

---

# Calculating Rényi Entropies with Neural Autoregressive Quantum States

---

**Zhaoyou Wang\***

Department of Applied Physics  
Stanford University  
Stanford, CA 94305  
zhaoyou@stanford.edu

**Emily J. Davis\***

Department of Physics  
Stanford University  
Stanford, CA 94305  
emilyjd@stanford.edu

## Abstract

We estimate generalized Rényi entropies  $S_n$  of an autoregressive neural network representation of a quantum state. A naïve “direct sampling” approach performs well for small Rényi order  $n$  but fails for larger orders when benchmarked on a 1D Heisenberg model. We therefore propose an improved “conditional sampling” method exploiting the autoregressive structure of the network ansatz, which outperforms direct sampling in both 1D and 2D Heisenberg models. The high order Rényi entropies calculated with conditional sampling allow for extraction of the largest eigenvalue of the reduced density matrix, and thus the single copy entanglement.

## 1 Introduction

Quantum entanglement is a fundamental property underpinning a diverse range of phenomena in condensed matter and gravitational systems. Entanglement entropy quantifies the entanglement across a cut in these systems, and reveals emergent behavior such as topological order [1; 2; 3] and quantum phase transitions [4; 5; 6]. The Rényi entropies

$$S_n(\rho_A) = \frac{1}{1-n} \ln \text{Tr} [\rho_A^n], n > 0, \quad (1)$$

are a generalization of the well-known von Neumann entropy. Here,  $\rho_A$  is the reduced density matrix for a bipartition of a pure state  $|\psi\rangle \in \mathcal{H}_A \otimes \mathcal{H}_B$  into subsystems  $A$  and  $B$ . The full set of Rényi entropies encode the entanglement spectrum [7; 8], whose low-lying eigenvalues dominate at high order  $n$  and may serve as order parameters [9; 10], or yield directly the single copy entanglement  $S_\infty$  [11].

Numerical methods have been developed to estimate entanglement entropy for quantum many-body systems because exact calculations are usually not feasible in the exponentially large Hilbert space. Such methods are often based on variational ansätze that represent quantum states with only polynomial number of parameters in the system size. Tensor networks work well for 1D systems and provide direct access to the entanglement spectrum [12; 13], but suffer from the #P-hard problem of exact tensor contraction in higher dimensions [14]. By contrast, Quantum Monte Carlo (QMC) techniques measure physical observables comparatively better in high dimensions, but entanglement entropy is only accessible for integer Rényi order  $n \geq 2$  via the replica trick [15].

Neural networks have recently been introduced as state ansätze in variational QMC and have successfully been trained to represent many-body ground states and to reconstruct quantum states from experimental data, both in 1D and higher dimensions [16; 17; 18; 19]. However, numerical study of entanglement entropy for neural network quantum states (NQS) has received limited attention,

---

\*Equal contribution

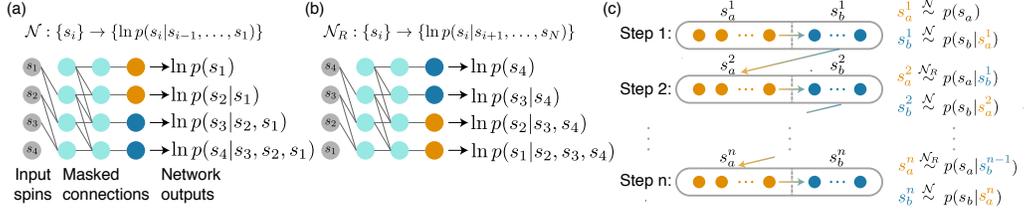


Figure 1: (a) The autoregressive network  $\mathcal{N}$  representing the probability distribution  $p(s)$  is bi-partitioned into subsystems  $A$  (orange) and  $B$  (blue). (b) The reverse net  $\mathcal{N}_R$  representing the same probability distribution  $p(s)$  is trained with a flipped conditional ordering. (c) Schematic illustration of conditional sampling sequence.

and only the Rényi entropy  $S_2$  has been computed for restricted Boltzmann machine (RBM) architectures [20]. To take advantage of state-of-the-art progress in machine learning, deeper and more complicated neural network architectures have been introduced as ansätze, but their potential for efficient entropy estimation has not yet been explored.

