

Restricted Boltzmann machine learning for solving strongly correlated quantum systemsYusuke Nomura,^{1,*} Andrew S. Darmawan,¹ Youhei Yamaji,^{1,2} and Masatoshi Imada¹¹*Department of Applied Physics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*²*JST, PRESTO, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*

(Received 11 September 2017; revised manuscript received 14 November 2017; published 29 November 2017)

We develop a machine learning method to construct accurate ground-state wave functions of strongly interacting and entangled quantum spin as well as fermionic models on lattices. A restricted Boltzmann machine algorithm in the form of an artificial neural network is combined with a conventional variational Monte Carlo method with pair product (geminal) wave functions and quantum number projections. The combination allows an application of the machine learning scheme to interacting fermionic systems. The combined method substantially improves the accuracy beyond that ever achieved by each method separately, in the Heisenberg as well as Hubbard models on square lattices, thus proving its power as a highly accurate quantum many-body solver.

DOI: [10.1103/PhysRevB.96.205152](https://doi.org/10.1103/PhysRevB.96.205152)**I. INTRODUCTION**

Obtaining accurate ground-state wave functions of many-body quantum Hamiltonians is one of the grand challenges in condensed matter physics. Great successes so far are, for example, Bardeen-Cooper-Schrieffer (BCS) wave functions for conventional superconductivity [1], Bethe-ansatz wave function for one-dimensional interacting systems [2], and Laughlin wave functions for fractional quantum Hall effect [3].

However, in order to construct the ground-state wave functions of many-body interacting systems and grasp the essential physics encoded in them, we often need to resort to numerical estimates. Currently, many numerical techniques are available such as the variational Monte Carlo (VMC) method [4–9], the density matrix renormalization group [10,11], tensor network methods [12,13], and the path-integral renormalization group [14]. Among them, the VMC method offers an accurate ground-state wave function for quantum spins as well as fermions on various lattices.

Recently, alternative approaches, based on machine learning, have attracted growing attention in many-body physics [15–48]. In particular, Carleo and Troyer [49] have proposed a machine-learning algorithm, which uses a restricted Boltzmann machine (RBM) as a variational wave function $|\Psi\rangle$ for representing the ground states of quantum spin systems. In this scheme, hidden artificial neurons are introduced on top of the physical degrees of freedom (quantum spins), to mediate entanglement in the state. The RBM variational wave functions are self-optimized through machine learning.

We can express a general quantum state $|\Psi\rangle$ by using the Fock space basis $\{|x\rangle\}$ in the form of a variational function as

$$|\Psi\rangle = \sum_x |x\rangle \mathcal{F}(x) \langle x | \phi_{\text{ref}} \rangle \quad (1)$$

with a correlation factor $\mathcal{F}(x)$ [50,51] and a reference state $|\phi_{\text{ref}}\rangle$. The RBM wave function in Ref. [49] is obtained by employing an RBM for $\mathcal{F}(x)$ and a product state for $|\phi_{\text{ref}}\rangle$ so that $\langle x | \phi_{\text{ref}} \rangle = 1$ is satisfied for any orthonormalized complete set x . The product state $|\phi_{\text{ref}}\rangle$ is not able to describe nonlocal quantum entanglement, although it is essential in

strongly correlated systems. Then the entanglement has to be represented solely by the RBM factor $\mathcal{F}(x)$. However, alternative choices of $|\phi_{\text{ref}}\rangle$ may already incorporate typical quantum correlations and can potentially allow $|\Psi\rangle$ to more efficiently capture ground state entanglement. In fact, in the many-variable VMC (mVMC) method [9], a pair-product (PP) wave function (or equivalently geminal wave function in quantum chemistry [52–54]) is chosen as $|\phi_{\text{ref}}\rangle$, which can efficiently capture a substantial part of the nonlocal entanglement in strongly correlated quantum systems by using many variational parameters.

In this paper, we propose a variational wave function for studying strongly correlated quantum systems called RBM+PP which combines a flexible and nonempirical correlation factor given by RBM and entangled reference state given by the PP wave function to inherit the advantages of both. When applied to the two-dimensional (2D) Heisenberg model on a square lattice, we show that our method significantly outperforms the original RBM method [49], which itself outperforms existing numerical techniques for finite lattices based on tensor networks.

The PP wave function can also flexibly incorporate nonlocal correlations in fermionic systems and account for the fermionic sign, allowing RBM+PP to be applied to interacting systems of fermions. When applied to the Hubbard model we show that the combined method achieves greater accuracy than either method applied separately. The RBM+PP method thus provides a powerful tool not only for quantum spins but also for highly entangled quantum states such as strongly correlated itinerant fermions.

The structure of the paper is as follows. In Sec. II, we explain the RBM+PP method after introducing individual RBM and PP wave functions. In Sec. IV, we apply the RBM+PP scheme to Heisenberg and Hubbard models and show significant improvement from RBM and mVMC results. The representability of the RBM+PP wave function is discussed in detail in Sec. V. Finally, we give a summary and present future perspectives in Sec. VI.

II. RBM+PP METHOD

In this section, we will define RBM and PP states and explain how these are combined in the RBM+PP method.

*nomura@ap.t.u-tokyo.ac.jp

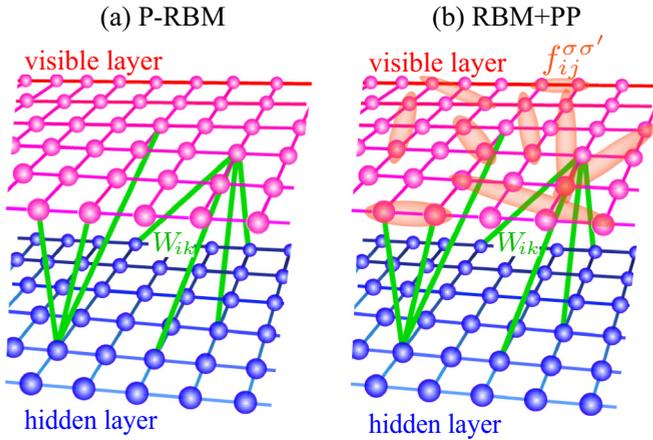


FIG. 1. Schematic illustration of (a) P-RBM and (b) RBM+PP to represent many-body wave function. Physical variables in the visible layer couple to artificial neurons in the hidden layer through W_{ik} interactions in Eq. (2). Whereas no entanglement among physical variables exists in the absence of the hidden layer in P-RBM, RBM+PP provides direct entanglement via $f_{ij}^{\sigma\sigma'}$ parameters in Eq. (3). For visibility, only a small portion of connections by W_{ik} and $f_{ij}^{\sigma\sigma'}$ are shown.

A. RBM wave function

The RBM state in Ref. [49] for spin Hamiltonians is given by setting $\mathcal{F}(x) = \mathcal{N}(x)$ with a neural-network correlation factor $\mathcal{N}(x)$ and $|\phi_{\text{ref}}\rangle$ to be the product state $|\phi_{\text{product}}\rangle$ ($\langle x|\phi_{\text{product}}\rangle = 1$) in Eq. (1), which from now we refer to as product-basis RBM (P-RBM) [Fig. 1(a)].

