{\mu_A\nu_A} = Z^{\text{cl}}(m, \Delta\tau)^{-1} \sum_{\boldsymbol{s}\backslash\{\mu_A,\nu_A\}} e^{-\tilde{E}(\boldsymbol{s})}, \tag{16}$$

where by $\boldsymbol{s}\backslash\{\mu_A, \nu_A\}$ we denote all the spins in the classical configuration which are not part of μ_A or ν_A. The reduced density matrix (2) is obtained from (16) by taking the continuum limit.

The sum $\sum_{s\backslash\{\mu_A,\nu_A\}} e^{-\tilde{E}(s)}$ appearing in (16) we shall, for simplicity, call the partition function and denote by $Z^{cl}_{\mu_A,\nu_A}(m,\Delta\tau)$. Except for very small system sizes, such a sum cannot be performed in practice due to the large number of classical spin configurations. In this manuscript, we propose to use autoregressive neural networks to estimate partition functions $Z^{cl}_{\mu_A,\nu_A}(m,\Delta\tau)$. We demonstrate that such partition functions can be estimated using one model of autoregressive generative neural networks for all μ_A and ν_A.

3. Method

3.1. Autoregressive neural networks

Given a configuration of spins s, an autoregressive neural network (ANN) [19, 20] can provide the probability $q_\theta(s)$ which is factorized into a product of conditional probabilities

$$q_\theta(s) = q\left(s^1\right) \prod_{i=2}^{N} q_\theta\left(s^i|s^1,s^2,\ldots,s^{i-1}\right) , \qquad (17)$$

where N is the number of spins. Given this decomposition, we can generate a sample using *ancestral sampling*. We start by generating s^1 with probability $q(s^1)$. This allows us to calculate the conditional probability $q(s^2|s^1)$ and hence generate spin s^2, which in turn allows us to calculate $q(s^3|s^1,s^2)$. We continue in this way until all the spins are generated.

The network is trained so that q_θ is as close as possible to the target distribution, which in our case is given by the Boltzmann probability distribution

$$p(s) = \frac{1}{Z} e^{-\tilde{E}(s)} . \qquad (18)$$

The training is done by minimizing the difference between q_θ and p as measured by the Kullback–Leibler divergence

$$D_{KL}(q_\theta|p) = \sum_s q_\theta(s) \log\left(\frac{q_\theta(s)}{p(s)}\right) . \qquad (19)$$

The decomposition of $q_\theta(s)$ into the product (17) can be done in an arbitrary order. In the original approach, called Variational Autoregressive Network (VAN) [20], the generation of spin configuration was performed row by row. In [21], we took advantage of this freedom as well as the Markov property of the Ising model, and proposed to generate spins hierarchically, using several neural networks instead of one. This approach, called Hierarchical Autoregressive Networks (HAN), provides faster training and considerably improves its efficiency. In what follows, we shall use HAN to sample

configurations needed for the calculation of sum (16). The architecture we propose in this work differs from the original HAN architecture by the addition of a row of spins (see Fig. 1) and the order of sampling the spins. There is still some freedom of ordering the spins left in HAN and we use it to order the spins included in μ_A and ν_A at the beginning: $\mu_A = \{s^1, \ldots, s^l\}$, $\nu_A = s^{l+1}, \ldots s^{2l}$ (denoted by numbers 1–6 in Fig. 1). If we fix those spins, we can train the neural network to approximate the conditional probability

$$q(\boldsymbol{s}\backslash\{\mu_A, \nu_A\}) \equiv \prod_{i=2l+1}^{N} q\left(s^i | s^1, \ldots, s^{2l}, \ldots, s^{i-1}\right). \tag{20}$$

In practice, we train the network on all the combinations of spins $s^1, \ldots, s^{2l}$ by drawing them at random from a uniform distribution (see Section 3.4).

3.2. Neural Importance Sampling (NIS)

For the estimation of partition functions

$$Z^{\text{cl}}_{\mu_A, \nu_A}(m, \Delta\tau) = \sum_{\boldsymbol{s}\backslash\{\mu_A, \nu_A\}} \mathrm{e}^{-\tilde{E}(\boldsymbol{s})}, \tag{21}$$

one can use the Neural Importance Sampling (NIS) [16]. To this end, we rewrite

$$
\begin{aligned}
Z^{\text{cl}}{}_{\mu_A, \nu_A}(m, \Delta\tau) &= \\
\sum_{\boldsymbol{s}\backslash\{\mu_A, \nu_A\}} &q_\theta(\boldsymbol{s}\backslash\{\mu_A, \nu_A\})\hat{w}_A(\boldsymbol{s}) \\
&\equiv \langle \hat{w}_A(\boldsymbol{s}) \rangle_{q_\theta},
\end{aligned}
\tag{22}
$$

where we defined importance weights

$$\hat{w}_A(\boldsymbol{s}) = \mathrm{e}^{-\tilde{E}(\boldsymbol{s})}\big/ q_\theta(\boldsymbol{s}\backslash\{\mu_A, \nu_A\}). \tag{23}$$

