


Efficient neural-network based variational Monte Carlo scheme for direct optimization of excited energy states in frustrated quantum systems

Tanja Đurić  and Tomislav Ševa

Department of Physics, Faculty of Science, University of Zagreb, Bijenička cesta 32, 10000 Zagreb, Croatia



(Received 24 February 2020; revised 9 June 2020; accepted 7 July 2020; published 3 August 2020)

We examine applicability of the valence bond basis correlator product state ansatz, equivalent to the restricted Boltzmann machine quantum artificial neural-network ansatz, and variational Monte Carlo method for direct optimization of excited energy states to study properties of strongly correlated and frustrated quantum systems. The energy eigenstates are found by stochastic minimization of the variational function for the energy eigenstates, which allows direct optimization of particular energy state without knowledge of the lower energy states. This approach combined with numerous tensor network or artificial neural-network ansatz wave functions then allows further insight into quantum phases and phase transitions in various strongly correlated models by considering properties of these systems beyond the ground-state properties. Also, the method is in general applicable to any dimension and has no sign instability. An example that we consider is the square lattice J_1 - J_2 antiferromagnetic Heisenberg model. The model is one of the most studied models in frustrated quantum magnetism since it is closely related to the disappearance of the antiferromagnetic order in the high- T_c superconducting materials and there is still no agreement about the properties of the system in the highly frustrated regime near $J_2/J_1 = 0.5$. For the J_1 - J_2 model, we write the variational ansatz in terms of the two site correlators and in the valence bond basis and calculate the lowest energy eigenstates in the highly frustrated regime near $J_2/J_1 = 0.5$ where the system has a paramagnetic phase. We find that our results are in good agreement with previously obtained results, which confirms applicability of the method to study frustrated spin systems.

DOI: [10.1103/PhysRevB.102.085104](https://doi.org/10.1103/PhysRevB.102.085104)

I. INTRODUCTION

Simulating frustrated quantum spin systems is among the most challenging computational tasks and is one of the central problems in condensed matter physics. Approximating the wave function of the system with a tensor network (TN) or artificial neural network (ANN) ansatz and employing Monte Carlo sampling to efficiently compute expectation values proved recently to be a very efficient approach to study a variety of strongly correlated models [1–23]. The approach can be applied to systems of any spatial dimensionality and is free of sign problems, and therefore overcomes the limitations of two other main numerical techniques used to simulate correlated quantum spin systems, the density-matrix renormalization group (DMRG) method [24–27] and quantum Monte Carlo (QMC) [28]. While the DMRG method gives very accurate results only in one dimension, the QMC suffers from the sign problem for frustrated (fermionic) quantum systems.

The tensor network states (TNS) and variational Monte Carlo (VMC) approach has so far been mostly used to study the ground-state properties of various quantum correlated systems. The main reason for this limitation was the lack of a robust and efficient excited-state variational principle analogous to the ground-state variational principle where the function that can be efficiently minimized is the energy $E(\psi) = \langle \psi | H | \psi \rangle / \langle \psi | \psi \rangle$. Recently, Zhao and Neuscamman proposed an efficient variational principle for the direct optimization of excited states [29–34], which can be used at polynomial cost with numerous approximate ansatz wave functions. The

method allows us to target a particular eigenstate without knowledge of the lower energy states by tuning the value of the energy shift parameter included in the variational function for the eigenstates and has so far been mostly used to study molecular excitations.

In this paper, we examine and confirm applicability of the mentioned variational principle combined with a suitable TNS or ANN ansatz to study properties of strongly correlated and frustrated quantum systems beyond one dimension for which other numerical techniques are limited and VMC-based calculations are one of the few remaining possibilities to study such systems. Although the proposed VMC scheme with flexible ansatz used in our calculations is in general applicable to systems with or without frustration and in any dimension, it is therefore the most relevant to demonstrate applicability of the scheme to mentioned numerically challenging cases not easily accessible with other numerical methods.

Specifically, we calculate the lowest energy eigenstates for the square lattice J_1 - J_2 antiferromagnetic Heisenberg model in the highly frustrated regime near $J_2/J_1 = 0.5$, where the system has a paramagnetic phase. The J_1 - J_2 model is one of the most studied models in frustrated quantum magnetism [27,35–46]. It is closely related to the disappearance of the antiferromagnetic order in high- T_c superconducting materials [47,48] and is therefore of great importance. The model has also been proposed as a possible model that supports topologically ordered chiral spin-liquid state [49–51] or Z_2 spin liquid state [43,52–56].

The model has so far been studied using several methods, among which are, for example, exact diagonalization (ED) [37,57,58], variational methods [12,42,45,59–61], DMRG [27], the Green's function Monte Carlo with stochastic re-configuration (GFMC SR) technique [41], and the cluster update algorithm for tensor product states (TPSs) [35]. Properties of the phase(s) in the highly frustrated regime near $J_2/J_1 = 0.5$ and presence of deconfined quantum critical point [16,35,62–66] at the transition from antiferromagnetic to paramagnetic phase have been debated for decades [27,35–45,49–56] and there is still no general agreement.