$\mathcal{N}(x)$ is defined by an artificial neural network (ANN) as

$$\mathcal{N}(x) = \sum_{\{h_k\}} \exp\left(\sum_i a_i \sigma_i + \sum_{i,k} W_{ik} \sigma_i h_k + \sum_k b_k h_k\right), \quad (2)$$

where $x = (\sigma_1, \sigma_2, \dots, \sigma_{N_{\text{visible}}})$ is a real space configuration of N_{visible} physical variables and σ_i is the i th discrete-valued physical variable (visible-layer spin variable). In the $S = \frac{1}{2}$ Heisenberg model, we take $\sigma_i = 2S_i^z = \pm 1$ with S_i^z being the z component of the $S = \frac{1}{2}$ spin at site i . Here N_{visible} is equal to the number of sites N_{site} . The auxiliary pseudospin variables $h_k = \pm 1$ are for the hidden neurons and $\{a_i, W_{ik}, b_k\}$ is a set of variational parameters. In this study, we take variational parameters to be real. Importantly, as there are no weights connecting hidden neurons, the sum over hidden variables can be evaluated exactly and Eq. (2) can be reduced to the form $\mathcal{N}(x) \equiv \prod_k 2 \cosh(b_k + \sum_i W_{ik} \sigma_i) \times e^{\sum_i a_i \sigma_i}$, which can be computed efficiently for each x .

In fermionic models we define a different RBM state which we refer to as F-RBM. The correlation factor $\mathcal{F}(x) = \mathcal{N}(x)$ is taken similarly but by slightly modifying the spin case. For $|\phi_{\text{ref}}\rangle$, our most primitive choice is a Fermi-sea state $|\phi_{\text{Fermi-sea}}\rangle$ rather than the product state. This is because the product state is too poor in representing fermionic entanglement. The Fermi sea is the ground state of the noninteracting fermion lattice models, much like product states are ground states of noninteracting spin models, and is thus able to account for the most primitive part of the fermionic entanglement and signs.

We can use the same form of $\mathcal{N}(x)$ as Eq. (2) by doubling the number of the visible-layer variables ($N_{\text{visible}} = 2N_{\text{site}}$) and mapping fermionic modes to these spins as $(\sigma_{2i}, \sigma_{2i-1}) = (2n_{i\uparrow} - 1, 2n_{i\downarrow} - 1)$, where $n_{i\sigma}$ is the number operator for the fermions at site i with spin σ .

As discussed in Ref. [49], the accuracy of the wave function can be controlled by the ‘‘hidden variable density’’ α , which is defined as $N_{\text{hidden}}/N_{\text{visible}}$ with the number of neurons N_{hidden} in the hidden layer and the number of physical variables N_{visible} in the visible layer.

B. PP wave function

Whereas the RBM states in the first step use the product and Fermi-sea states for $|\phi_{\text{ref}}\rangle$, more sophisticated choices of $|\phi_{\text{ref}}\rangle$ are able to incorporate more involved entanglement directly into the state. In this work we will use the pair-product (PP) state $|\phi_{\text{pair}}\rangle$ as $|\phi_{\text{ref}}\rangle$. The PP wave function is given by

$$|\phi_{\text{pair}}\rangle = \left(\sum_{i,j=1}^{N_{\text{site}}} \sum_{\sigma,\sigma'=\uparrow,\downarrow} f_{ij}^{\sigma\sigma'} c_{i\sigma}^\dagger c_{j\sigma'}^\dagger \right)^{N_e/2} |0\rangle, \quad (3)$$

where N_e is the number of electrons, $f_{ij}^{\sigma\sigma'}$ are variational parameters, and $c_{i\sigma}^\dagger$ is the operator creating a σ -spin electron at site i . For a given real space configuration x , $\phi_{\text{pair}}(x) = \langle x|\phi_{\text{pair}}\rangle$ can be expressed as the Pfaffian of a matrix, and can be calculated efficiently, much like $\mathcal{N}(x)$. Note that an accurate description of the node position is crucially important for fermionic wave function while the simple product state with the positive definite coefficients does not describe the node. In contrast, the Pfaffian wave function is able to optimize the nodal structure within the framework of the Pfaffian wave function. Therefore, the PP wave function can account for typical nonlocal entanglement not only in nonfrustrated spin systems but also in frustrated spin and fermionic systems.

In spin models, to prohibit the double occupation, PP wave function is supplemented by the Gutzwiller factor $\mathcal{P}_G^\infty = \prod_i (1 - n_{i\uparrow} n_{i\downarrow})$. This form of reference function $|\phi_{\text{ref}}\rangle = \mathcal{P}_G^\infty |\phi_{\text{pair}}\rangle$ is able to represent resonating valence bond (RVB) wave functions [55,56].

C. RBM+PP wave function

In this paper, we will study the combined wave function RBM+PP [Fig. 1(b)] with $\mathcal{F}(x) = \mathcal{N}(x)$ and $|\phi_{\text{ref}}\rangle = |\phi_{\text{pair}}\rangle$ (itinerant fermions) or $\mathcal{P}_G^\infty |\phi_{\text{pair}}\rangle$ (spins) in Eq. (1). We remark that the RBM+PP wave function is similar to that used in the mVMC method [9], except that the mVMC method uses an empirical form of $\mathcal{F}(x)$ instead of more flexible and unbiased neural-network factor $\mathcal{N}(x)$. Specifically, in the mVMC method, $\mathcal{F}(x)$ is given by $\mathcal{F}(x) = \langle x|\mathcal{P}_G \mathcal{P}_J|x\rangle$ with Gutzwiller \mathcal{P}_G (controlling double occupancy) [51] and Jastrow \mathcal{P}_J (for long-ranged charge-charge correlation) [50] factors. In Appendix A, we show that the neural-network factor $\mathcal{N}(x)$ is indeed more flexible than the empirical factors by showing that $\mathcal{N}(x)$ can represent both the Gutzwiller and Jastrow factors. The $\mathcal{N}(x)$ factor can also represent many-body (more than two-body) correlations [57] at the same time.

Various symmetries can be imposed on the wave function to improve accuracy and reduce computational cost [58]. In this study, as in Ref. [49], we impose translational symmetry

in the variational parameters $\{a_i, W_{ik}, b_k\}$ in the RBM. In the antiferromagnetic Heisenberg and half-filled Hubbard models, because $\sum_i \sigma_i = 0$ holds in the ground state, the translationally-invariant bias term $a_i = a$ becomes irrelevant. Therefore, we neglect it. Furthermore, as the ground states of the Hamiltonians considered have total spin $S = 0$ and momentum $K = 0$, we apply the projections onto these subspaces, respectively, $\mathcal{L}^{S=0}$ and $\mathcal{L}^{K=0}$, to $|\phi_{\text{pair}}\rangle$ or $\mathcal{P}_G^\infty |\phi_{\text{pair}}\rangle$ to improve accuracy. If we apply both $\mathcal{L}^{S=0}$ and $\mathcal{L}^{K=0}$, the reference state becomes $|\phi_{\text{ref}}\rangle = \mathcal{L}^{K=0} \mathcal{L}^{S=0} |\phi_{\text{pair}}\rangle$ for itinerant fermions and $|\phi_{\text{ref}}\rangle = \mathcal{L}^{K=0} \mathcal{L}^{S=0} \mathcal{P}_G^\infty |\phi_{\text{pair}}\rangle$ for spins.