Equation (22) can be treated as a standard Monte Carlo average; therefore, we can estimate it using some relatively small number (compared to the number of states) of samples

$$
\begin{aligned}
Z^{\text{cl}}_{\mu_A, \nu_A}(m, \Delta\tau) &\approx \frac{1}{N_s} \sum_{i=1}^{N_s} \hat{w}_A(\boldsymbol{s}_i), \\
\boldsymbol{s}_i\backslash\{\mu_A, \nu_A\} &\sim q_\theta(\boldsymbol{s}\backslash\{\mu_A, \nu_A\}),
\end{aligned}
\tag{24}
$$

where we denoted that the part of the configuration which is not μ_A or ν_A is drawn from the distribution $q_\theta(\boldsymbol{s}\backslash\{\mu_A, \nu_A\})$.

3.3. Sampling

The sampling from the conditional probability distribution $q_\theta(\boldsymbol{s} \backslash \{\mu_A, \nu_A\})$ proceeds as follows:

1. Given two product states of subsystem A, $|\mu_A\rangle$ and $|\nu_A\rangle$, we fix the classical spins in the first and last row (denoted as numbers 1 to 6 in Fig. 1).

2. We draw conditionally the spins from part B (numbers 7–11 in Fig. 1) using the first autoregressive network from HAN. The probability of spins in part B depends on spins A. Spins from part B are copied to the first and last row of the configuration.

3. Once the upper and lower rows are fixed, the interior of the configuration is filled. We make use of the Hammersley–Clifford theorem [22], which assures that once a closed loop of spins is fixed, the probability of loop-interior spins does not depend on the exterior. This is the key observation for the HAN algorithm [21], which is applied here. We use a hierarchy of networks to fix groups of spins; in Fig. 1, we schematically show such a division with different marks.

   We choose the vertical size m (where, $m = \beta/\Delta\tau$) of the configuration to be a multiple of L: $m = kL$, $k = 2, 3, 4, \ldots$ To form loops of spins, we divide the interior of the configuration into squares of size $L \times L$ by fixing spins on horizontal lines (denoted by "$*$" in Fig. 1) — this is done iteratively: if k is even, we split the configuration in half by a line of spins, and if k is odd, we effectively reduce it by one, "cutting off" one square of size $L \times L$ using the line of spins. This procedure is continued recursively until only squares remain.

   Fixing the spins is done with further networks from the hierarchy, which sample spins using probabilities depending on the spins fixed previously.

4. At the next step, $\times$-spins are drawn using yet another network and they depend only on their surrounding spins (16 in total for $L = 8$). Since the functional dependencies on the outside spins are the same for each remaining square of the configuration, we can use the same network, run in parallel, to fix "$\times$"-spins in each square. The last network from the hierarchy can be used to fix "$+$"-spins. The remaining "$\bullet$"-spins are drawn using the heatbath algorithm. This hierarchy is extended for the case of L larger than 8.

3.4. Training of HAN networks

We follow the original VAN/HAN approach and apply backward train-
ing, namely, the samples for training are provided by the network itself.
Since we calculate partition functions (24) for each of the states μ_A and
ν_A independently, we could, in principle, train a separate HAN for each of
the matrix elements. This would, however, increase the numerical cost of
the whole simulation, so instead, we train one HAN for all the matrix ele-
ments. To be specific, the batch of samples for training is obtained by: *(i)*
drawing (with probability 1/2) all spins for states μ_A and ν_A, *(ii)* sampling
other spins from $q_\theta(\boldsymbol{s}\backslash\{\mu_A, \nu_A\})$ according to the algorithm explained in the
previous section.

The training is performed by minimizing the D_{KL}-based loss function,
which is calculated using a batch of configurations sampled from $q_\theta(\boldsymbol{s}\backslash\{\mu_A, \nu_A\})$
as

$$F_q = \frac{1}{N_{\mathrm{batch}}} \sum_{k=1}^{N_{\mathrm{batch}}} \left[\tilde{E}(\boldsymbol{s}_k) + \log q_\theta(\boldsymbol{s}_k\backslash\{\mu_A, \nu_A\}) \right] . \qquad (25)$$

Since the batch contains configurations with different states μ_A and ν_A, the
network is trained to generate configurations for all reduced density matrix
elements (16).

Some additional details on the neural network architecture and training
quality are described in Appendix A.