Here we demonstrate that VMC method for direct optimization of excited energy states combined with appropriate TN or ANN ansatz can provide further insight into quantum phases and phase transitions in complex models such as the frustrated Heisenberg J_1 - J_2 model. The method is applicable in any dimension, allows studying properties of the system beyond the ground-state properties, and in general has no sign instability.

Our calculations are performed with correlator product state (CPS) ansatz [9,10,14,15] in the valence bond (VB) basis [42–45,47,59,67–72] as an ansatz for the energy eigenstates. Here the CPS ansatz is built from two-site correlators associated with the pairs of sites. The ansatz is equivalent to the restricted Boltzmann machine (RBM) representation of the wave function where the number of hidden units equals the number of different pairs of sites. RBMs are types of generative stochastic artificial neural networks (ANNs) [5,6,8,17–21] that can learn a distribution over the set of their inputs. The inputs here are spin configurations and the wave function corresponds to the complex probability distribution that the network tries to approximate. Correlations in ANNs are included by hidden units and are nonlocal in space. Because of their nonlocal geometry, ANNs can describe some of the states that cannot be described by traditional TNSs, for example, chiral spin liquid states or lattice fractional quantum Hall states.

Variational function for the energy eigenstates is minimized using a stochastic optimization scheme [10,14–16,44,73,74] which requires knowledge only of the first derivatives of the variational function with respect to the variational parameters in the CPS ansatz. We calculate the lowest energy eigenstates in the total spin zero sector for the system sizes with $N = 36$ and 64 lattice sites and with periodic boundary conditions. Our results for the energy gap between the first excited energy state and the ground state in the total spin zero sector is in good agreement with results obtained previously with other methods. This demonstrates general applicability of the method to study properties of complex interacting many-body systems.

We also note that statistical error present in the stochastic algorithm can result in significant error for energy eigenstates for smaller system sizes. However, influence of the error can be controlled by increasing the system size and does not affect results for the energy gap where the error cancels when subtracting values of the eigenenergies.

The paper is organized as follows. In Sec. II, we define CPS variational ansatz in terms of two-site correlators and in the VB basis. In Sec. III, we describe the stochastic optimization scheme for efficient minimization of the variational function

for energy eigenstates. Our numerical results for the system sizes $N = L \times L$ lattice sites with $L = 6$ and 8 and periodic boundary conditions are presented in Sec. IV. In the final section, Sec. V, we draw our conclusions, summarize results, and discuss possible directions for future research.

II. VALENCE BOND BASIS CORRELATOR PRODUCT STATES

We consider the square lattice J_1 - J_2 antiferromagnetic Heisenberg model

$$H = J_1 \sum_{\langle i,j \rangle} \vec{S}_i \vec{S}_j + J_2 \sum_{\langle\langle i,j \rangle\rangle} \vec{S}_i \vec{S}_j, \quad (1)$$

where \vec{S}_i are spin-1/2 operators, and J_1 and J_2 are antiferromagnetic couplings for neighboring and next-neighboring sites, respectively.

In general, in the usual basis of the $S_z^{\text{tot}} = \sum_{i=1}^N S_i^z$ eigenstates, an eigenstate of the Hamiltonian (1) can be written in the form

$$|\psi\rangle = \sum_{\{\vec{s}\}} W(\{\vec{s}\}) |\{\vec{s}\}\rangle, \quad (2)$$

where $|\{\vec{s}\}\rangle = |s_1, \dots, s_N\rangle$ and $s_i \in \{-1/2, 1/2\}$ are eigenvalues of the local S_i^z operator. Since the total magnetization along the z axis is a good quantum number, i.e., the Hamiltonian (1) commutes with the operator $M_z = S_z^{\text{tot}}$,

$$[H, M_z] = 0, \quad (3)$$

all calculations can be performed in the canonical ensemble, i.e., for a chosen M_z sector. The eigenstates can then also be written in the form

$$|\psi\rangle_{M_z} = \sum_{\{\vec{s}\}} W(\{\vec{s}\}) P_{M_z} |\{\vec{s}\}\rangle, \quad (4)$$

where the projection operator P_{M_z} projects to the spin configurations with $\sum_{i=1}^N S_i^z = M_z$.