Here, we estimate generalized Rényi entropies  $S_n$  for an autoregressive neural network representation of a quantum state using both “direct sampling” and an improved “conditional sampling” method that exploits the autoregressive structure of the network ansatz. Benchmarking with 1D and 2D Heisenberg models shows that the conditional sampling outperforms direct sampling for large Rényi order  $n$ . The significant variance reduction from conditional sampling allows access to Rényi entropies for  $n > 30$  as well as the largest eigenvalue of the reduced density matrix  $\rho_A$ .

## 2 Methods

### 2.1 Neural Autoregressive Quantum States

The wavefunction of a  $N$ -spin quantum state in the computational basis  $s = \{s_1, \dots, s_N\}$ ,  $s_i = \pm 1$  can be decomposed as  $\psi(s) = \sqrt{p(s)}e^{i\phi(s)}$ , where  $p(s)$  gives the probability for each spin configuration  $s$  and  $\phi(s)$  describes the corresponding phase factor. Inspired by the work of Sharir et al. [21] and Wu et al. [22], we choose an autoregressive network  $\mathcal{N}$  to model  $p(s)$ , and train a separate fully-connected network for the phase; together, these comprise our neural autoregressive quantum state (NAQS). To represent a many-body ground state, the network parameters can be trained by minimization of the energy [16; 21] if the Hamiltonian is sparse in the computational basis.

Autoregressive networks were first introduced in the machine learning community as generative models which express a high-dimensional joint probability distribution as a product of conditional probabilities  $p(s_1, \dots, s_N) = \prod_{i=1}^N p(s_i | s_{i-1}, \dots, s_1)$  [23; 24]. Similarly, our probability network  $\mathcal{N}$  takes a spin configuration  $s$  as input and outputs the logarithm of  $N$  conditional probabilities (Fig. 1(a)), which are summed to obtain  $\ln[p(s)]$ . Having these conditionals allows efficient generation of i.i.d. samples  $\bar{s} = (\bar{s}_1, \dots, \bar{s}_N)$  from the state distribution by sequentially drawing  $\bar{s}_1 \sim p(s_1)$ ,  $\bar{s}_2 \sim p(s_2 | \bar{s}_1)$ , ...,  $\bar{s}_N \sim p(s_N | \bar{s}_{N-1}, \dots, \bar{s}_1)$ . Compared with Metropolis sampling with Markov chain Monte Carlo required for e.g. RBMs [16], sampling from  $\mathcal{N}$  is both faster and guaranteed to be i.i.d..

### 2.2 Direct Sampling

Once the network is trained to represent the desired target state  $\psi$ , samples from  $p(s)$  can be used to estimate the Rényi entropies for integer  $n \geq 2$  with the replica trick [6; 15]. More explicitly, the quantity  $\text{Tr}[\rho_A^n]$  can be computed as

$$\text{Tr}[\rho_A^n] = \sum_{\{s_a^i, s_b^i\}} \prod_{i=1}^n \psi(s_a^i, s_b^i) \prod_{i=1}^n \psi^*(s_a^{i+1}, s_b^i) \equiv \sum_{\mathbf{s}_a, \mathbf{s}_b} \Omega(\mathbf{s}_a, \mathbf{s}_b), \quad (2)$$

where we have defined  $\Omega(\mathbf{s}_a, \mathbf{s}_b) \equiv \prod_{i=1}^n \psi(s_a^i, s_b^i) \prod_{i=1}^n \psi^*(s_a^{i+1}, s_b^i)$ ,  $s_a^{n+1} \equiv s_a^1$ , and  $\mathbf{s}_{a(b)} \equiv \{s_{a(b)}^i, i = 1, \dots, n\}$  with each  $s_{a(b)}^i$  representing a sum over all basis vectors in  $\mathcal{H}_{A(B)}$ .