4. Results

In the previous section, we described the method of sampling spin config-
urations, which gives the density matrix at finite $\Delta\tau$ for a given temperature
β (given m). In this section, we shall present numerical results for the quan-
tum Ising model (5). We choose $J = h = 1$, so that the system is at the
critical point. All the results we shall present are obtained for a spin chain
of size $L = 32$ spins.

For simplicity, we consider only several values of m: $m = kL$ and
$k = 2, 3, \ldots, 8$, which are then extrapolated to $k = \infty$. We evaluate the
density matrix elements at 5 linearly spaced values of $\Delta\tau$: $\Delta\tau = 0.4$, 0.35,
0.3, 0.25, 0.2, which allows us to make an extrapolation $\Delta\tau \to 0$. For each
choice of parameters $(l, m, \Delta\tau)$, we train a separate hierarchy of networks.

4.1. Reduced density matrix

We start with calculating the reduced density matrix (16) for $k = 8$ and
$\Delta\tau = 0.2$. The classical spin configurations (for $L = 32$) are of size 32×256
spins $(k = 8)$. We choose the subsystem A to be of size $l = 5$, for which the
reduced density matrix is of size 32×32. In Fig. 2, we show the result of

the estimation of density matrix elements. On the horizontal and vertical axes, we denoted 32 product states; for clarity, we introduced notation: 1 — spin has value 1, 0 — spin has value -1. To better visualize the order of magnitude of matrix elements, we plotted the logarithm of their values (this is possible as the values of the reduced density matrix elements are real and positive in this representation).
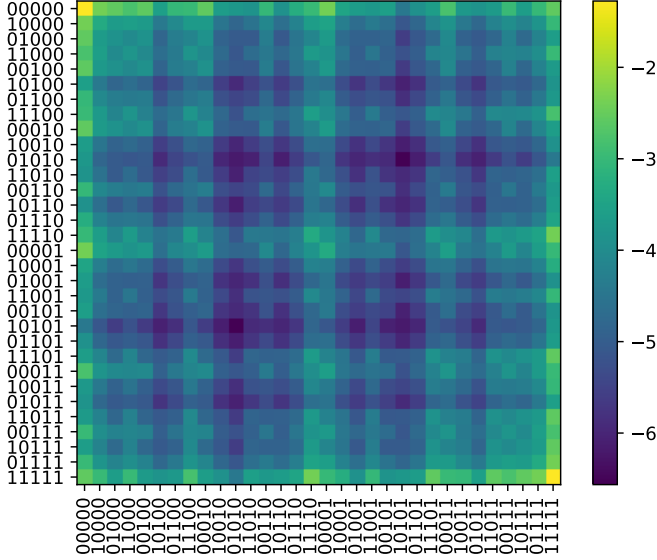

Fig. 2.   Natural logarithm of the density matrix elements in the product base $\{|s_1\rangle \otimes |s_2\rangle \otimes \ldots |s_5\rangle\}$. Plotted for $L = 32$, $l = 5$, $k = 8$, $\Delta\tau = 0.2$. On the axes, we introduced notation: 1 — spin has value $+1$, 0 — spin has value -1.

We first note that the density matrix is symmetric w.r.t. the diagonal and antidiagonal, which reflects the hermiticity of ρ_A and Z_2 symmetry of the Ising model and the specific choice of our basis of the Hilbert space. Looking at the values of the matrix elements, one sees that the biggest values correspond to elements

$$\langle 1, 1, 1, 1, 1 \ |\rho_A| \ 1, 1, 1, 1, 1 \rangle,$$
$$\langle -1, -1, -1, -1, -1 \ |\rho_A| \ -1, -1, -1, -1, -1 \rangle,$$

and the smallest to

$$\langle -1, 1, -1, 1, -1 \ |\rho_A| \ 1, -1, 1, -1, 1 \rangle,$$
$$\langle 1, -1, 1, -1, 1 \ |\rho_A| \ -1, 1, -1, 1, -1 \rangle.$$