The ground state and the lowest excited energy state in the highly frustrated regime near $J_2/J_1 = 0.5$ are in the $M_z = 0$ sector [36,37,45] and we therefore restrict our calculations to $M_z = 0$. Then the most suited basis for the spin rotationally invariant Hamiltonians is the VB basis [42–45,47,67–72]. A VB configuration $|\alpha\rangle$

$$|\alpha\rangle = |(i_1^\alpha, j_1^\alpha)(i_2^\alpha, j_2^\alpha) \dots (i_{N/2}^\alpha, j_{N/2}^\alpha)\rangle \quad (5)$$

is a product of two-spin singlets

$$(i, j) = \frac{1}{\sqrt{2}} (|\uparrow_i \downarrow_j\rangle - |\downarrow_i \uparrow_j\rangle), \quad (6)$$

and any total singlet state with $S_z^{\text{tot}} = 0$ can be written in the VB basis

$$|\psi\rangle = \sum_{\alpha} W(\alpha) |\alpha\rangle, \quad (7)$$

where VB configurations α correspond to all possible pairings of N spins into $N/2$ valence bonds.

The VB basis is an overcomplete basis and the overlap between the VB states is [42–44,47,69–71]

$$\langle \alpha | \beta \rangle = (-1)^{n_\alpha + n_\beta} 2^{N_i - N/2} = \pm 2^{N_i - N/2}, \quad (8)$$

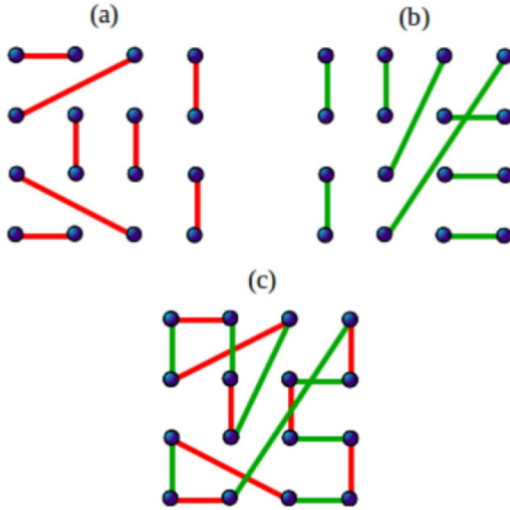


FIG. 1. Transposition graph (c) $\langle \alpha | \beta \rangle$ of two valence bond states (a) $|\alpha\rangle$ and (b) $|\beta\rangle$. The valence bond basis (5) is an overcomplete basis and the overlap between two valence bond states is $\langle \alpha | \beta \rangle = \pm 2^{N_l - N/2}$, where N_l is the number of loops in the transposition graph and N is the number of lattice sites. Here $N = 16$ and $N_l = 2$.

where N_l is the number of loops in the transposition graph obtained when the VB states $|\alpha\rangle$ and $|\beta\rangle$ are superimposed (Fig. 1) and the overall sign depends on the convention for assigning directions to the bonds (n_α and n_β denote the number of valence bonds in $|\alpha\rangle$ and $|\beta\rangle$ in the opposite direction than defined to be + direction). Because of the overcompleteness of the VB basis, the expansion coefficients $W(\alpha)$ in Eq. (7) are not unique. This is, however, not a problem for any practical calculations.

Also, the trial wave function does not have to be constructed from the largest possible set of VB states in which all spins are joined by valence bonds in all possible ways. A more restricted (and still overcomplete) basis can be obtained by dividing the system into two groups of sites, A and B, and keeping only VB states with bipartite bonds which connect sites from different groups A and B. The overcompleteness property can be written as $(i, k)(j, l) \rightarrow (i, j)(k, l) - (i, l)(k, j)$ [68] and VB states with nonbipartite AA and BB bonds can be therefore written in terms of VB states with AB bonds.

For a bipartite lattice, the typical choice of the sites A and B corresponds to the two sublattices in the bipartite lattice. For a square lattice, AB labels are usually assigned to form checkerboard or collinear patterns. If in addition the direction of each singlet (i, j) in a VB state is fixed such that $i \in A$ and $j \in B$, it can be shown that all expansion coefficients $W(\alpha)$ can be taken to be real and positive [67,69]. That corresponds to Marshall's sign rule [75,76] in the absence of frustration when the wave function is written in the standard basis of eigenstates of the S_i^z operator (2).

The Marshall's sign rule exists in two limits, $J_2/J_1 = 0$ and $J_2/J_1 = \infty$, and it can be shown that the sign rule survives the frustration in the J_1 - J_2 model on the square lattice for a relatively large range of the parameter J_2/J_1 values away from the points $J_2/J_1 = 0$ and $J_2/J_1 = \infty$ [77,78]. Also, for any J_2/J_1 exists in principle a positive-definite expansion of

the wave function in the VB basis, since $W(\alpha)$ can be made positive by simply reversing the order of the indices in one singlet in that particular state. However, in general there is no practically useful rule for fixing the order.