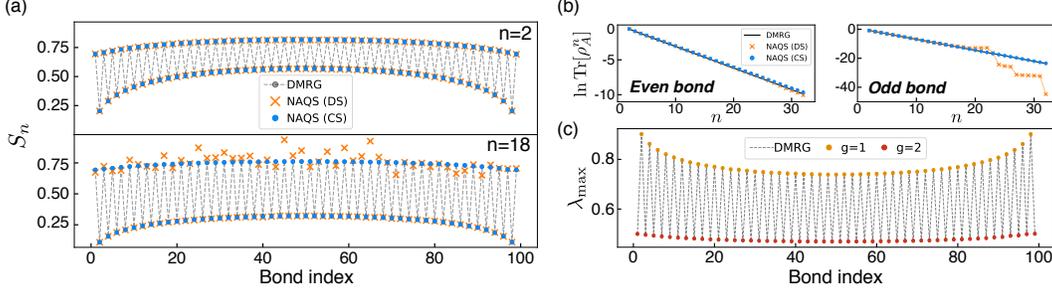


Figure 2: (a) Rényi entropies  $S_2$  (top) and  $S_{18}$  (bottom) for the ground state of the 1D AFM Heisenberg model with 100 spins and open boundary conditions. For small Rényi order  $n = 2$ , direct sampling (orange crosses) and conditional sampling (blue dots) both match the DMRG results (gray dots; gray dashed line is a guide to the eye). For larger Rényi order  $n = 18$ , direct sampling becomes comparatively noisier at the odd bonds. (b)  $\ln \text{Tr}[\rho_A^n]$  for representative even bond index 50 and odd bond index 51 for integer Rényi orders  $2 \leq n \leq 32$ . The conditional sampling consistently gives close results to DMRG while direct sampling gets worse for larger  $n$  at odd bonds. (c) Largest eigenvalue  $\lambda_{\max}$  and its degeneracy  $g$  extracted from fitting the entropy data; both agree well with the exact values. Number of samples for all plots (a-c) are about 50,000 for conditional sampling and 100,000 for direct sampling.

Using a straightforward direct sampling (DS) scheme, Eq. 2 can be estimated in QMC as  $\text{Tr}[\rho_A^n] = \langle f_{DS} \rangle \equiv \langle \Omega / P_{DS} \rangle_{(s_a, s_b) \sim P_{DS}}$ , with

$$P_{DS}(s_a, s_b) = \prod_{i=1}^n p(s_a^i, s_b^i). \quad (3)$$

For each Monte Carlo step,  $n$  samples  $\{(\bar{s}_a^i, \bar{s}_b^i), i = 1, \dots, n\}$  are drawn independently from the state distribution  $p(s_a, s_b)$ , then permuted as  $\{(\bar{s}_a^{i+1}, \bar{s}_b^i), i = 1, \dots, n\}$  to evaluate the estimator  $f_{DS}$ . Our benchmarking shows that  $f_{DS}$  usually has a large variance that worsens with higher Rényi order  $n$ . A “ratio trick” may be applied to control the variance of the entropy estimator [15], but it requires the valence bond basis to make sure all relevant weights are positive which is not true for the NAQS ansatz.

### 2.3 Conditional Sampling

To solve this variance problem and access higher order Rényi entropies, we propose an improved conditional sampling (CS) method, which generates a batch of correlated rather than independent samples. Now, the sum in Eq. 2 is estimated as  $\text{Tr}[\rho_A^n] = \langle f_{CS} \rangle \equiv \langle \Omega / P_{CS} \rangle_{(s_a, s_b) \sim P_{CS}}$ , with

$$P_{CS}(s_a, s_b) = p(s_a^1)p(s_b^1|s_a^1)p(s_a^2|s_b^1)p(s_b^2|s_a^2) \cdots p(s_b^n|s_a^n). \quad (4)$$