D. Machine learning of variational parameters

The form of the wave function in Eq. (1) allows calculating physical quantities and derivatives with respect to variational parameters to be approximated efficiently using Markov chain Monte Carlo sampling over the probability distribution $p(x) = \langle \Psi | x \rangle \langle x | \Psi \rangle / \langle \Psi | \Psi \rangle$. We use a machine learning method (called stochastic reconfiguration in Ref. [8] and natural gradient in Refs. [59,60]) to optimize the variational parameters in the wave function with respect to the energy. The computational cost of the optimization scales as $\mathcal{O}(N_{\text{site}}^3)$ for RBM+PP, compared to $\mathcal{O}(\alpha N_{\text{site}}^2)$ for P-RBM [49]. Thus, the improved accuracy of RBM+PP over P-RBM comes at some additional computational cost. Details and comparisons of the variational wave functions we introduced in this section are listed in Appendix B.

III. MODELS

We apply the RBM+PP scheme to calculate the ground states of 2D $S = \frac{1}{2}$ antiferromagnetic (AFM) Heisenberg and 2D Hubbard models on the square lattice. Their Hamiltonians are defined as follows:

$$\mathcal{H}_{\text{Heisenberg}} = J \sum_{(i,j)} \mathbf{S}_i \cdot \mathbf{S}_j \quad (J > 0), \quad (4)$$

$$\mathcal{H}_{\text{Hubbard}} = -t \sum_{(i,j)\sigma} c_{i\sigma}^\dagger c_{j\sigma} + U \sum_i n_{i\uparrow} n_{i\downarrow}. \quad (5)$$

The sum over sites i, j is restricted to nearest-neighbor pairs. We take the exchange constant J and hopping t as an energy unit in each case. The onsite repulsion U controls the strength of correlation in the Hubbard model.

In our calculations, fully periodic boundary conditions are imposed for the Heisenberg model, while periodic (in the x direction) and antiperiodic (in the y direction) (P-AP) boundary conditions are imposed for the Hubbard model. Further details of the computation conditions are available in Appendix C.

IV. RESULTS

A. Heisenberg model

Figure 2 shows the RBM+PP ground-state energy for the 8×8 2D Heisenberg model compared to quantum Monte Carlo calculations using the stochastic series expansion (SSE-QMC) at sufficiently low temperature T of $1/T = 64$ [61], which gives practically exact ground state energy. For com-

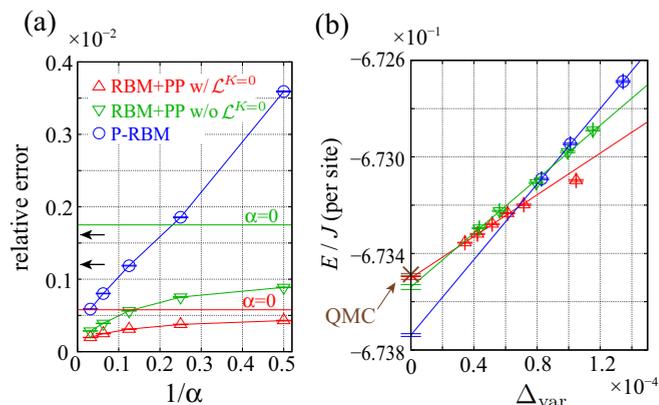


FIG. 2. RBM+PP results for energy of 2D AFM Heisenberg model defined on the 8×8 square lattice with fully periodic boundary condition. (a) Relative error of energy to SSE-QMC energy ($E/J = -0.673487(4)$) [61] as a function of $1/\alpha$ (α : hidden variable density). Δ (∇) symbol: RBM+PP $|\Psi\rangle = \mathcal{N} \mathcal{L}^{K=0} \mathcal{L}^{S=0} \mathcal{P}_G^\infty |\phi_{\text{pair}}\rangle$ ($|\Psi\rangle = \mathcal{N} \mathcal{L}^{S=0} \mathcal{P}_G^\infty |\phi_{\text{pair}}\rangle$). \circ symbol: P-RBM $|\Psi\rangle = \mathcal{N} |\phi_{\text{product}}\rangle$. $\alpha = 0$ (solid horizontal lines) corresponds to the mVMC results (red: $|\Psi\rangle = \mathcal{L}^{K=0} \mathcal{L}^{S=0} \mathcal{P}_G^\infty |\phi_{\text{pair}}\rangle$, green: $|\Psi\rangle = \mathcal{L}^{S=0} \mathcal{P}_G^\infty |\phi_{\text{pair}}\rangle$). Two arrows (from top to bottom) indicate the result of entangled-plaquette states (EPS) [62] and variational QMC to evaluate the projected entangled pair states (PEPS) for virtual bond dimensions of 16 [63]. (b) Variance Δ_{var} extrapolation of energy. The cross (\times) on the ordinate shows the SSE-QMC energy. The data points plotted in this variance range are $\alpha = 0, 2, 4, 8, 16, 32$ (from right to left) for red Δ symbols, $\alpha = 2, 4, 8, 16, 32$ for green ∇ symbols, and $\alpha = 8, 16, 32$ for blue \circ symbols, respectively. Linear fit and extrapolation to $\Delta_{\text{var}} \rightarrow 0$ is shown as solid lines. All the data points in this range except $\alpha = 0$ data in red are used in the fit. Error bars show standard errors of Monte Carlo measurements of energy [and also variance in case of (b)] for the optimized variational wave function.

parison, the mVMC results [9] and the P-RBM wave function employed in Ref. [49] are shown. Here, mVMC is equivalent to $\alpha = 0$ of RBM+PP. This is because the Gutzwiller factor \mathcal{P}_G is fixed to freeze charge degrees of freedom and Jastrow factor \mathcal{P}_J becomes irrelevant in the absence of charge degrees of freedom.

As discussed in Refs. [64,65], the resonating valence bond (RVB) wave function is known to provide a highly accurate description of the 2D Heisenberg model. The relative error in the result of the mVMC function, which can represent the RVB wave function, is indeed less than 0.2 percent. We see that the nonempirical P-RBM wave function is also powerful, giving a comparable accuracy to the mVMC results. The RBM+PP wave functions, which take advantages of the above two, substantially improve the accuracy of independent mVMC and P-RBM schemes.

It is interesting to note that each of the three curves in Fig. 2 corresponds to the results with the very same form of correlation factor $\mathcal{N}(x)$ but with different reference wave functions $|\phi_{\text{ref}}\rangle$. The lowest energy is obtained when $|\phi_{\text{ref}}\rangle = \mathcal{L}^{K=0} \mathcal{L}^{S=0} \mathcal{P}_G^\infty |\phi_{\text{pair}}\rangle$ followed by $|\phi_{\text{ref}}\rangle = \mathcal{L}^{S=0} \mathcal{P}_G^\infty |\phi_{\text{pair}}\rangle$, with the product state $|\phi_{\text{ref}}\rangle = |\phi_{\text{product}}\rangle$ having the highest energy. Therefore, improving reference function helps the RBM to learn the ground state more efficiently.

TABLE I. P-RBM and RBM+PP results for spin structure factor $S(\pi, \pi)$ in 2D Heisenberg model. For comparison, $S(\pi, \pi)/N_{\text{site}} = 0.05986(3)$ in mVMC and $0.059280(3)$ in SSE-QMC [61] results.