Within the CPS approach [10,14,15], the coefficients in Eq. (2) or Eq. (7) are written in terms of correlator coefficients associated with groups of sites. The CPS ansatz can then be used as a basis for VMC simulations where the coefficients are optimized using one of the efficient optimization methods, for example, the stochastic optimization scheme described in the following section, which requests only the first energy derivatives.

Here we consider two-site CPS where a correlator is associated with a pair of sites

$$W(\alpha) = (-1)^{n_\alpha} \prod_{i,j} C_{ij}. \quad (9)$$

In the previous calculations for the J_1 - J_2 model, computational cost was reduced by assuming symmetries for the coefficients in an ansatz wave function, for example, 2×2 sublattice structure and translational invariance of the coefficients in terms of the sublattice period [45]. Since it is better to use a flexible ansatz wave function without any constraint on coefficients C_{ij} we do not take into consideration any symmetries and we also do not impose any constraint between C_{ij} and C_{ji} . Amplitudes C_{ij} can then be taken to be real and positive. We consider $L \times L$ square lattices with periodic boundary conditions where $i = (x, y)$ and $x, y = 1, \dots, L$. For the $L \times L$ square lattice, with $N = L^2$ lattice sites, the number of coefficients W_α is $N!$, and the two-site CPS ansatz (9) leads to $N(N-1)$ variational parameters.

CPS states in VB basis can be used to describe both ordered and disordered phases [69,70]. The flexible wave function without any constraint used in our calculations can describe antiferromagnetic, valence-bond-crystal as well as spin-liquid states, and excitations above these ground states in total magnetization $M_z^{\text{tot}} = 0$ sector, on equal footing. A Neel-ordered ground state with spin correlations decaying with a power law as a function of distance requires long-range amplitudes C_{ij} while a disordered state requires larger decay rate of amplitudes C_{ij} with distance $|i-j|$.

Namely, power-law correlations and low-energy nonzero spin excitations in a state with nonzero staggered magnetization can be generated only if longer bond amplitudes are taken into account. Low-energy excitations arise from breaking valence bonds or simply rearranging the bonds. While the excitations that rearrange valence bonds can have low energy even in the short-range RVB state, the excitations that break valence bonds have low energy only for RVB states with long-range valence bonds. If two spins that are far apart form a valence bond, those spins are only weakly bound into a singlet that can be broken into free spins with relatively little energy and can therefore be more easily excited into a state with nonzero spin. Longer bonds correspond to weaker singlet pairing energies. Therefore, the states that have a significant weight from long-range valence bonds have more low-energy spin excitations than states with mainly short-range valence bonds.

For example, for $J_2 = 0$, Liang *et al.* demonstrated how tuning decay rate of bond amplitudes that decrease with

distance can smoothly transform a Néel state into a disordered state for $J_2 = 0$ [69]. Their results showed that short-bond wave functions have short-range spin-spin correlations with an exponential decay at large distances. An algebraic decay of correlation functions and a nonzero staggered magnetization can be generated only if longer bond amplitudes are taken into account in the trial wave function, and a Néel ordered ground state therefore requires an algebraic decay of the bond length probability.

Also, this CPS ansatz is equivalent to the RBM representation of the wave function with $M = N(N - 1)/2$ hidden units [5,21]. Boltzmann machines are types of generative stochastic artificial neural networks that can learn a distribution over the set of their inputs [5,6,8,17–21]. Here the network inputs are spin configurations and the wave function corresponds to a complex probability distribution that the network tries to approximate.

III. STOCHASTIC OPTIMIZATION OF THE VARIATIONAL FUNCTION FOR ENERGY EIGENSTATES

So far, variational calculations have been mostly restricted to studying the ground-state properties of various systems with the energy as a function that can be efficiently minimized,

$$E(\psi) = \frac{\langle \psi | H | \psi \rangle}{\langle \psi | \psi \rangle}, \quad (10)$$

where H is the Hamiltonian of the system. Optimization of an excited energy state would then require knowledge of all energy states with energy lower than the energy of the chosen excited state.

However, Zhao and Neuscamman [30–34] recently introduced an efficient variational principle for direct optimization of excited states that does not require knowledge of lower energy states. They have defined a function

$$\Omega(\psi, \omega) = \frac{\langle \psi | (\omega - H) | \psi \rangle}{\langle \psi | (\omega - H)^2 | \psi \rangle} = \frac{\omega - E}{(\omega - E)^2 + \sigma^2}, \quad (11)$$

where

$$\sigma^2 = \frac{\langle \psi | (H - E)^2 | \psi \rangle}{\langle \psi | \psi \rangle} \quad (12)$$

is the variance, whose global minimum is an excited energy state with the energy immediately above the energy shift ω that is placed between distinct eigenvalues of H .

