Supplemental Materials: Continuous-mixture autoregressive networks learning the Kosterlitz-Thouless transition

Lingxiao Wang,[1,2,*] Yin Jiang,[3,†] Lianyi He,[2,‡] and Kai Zhou[1,§]

[1]*Frankfurt Institute for Advanced Studies, Ruth-Moufang-Str. 1, 60438 Frankfurt am Main, Germany*
[2]*State Key Laboratory of Low-Dimensional Quantum Physics and Department of Physics, Tsinghua University, Beijing 100084, China*
[3]*Department of Physics, Beihang University, Beijing 100191, China.*

(Dated: November 19, 2022)

I. SAMPLING USING AUTOREGRESSIVE NETWORKS

For a well-trained autoregressive network, it's direct to sample configurations from it, while the unavoidable numerically non-zero loss implies that the samples follow a distribution $q_\theta(\mathbf{s})$ which approximates but is not exactly the same as $p(\mathbf{s})$. To reach unbiased estimations of the thermodynamic observables, we adopt the importance sampling(IS) approach [1, 2], and constructs an unbiased estimator for each observable $O(\mathbf{s})$ as following,

$$\langle O\rangle \simeq \sum_{\mathbf{s}\sim q_\theta(\mathbf{s})} O(\mathbf{s})w(\mathbf{s}), \tag{1}$$

where the weight is $w(\mathbf{s}) = (p(\mathbf{s})/q_\theta(\mathbf{s}))/A$ with a normalization factor $A = \sum_{\mathbf{s}\sim q_\theta(\mathbf{s})} p(\mathbf{s})/q_\theta(\mathbf{s})$, and the configurations $\{\mathbf{s}\}$ sampled from the trained autoregressive network are following $q_\theta(\mathbf{s})$ which can be well evaluated. The above importance sampling corrected estimator can be directly derived by inserting the reference distribution of $q_\theta(\mathbf{s})$ into the evaluation of the observable, $\langle O\rangle = \int O(\mathbf{s})(p(\mathbf{s})/q_\theta(\mathbf{s}))q_\theta(\mathbf{s})d\mathbf{s}$.

After insertion the estimation can be viewed as a weighted integration with respect to $q_\theta(\mathbf{s})$, and the integral can naturally be evaluated with the summation of all weighted observables as Eq.(1) shown. In the main text, we denote calculations from the CANs together with importance sampling to be "CANs-IS", while use "CANs" to denote calculations with direct sampling purely from CANs.

II. TRAINING AND PARAMETER SET-UPS

Before more detailed comparisons, the information on the training process should be added here. We adopt the Adam algorithm as the optimizer with parameters $\epsilon = 10^{-8}, \beta_1 = 0.5, \beta_2 = 0.999$. The initial learning rate is set as $\eta = 10^{-3}$ and adjusts in *ReduceLROnPlateau* scheme. Besides, the annealing strategy for inverse temperature was also adopt with a rate of 0.998 at step-t, $\beta_t = \beta * (1-0.998^t)$. The gradient clip was also applied to prevent exploding gradients. The stopping condition should be when variational free energy is converging, but in our practical computations, they all converge before epoch$= 10,000$. Thus, we trained the CANs at different inverse temperatures with 10,000 steps.

The advantage of the variational CANs is that the free energy per site can be directly estimated. It is presented Fig. 1 for three different lattice sizes, $L = 4, 8, 16$. The results for $L = 8$ and $L = 16$ indicate that the free energy converges rapidly with increasing lattice sizes, which ensures that the size effect can be avoided.

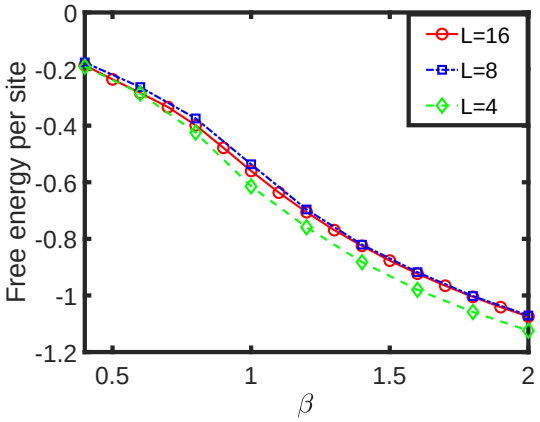

FIG. 1: The variational free energy F_θ per site on a square lattice from CANs.

For the MCMC calculations mentioned in the main text, the equilibration steps is set to be 50,000 for one Markov Chain, and the production steps is 50,000,000 from where we take sample per 10,000 steps. We adopt the classical Metropolis-Hastings algorithm to update the configurations which means the proposal for each step is that randomly rotating all spins iteratively, see more details in Refs [3, 4].

*Electronic address: lwang@fias.uni-frankfurt.de
†Electronic address: jiang_y@buaa.edu.cn
‡Electronic address: lianyi@mail.tsinghua.edu.cn
§Electronic address: zhou@fias.uni-frankfurt.de

TABLE I: The energy and the number of free vortices (or anti-vortices) per site extracted from CANs and MCMC. The results from CANs are obtained from an ensemble average with 1000 configurations.

	β	0.4	0.6	0.8	1.0	1.2
Energy	MCMC	**-0.424**	**-0.682**	**-0.996**	**-1.336**	**-1.502**
	CANs(2)	-0.231	-0.585	-1.165	-1.395	-1.513
	CANs	-0.386	-0.589	-1.037	**-1.358**	**-1.496**
	CANs+IS	**-0.423**	**-0.672**	**-1.022**	-1.387	-1.518
Vortices	MCMC	**0.114**	**0.080**	**0.042**	**0.010**	**0.002**
	CANs(2)	0.138	0.087	0.012	0.001	0.000
	CANs	0.121	0.094	0.033	0.005	0.001
	CANs+IS	**0.114**	**0.081**	**0.040**	**0.008**	**0.002**

Bibliography

[1] T. Müller, B. McWilliams, F. Rousselle, M. Gross, and J. Novák, ACM Transactions on Graphics (TOG) **38**, 1 (2019).

[2] K. Nicoli, P. Kessel, N. Strodthoff, W. Samek, K.-R. Müller, and S. Nakajima, arXiv preprint arXiv:1903.11048 (2019).

[3] J. Tobochnik and G. V. Chester, Phys. Rev. B **20**, 3761 (1979).

[4] S. Teitel and C. Jayaprakash, Phys. Rev. B **27**, 598 (1983).