The density matrix elements themselves depend on the basis of states; on the other hand, full physical information is encoded in the eigenvalues of ρ. In Fig. 3, we plot all 32 eigenvalues of the reduced density matrix discussed above. We obtained them by direct diagonalization of the density matrix using the numpy package. The eigenvalues were sorted and plotted in terms of their ordinal number. Errors were obtained using the bootstrap method with 800 samples. We see that only the 6 biggest eigenvalues are nonzero within the errors. The values decrease roughly exponentially for the biggest values. The first eigenvalue is of the order 1, whereas the 6th one is already of the order 10^{-3}.
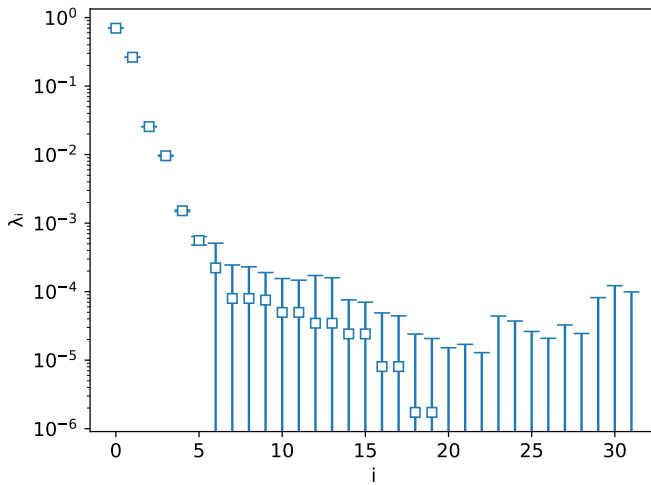
Fig. 3. Eigenvalues of reduced density matrix $\{\lambda_i\}_{i=0,\ldots,2^l-1}$ obtained for $L = 32$, $l = 5$, $k = 8$, $\Delta\tau = 0.2$. They were plotted in terms of ordinal numbers (from biggest to smallest).

4.2. Entanglement entropies

In this section, we present results for von Neumann and Rényi entropies. They are obtained from eigenvalues of the reduced density matrix $\{\lambda_i\}_{i=0,\ldots,2^l-1}$ as

$$S(A) = \sum_{i=0}^{2^l-1} \lambda_i \log \lambda_i \,, \tag{26}$$

$$S_n(A) = \frac{1}{1-n} \log \sum_{i=0}^{2^l-1} \lambda_i^n \,. \tag{27}$$

The uncertainties of entanglement entropies were obtained using the bootstrap method.

Results presented in this section are obtained after taking the zero temperature limit, $\beta \to \infty$, in order to study the entanglement entropy of the ground state. In this case, we perform the limit by first taking $k \to \infty$ and then $\Delta\tau \to 0$. Below, we present the details of this procedure.

4.2.1. Zero temperature and continuum extrapolations

In Fig. 4, we show the von Neumann entanglement entropy as functions of k for several values of $\Delta\tau$ (points plotted with different colors). The corrections coming from finite k decay exponentially with k [14], hence the extrapolation $k \to \infty$ is performed by fitting a function of the form $y_{\Delta\tau}(k) = a_{\Delta\tau} + b_{\Delta\tau} \exp(-c_{\Delta\tau}k)$ (curves in Fig. 4). For each $\Delta\tau$, we therefore extract the value of entanglement entropy at $k = \infty$, which is the parameter $a_{\Delta\tau}$ in our fits.
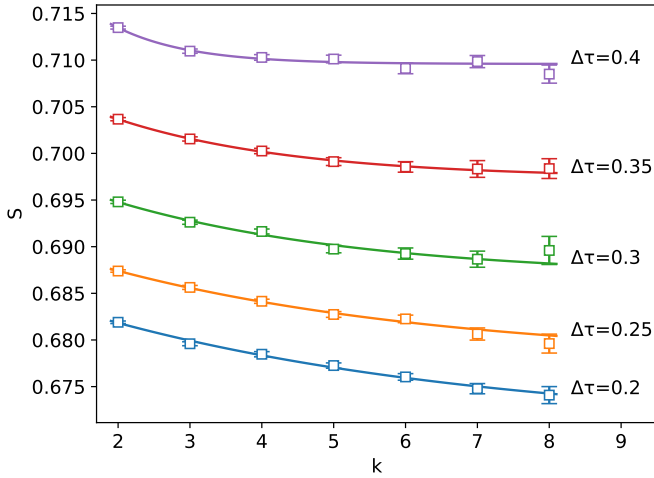

Fig. 4. The von Neuman entanglement entropy as a function of size of the temperature dimension, $k = \beta/(\Delta\tau L)$, plotted for several values of the discretization parameter $\Delta\tau$: 0.2, 0.25, 0.3, 0.35, 0.4 (denoted by points with different colors); $L = 32$ and $l = 3$. For each $\Delta\tau$, we fit the function $y_{\Delta\tau_i}(k)$, see Eq. (28), which we denote by curves.

In Fig. 5, we show fitted values of $y_{\Delta\tau}(k = \infty) = a_{\Delta\tau}$ as a function of the square of parameter $\Delta\tau$. Since the behavior we observe seems to be linear, we postulate the $\sim (\Delta\tau)^2$ behavior of the leading corrections coming from the finite discretization constant $\Delta\tau$.