Iteratively, after drawing  $\bar{s}_a^i$ , we sample  $\bar{s}_b^i \sim p(s_b|\bar{s}_a^i)$  followed by  $\bar{s}_a^{i+1} \sim p(s_a|\bar{s}_b^i)$ , which generates the sample sequence  $\bar{s}_a^1 \rightarrow \bar{s}_b^1 \rightarrow \bar{s}_a^2 \rightarrow \bar{s}_b^2 \rightarrow \cdots \rightarrow \bar{s}_a^n \rightarrow \bar{s}_b^n$  (Fig. 1(c)). Assuming the predetermined sampling order of the probability network  $\mathcal{N}$  is  $s_a \rightarrow s_b$ , then sampling in the other direction  $s_b \rightarrow s_a$  requires a “reverse network”  $\mathcal{N}_R$  which models the same probability distribution  $p(s)$  as  $\mathcal{N}$ , but outputs conditionals in the reverse order (Fig. 1(b)). An analytic derivation of the parameters for  $\mathcal{N}_R$  is difficult, so we instead train it as a separate autoregressive network by minimizing its KL divergence with  $\mathcal{N}$ .

## 3 Results and Discussion

### 3.1 1D Heisenberg Model

We first benchmark the sampling methods on a network trained to find the ground state of a 1D antiferromagnetic (AFM) Heisenberg model  $H = \sum \vec{S}_i \cdot \vec{S}_{i+1}$  for  $N = 100$  spins with open boundary conditions. After training the NAQS representation by minimizing  $\langle H \rangle$ , we draw independent samples

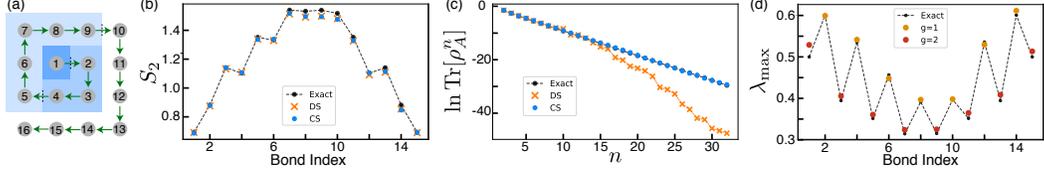


Figure 3: (a) Schematic for spiral mapping from 2D grid to 1D string. (b) Rényi entropy  $S_2$  for the ground state of the 2D AFM Heisenberg model on a  $4 \times 4$  grid with periodic boundary conditions. DS (orange crosses) and CS (blue dots) data both match exact values (black); the error comes mainly from the infidelity of the NAQS ansatz. (c)  $\ln \text{Tr}[\rho_A^n]$  at bond index 10 for integer Rényi orders  $2 \leq n \leq 32$ . The DS data (orange crosses) is about an order of magnitude worse than the CS data (blue dots) at large  $n$ . (d) The maximum eigenvalue  $\lambda_{\max}$  and degeneracy  $g$  are extracted from a linear fit to the CS data for  $\ln \text{Tr}[\rho_A^n]$ , with  $g$  restricted to an integer value. Number of samples for DS and CS data in (b-d) is 150,000.

$\{(s_a^i, s_b^i), i = 1, \dots, n\}$  from  $P_{\text{DS}}$  to estimate  $\text{Tr}[\rho_A^n]$  using direct sampling. As shown in Figure 2(a), the direct sampling method (orange crosses) yields accurate results for  $n = 2$  compared with an exact DMRG computation (gray dots with dashed line) [25]. However, for  $n = 18$  the entropy estimate across “odd bonds”, which partition the system such that  $A$  and  $B$  have an odd number of spins, starts to show larger variance. Figure 2(b) shows that the agreement between direct sampling and DMRG is consistently close at even bonds but worsens by an order of magnitude at odd bonds for large  $n$ .

We then use conditional sampling to estimate the entropy, which indeed removes the larger variance at higher  $n$ . As shown in Figure 2(b), we find good agreement with DMRG up to  $n = 32$ , and the error mainly comes from the infidelity of the NAQS state itself. We should emphasize that the success of conditional sampling is closely related to the existence of large classical mutual information between regions  $A$  and  $B$ , which is indeed true for the singlet ground state of the AFM Heisenberg model.