Wave function	$S(\pi, \pi)/N_{\text{site}} \times 10^2$		
	$\alpha = 2$	$\alpha = 8$	$\alpha = 32$
$\mathcal{N} \phi_{\text{product}}\rangle$	6.017(2)	5.955(2)	5.946(2)
$\mathcal{N}\mathcal{L}^{K=0}\mathcal{L}^{S=0}\mathcal{P}_G^\infty \phi_{\text{pair}}\rangle$	5.969(2)	5.956(2)	5.944(2)

In Fig. 2(b), we plot the total energy as a function of its variance $\Delta_{\text{var}} = (\langle \mathcal{H}^2 \rangle - \langle \mathcal{H} \rangle^2) / \langle \mathcal{H} \rangle^2$. The variance is zero in the case of an exact ground state (or more generally, an exact eigenstate of Hamiltonian). By the linear fit of the energy as a function of the variance and extrapolating to $\Delta_{\text{var}} = 0$, we can obtain a more accurate estimate of the ground state energy [8, 14, 66–68]. The variance extrapolation works better for the RBM+PP than the P-RBM wave function because the variance is already small. The accuracy of the extrapolated energy for RBM+PP wave functions with spin and momentum quantum projections $\mathcal{N}\mathcal{L}^{K=0}\mathcal{L}^{S=0}\mathcal{P}_G^\infty|\phi_{\text{pair}}\rangle$ [red line in Fig. 2(b)] reaches an order of 10^{-5} (0.001 percent) in the relative error, which is comparable to the size of error bars of SSE-QMC calculations [61].

The RBM+PP method can also be used to accurately calculate other physical quantities, besides the energy. For instance, we also measure spin structure factor $S(\mathbf{q}) = \frac{1}{3N_{\text{site}}} \sum_{i,j} \sum_{\alpha=x,y,z} \langle S_i^\alpha S_j^\alpha \rangle e^{i\mathbf{q}\cdot(\mathbf{r}_i - \mathbf{r}_j)}$. The result for $S(\mathbf{q}_{\text{peak}}) = S(\pi, \pi)$ is shown in Table I, indicating high accuracy of the correlation functions by the RBM+PP.

B. Hubbard model

Figures 3(a) and 3(b) show the RBM+PP result for the ground state energy of 8×8 Hubbard model at half filling for $U/t = 4$ and 8, respectively. The relative error to the auxiliary-field quantum Monte Carlo (AF-QMC), which gives practically the exact results within the error bars, is plotted. For comparison, the mVMC results using Gutzwiller-Jastrow correlation factors and the F-RBM results are shown as well.

In both cases ($U/t = 4$ and $U/t = 8$), the F-RBM has an error of several percent. The RBM+PP method, in which the variational parameters in both the \mathcal{N} and $|\phi_{\text{pair}}\rangle$ are optimized, achieves significantly higher accuracy. We see the same trend as the Heisenberg model; namely, the accuracy is improved by choosing a better reference wave function $|\phi_{\text{ref}}\rangle$.

The RBM+PP wave function also surpasses the accuracy of the mVMC wave function, which indicates superiority of more unbiased neural-network factor $\mathcal{N}(x)$ to the empirical Gutzwiller-Jastrow factors $\mathcal{P}(x) = \langle x | \mathcal{P}_G \mathcal{P}_J | x \rangle$. We expect that the advantage of a self-optimized neural network will be more substantial for more complicated Hamiltonians than the single-band Hubbard model. In more complex Hamiltonians, more flexible forms for correlation factors are likely to be necessary. Another advantage of the RBM+PP to the mVMC methods is that the accuracy improves as U increases (see Appendix D and Fig. 4), while the mVMC results show the opposite trend [9].

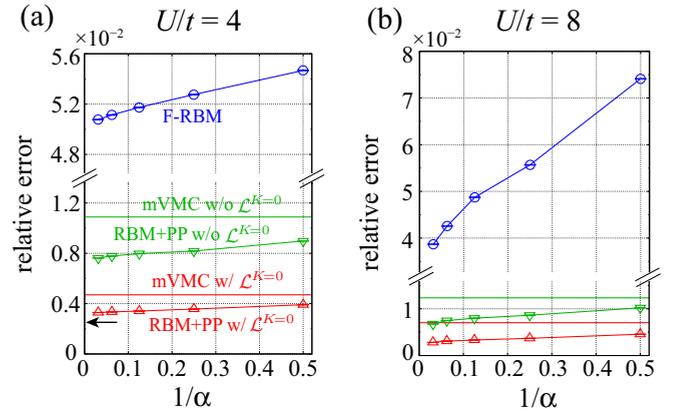


FIG. 3. RBM+PP result for energy of 2D Hubbard model at half filling with (a) $U/t = 4$ and (b) $U/t = 8$ on 8×8 square lattice with P-AP boundary condition. Relative error of the RBM+PP energy to the AF-QMC energy ($E/t = -0.8642(2)$ and $-0.5259(3)$ for $U/t = 4$ and $U/t = 8$, respectively) [69] as a function of $1/\alpha$ (α : hidden variable density) is shown. Δ (∇) symbol: RBM+PP $|\Psi\rangle = \mathcal{N}\mathcal{L}^{K=0}|\phi_{\text{pair}}\rangle$ ($|\Psi\rangle = \mathcal{N}|\phi_{\text{pair}}\rangle$). \circ symbol: F-RBM $|\Psi\rangle = \mathcal{N}|\phi_{\text{Fermi-sea}}\rangle$. Solid horizontal red (green) lines: results of mVMC functions $|\Psi\rangle = \mathcal{P}_G \mathcal{P}_J \mathcal{L}^{K=0}|\phi_{\text{pair}}\rangle$ ($|\Psi\rangle = \mathcal{P}_G \mathcal{P}_J |\phi_{\text{pair}}\rangle$). The arrow in (a) indicates the TNVMC result [70]. Error bars show standard errors of Monte Carlo measurements of energy for the optimized variational wave function.

At $U/t = 4$ [Fig. 3(a)], the best RBM+PP accuracy is comparable to the accuracy of TNVMC (tensor network combined with mVMC), where the relative error is ~ 0.25 percent as obtained at available maximum tensor bond dimension $D = 16$, which is the computationally practical upper limit [70]. In the TNVMC method, the mVMC wave function is supplemented by the tensor network enabling one of the most accurate schemes among existing numerical methods [70]. Compared to the TNVMC method, the RBM+PP has an advantage, because it can be applied flexibly and easily to any kind of lattice and does not require an involved contraction procedure [13] in contrast to the TNVMC method.

The spin structure factor $S(\pi, \pi)$ is listed in Table II. At both $U/t = 4$ and $U/t = 8$, with increasing α , the value becomes closer to the exact AF-QMC value.

V. DISCUSSION

While the physical properties of the RBM have only recently started being discussed in condensed matter physics

TABLE II. RBM+PP ($\mathcal{N}\mathcal{L}^{K=0}|\phi_{\text{pair}}\rangle$) results for spin structure factor $S(\pi, \pi)$ computed for 2D Hubbard model at half filling. For comparison, mVMC ($\mathcal{P}_G \mathcal{P}_J \mathcal{L}^{K=0}|\phi_{\text{pair}}\rangle$) and AF-QMC [71] results are also listed.