To reduce the uncertainty of obtained results for the entropy at $k = \infty$ and $\Delta\tau = 0$, we use a combined fit in k for all $\Delta\tau$, where $a_{\Delta\tau}$ is quadratic in $\Delta\tau$. To be more specific, we employ the following parametrization of data for entanglement entropy:
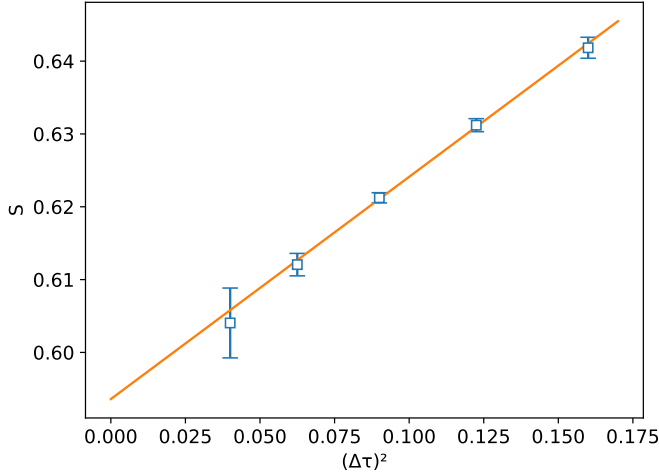
Fig. 5. The von Neumann entropy extrapolated to $k = \infty$ as a function of $(\Delta\tau)^2$ for $l = 2$.

$$\begin{cases} y_{\Delta\tau_1}(k) = S + a_1(\Delta\tau_1)^2 + b_{\Delta\tau_1} \exp\left(-c_{\Delta\tau_1}k\right) \\ y_{\Delta\tau_2}(k) = S + a_1(\Delta\tau_2)^2 + b_{\Delta\tau_2} \exp\left(-c_{\Delta\tau_2}k\right) \\ \dots \\ y_{\Delta\tau_p}(k) = S + a_1(\Delta\tau_p)^2 + b_{\Delta\tau_p} \exp\left(-c_{\Delta\tau_p}k\right) \end{cases}, \qquad (28)$$

where S corresponds to the entanglement entropy in the zero temperature and continuum limit, and p is the number of values of $\Delta\tau$ (in our case $p = 5$).

To assess the systematic uncertainty of the extrapolated value, we also consider including terms proportional to $(\Delta\tau)^3$ in (28) and perform a second fit[1]. The difference between the values of S obtained with the two functional forms is treated as the systematic uncertainty of our result, which is added in quadrature to the statistical uncertainty.

4.2.2. Ground-state entanglement

We present the von Neumann and Rényi entanglement entropies as functions of l, the size of the subsystem A. We keep the size of the whole system as $L = 32$. One expects that $\lim_{n\to 1} S_n(A) = S(A)$; therefore, we denote a value of the von Neumann entropy as $n = 1$. In Fig. 6, we show $S(l)$ and $S_n(l)$ for $n = 2, 3, 4, 9$ (points) — note that we skip some values of n for the brevity of the plot. As we noted before, we limit ourselves to $l \leq 5$, so that the reduced density matrices are relatively small. The uncertainties are at the level of one percent or less.

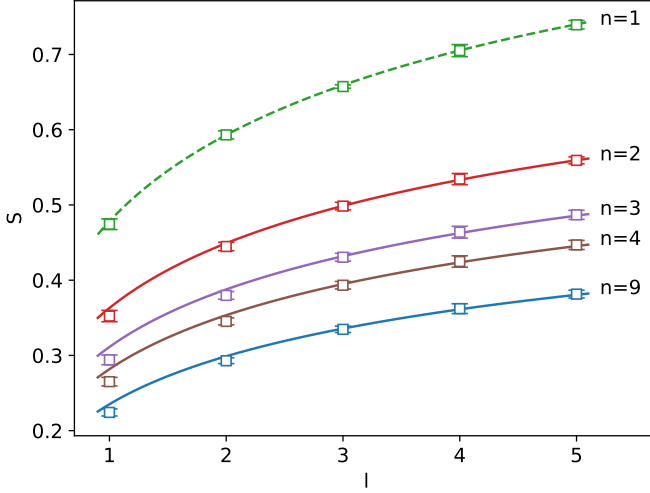[1] We have also checked the terms $\sim (\Delta\tau)^4$, and the error is similar.

Fig. 6. The von Neumann ($n = 1$) and Rényi entropies ($n \geq 2$) as functions of the subsystem of size l for spin chain with $L = 32$ spins at the zero temperature (points). The dashed line corresponds to the conformal field theory prediction [23], see Eq. (29). Solid lines represent fits of Eq. (30) to our results for $3 \leq l \leq 5$.

