Wave-function positivization via automatic differentiation

Giacomo Torlai [●],[1] Juan Carrasquilla,[2,3] Matthew T. Fishman,[1] Roger G. Melko,[3,4] and Matthew P. A. Fisher[5]

[1]*Center for Computational Quantum Physics, Flatiron Institute, New York, New York 10010, USA*
[2]*Vector Institute for Artificial Intelligence, MaRS Centre, Toronto, Ontario, Canada M5G 1M1*
[3]*Department of Physics and Astronomy, University of Waterloo, Ontario, Canada N2L 3G1*
[4]*Perimeter Institute for Theoretical Physics, Waterloo, Ontario, Canada N2L 2Y5*
[5]*Department of Physics, University of California, Santa Barbara, California 93106, USA*

We introduce a procedure to systematically search for a local unitary transformation that maps a wave function with a nontrivial sign structure into a positive-real form. The transformation is parametrized as a quantum circuit compiled into a set of one- and two-qubit gates. We design a cost function that maximizes the average sign of the output state and removes its complex phases. The optimization of the gates is performed through automatic differentiation algorithms, widely used in the machine learning community. We provide numerical evidence for significant improvements in the average sign for a two-leg triangular Heisenberg ladder with next-to-nearest-neighbor and ring-exchange interactions. This model exhibits phases where the sign structure can be removed by simple local one-qubit unitaries, but also an exotic Bose-metal phase whose sign structure induces "Bose surfaces" with a fermionic character and a higher entanglement that requires deeper circuits.

I. INTRODUCTION

The most striking contrast between the classical and quantum world is the fact that quantum wave functions contain "probability" amplitudes that are not strictly real and positive. This so-called *sign* (or phase) structure is an essential feature of a variety of quantum phenomena with no classical counterpart, such as the Pauli exclusion principle, entanglement, and quantum interference. It lies at the heart of any algorithm for quantum computing [1].

A sign structure often hinders the simulation of quantum many-body states by means of classical resources, and it essentially defines the threshold for what can be considered truly quantum mechanical. Indeed, there is a one-to-one mapping between a real, *non-negative* wave function and a classical probability distribution, formulated explicitly by the Born rule. However, the sign structure is not a universal feature of a quantum state, since it strongly depends on the choice of basis. As such, for a given state it is only natural to wonder: Is there a *local* change of basis that removes the sign structure, leading to a non-negative wave function?

Given a preferred "computational basis," finding and applying a change of basis involves implementing a unitary transformation. The resources required for this task can however be nontrivial. For example, any ground state becomes non-negative in the energy eigenbasis, but finding the

corresponding (nonlocal) unitary transformation is equivalent to diagonalization, with a complexity that scales exponentially in the number of qubits. The question becomes, can a change of basis be discovered with a transformation represented by a local unitary circuit of small depth?

Such transformations are typically identified based on simple physical principles related to the structure of the Hamiltonian and its symmetries. The most notable example is the Marshall sign rule [2], eliminating the sign structure from the ground states of quantum antiferromagnets on bipartite lattices. The resulting theoretical insight means that new bases that simplify the sign structure for a specific frustrated magnet or fermion model are routinely discovered [3–6]. In turn, in a few instances it has also been rigorously proven that efficient transformations do not exist [7,8], rendering the sign structure "intrinsic." However, if no obvious transformation is known, it is generally unclear whether the offending sign structure is intrinsic or whether it only persists due to a lack of physical insight. An automated procedure to search for relevant transformations is therefore highly desirable.