	$S(\pi, \pi)/N_{\text{site}} \times 10^2$				
	$\alpha = 2$	$\alpha = 8$	$\alpha = 32$	mVMC	AF-QMC
$U/t = 4$	3.078(5)	3.057(5)	3.021(5)	3.107(4)	2.92(2)
$U/t = 8$	5.233(9)	5.206(9)	5.20(1)	5.30(1)	5.0(1)

[72–80], more general discussion of representability can be traced back to earlier studies [81–83], which show that the RBM is able to describe any smooth function, if arbitrarily large α is allowed. In the present 2D Heisenberg model, a gauge transformation can make probability amplitude of the exact ground-state wave function $|\Psi_{\text{GS}}\rangle$ positive ($\langle x|\Psi_{\text{GS}}\rangle > 0$ for all x). Therefore, the exact ground state can be represented by real-variable RBM with infinite α . Accordingly, the relative error should go to zero as $1/\alpha \rightarrow 0$. Indeed, in Fig. 2(a), it looks that the RBM+PP energy curves (red and green) start bending toward 0 as $1/\alpha$ decreases. It is likely that a better reference state makes $\langle x|\Psi_{\text{GS}}\rangle/\langle x|\phi_{\text{ref}}\rangle$ [to be represented by $\mathcal{N}(x)$] smoother and helps to reach faster convergence at small α .

For fermionic problems, nodal structure of wave functions is crucial [84], which is beyond the representability of the real-variable RBM giving positive $\mathcal{N}(x)$ [79]. In the RBM+PP, $|\phi_{\text{pair}}\rangle$ is expected to accurately reproduce the nodal structure. Then, $\langle x|\Psi_{\text{GS}}\rangle/\langle x|\phi_{\text{pair}}\rangle$ may become smooth enough so that with moderate α , $\mathcal{N}(x)$ can represent a quick convergence to the exact value at $\alpha \rightarrow \infty$. However, rigorously speaking, the nodal structure of $|\phi_{\text{pair}}\rangle$ is likely to be different from the exact one even when the variational parameters contained in $|\phi_{\text{pair}}\rangle$ are ideally optimized. Therefore, introduction of complex variational parameters in the RBM part may be useful to adjust the nodes to the exact positions beyond the framework of the Pfaffian wave function. Although it is an interesting open question whether the energy curve as a function of $1/\alpha$ in the Hubbard model (Fig. 3) goes to 0 as $1/\alpha \rightarrow 0$, in practical computations, the optimization at larger α becomes more and more difficult, which might hamper the expected convergence.

VI. SUMMARY AND PERSPECTIVES

In this paper, we have proposed a variational ansatz for studying the ground states of many-body interacting quantum systems. Our variational wave function, which we call RBM+PP, combines the RBM based on the ANN and the mVMC methods. We have shown that, in both 2D Heisenberg and Hubbard models, the RBM+PP results show a dramatic improvement of accuracy over simple neural network wave functions (the P-RBM and F-RBM wave functions). We also see the superiority of the RBM+PP to the mVMC method. Since the RBM+PP method can be flexibly applied not only to bosonic (or spin) systems but also to fermionic problems, the RBM+PP method offers a wide range of applications with high accuracy and a reasonable computational cost.

As a future perspective, it would be interesting to go beyond the RBM structure and introduce a second hidden layer [deep Boltzmann machine (DBM)]. DBM is argued to have more efficient representation of certain many-body wave functions than RBM [75]. In DBM, the spin variables of the neurons distributed in more than one hidden layer cannot be traced out analytically, thus we need to introduce additional Monte Carlo samplings for hidden spins.

ACKNOWLEDGMENTS

We acknowledge useful discussions with G. Carleo, T. Ohgoe, H.-H. Zhao, and K. Ido. We also thank S. Zhang

for providing us with the auxiliary-field quantum Monte Carlo result of energy and spin structure factor in the 2D Hubbard model. The implementation of the RBM+PP scheme is done based on the mVMC package [85]. The computation was mainly done at Supercomputer Center, Institute for Solid State Physics, University of Tokyo. This work was financially supported by Grant-in-Aids for Scientific Research (JSPS KAKENHI) (No. 16H06345 and No. 17K14336) from Ministry of Education, Culture, Sports, Science and Technology (MEXT), Japan. This work was also supported in part by MEXT as a social and scientific priority issue [creation of new functional devices and high-performance materials to support next-generation industries (CDMSI)] to be tackled by using post-K computer. We also thank the support by the RIKEN Advanced Institute for Computational Science through the HPCI System Research project (hp160201, hp170263) supported by MEXT. Y.Y. was also supported by PRESTO, JST (JPMJPR15NF).

APPENDIX A: ABILITY OF RBM TO REPRESENT GUTZWILLER AND JASTROW FACTORS

The Gutzwiller factor $\mathcal{P}_G = \exp(-gn_{i\uparrow}n_{i\downarrow})$ at site i can be recast as (except for trivial constant factor and one-body potential)

$$\mathcal{P}_G = \exp\left(-\frac{g}{4}\sigma_{2i}\sigma_{2i-1}\right), \quad (\text{A1})$$

where $(\sigma_{2i}, \sigma_{2i-1}) = (2n_{i\uparrow} - 1, 2n_{i\downarrow} - 1)$ are physical variables in the visible layer defined in Sec. II A. This interaction between physical variables can be mediated by adding one hidden neuron h as (except for trivial constant factor)

$$\begin{aligned} \mathcal{P}_G &= \sum_{h=\pm 1} \exp(W_1\sigma_{2i}h + W_2\sigma_{2i-1}h) \\ &= 2 \cosh(W_1\sigma_{2i} + W_2\sigma_{2i-1}) \end{aligned} \quad (\text{A2})$$

with $W_1 = \frac{1}{2}\text{arcosh}(\exp(|g|/2))$ and $W_2 = -\text{sgn } g \times W_1$. This form is consistent with the neural-network factor \mathcal{N} defined in Eq. (2). In the very same way, we can show that the neural-network correlation factor \mathcal{N} can represent the Jastrow factor $\mathcal{P}_J = \exp(-\frac{1}{2}\sum_{i,j(i\neq j)} v_{ij}n_i n_j)$.

APPENDIX B: LIST OF VARIATIONAL WAVE FUNCTIONS EMPLOYED IN SIMULATIONS

Tables III and IV summarize the forms of the wave functions employed in the present study for solving the Heisenberg and Hubbard models, respectively.

APPENDIX C: DETAILS OF CALCULATION CONDITIONS

1. Stabilization factor

In the present study, the parameters are optimized by the stochastic reconfiguration (SR) method [8]. The same optimization scheme is called natural gradient in Refs. [59,60]. This optimization is equivalent to the imaginary-time evolution $e^{-\tau\mathcal{H}}|\Psi\rangle$ of the wave function $|\Psi\rangle$ to reach the ground state for sufficiently large imaginary time τ in the truncated Hilbert space which is spanned by the variational wave function. In the SR optimization, the variational parameters

TABLE III. List of wave functions used in the analysis of the 8×8 Heisenberg model ($N_{\text{site}} = 64$). To describe the singlet state, the σ and σ' spins in the $f_{ij}^{\sigma\sigma'}$ parameters in the PP wave function are always set to be a pair of \uparrow and \downarrow spins. We impose 2×2 sublattice structure in $f_{ij}^{\uparrow\downarrow}$ parameters in the PP (geminal) wave function. In this case, the number of independent $f_{ij}^{\uparrow\downarrow}$ parameters becomes $2 \times 2 \times N_{\text{site}} = 4N_{\text{site}}$, and the other $f_{ij}^{\uparrow\downarrow}$ parameters are determined by using spatial translation operations. The RBM part ($\{b_k, W_{ik}\}$) is taken to be fully translationally invariant (1×1 sublattice structure) [49]. When the double occupancy is completely prohibited by the Gutzwiller factor $\mathcal{P}_G^\infty = \prod_i (1 - n_{i\uparrow}n_{i\downarrow})$, the onsite $f_{ii}^{\uparrow\downarrow}$ parameters become completely irrelevant, i.e., the wave function and the energy do not depend on $f_{ii}^{\uparrow\downarrow}$ at all. Thus the number of $f_{ij}^{\uparrow\downarrow}$ parameters in the table are reduced from $4N_{\text{site}}$ to $4(N_{\text{site}} - 1)$.