We shall now compare our results with the prediction of CFT. The von Neumann entropy is obtained as [24, 25]

$$S^{\mathrm{CFT}}(l) = \frac{1}{6} \ln \left[\frac{L}{\pi} \sin \left(\frac{\pi l}{L} \right) \right] + b_1 \,, \tag{29}$$

where the constant coefficient $b_1 \approx 0.478558$ is known with numerical precision [23, 26, 27]. In Fig. 6, we plotted the CFT result (29) using a dashed curve. The agreement is very good — the χ^2 per degree of freedom ($\chi^2/\mathrm{d.o.f.}$) is 0.3. This comparison serves as a cross-check of our results.

The Rényi entropy in the CFT is given by [25]

$$S_n^{\mathrm{CFT}}(l) = \frac{1}{12} \left(1 + \frac{1}{n} \right) \ln \left[\frac{L}{\pi} \sin \left(\frac{\pi l}{L} \right) \right] + b_n \,, \tag{30}$$

where b_n are constants depending only on n when $1 \ll l \ll L$. To our knowledge, for the Rényi entropies, b_n are unknown, hence, we cannot make a direct comparison with our results. However, we perform fits to our data of equation (30) for each $n \geq 2$ separately with only one free parameter, b_n. Taking all values, at $l = 1, 2, 3, 4, 5$, we get $\chi^2/\mathrm{d.o.f.}$ between 0.5 and 1.7, which means that our data are not exactly described by (30). This is not a surprise [28–30] as the condition $1 \ll l$ is not fulfilled. Taking values for

$l = 3, 4, 5$ already leads to very low $\chi^2/\text{d.o.f.}$ between 0.01 and 0.09. In Fig. 6, we showed such curves obtained with the fit to only 3 points — this does not provide strong statistical evidence that Eq. (30) is fulfilled, especially because we expect large corrections due to the smallness of l.

In Fig. 7, we show a comparison of the Rényi entropies S_n for $n = 2, 3, \ldots 8$ and the von Neumann entropy S (at $n = 1$). The values were obtained for $l = 5$. By looking only at $n \geq 2$ values, we have not found a simple form of the function which reproduces the value at $n = 1$.
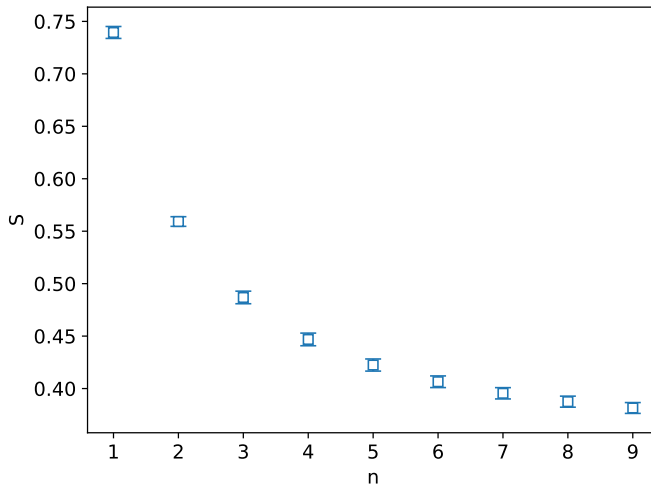
Fig. 7. The Rényi entropies of order $n = 2, 3, \ldots 9$ compared with von Neumann entropy, plotted at $n = 1$. The result was obtained for spin chain with $L = 32$ spins at zero temperature, where the size of the subsystem A is $l = 5$ spins.

5. Conclusions and outlook

In this paper, we presented a method of calculating the reduced density matrix for spin systems and demonstrated it using the 1D quantum Ising model with $L = 32$ spins. The method uses the path integral formalism to relate the 1D quantum model to a 2D classical model, which is then simulated using Monte Carlo methods based on autoregressive neural networks and the Neural Importance Sampling algorithm. The key feature of autoregressive networks is explicit access to the conditional probabilities of the spins, which allows us to use a single hierarchy of neural networks for the calculation of all density matrix elements at fixed time discretization and lattice volume.

The reduced density matrices are obtained for sizes of the subsystem equal to $l = 1, 2, 3, 4, 5$. Those matrices were later diagonalized, and the eigenvalues were used to calculate the von Neumann and Rényi entanglement entropies. Extrapolations allowed us to obtain results for entropies at zero temperature and in the continuum limit, where numerous conformal field theory (CFT) predictions are available. We found excellent agreement within our uncertainties for the von Neumann entanglement entropy. For the Rényi entanglement entropy, we fitted the dependence on l predicted by CFT; however, we observed significant final-size effects for small values of l.

Our method is not limited to zero temperature and can be applied to thermal states as well. Extraction of the entanglement entropies at some given temperature requires appropriate tuning of the size of the temporal extent m as the lattice spacing $\Delta\tau$ is decreased; taking the additional limit $T \to 0$ is not required.