In the limit as Rényi order  $n \rightarrow \infty$ , the main contribution to the entropy comes from the largest eigenvalue  $\lambda_{\max}$  of  $\rho_A$ , which yields the single copy entanglement  $S_\infty = -\ln \lambda_{\max}$  [11]. From a linear fit to the CS data for  $\ln \text{Tr}[\rho_A^n]$  in the range  $10 \leq n \leq 32$ , we extract  $\lambda_{\max}$  from the slope and its degeneracy  $g \in \mathbb{N}$  from the intercept (Fig. 2(b)). The results are plotted in Figure 2(c); the differences between the extracted  $\lambda_{\max}$  and DMRG results are within 0.005. Due to the  $\text{SU}(2)$  symmetry of the singlet ground state, the largest eigenvalue  $\lambda_{\max}$  of  $\rho_A$  has degeneracy  $g = 2$  or  $1$  at the odd or even bonds respectively. The marker colors in Figure 2(c) indicate the fitted degeneracies, which also match the exact values.

### 3.2 2D Results

Finally, we benchmark entropy calculations for the ground state of the 2D Heisenberg model  $H = \sum_{\langle i,j \rangle} \vec{S}_i \cdot \vec{S}_j$  on a  $4 \times 4$  square lattice with periodic boundary conditions. Although the NAQS ansatz has an intrinsic 1D ordering corresponding to the decomposition of the joint probability distribution into conditionals, it can still efficiently represent quantum states in higher dimensions [22; 21], which by contrast would require infeasibly large bond dimension for 1D tensor network states. In order to estimate  $S_n$  for square regions of increasing area using conditional sampling, we choose a spiral ordering for the 2D grid (Fig. 3(a)). Figure 3(b) shows that the second Rényi entropy  $S_2$  for both DS and CS agrees well with exact diagonalization [26], with systematic error caused by the NAQS infidelity. As in the 1D system,  $\ln \text{Tr}[\rho_A^n]$  converges up to  $n = 32$  for CS but not for DS (Fig. 3(c)). The clean CS data allows the maximum eigenvalue of  $\rho_A$  and its degeneracy to be extracted from a linear fit (Fig. 3(d)).

## 4 Conclusion and Outlook

We have demonstrated a numerical calculation of Rényi entropies using NAQS up to  $n = 32$  as well as extraction of the largest eigenvalue of the reduced density matrix. Currently, we are testing our methods for larger system sizes to understand their scaling in 2D. In future work, better data quality will allow reconstruction of more eigenvalues, and thus the Schmidt gap [10]. Improvements to the

NAQS ansatz incorporating e.g. symmetries or different network architectures would be advantageous for training larger system sizes. More generally, it may prove fruitful to design importance sampling methods suitable for a given neural network structure, or to tailor network architectures to a particular sampling scheme for variance reduction.