Method	Wave function	Symbol in Fig. 2	Variational parameters	Number of variational parameters
P-RBM	$\mathcal{N} \phi_{\text{product}}\rangle$	open circle (blue)	$\{b_k, W_{ik}\}$	$\alpha(N_{\text{site}} + 1) = 65\alpha$
RBM+PP	$\mathcal{N}\mathcal{L}^{S=0}\mathcal{P}_G^\infty \phi_{\text{pair}}\rangle$	down-pointing triangle (green)	$\{b_k, W_{ik}, f_{ij}^{\uparrow\downarrow}\}$	$\alpha(N_{\text{site}} + 1) + 4(N_{\text{site}} - 1) = 65\alpha + 252$
RBM+PP	$\mathcal{N}\mathcal{L}^{K=0}\mathcal{L}^{S=0}\mathcal{P}_G^\infty \phi_{\text{pair}}\rangle$	up-pointing triangle (red)	$\{b_k, W_{ik}, f_{ij}^{\uparrow\downarrow}\}$	$\alpha(N_{\text{site}} + 1) + 4(N_{\text{site}} - 1) = 65\alpha + 252$
mVMC	$\mathcal{L}^{S=0}\mathcal{P}_G^\infty \phi_{\text{pair}}\rangle$	green solid horizontal line	$\{f_{ij}^{\uparrow\downarrow}\}$	$4(N_{\text{site}} - 1) = 252$
mVMC	$\mathcal{L}^{K=0}\mathcal{L}^{S=0}\mathcal{P}_G^\infty \phi_{\text{pair}}\rangle$	red solid horizontal line	$\{f_{ij}^{\uparrow\downarrow}\}$	$4(N_{\text{site}} - 1) = 252$

γ_m ($m = 1, 2, \dots, N_v$) at the p th iteration are updated as

$$\gamma_m^{(p+1)} = \gamma_m^{(p)} + \Delta\gamma_m^{(p)}, \quad (\text{C1})$$

where the difference in the update $\Delta\gamma_m^{(p)}$ is given by

$$\sum_{n=1}^{N_v} S_{mn}^{(p)} \Delta\gamma_n^{(p)} = -\Delta\tau g_m^{(p)} \quad (\text{C2})$$

with a small imaginary time step $\Delta\tau$. Here, S is a positive definite matrix given by

$$S_{mn} = \langle \partial_{\gamma_m} \bar{\Psi} | \partial_{\gamma_n} \bar{\Psi} \rangle \quad (\text{C3})$$

with a normalized variational wave function $|\bar{\Psi}\rangle = |\Psi\rangle / \sqrt{\langle \Psi | \Psi \rangle}$ and $|\partial_{\gamma_m} \bar{\Psi}\rangle = \frac{\partial}{\partial \gamma_m} |\bar{\Psi}\rangle$. The g vector is the gradient of energy with respect to the γ parameters:

$$g_m = \frac{\partial E}{\partial \gamma_m} = \frac{\partial}{\partial \gamma_m} \langle \bar{\Psi} | \mathcal{H} | \bar{\Psi} \rangle. \quad (\text{C4})$$

To stabilize the optimization, we introduce the stabilization factor to the diagonal elements of the S matrix as

$$S_{mm}^{(p)} \rightarrow S_{mm}^{(p)} (1 + \epsilon_1^{(p)}) + \epsilon_2^{(p)} \max_m \{S_{mm}^{(p)}\}. \quad (\text{C5})$$

Here, ϵ_1 scales the diagonal elements and ϵ_2 gives a constant shift to the S matrix. Although ϵ_2 strongly stabilizes the optimization, it sometimes makes convergence slower, so

we typically put a small number to ϵ_2 : $\epsilon_2^{(p=0)} \sim 10^{-3}$ and we gradually decrease to $\epsilon_2^\infty \sim 10^{-7} - 10^{-6}$ through the first several hundred iterations of optimization. As for ϵ_1 , we found that a smaller ϵ_1 ($\sim 10^{-5} - 10^{-4}$) factor for neural-network-related variational parameters $\{b_k, W_{ik}\}$ sometimes helps to lower the energy. Meanwhile, we also found that a small ϵ_1 ($\lesssim 10^{-3}$) makes the optimization of $f_{ij}^{\uparrow\downarrow}$ parameters in the pair-product (geminal) wave function unstable. To overcome this problem, we use parameter dependent ϵ_1 factor, namely, $\epsilon_1^{(p)}(\{f_{ij}^{\uparrow\downarrow}\}) = \epsilon_1^{(p)}(\{b_k, W_{ik}\}) + \Delta\epsilon_1$ with $\Delta\epsilon_1 \sim 10^{-2}$. With this condition, we typically use $\epsilon_1^{(p=0)}(\{b_k, W_{ik}\}) \sim 10^{-2} - 10^2$ and gradually decrease it to $\epsilon_1^\infty(\{b_k, W_{ik}\}) \sim 10^{-5} - 10^{-4}$ in the first several hundred iterations of the optimization.

As for the initial $\{b_k, W_{ik}\}$ parameters, we use small random numbers. We run several calculations with different seeds for generating random numbers and adopt the wave function that has the lowest energy.

2. Particle-hole transformation of Hubbard model

When we analyze many-body Hamiltonians by theoretical or numerical solvers, we can utilize transformations of the Hamiltonians to find a representation suitable for the solver in hand [86]. In the conventional variational Monte Carlo method, many-body interacting models defined in the real-space basis

TABLE IV. List of wave functions used in the analysis of the 8×8 Hubbard model ($N_{\text{site}} = 64$). As in the case of the Heisenberg model, the σ and σ' spins in the $f_{ij}^{\sigma\sigma'}$ parameters are set to be always antiparallel. In the mVMC method, we use Gutzwiller \mathcal{P}_G and Jastrow \mathcal{P}_J factors, whose forms are $\mathcal{P}_G = \exp(-\sum_i g_i n_{i\uparrow} n_{i\downarrow})$ and $\mathcal{P}_J = \exp(-\frac{1}{2} \sum_{i,j(i \neq j)} v_{ij} n_i n_j)$, respectively. All the g_i , v_{ij} , and $f_{ij}^{\uparrow\downarrow}$ parameters are taken to be independent (8×8 or full sublattice structure). On the other hand, the RBM part is taken to be translationally invariant (1×1 sublattice structure) [49] to save computational cost. We have confirmed that, in the 4×4 Hubbard model, taking full sublattice structure in the RBM part does not help much to lower the energy compared to the 1×1 sublattice case, although it drastically increases the number of variational parameters and hence increases the computational cost.