There are other future directions that may be explored in using our approach. For example, it is relatively straightforward to explore different Hamiltonians (also including defects) and boundary conditions of spin systems. Such studies are interesting from the CFT point of view [31, 32]. Going beyond 1D quantum (2D classical) systems is also possible — the first generalization of hierarchical autoregressive networks to 3D classical systems is available [33]. The most important bottleneck for developing our method is its scaling with the system size L, hence, further development of neural architectures is necessary.

Appendix A

Neural network details

The HAN algorithm can, in principle, be implemented using various architectures of neural networks. Our implementation is based on the dense networks with some connections masked (weights multiplied by 0). We used two masked linear layers, the first one with PreLU activation, and the second one with Sigmoid.

The networks were trained in four stages. In all of them, the batch size was set to 2048, and the learning rate was 0.003, 0.001, 0.0001, and 0.00001 with Adam optimizer. Each stage lasted 30000 epochs. Depending on the size of the sample, the whole training for one set of parameters lasted up to three hours (on Nvidia A100 GPU).

The effective sample size (ESS) is defined as

$$\text{ESS} = \frac{\langle \hat{w} \rangle_{q_\theta}^2}{\langle \hat{w}^2 \rangle_{q_\theta}} \approx \frac{\left(\sum_{i=1}^{\mathcal{N}} \hat{w}(\boldsymbol{s}_i) \right)^2}{\mathcal{N} \sum_{i=1}^{\mathcal{N}} \hat{w}^2(\boldsymbol{s}_i)} . \tag{A.1}$$

The ESS $\in (0,1]$ is an indicator of how well the networks were trained [34] — the higher ESS, the closer q_θ is to p.

The ESS of the samples generated by the trained networks was in the range of 0.0001–0.5, depending mainly on k, but also on $\Delta\tau$. Small values of ESS imply a large variance of the weights $\hat{w}_A(\mathbf{s})$, but we have gathered enough statistics to keep the errors small.

For each density matrix element, there were 16–160 mln samples generated. The generation of 8 mln samples with a batch size set to 2^{14} lasted from 90 s to 300 s, depending on k.

REFERENCES

[1] P. Calabrese, J. Cardy, E. Tonni, «Entanglement Negativity in Quantum Field Theory», *Phys. Rev. Lett.* **109**, 130502 (2012), arXiv:1206.3092 [cond-mat.stat-mech].

[2] G. Torlai, K.D. McAlpine, G. De Chiara, «Schmidt gap in random spin chains», *Phys. Rev. B* **98**, 085153 (2018).

[3] G.S. Thekkadath *et al.*, «Direct Measurement of the Density Matrix of a Quantum System», *Phys. Rev. Lett.* **117**, 120401 (2016).

[4] C. Li *et al.*, «Direct measurement of density-matrix elements with a phase-shifting technique on a quantum photonic chip», *Phys. Rev. A* **105**, 062414 (2022).

[5] L.A. Peña Ardila, M. Heyl, A. Eckardt, «Measuring the Single-Particle Density Matrix for Fermions and Hard-Core Bosons in an Optical Lattice», *Phys. Rev. Lett.* **121**, 260401 (2018).

[6] B.-B. Mao *et al.*, «Sampling reduced density matrix to extract fine levels of entanglement spectrum and restore entanglement Hamiltonian», *Nat. Commun.* **16**, 2880 (2025), arXiv:2310.16709 [quant-ph].

[7] U. Schollwöck, «The density-matrix renormalization group in the age of matrix product states», *Ann. Phys.* **326**, 96 (2011), arXiv:1008.3477 [cond-mat.str-el].

[8] P. Calabrese, J. Cardy, «On the definition of a unique effective temperature for non-equilibrium critical systems», *J. Stat. Mech.: Theory Exp.* **2004**, 06002 (2004), arXiv:hep-th/0405152.

[9] M.B. Hastings, I. González, A.B. Kallin, R.G. Melko, «Measuring Renyi Entanglement Entropy in Quantum Monte Carlo Simulations», *Phys. Rev. Lett.* **104**, 157201 (2010).

[10] S. Humeniuk, T. Roscilde, «Quantum Monte Carlo calculation of entanglement Rényi entropies for generic quantum systems», *Phys. Rev. B* **86**, 235116 (2012), arXiv:1203.5752 [cond-mat.str-el].

[11] J. D'Emidio, «Entanglement Entropy from Nonequilibrium Work», *Phys. Rev. Lett.* **124**, 110602 (2020).

[12] J. Zhao *et al.*, «Measuring Rényi entanglement entropy with high efficiency and precision in quantum Monte Carlo simulations», *npj Quantum Materials* **7**, 69 (2022).