To efficiently evaluate and optimize $\Omega(\psi, \omega)$ using the variational Monte Carlo scheme, $\Omega(\psi, \omega)$ is rewritten in the form

$$\Omega(\psi, \omega) = \frac{\sum_m P(m) w_m}{\sum_m P(m) w_m^2}, \quad (13)$$

where $m = (\alpha, \alpha')$ denotes a pair of two VB configurations $|\alpha\rangle$ and $|\alpha'\rangle$,

$$w_m \equiv \frac{\langle \alpha' | (\omega - H) | \alpha \rangle}{\langle \alpha' | \alpha \rangle} \quad (14)$$

and

$$P(m) = \frac{W_{\alpha'} W_{\alpha} \langle \alpha' | \alpha \rangle}{\sum_{\alpha, \alpha'} W_{\alpha'} W_{\alpha} \langle \alpha' | \alpha \rangle}. \quad (15)$$

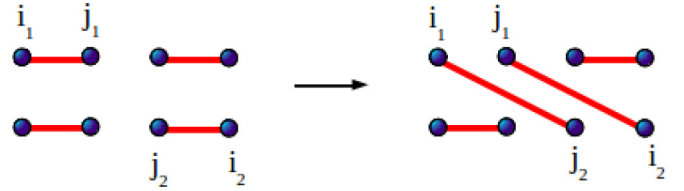


FIG. 2. Local two-bond updates. First site i_1 is randomly chosen, then one of four diagonal neighbor sites i_2 is chosen randomly, and the ends of bonds are exchanged: $(i_1, j_1)(i_2, j_2) \rightarrow (i_1, j_2)(i_2, j_1)$. The reconfiguration is accepted with probability (19).

$P(m)$ is always positive since the coefficients C_{ij} are taken here to be real and positive.

Similarly to the overlap of two VB configurations (8), matrix elements of relevant operators can be typically calculated by considering loops in the transposition graph [67–70]. The spin-spin correlations and $\langle \alpha' | H | \alpha \rangle / \langle \alpha' | \alpha \rangle$ can be computed from the relation

$$\frac{\langle \alpha' | \vec{S}_i \vec{S}_j | \alpha \rangle}{\langle \alpha' | \alpha \rangle} = \frac{3}{4} \phi_i \phi_j \delta_{\lambda_i, \lambda_j}, \quad (16)$$

where $\phi_i = +1$ if sites i and j belong to two different groups of sites (A and B), $\phi_i = -1$ if sites i and j belong to the same group of sites (A or B), and $\delta_{\lambda_i, \lambda_j} = 1$ if $\lambda_i = \lambda_j$ and zero otherwise. Here λ_i is a label for the loop to which site i belongs, and matrix element (16) vanishes if sites i and j belong to different loops.

Within the VMC scheme, the phase space considered as ensemble of pairs $m = (\alpha, \alpha')$ is summed over according to probability distribution $P(m)$ and

$$\Omega_{\text{VMC}}(\psi, \omega) = \frac{\sum_{m \in \zeta} w_m}{\sum_{m \in \zeta} w_m^2}, \quad (17)$$

where elements of ζ are sampled via a METROPOLIS walk. In evaluating the sums in Eq. (17), a new pair of valence bonds m' is generated starting from a valence bond pair m with a chosen update scheme, and METROPOLIS acceptance probability for such update is

$$P_{\text{acc}} = \min \left[\frac{W_{m'}}{W_m} 2^{\Delta N_l}, 1 \right], \quad (18)$$

where $W_m = W_{\alpha'} W_{\alpha}$ and $\Delta N_l = N_l(m') - N_l(m)$ denotes change in the number of loops in the transposition graph (Fig. 1). Here Monte Carlo sampling is performed by nonlocal bond-loop updates [70] instead with local two-bond updates [69,70], since nonlocal updates sample the phase space in the highly frustrated regime much more efficiently. In a local two-bond update, a new configuration in the MC move is generated by exchanging the ends of two bonds (i_1, j_1) and (i_2, j_2) as illustrated in Fig. 2. First site i_1 is randomly chosen, then one of four diagonal neighbor sites i_2 is chosen randomly, and the ends of bonds are exchanged $(i_1, j_1)(i_2, j_2) \rightarrow (i_1, j_2)(i_2, j_1)$. The METROPOLIS acceptance probability is then

$$P_{\text{acc}} = \min \left[\frac{C_{i_1, j_1} C_{i_2, j_2}}{C_{i_1, j_2} C_{i_2, j_1}} 2^{\Delta N_l}, 1 \right] \quad (19)$$

and ΔN_l can be $+1$, -1 , or 0 corresponding to the cases when two loops join in one, one loop splits into two, or number of