In this Rapid Communication, we propose an algorithm to tackle this question which combines tensor networks and differentiable programming. We formulate the search for the local basis as an optimization task over quantum circuits compiled into a set of local quantum gates. By optimizing a suitable cost function, a quantum circuit is used to positivize a quantum state with a sign structure. We show how this procedure can be realized in practice by adopting a tensor network representation of the quantum circuit, and applying automatic differentiation to obtain a "learning signal" for each quantum gate. We present a proof-of-principle demonstration for a two-leg triangular Heisenberg ladder with a four-spin ring-exchange interaction, which harbors a sign structure of
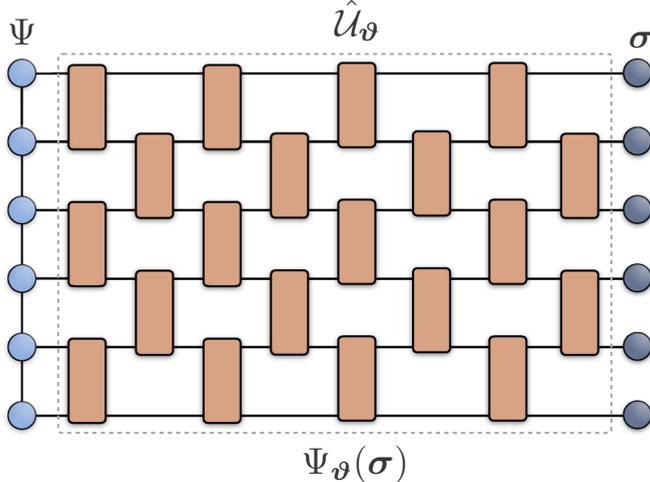

FIG. 1. Projection of the wave function on the basis state $\langle\boldsymbol{\sigma}|$ after the application of the quantum circuit $\hat{\mathcal{U}}_{\boldsymbol{\vartheta}}$ implementing the change of basis. Here, the circuit is compiled into a set of local two-qubit gates.

tunable complexity, including that of an exotic highly entangled spin Bose-metal phase.

II. LEARNING THE SIGN STRUCTURE

We study a system composed of N qubits described by a wave function $|\Psi\rangle$. For a given choice of basis of the many-body Hilbert space $|\boldsymbol{\sigma}\rangle = |\sigma_1, \ldots, \sigma_N\rangle$, we assume that the wave function has a sign structure, i.e., the coefficients $\Psi(\boldsymbol{\sigma}) = \langle\boldsymbol{\sigma}|\Psi\rangle$ appear with both positive and negative signs. We note that, while we restrict to real wave functions, the following approach identically applies to the case where the wave function is complex-valued.

Given the sign structure $\text{sign}[\Psi(\boldsymbol{\sigma})]$, how can we run an automated search for a local unitary transformation $\hat{\mathcal{U}}$ generating a non-negative wave function? For this purpose, it is natural to express the unitary as a quantum circuit, where locality can be imposed at the level of the quantum gates (Fig. 1). Because of their universality [1], we can restrict to single- and two-qubit gates acting on pairs of adjacent sites. Then, the unitary transformation is written in terms of parameters $\boldsymbol{\vartheta} = \{\boldsymbol{\vartheta}^{[1]}, \boldsymbol{\vartheta}^{[2]}, \ldots\}$, where $\boldsymbol{\vartheta}^{[k]}$ are a set of real and continuous parameters characterizing each single gate.

Starting from a wave function $\Psi(\boldsymbol{\sigma})$ displaying a sign structure, provided as input to the quantum circuit, the goal is to discover a set of gates such that the output state is non-negative. We choose to phrase this problem as an optimization task, where the non-negativity of the output state is enforced upon minimizing a suitable *cost function* $\mathcal{C}(\boldsymbol{\vartheta})$. More precisely, the optimal set of parameters $\boldsymbol{\vartheta}^* = \text{argmin}_{\boldsymbol{\vartheta}} \, \mathcal{C}(\boldsymbol{\vartheta})$ should satisfy $\Psi_{\boldsymbol{\vartheta}^*}(\boldsymbol{\sigma}) \geqslant 0 \; \forall|\boldsymbol{\sigma}\rangle$, where $\Psi_{\boldsymbol{\vartheta}^*}(\boldsymbol{\sigma}) = \langle\boldsymbol{\sigma}|\hat{\mathcal{U}}_{\boldsymbol{\vartheta}^*}|\Psi\rangle$. Given some initial configuration of the circuit, the optimization is solved by iteratively updating the gates according to the gradient of the cost function, $\boldsymbol{\vartheta} \to \boldsymbol{\vartheta} - \eta \, \mathcal{G}(\boldsymbol{\vartheta})$, where $\mathcal{G}(\boldsymbol{\vartheta}) = \boldsymbol{\nabla}_{\boldsymbol{\vartheta}} \mathcal{C}(\boldsymbol{\vartheta})$ and η is the step size of the update (often called *learning rate*). More sophisticated algorithms developed within the machine learning community can also be

implemented, such as the adaptive learning rates [9,10] or higher-order gradients [11].