Method	Wave function	Symbol in Fig. 3	Variational parameters	Number of variational parameters
F-RBM	$\mathcal{N} \phi_{\text{Fermi-sea}}\rangle$	open circle (blue)	$\{b_k, W_{ik}\}$	$\alpha(2N_{\text{site}} + 1) = 129\alpha$
RBM+PP	$\mathcal{N} \phi_{\text{pair}}\rangle$	down-pointing triangle (green)	$\{b_k, W_{ik}, f_{ij}^{\uparrow\downarrow}\}$	$\alpha(2N_{\text{site}} + 1) + 64N_{\text{site}} = 129\alpha + 4096$
RBM+PP	$\mathcal{N}\mathcal{L}^{K=0} \phi_{\text{pair}}\rangle$	up-pointing triangle (red)	$\{b_k, W_{ik}, f_{ij}^{\uparrow\downarrow}\}$	$\alpha(2N_{\text{site}} + 1) + 64N_{\text{site}} = 129\alpha + 4096$
mVMC	$\mathcal{P}_G\mathcal{P}_J \phi_{\text{pair}}\rangle$	green solid horizontal line	$\{g_i, v_{ij}, f_{ij}^{\uparrow\downarrow}\}$	$N_{\text{site}} + 32(N_{\text{site}} - 1) + 64N_{\text{site}} = 6176$
mVMC	$\mathcal{P}_G\mathcal{P}_J\mathcal{L}^{K=0} \phi_{\text{pair}}\rangle$	red solid horizontal line	$\{g_i, v_{ij}, f_{ij}^{\uparrow\downarrow}\}$	$N_{\text{site}} + 32(N_{\text{site}} - 1) + 64N_{\text{site}} = 6176$

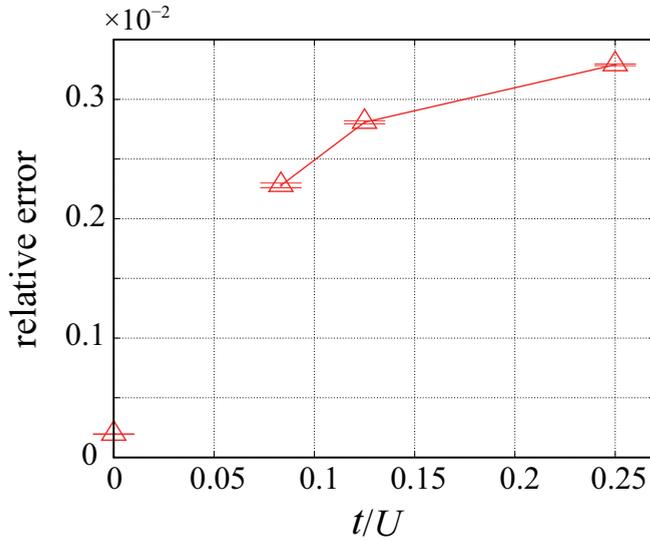


FIG. 4. U dependence of RBM+PP ($\mathcal{N}\mathcal{L}^{K=0}|\phi_{\text{pair}}\rangle$ with $\alpha = 32$) energy of 2D Hubbard model defined on the 8×8 square lattice with P-AP boundary condition. Relative error of the RBM+PP energy to that obtained by the AF-QMC [69,71] is shown. For comparison, at $t/U = 0$, we show the RBM+PP ($\mathcal{N}\mathcal{L}^{K=0}\mathcal{L}^{S=0}\mathcal{P}_G^\infty|\phi_{\text{pair}}\rangle$ with $\alpha = 32$) result for the 8×8 Heisenberg model. Error bars show standard errors of Monte Carlo measurements of energy for the optimized variational wave function.

such as the Hubbard model is analyzed as they are, because the empirical form of the correlation factors is also defined in the real-space basis. On the other hand, in cases where we employ machine learning technique, because the neural network will find a way to lower the energy even when the form of the Hamiltonian is complicated, we could think of the “best” transformation of the Hamiltonian such that the neural network can lower the energy with a small number of parameters.

Though it would not be the best of the best, we find that performing a staggered particle-hole transformation ($c_{i\downarrow} \rightarrow (-1)^i c_{i\downarrow}^\dagger$ and $c_{i\uparrow}^\dagger \rightarrow (-1)^i c_{i\uparrow}$) and mapping onto the attractive Hubbard model help the RBM+PP wave function to lower the energy of the Hubbard model. This transformation is also helpful for mVMC calculations to get a better energy [87]. Thus, we have solved the Hubbard model with this transformation. Finding a better transformation (by again employing machine learning technique) is an interesting future problem.

APPENDIX D: U DEPENDENCE OF ENERGY IN RBM+PP

Figure 4 shows the U dependence of RBM+PP ($\mathcal{N}\mathcal{L}^{K=0}|\phi_{\text{pair}}\rangle$ with $\alpha = 32$) energy of the 2D 8×8 Hubbard model. The result shows that the error decreases with increasing U/t .

- [1] J. Bardeen, L. N. Cooper, and J. R. Schrieffer, *Phys. Rev.* **108**, 1175 (1957).
- [2] E. H. Lieb and F. Y. Wu, *Phys. Rev. Lett.* **20**, 1445 (1968).
- [3] R. B. Laughlin, *Phys. Rev. Lett.* **50**, 1395 (1983).
- [4] W. L. McMillan, *Phys. Rev.* **138**, A442 (1965).
- [5] D. Ceperley, G. V. Chester, and M. H. Kalos, *Phys. Rev. B* **16**, 3081 (1977).
- [6] H. Yokoyama and H. Shiba, *J. Phys. Soc. Jpn.* **56**, 3582 (1987).
- [7] H. Yokoyama and H. Shiba, *J. Phys. Soc. Jpn.* **56**, 1490 (1987).
- [8] S. Sorella, *Phys. Rev. B* **64**, 024512 (2001).
- [9] D. Tahara and M. Imada, *J. Phys. Soc. Jpn.* **77**, 114701 (2008).
- [10] S. R. White, *Phys. Rev. Lett.* **69**, 2863 (1992).
- [11] S. R. White, *Phys. Rev. B* **48**, 10345 (1993).
- [12] F. Verstraete, V. Murg, and J. Cirac, *Adv. Phys.* **57**, 143 (2008).
- [13] R. Orús, *Ann. Phys.* **349**, 117 (2014).
- [14] M. Imada and T. Kashima, *J. Phys. Soc. Jpn.* **69**, 2723 (2000).
- [15] J. Carrasquilla and R. G. Melko, *Nat. Phys.* **13**, 431 (2017).
- [16] E. P. L. van Nieuwenburg, Y.-H. Liu, and S. D. Huber, *Nat. Phys.* **13**, 435 (2017).
- [17] Y. Zhang and E.-A. Kim, *Phys. Rev. Lett.* **118**, 216401 (2017).
- [18] T. Ohtsuki and T. Ohtsuki, *J. Phys. Soc. Jpn.* **85**, 123706 (2016).
- [19] T. Ohtsuki and T. Ohtsuki, *J. Phys. Soc. Jpn.* **86**, 044708 (2017).
- [20] H. Saito, *J. Phys. Soc. Jpn.* **86**, 093001 (2017).
- [21] W. Hu, R. R. P. Singh, and R. T. Scalettar, *Phys. Rev. E* **95**, 062122 (2017).
- [22] A. Tanaka and A. Tomiya, *J. Phys. Soc. Jpn.* **86**, 063001 (2017).
- [23] F. Schindler, N. Regnault, and T. Neupert, *Phys. Rev. B* **95**, 245134 (2017).
- [24] P. Mehta and D. J. Schwab, [arXiv:1410.3831](https://arxiv.org/abs/1410.3831).
- [25] S. J. Wetzel, *Phys. Rev. E* **96**, 022140 (2017).
- [26] S. J. Wetzel and M. Scherzer, *Phys. Rev. B* **96**, 184410 (2017).
- [27] L. Huang and L. Wang, *Phys. Rev. B* **95**, 035105 (2017).
- [28] G. Torlai and R. G. Melko, *Phys. Rev. Lett.* **119**, 030501 (2017).
- [29] P. Broecker, J. Carrasquilla, R. G. Melko, and S. Trebst, *Sci. Rep.* **7**, 8823 (2017).
- [30] P. Broecker, F. F. Assaad, and S. Trebst, [arXiv:1707.00663](https://arxiv.org/abs/1707.00663).
- [31] K. Ch’ng, J. Carrasquilla, R. G. Melko, and E. Khatami, *Phys. Rev. X* **7**, 031038 (2017).
- [32] K. Ch’ng, N. Vazquez, and E. Khatami, [arXiv:1708.03350](https://arxiv.org/abs/1708.03350).
- [33] A. Morningstar and R. G. Melko, [arXiv:1708.04622](https://arxiv.org/abs/1708.04622).
- [34] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, *Nature (London)* **549**, 195 (2017).
- [35] K. Mills and I. Tamblyn, [arXiv:1706.09779](https://arxiv.org/abs/1706.09779).
- [36] M. Bukov, A. G. Day, D. Sels, P. Weinberg, A. Polkovnikov, and P. Mehta, [arXiv:1705.00565](https://arxiv.org/abs/1705.00565).
- [37] G. Torlai and R. G. Melko, *Phys. Rev. B* **94**, 165134 (2016).
- [38] J. Liu, H. Shen, Y. Qi, Z. Y. Meng, and L. Fu, *Phys. Rev. B* **95**, 241104 (2017).
- [39] X. Y. Xu, Y. Qi, J. Liu, L. Fu, and Z. Y. Meng, *Phys. Rev. B* **96**, 041119 (2017).
- [40] Y. Nagai, H. Shen, Y. Qi, J. Liu, and L. Fu, *Phys. Rev. B* **96**, 161102 (2017).
- [41] L. Wang, *Phys. Rev. E* **96**, 051301(R) (2017).
- [42] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, [arXiv:1703.05334](https://arxiv.org/abs/1703.05334).
- [43] N. C. Costa, W. Hu, Z. J. Bai, R. T. Scalettar, and R. R. P. Singh, [arXiv:1708.04762](https://arxiv.org/abs/1708.04762).