[13] A. Bulgarelli, M. Panero, *J. High Energy Phys.* **2023**, 030 (2023), arXiv:2304.03311 [quant-ph].

[14] P. Białas, P. Korcyl, T. Stebel, D. Zapolski, «Rényi entanglement entropy of a spin chain with generative neural networks», *Phys. Rev. E* **110**, 044116 (2024), arXiv:2406.06193 [cond-mat.stat-mech].

[15] A. Bulgarelli *et al.*, «Flow-Based Sampling for Entanglement Entropy and the Machine Learning of Defects», *Phys. Rev. Lett.* **134**, 151601 (2025), arXiv:2410.14466 [quant-ph].

[16] K.A. Nicoli *et al.*, «Asymptotically unbiased estimation of physical observables with neural samplers», *Phys. Rev. E* **101**, 023304 (2020), arXiv:1910.13496 [cond-mat.stat-mech].

[17] K.A. Nicoli *et al.*, «Estimation of Thermodynamic Observables in Lattice Field Theories with Deep Generative Models», *Phys. Rev. Lett.* **126**, 032001 (2021).

[18] E. Fradkin, L. Susskind, «Order and disorder in gauge systems and magnets», *Phys. Rev. D* **17**, 2637 (1978).

[19] M. Germain, K. Gregor, I. Murray, H. Larochelle, «MADE: Masked Autoencoder for Distribution Estimation», arXiv:1502.03509 [1502.03509].

[20] D. Wu, L. Wang, P. Zhang, «Solving Statistical Mechanics Using Variational Autoregressive Networks», *Phys. Rev. Lett.* **122**, 080602 (2019), `arXiv:1809.10606 [cond-mat.stat-mech]`.

[21] P. Białas, P. Korcyl, T. Stebel, «Hierarchical autoregressive neural networks for statistical systems», *Comput. Phys. Commun.* **281**, 108502 (2022), `arXiv:2203.10989 [cond-mat.stat-mech]`.

[22] P.Clifford, J.M. Hammersley, «Markov fields on finite graphs and lattices», 1971.

[23] B.Q. Jin, V.E. Korepin, «Quantum Spin Chain, Toeplitz Determinants and the Fisher–Hartwig Conjecture», *J. Stat. Phys.* **116**, 79 (2004), `arXiv:quant-ph/0304108`.

[24] G. Vidal, J.I. Latorre, E. Rico, A. Kitaev, «Entanglement in Quantum Critical Phenomena», *Phys. Rev. Lett.* **90**, 227902 (2003).

[25] P. Calabrese, J. Cardy, «Entanglement entropy and conformal field theory», *J. Phys. A: Math. Theor.* **42**, 504005 (2009).

[26] J.L. Cardy, O.A. Castro-Alvaredo, B. Doyon, «Form Factors of Branch-Point Twist Fields in Quantum Integrable Models and Entanglement Entropy», *J. Stat. Phys.* **130**, 129 (2008), `arXiv:0706.3384 [hep-th]`.

[27] H. Tian, T. He, X. Wu, «Finite-size scaling of half-chain entanglement entropy in the one-dimensional transverse field Ising model and the model», *Phys. Rev. B* **111**, 104437 (2025).

[28] B. Nienhuis, M. Campostrini, P. Calabrese, «Entanglement, combinatorics and finite-size effects in spin chains», *J. Stat. Mech.: Theory Exp.* **2009**, P02063 (2009), `arXiv:0808.2741 [cond-mat.stat-mech]`.

[29] J. Cardy, P. Calabrese, «Unusual corrections to scaling in entanglement entropy», *J. Stat. Mech.* **1004**, P04023 (2010), `arXiv:1002.4353 [cond-mat.stat-mech]`.

[30] T. Chanda *et al.*, «Spectral properties of the critical (1+1)-dimensional Abelian-Higgs model», *Phys. Rev. B* **109**, 045103 (2024), `arXiv:2304.01030 [hep-th]`.

[31] I. Affleck, N. Laflorencie, E.S. Sørensen, «Entanglement entropy in quantum impurity systems and systems with boundaries», *J. Phys. A: Math. Gen.* **42**, 504009 (2009), `arXiv:0906.1809 [cond-mat.stat-mech]`.

[32] A. Roy, H. Saleur, «Entanglement Entropy in the Ising Model with Topological Defects», *Phys. Rev. Lett.* **128**, 090603 (2022).

[33] P. Białas *et al.*, «Hierarchical autoregressive neural networks in three-dimensional statistical system», *Comput. Phys. Commun.* **318**, 109892 (2026), `arXiv:2503.08610 [cond-mat.stat-mech]`.

[34] J.S. Liu, «Metropolized independent sampling with comparisons to rejection sampling and importance sampling», *Stat. Comput.* **6**, 113 (1996).