The cost function is the most crucial ingredient. On one hand, it needs to correctly capture the objective of the optimization. On the other hand, the sign structure is a global property of the quantum state, and thus the calculation of the cost function (and its gradients) should also remain scalable with the number of qubits. For the latter, it is prudent to express $\mathcal{C}(\boldsymbol{\vartheta})$ as an expectation value over the probability distribution underlying the quantum state at the output of the circuit,

$$\mathcal{C}(\boldsymbol{\vartheta}) = \sum_{\boldsymbol{\sigma}} |\Psi_{\boldsymbol{\vartheta}}(\boldsymbol{\sigma})|^2 \mathcal{C}_{\boldsymbol{\vartheta}}(\boldsymbol{\sigma}). \tag{1}$$

In fact, provided one can sample the distribution $p_{\boldsymbol{\vartheta}}(\boldsymbol{\sigma}) = |\Psi_{\boldsymbol{\vartheta}}(\boldsymbol{\sigma})|^2$, the expectation value of Eq. (1) can be approximated with a sum over a finite number of configurations $\{\boldsymbol{\sigma}_j\}$ drawn from $p_{\boldsymbol{\vartheta}}(\boldsymbol{\sigma})$. Now, the only task that remains is designing an appropriate function $\mathcal{C}_{\boldsymbol{\vartheta}}(\boldsymbol{\sigma})$.

Besides the sign of the wave function, an additional constraint that should be taken into account is that the complex phases, necessarily accumulated by a universal gate set, are eliminated by the end of the unitary evolution. To capture both conditions on the imaginary part and the sign, it is convenient to split the cost function into a convex sum of two contributions:

$$\mathcal{C}_{\boldsymbol{\vartheta}}(\boldsymbol{\sigma}) = \gamma|\text{Im}[\Psi_{\boldsymbol{\vartheta}}(\boldsymbol{\sigma})]| - (1-\gamma)\text{sign}[\text{Re}[\Psi_{\boldsymbol{\vartheta}}(\boldsymbol{\sigma})]], \tag{2}$$

where $\gamma \in [0, 1]$. By tuning the parameters according to the gradient $\mathcal{G}(\boldsymbol{\vartheta})$, the quantum circuit will try to increase the sign of the real part of the wave function, while forcing the imaginary part to be zero. Note that the initial average sign can always be set to a positive value by an appropriate global transformation.

III. DIFFERENTIABLE PROGRAMMING

Next, in order to evaluate the gradients of the cost function we need to adopt a representation of the input quantum state and the quantum circuit amenable to scalable simulations. To this end, we assume that the initial state admits an efficient matrix product state (MPS) representation, and obtain the final state by contracting the MPS with the various gates in the circuit. At each intermediate step, provided the circuit depth is not too large, the quantum state can be restored into an MPS form by means of singular value decompositions.

The calculation of the gradients is the most involved step in the procedure, and analytical approaches would clearly be intractable. We leverage *automatic differentiation* (AD) techniques [12], routinely used in machine learning applications to train neural-network architectures [13] and recently applied to optimize tensor network states [14]. The core object in AD is the computational graph implementing the set of elementary computations (edges) acting on the variables (nodes). We specifically implement *reverse-accumulation* AD, where a forward pass first calculates the output of the graph, and derivatives are calculated starting from the output, and backpropagated through the graph using a sequence of Jacobianvector products.
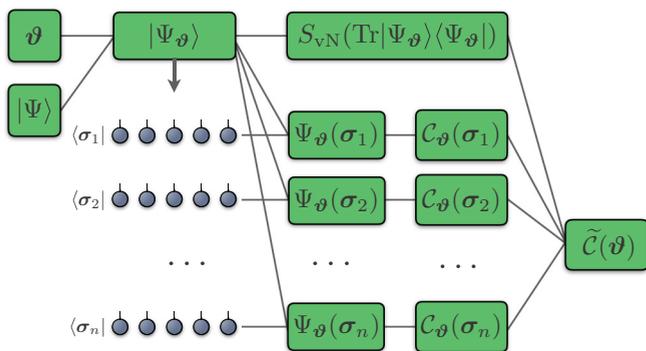
FIG. 2. Schematic of the computational graph for the calculation of the cost function. The output quantum state $|\Psi_\vartheta\rangle$, obtained by contracting the circuit tensor network, is sampled to generate the configurations $\{\sigma_j\}$, which are used to compute the cost function. In addition, the entanglement entropy of the output state is added as a regularization to mitigate the growth of entanglement in deep circuits.