## References

- [1] Alexei Kitaev and John Preskill. Topological entanglement entropy. Physical review letters, 96(11):110404, 2006.
- [2] Michael Levin and Xiao-Gang Wen. Detecting topological order in a ground state wave function. Physical review letters, 96(11):110405, 2006.
- [3] Steven T Flammia, Alioscia Hamma, Taylor L Hughes, and Xiao-Gang Wen. Topological entanglement rényi entropy and reduced density matrix structure. Physical review letters, 103(26):261601, 2009.
- [4] Tobias J Osborne and Michael A Nielsen. Entanglement in a simple quantum phase transition. Physical Review A, 66(3):032110, 2002.
- [5] Guifre Vidal, José Ignacio Latorre, Enrique Rico, and Alexei Kitaev. Entanglement in quantum critical phenomena. Physical review letters, 90(22):227902, 2003.
- [6] Pasquale Calabrese and John Cardy. Entanglement entropy and quantum field theory. Journal of Statistical Mechanics: Theory and Experiment, 2004(06):P06002, 2004.
- [7] Pasquale Calabrese and Alexandre Lefevre. Entanglement spectrum in one-dimensional systems. Physical Review A, 78(3):032329, 2008.
- [8] Hui Li and F Duncan M Haldane. Entanglement spectrum as a generalization of entanglement entropy: Identification of topological order in non-abelian fractional quantum hall effect states. Physical review letters, 101(1):010504, 2008.
- [9] R Thomale, A Sterdyniak, N Regnault, and B Andrei Bernevig. Entanglement gap and a new principle of adiabatic continuity. Physical review letters, 104(18):180502, 2010.
- [10] G De Chiara, L Lepori, M Lewenstein, and A Sanpera. Entanglement spectrum, critical exponents, and order parameters in quantum spin chains. Physical review letters, 109(23):237208, 2012.
- [11] Jens Eisert and Marcus Cramer. Single-copy entanglement in critical quantum spin chains. Physical Review A, 72(4):042112, 2005.
- [12] Guifré Vidal. Efficient classical simulation of slightly entangled quantum computations. Physical review letters, 91(14):147902, 2003.
- [13] Guifré Vidal. Efficient simulation of one-dimensional quantum many-body systems. Physical review letters, 93(4):040502, 2004.
- [14] Norbert Schuch, Michael M Wolf, Frank Verstraete, and J Ignacio Cirac. Computational complexity of projected entangled pair states. Physical review letters, 98(14):140506, 2007.
- [15] Matthew B Hastings, Iván González, Ann B Kallin, and Roger G Melko. Measuring renyi entanglement entropy in quantum monte carlo simulations. Physical review letters, 104(15):157201, 2010.
- [16] Giuseppe Carleo and Matthias Troyer. Solving the quantum many-body problem with artificial neural networks. Science, 355(6325):602–606, 2017.
- [17] Zi Cai and Jinguo Liu. Approximating quantum many-body wave functions using artificial neural networks. Physical Review B, 97(3):035116, 2018.
- [18] Giacomo Torlai, Guglielmo Mazzola, Juan Carrasquilla, Matthias Troyer, Roger Melko, and Giuseppe Carleo. Neural-network quantum state tomography. Nature Physics, 14(5):447, 2018.

- [19] Juan Carrasquilla, Giacomo Torlai, Roger G Melko, and Leandro Aolita. Reconstructing quantum states with generative models. Nature Machine Intelligence, 1(3):155, 2019.
- [20] Ivan Glasser, Nicola Pancotti, Moritz August, Ivan D Rodriguez, and J Ignacio Cirac. Neural-network quantum states, string-bond states, and chiral topological states. Physical Review X, 8(1):011006, 2018.
- [21] Or Sharir, Yoav Levine, Noam Wies, Giuseppe Carleo, and Amnon Shashua. Deep autoregressive models for the efficient variational simulation of many-body quantum systems. arXiv preprint arXiv:1902.04057, 2019.
- [22] Dian Wu, Lei Wang, and Pan Zhang. Solving statistical mechanics using variational autoregressive networks. Physical review letters, 122(8):080602, 2019.
- [23] Mathieu Germain, Karol Gregor, Iain Murray, and Hugo Larochelle. Made: Masked autoencoder for distribution estimation. In International Conference on Machine Learning, pages 881–889, 2015.
- [24] Aaron van den Oord, Nal Kalchbrenner, and Koray Kavukcuoglu. Pixel recurrent neural networks. arXiv preprint arXiv:1601.06759, 2016.
- [25] Johannes Hauschild and Frank Pollmann. Efficient numerical simulations with Tensor Networks: Tensor Network Python (TeNPy). SciPost Phys. Lect. Notes, page 5, 2018. Code available from <https://github.com/tenpy/tenpy>.
- [26] Phillip Weinberg and Marin Bukov. QuspIn: a python package for dynamics and exact diagonalisation of quantum many body systems part i: spin chains. SciPost Phys, 2(1), 2017.