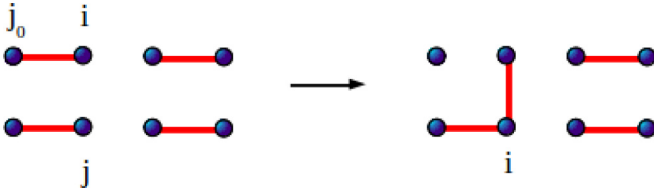


FIG. 3. Nonlocal bond-loop updates. The first step in a bond-loop update discussed in the text creates two defects (one site with two bonds and one empty site). Since such defects cannot be present in a valence bond state, one of these defects is further moved by subsequent bond moves until it annihilates with the second defect and the loop closes.

loops is preserved. However, local two-bond updates usually involve long bonds through reconfiguration and the sampling process becomes inefficient when long bond amplitudes have small weights. Since the values of longer bond amplitudes decrease with decrease of antiferromagnetic ordering, nonlocal bond updates can be much more efficient sampling schemes, particularly in the highly frustrated regime where the disappearance of antiferromagnetic ordering is predicted. Within a bond-loop update scheme, one end of a randomly chosen bond is moved, resulting in two defects that correspond to an empty site and one site with two bonds (Fig. 3). Since such defects cannot be present in a VB state, one of these defects is further moved by subsequent bond moves until it annihilates with the second defect and the loop closes.

In the algorithm, the bonds are represented by an array of links between sites, here denoted by v , such that if sites i and j are connected by a VB then $v(i) = j$ and $v(j) = i$. In the bond-loop update algorithm, starting from a randomly selected lattice site j_0 with $v(j_0) = i$, a new lattice site j is chosen according to probability distribution

$$P_{ij} = \frac{C_{ij}}{\sum_j C_{ij}}, \quad (20)$$

proportional to corresponding weight C_{ij} . After the bond emerging from i is moved to an acceptable j , $v(i)$ changes from $v(i) = j_0$ to $v(i) = j$, and the original link between sites i and j_0 is destroyed and no longer needed.

The original site j_0 now has no bond attached to it (unless $j = j_0$, which immediately terminates bond-loop update and a new loop update starts from a different randomly chosen site) and the new site j has two VBs on it, corresponding to two defects in VB configuration. To remove such defects which cannot be present in a VB state, the end of the old bond i is moved by repeating the same steps as for the initial bond move only with j_0 replaced by j . This procedure is repeated until it happens that $j = j_0$, which results in annihilation of the double-bond and no-bond defects and closing of the loop. Since the loops can be large, bond-loop updates can be much more efficient than local two-bond updates.

To optimize $\Omega(\psi, \omega)$, we further use a stochastic optimization scheme [10,14–16,44,73,74] which requests only knowledge of the signs of the first derivatives of $\Omega(\psi, \omega)$ with respect to parameters C_{ij} which are updated according to

$$\ln C_{ij} \rightarrow \ln C_{ij} - r\delta(k)\text{sign}\left(\frac{\partial\Omega}{\partial C_{ij}}\right), \quad (21)$$

where $r \in [0, 1)$ is a random number and $\delta(k)$ is the optimization step for given iteration k . Without a random number, this kind of update scheme is known as Manhattan learning [79,80], previously introduced in the context of neural networks. Here, a random number is introduced because it was shown that it speeds up the convergence [44].

Similarly as in simulated annealing methods [81], the optimization step $\delta(k)$ is reduced in each iteration k to reach the optimum solution. Here, the annealing scheme that ensures convergence of the method is

$$\delta(k) = \delta_0 \frac{1}{k^\nu} \quad (22)$$

with $0.5 < \nu < 1$, as demonstrated previously for the energy minimization [44]. We find that taking $\delta_0 = 0.5$ and $\nu = 0.75$ works well. Alternatively, a geometric form $\delta(k) = \delta_0 v^k$ for the annealing scheme can also be used with $\nu = 1 - \epsilon$ and $\epsilon \ll 1$ [44].

First derivatives $\partial\Omega/\partial C_{ij}$ can be efficiently evaluated using the variational Monte Carlo scheme using the following expression,

$$\frac{\partial\Omega}{\partial C_{ij}} = \frac{1}{\langle\mathcal{O}_2\rangle} \frac{\partial\langle\mathcal{O}_1\rangle}{\partial C_{ij}} - \frac{\langle\mathcal{O}_1\rangle}{\langle\mathcal{O}_2\rangle^2} \frac{\partial\langle\mathcal{O}_2\rangle}{\partial C_{ij}}, \quad (23)$$

where $\langle\mathcal{O}_1\rangle = \langle\psi|\omega - H|\psi\rangle/\langle\psi|\psi\rangle \equiv \langle\omega - H\rangle$ and $\langle\mathcal{O}_2\rangle = \langle(\omega - H)^2\rangle$,

$$\frac{\partial\langle\mathcal{O}_k\rangle}{\partial C_{ij}} = \langle\Delta_{ij}\mathcal{O}_k\rangle - \langle\Delta_{ij}\rangle\langle\mathcal{O}_k\rangle \quad (24)$$

for $k = 1$ or 2 , and

$$\Delta_{ij} = \frac{1}{C_{ij}} \frac{\partial W_m}{\partial C_{ij}} = \frac{b_{ij}}{C_{ij}}. \quad (25)$$