The computational graph implementing the positivization is divided into three stages (Fig. 2). First, the circuit $\hat{\mathcal{U}}_\vartheta$ is applied to the input state through a series of tensor contractions. The resulting output quantum state $|\Psi_\vartheta\rangle = \hat{\mathcal{U}}_\vartheta|\Psi\rangle$ is then sampled to generate a set of n configurations $\{\sigma_j\}$ approximating the sum in Eq. (1). The projections of $|\Psi_\vartheta\rangle$ into these configurations are used to estimate the sample-wise cost function

$$\widetilde{\mathcal{C}}(\vartheta) = \frac{1}{n}\sum_{j=1}^{n}\mathcal{C}_\vartheta(\sigma_j) + \alpha S_{\text{vN}}(\hat{\rho}_A). \qquad (3)$$

Note that we have also added a term proportional to the entanglement entropy $S_{\text{vN}}(\hat{\rho}_A) = -\text{Tr}(\hat{\rho}_A \log \hat{\rho}_A)$, where $\hat{\rho}_A$ is the reduced density matrix for a equal bipartition of the qubits and α is a small weight. We introduce this type of regularization to the cost function to limit the growth of entanglement generated by the application of the gates, particularly relevant in the optimization of deep quantum circuits. Once the computational graph is compiled, the reverse-accumulation step evaluates the derivatives with respect to each gate parameter in the circuit (see Supplemental Material [15] for more details).

IV. RESULTS

We focus on the ground-state wave functions of a two-leg triangular ladder with the Hamiltonian

$$\hat{H} = J_1 \sum_j \hat{\mathbf{S}}_j \cdot \hat{\mathbf{S}}_{j+1} + J_2 \sum_j \hat{\mathbf{S}}_j \cdot \hat{\mathbf{S}}_{j+2}$$
$$+ \frac{J_r}{2}\sum_j \hat{\mathbf{P}}_{j,j+1,j+3,j+2} + \hat{\mathbf{P}}_{j,j+1,j+3,j+2}^\dagger, \qquad (4)$$

where $\hat{\mathbf{S}}_j$ are spin-1/2 operators. Here, the ring-exchange term corresponds to the cyclic exchange of spin states, $\hat{\mathbf{P}}_{i,j,k,l}|S_i^z, S_j^z, S_k^z, S_l^z\rangle = |S_l^z, S_i^z, S_j^z, S_k^z\rangle$, and the couplings are $J_1 = 1$, $J_2, J_r > 0$. The model in Eq. (4) exhibits a range of ground states with a sign structure of tunable complexity, so it serves as a representative test bed for our experiments.
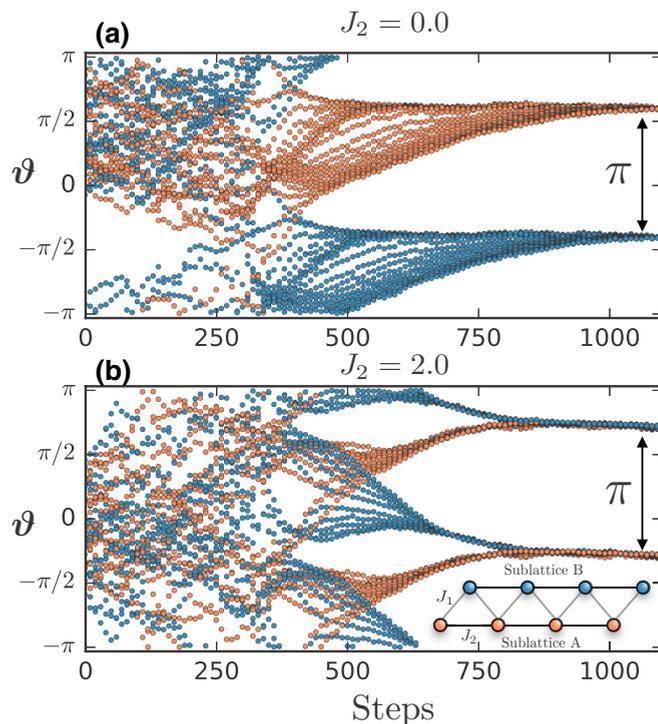

FIG. 3. Dynamics of the parameters ϑ during training for a ladder of $N = 40$ spins ($J_r = 0$) with (a) $J_2 = 0.0$ and (b) $J_2 = 2.0$. The quantum circuit contains only one-qubit rotations around the z axis.