- [44] H. Fujita, Y. O. Nakagawa, S. Sugiura, and M. Oshikawa, [arXiv:1705.05372](#).
- [45] P. Zhang, H. Shen, and H. Zhai, [arXiv:1708.09401](#).
- [46] W.-J. Rao, Z. Li, Q. Zhu, M. Luo, and X. Wan, [arXiv:1709.02597](#).
- [47] N. Yoshioka, Y. Akagi, and H. Katsura, [arXiv:1709.05790](#).
- [48] H. Saito and M. Kato, [arXiv:1709.05468](#).
- [49] G. Carleo and M. Troyer, *Science* **355**, 602 (2017).
- [50] R. Jastrow, *Phys. Rev.* **98**, 1479 (1955).
- [51] M. C. Gutzwiller, *Phys. Rev. Lett.* **10**, 159 (1963).
- [52] A. C. Hurley, J. E. Lennard-Jones, and J. A. Pople, *Proc. R. Soc. (London) A* **220**, 446 (1953).
- [53] V. A. Rassolov, *J. Chem. Phys.* **117**, 5978 (2002).
- [54] Y. Shao, L. F. Molnar, Y. Jung, J. Kussmann, C. Ochsenfeld, S. T. Brown, A. T. B. Gilbert, L. V. Slipchenko, S. V. Levchenko, D. P. O'Neill *et al.*, *Phys. Chem. Chem. Phys.* **8**, 3172 (2006).
- [55] P. Anderson, *Mater. Res. Bull.* **8**, 153 (1973).
- [56] P. W. Anderson, *Science* **235**, 1196 (1987).
- [57] K. Schmidt, M. H. Kalos, M. A. Lee, and G. V. Chester, *Phys. Rev. Lett.* **45**, 573 (1980).
- [58] T. Mizusaki and M. Imada, *Phys. Rev. B* **69**, 125110 (2004).
- [59] S.-I. Amari, K. Kurata, and H. Nagaoka, *IEEE Transactions on Neural Networks* **3**, 260 (1992).
- [60] S.-I. Amari, *Neural Comput.* **10**, 251 (1998).
- [61] A. W. Sandvik, *Phys. Rev. B* **56**, 11678 (1997).
- [62] F. Mezzacapo, N. Schuch, M. Boninsegni, and J. I. Cirac, *New J. Phys.* **11**, 083026 (2009).
- [63] L. Wang, I. Pižorn, and F. Verstraete, *Phys. Rev. B* **83**, 134421 (2011).
- [64] S. Liang, B. Douçot, and P. W. Anderson, *Phys. Rev. Lett.* **61**, 365 (1988).
- [65] S. Liang, *Phys. Rev. B* **42**, 6555 (1990).
- [66] T. Kashima and M. Imada, *J. Phys. Soc. Jpn.* **70**, 2287 (2001).
- [67] Y. Kwon, D. M. Ceperley, and R. M. Martin, *Phys. Rev. B* **58**, 6800 (1998).
- [68] T. Mizusaki and M. Imada, *Phys. Rev. C* **65**, 064319 (2002).
- [69] M. Qin, H. Shi, and S. Zhang, *Phys. Rev. B* **94**, 085103 (2016).
- [70] H.-H. Zhao, K. Ido, S. Morita, and M. Imada, *Phys. Rev. B* **96**, 085103 (2017).
- [71] S. Zhang (private communication, 2017).
- [72] D.-L. Deng, X. Li, and S. Das Sarma, *Phys. Rev. X* **7**, 021021 (2017).
- [73] D.-L. Deng, X. Li, and S. Das Sarma, [arXiv:1609.09060](#).
- [74] Y. Levine, D. Yakira, N. Cohen, and A. Shashua, [arXiv:1704.01552](#).
- [75] X. Gao and L.-M. Duan, *Nat. Commun.* **8**, 662 (2017).
- [76] J. Chen, S. Cheng, H. Xie, L. Wang, and T. Xiang, [arXiv:1701.04831](#).
- [77] Y. Huang, and J. E. Moore, [arXiv:1701.06246](#).
- [78] W.-C. Gan and F.-W. Shu, [arXiv:1705.05750](#).
- [79] Z. Cai, [arXiv:1704.05148](#).
- [80] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, and J. I. Cirac, [arXiv:1710.04045](#).
- [81] K. Hornik, *Neural Networks* **4**, 251 (1991).
- [82] G. Cybenko, *Math. Control Signals Syst.* **2**, 303 (1989).
- [83] N. L. Roux and Y. Bengio, *Neural Comput.* **20**, 1631 (2008).
- [84] D. M. Ceperley and B. J. Alder, *Phys. Rev. Lett.* **45**, 566 (1980).
- [85] <https://github.com/issp-center-dev/mVMC>.
- [86] H. Yokoyama and H. Shiba, *J. Phys. Soc. Jpn.* **57**, 2482 (1988).
- [87] T. Ohgoe (private communication, 2017).