Here, b_{ij} denotes the number of times the coefficient C_{ij} appears in the product $W_m = W_\alpha W_{\alpha'}$, where the amplitude W_α for the VB configuration $|\alpha\rangle$ is given by Eq. (9). Since we do not include any symmetries, a coefficient C_{ij} appears in each VB configuration only once. Therefore $b_{ij} = 1$ if C_{ij} appears in only one of the VB configurations (α or α') or $b_{ij} = 2$ if C_{ij} appears in both VB configurations (α and α'). The first derivatives can be calculated from the same sample as $\Omega_{\text{VMC}}(\psi, \omega)$ (17) obtained by a Markov chain in the METROPOLIS algorithm.

The variational algorithm starts from randomly chosen values for the coefficients C_{ij} (between 0 and 1), then Ω and its gradient vector are evaluated for particular value of the parameter ω , and all coefficients C_{ij} are updated according to (21). In each iteration, the same procedure is repeated, starting from the coefficients C_{ij} from the previous iteration, until convergence of Ω is reached. The value of energy E is calculated using coefficients obtained by minimizing Ω .

In each iteration k , the variational function Ω and its derivative are estimated from $F(k) \times N$ sampled values, where N is the number of lattice sites. $F(k)$ is called the number of sweeps per sample. In each sweep, a random lattice site is chosen and a move to a new configuration obtained by a bond-loop update is proposed N times. In addition to careful tuning of the gradient step $\delta(k)$ to achieve the convergence, the number of sweeps $F(k)$ per iteration is increased to reduce effects of noise on the calculation of the first derivatives.

Namely, the derivatives become smaller as the Ω minimum is approached and require more sampled values in order not to be dominated by noise. Here, the number of sweeps is increased linearly for each iteration, $F(k) = F_0 \times k$.

Also, the procedure of evaluating Ω and updating the coefficients C_{ij} (21) is repeated $G(k) = G_0 \times k$ times, where increasing G corresponds to a slower cooling rate. Here, we take $G_0 = 20$, $F_0 = 20$ for $L = 6$, and $F_0 = 10$ for $L = 8$. The minimization routine is performed for 100 iterations, and after the minimization is complete the values of Ω , and corresponding energy E for each value of ω , are calculated by repeating the procedure for a single iteration with zero step size and large F and G to obtain more accurate estimates of Ω and E .

IV. NUMERICAL RESULTS FOR THE SQUARE LATTICE J_1 - J_2 ANTIFERROMAGNETIC HEISENBERG MODEL

Before proceeding to numerical results for the square lattice J_1 - J_2 antiferromagnetic Heisenberg model, it is important to clarify consequences of using approximate ansatz for the wave functions that correspond to the energy eigenstates. For the exact wave function, $\sigma_n = 0$ and $\Omega(\psi, \omega)$ diverges at $\omega_n = E_n$. However, for an ansatz wave function, variance σ has a nonzero value and the optimized ansatz, its energy, and value of $\Omega(\psi, \omega)$ depend on particular choice of ω .

If the approximate ansatz wave function is very close to the exact wave function, energy dependence on a precise choice of ω is small and the function $\Omega(\psi, \omega)$ has a finite minimum near $\omega_n = E_n - \sigma_n$ for the n th eigenstate with energy E_n [29]. Value of $\omega_n = E_n - \sigma_n$ corresponds to the analytic solution for the minimum of $\Omega(\psi, \omega)$, when E and σ are held fixed. Consequently, for a finite and small σ_n , the value of ω_n at which minimum of $\Omega(\psi, \omega)$ changes states no longer occurs at the energy of lower energy eigenstate and is shifted downward to the value of ω close to $\omega_n = E_n - \sigma_n$. However, in the cases where the wave-function approximation leads to a bit larger value of variance σ , as happens in our many-body calculation, a stronger dependence of energy and Ω on ω may arise and $\Omega(\psi, \omega)$ no longer simply has a form with finite minimum near $\omega_n = E_n - \sigma_n$. In any case, as proposed by Zhao and Neuscamman [29], to calculate energy of the n th eigenstate we choose value of ω within the range $E_{n-1} \leq \omega \leq E_n$ for which the smallest value of Ω is obtained with minimization procedure for each ω in that range.

To summarize, there are three steps in our calculations to find energy eigenvalues E_n above the ground state: (1) for each ω , minimize $\Omega(\psi^{\text{ansatz}}, \omega)$ and from the optimized ansatz calculate corresponding energy $E = \langle \psi^{\text{ansatz}} | H | \psi^{\text{ansatz}} \rangle$; variance σ can then be calculated from the values of E and Ω ; (2) find ω_n^{min} within the range $E_{n-1} \leq \omega \leq E_n$ for which $\Omega(\psi^{\text{ansatz}}, \omega)$ has the smallest value; and (3) find n th energy eigenvalue that corresponds to the energy obtained from the optimized ansatz at ω_n^{min} .