Whereas for $J_1 = J_r = 0$ or $J_2 = J_r = 0$ the sign of the ground-state wave function can be eliminated via a unitary transformation acting on single spins [16], for $J_r/J_1 \gg 1$ and $J_2/J_1 \ll 1$ the model displays an exotic spin Bose-metal (SBM) phase endowed with a complex sign structure associated with the presence of singular wave vectors or "Bose surfaces" [17]. After obtaining the ground-state MPS using standard density matrix renormalization group techniques [18,19], we implement the AD graph using the machine learning library TENSORFLOW [20].

We first consider the case of $J_r = 0$, corresponding to the one-dimensional J_1-J_2 model. In the limit of $J_2 = 0$ we recover the Heisenberg model, where the sign structure of the ground state $\Psi(\mathbf{S}^z)$ in the Ising basis $|\mathbf{S}^z\rangle = |S_1^z, \ldots, S_N^z\rangle$ follows the Marshall sign rule [2,16]. The transformation removing this sign structure can be composed as a set of $N/2$ rotations of angle π about the z axis, corresponding to a depth-one quantum circuit. To check if this can be recovered by our procedure, we construct the variational quantum circuit $\hat{\mathcal{U}}_\vartheta$ using one layer of single-qubit rotations around the z axis. We run the positivization procedure for a chain containing $N = 40$ spins. After randomly initializing the circuit parameters (i.e., $N = 40$ angles) we train the circuit to minimize the cost function using $n_S = 10^3$ configurations sampled from the final MPS distribution $|\Psi_\vartheta(\mathbf{S}^z)|^2$ [21,22], and update the parameters using the Adam optimizer [10].

We show the behavior of the positivization algorithm in Fig. 3, where we plot the values of each single rotation angle as a function of the training iteration. For $J_2 = 0$ [Fig. 3(a)] we observe that all angles corresponding to rotations on sublattice

A (B) converge to the value $\vartheta^{[k]} = \pi/2$ ($-\pi/2$), equivalent to the Marshall sign rule. We then repeat the optimization for an initial ground state obtained by setting $J_2 = 2.0$ [Fig. 3(b)]. Here, we observe that the angles converge to two values separated by π, but now the rotations on sites from different sublattices are mixed together. It is easy to see that this circuit implements the Marshall sign rule in the limit of $J_1 = 0$, corresponding to two decoupled Heisenberg chains. In both cases we measure an average sign of about 0.99.

Although the relationship is not fully understood, the sign structure of a quantum state is related to its entanglement. For example, a typical random positive wave function exhibits a constant law for Renyi entanglement entropies with Renyi index $n > 1$, while states with Renyi entropy scaling as a volume law will have a complex sign structure [23], suggesting a nonlocal positivization transformation. It therefore stands to reason that circuits of large depths may be required to remove the sign structure when the entanglement needs significant modification.

In order to increase the entanglement of the starting state, we turn to the exotic spin Bose-metal (SBM) phase which contains significant entanglement due to the presence of a Bose surface [17]. We set $J_2 = 0$ and examine different initial ground-state MPSs obtained for $J_r \in [0, 1]$, which spans the phase transition into the SBM phase. We optimize circuits with different depths, where a single layer consists of a set of simultaneous commuting two-qubit gates (Fig. 1). In all simulations, the truncation error in the singular value decompositions performed to restore the MPS representation of the quantum state was kept below 10^{-6}.

We first examine a spin ladder with $N = 20$ sites, and optimize circuits of increasing depth for initial ground states obtained at different values of J_r. We plot the average sign (circles) and imaginary part (triangles) in Fig. 4(a). As expected, a larger depth systematically increases the effectiveness of the positivization, which becomes significantly harder as the system is driven into the SBM phase ($J_r \approx 0.6$). The transition in complexity is highlighted in Fig. 4(b), where we show the scalings with the system size for different values of J_r near the critical point for a circuit of fixed depth. In the Bethe phase (small J_r), the sign remains sufficiently high as the system size is increased, while the positivization becomes ineffective for larger N in the SBM phase. Finally, we show the scaling against the circuit depth for several sizes N in the two phases of the spin ladder [Figs. 4(c) and 4(d)]. The results confirm that in the SBM phase, in contrast to the Bethe phase, the depth required to achieve a given average sign increases with the number of spins N. In all instances, the optimization succeeds in producing quantum states with real coefficients to a good approximation.

V. CONCLUSIONS

We have introduced a procedure to systematically search for a local unitary transformation that maps a wave function with a sign structure into a non-negative form. The transformation is parametrized as a universal quantum circuit, and the gates are optimized through automatic differentiation algorithms, widely adopted in the machine learning community and implemented with TENSORFLOW [20]. We demonstrated
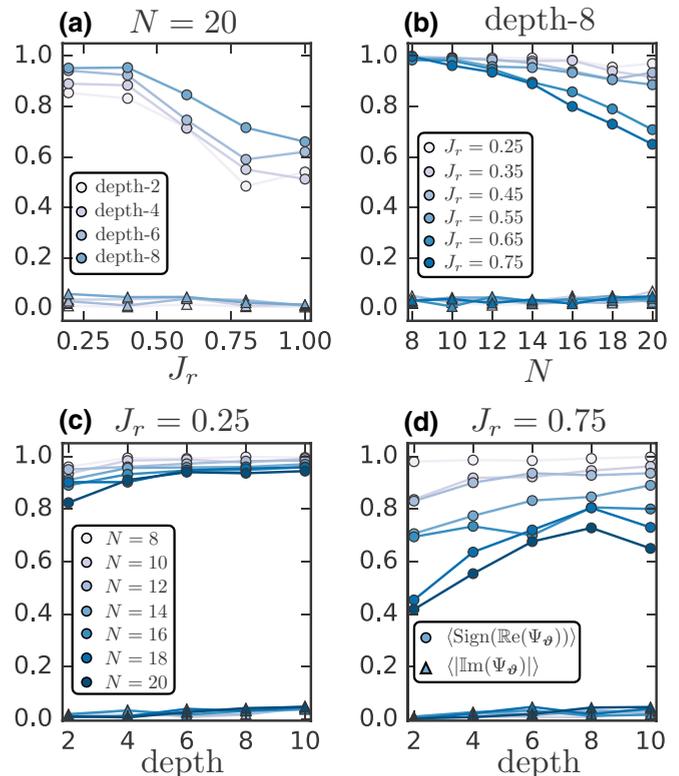

FIG. 4. Optimization performance for the two-leg triangular Heisenberg chain with (a) different depths as a function of J_r, (b) different values of J_r as a function of the number of spins, and different system sizes as a function of the depth for (c) $J_r = 0.25$ and (d) $J_r = 0.75$. The transition into the SBM phase occurs at $J_r \approx 0.6$.

this technique for ground states of a triangular spin ladder with Heisenberg interactions. For the limit of the J_1-J_2 model, we have shown that the optimization is capable of removing the sign structure, recovering the well-known Marshall sign rule. In the presence of a ring-exchange interaction, we observed that the SBM phase demands circuits where the depths scale with the size of the ladders.

The ability to discover a local basis where the average sign of a quantum state becomes substantially higher is particularly relevant for the alleviation of the sign problem in quantum Monte Carlo simulations [24–27]. In this context, our positivization algorithm could be repurposed to increase the "stoquasticity" of a target Hamiltonian [28]. This would require the optimization of a suitably modified cost function, where the input is a matrix product operator representation of the Hamiltonian. This opens interesting prospects for path integrals and projective quantum Monte Carlo simulations, which should be explored in future studies.

The non-negativity of a wave function in a local basis also has direct implications for the data-driven reconstruction of quantum states, which is becoming increasingly important for validating noisy-intermediate-scale quantum hardware [29]. In fact, for wave functions with positive amplitudes, experimental data from a single measurement basis is sufficient for the quantum reconstruction of the state, with a particularly favorable scaling with both the system size and the number of measurements [30,31].

Finally, our procedure provides a universal, automated method of assessing the complexity associated with the sign problem of a given wave function. Ultimately, by performing a systematic finite-size scaling analysis of the resources required to achieve a given average sign, this procedure could be used to determine the complexity class associated with removing the sign structure in various cases, including gapped, critical, or fermionic wave functions. In the future, automated numerical methods based on machine learning technology may be the most promising route to determining the relative "difficulty" of various sign structures, and will play a crucial role in formulating a complete theory relating a wave function's sign to its entanglement structure and simulation complexity.

ACKNOWLEDGMENTS

[1] G. Benenti, G. Casati, and G. Strini, *Principles of Quantum Computation and Information* (World Scientific, Singapore, 2004).

[2] W. Marshall and R. E. Peierls, Proc. R. Soc. London, Ser. A **232**, 48 (1955).

[3] Z.-X. Li, Y.-F. Jiang, and H. Yao, Phys. Rev. B **91**, 241117(R) (2015).

[4] R. K. Kaul, R. G. Melko, and A. W. Sandvik, Annu. Rev. Condens. Matter Phys. **4**, 179 (2013).

[5] S. Wessel, B. Normand, F. Mila, and A. Honecker, SciPost Phys. **3**, 005 (2017).

[6] A. Honecker, S. Wessel, R. Kerkdyk, T. Pruschke, F. Mila, and B. Normand, Phys. Rev. B **93**, 054408 (2016).

[7] M. B. Hastings, J. Math. Phys. **57**, 015210 (2016).

[8] Z. Ringel and D. L. Kovrizhin, Sci. Adv. **3**, e1701758 (2017).

[9] M. D. Zeiler, arXiv:1212.5701.

[10] D. P. Kingma and J. Ba, arXiv:1412.6980.

[11] S. Amari, in *Advances in Neural Information Processing Systems*, edited by M. I. Jordan, Y. LeCun, and S. A. Solla (MIT Press, Cambridge, MA, 1997), pp. 127–133.

[12] M. Bartholomew-Biggs, S. Brown, B. Christianson, and L. Dixon, J. Comput. Appl. Math. **124**, 171 (2000).

[13] D. E. Rumelhart, G. E. Hinton, and R. J. Williams, Nature (London) **323**, 533 (1986).

[14] H.-J. Liao, J.-G. Liu, L. Wang, and T. Xiang, Phys. Rev. X **9**, 031041 (2019).

[15] See Supplemental Material at http://link.aps.org/supplemental/10.1103/PhysRevResearch.2.032060 for details on the auto-

matic differentiation scheme used to optimize the quantum circuits.

[16] L. Capriotti, Int. J. Mod. Phys. B **15**, 1799 (2001).

[17] D. N. Sheng, O. I. Motrunich, and M. P. A. Fisher, Phys. Rev. B **79**, 205112 (2009).

[18] S. R. White, Phys. Rev. Lett. **69**, 2863 (1992).

[19] ITensor Library (version 2.0.11), http://itensor.org.

[20] M. Abadi *et al.*, TensorFlow: Large-scale machine learning on heterogeneous systems, 2015, software available from http://tensorflow.org.

[21] S. R. White, Phys. Rev. Lett. **102**, 190601 (2009).

[22] A. J. Ferris and G. Vidal, Phys. Rev. B **85**, 165146 (2012).

[23] T. Grover and M. P. A. Fisher, Phys. Rev. A **92**, 042308 (2015).

[24] M. Marvian, D. A. Lidar, and I. Hen, Nat. Commun. **10**, 1571 (2019).

[25] J. Klassen and B. M. Terhal, Quantum **3**, 139 (2019).

[26] D. Hangleiter, I. Roth, D. Nagaj, and J. Eisert, Sci. Adv. **6**, eabb8341 (2020).

[27] L. Gupta and I. Hen, arXiv:1910.13867.

[28] S. Bravyi, D. P. DiVincenzo, R. I. Oliveira, and B. M. Terhal, Quant. Inf. Comp. **8**, 0361 (2008).

[29] J. Preskill, Quantum **2**, 79 (2018).

[30] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Nat. Phys. **14**, 447 (2018).

[31] G. Torlai and R. G. Melko, Annu. Rev. Condens. Matter Phys. **11**, 325 (2020).