For the ground state, it is important to point out that optimizing the energy $E = \langle \psi | H | \psi \rangle / \langle \psi | \psi \rangle$ and $\Omega(\psi, \omega)$ is different and leads to a different optimized ansatz wave function since for the Ω minimization the wave function depends both on energy and its variance. Therefore, to calculate the energy difference $E_n - E_0$ even for the ground state, we need

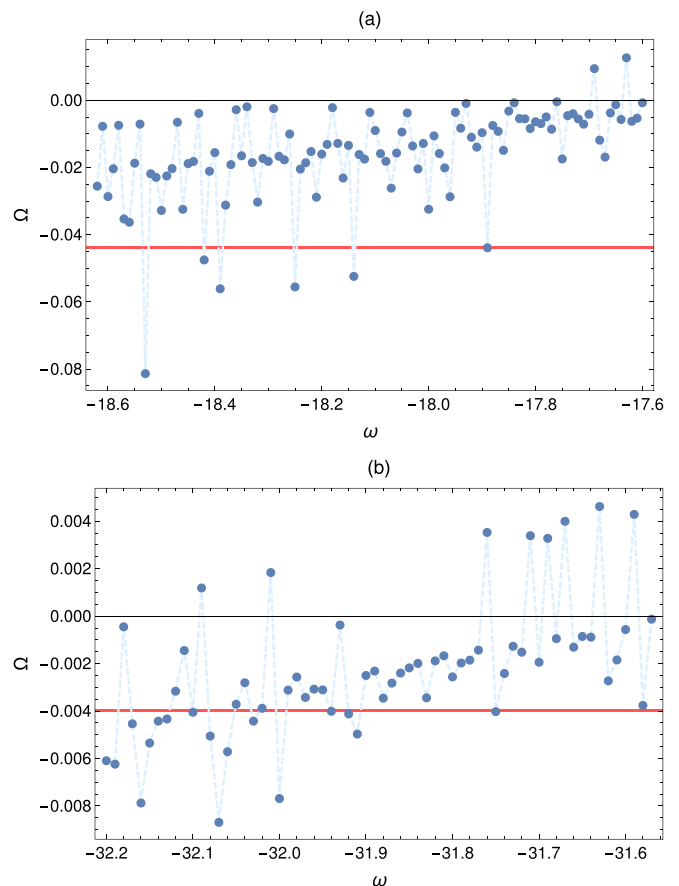


FIG. 4. Ω as a function of the energy shift parameter ω for the ground state and the first excited energy state ($M_z = \sum_{i=1}^N S_i^z = 0$ sector) for $J_2/J_1 = 0.5$ (J_1 is set to $J_1 = 1$) and the system sizes $N = L \times L$ with (a) $L = 6$, (b) $L = 8$, and with periodic boundary conditions. Here, $|\psi^{\text{ansatz}}\rangle$ is optimized using a stochastic optimization method to minimize $\Omega(\psi^{\text{ansatz}}, \omega)$ for each value of ω . Red lines denote minimal value of $\Omega(\omega)$, Ω_1^{min} , for ω within the range $E_0 \leq \omega \leq E_1$. Energy of the first excited energy state obtained by optimization procedure is calculated using the optimized ansatz that minimized Ω at ω_1^{min} , where $\Omega(\omega_1^{\text{min}}) = \Omega_1^{\text{min}}$.

to take the value of E_0 obtained by minimizing Ω at $\omega = \omega_0$ such that $\Omega(\psi_0, \omega_0) \approx \Omega(\psi_1, \omega_1)$, where ω_0 is in the range of values $\omega \leq E_0$. In this many-body case, where a stronger dependence of Ω on ω appears, such choice of ω_0 is equivalent to the choice of ω_0 close to the value where Ω has a finite minimum in the paper written by Zhao and Neuscamman [29] where in their calculation of molecular excitations energy dependence on particular choice of ω was very small. The value of E_0 obtained by minimizing E corresponds to the minimum of $\Omega(\psi, \omega \rightarrow \infty)$.

Our results obtained by minimizing Ω for $J_2/J_1 = 0.5$ (where J_1 is set to $J_1 = 1$) and the system sizes $N = L \times L$ with $L = 6, 8$ and periodic boundary conditions are shown in Figs. 4–6. Statistical error present in the stochastic algorithm is controlled by increasing the system size [10,15] since having a larger number of parameters allows the optimization method more freedom in finding the minimum of $\Omega(\psi, \omega)$ for a given value of ω and consequently better estimates for the energy eigenvalues. It is therefore difficult to obtain good

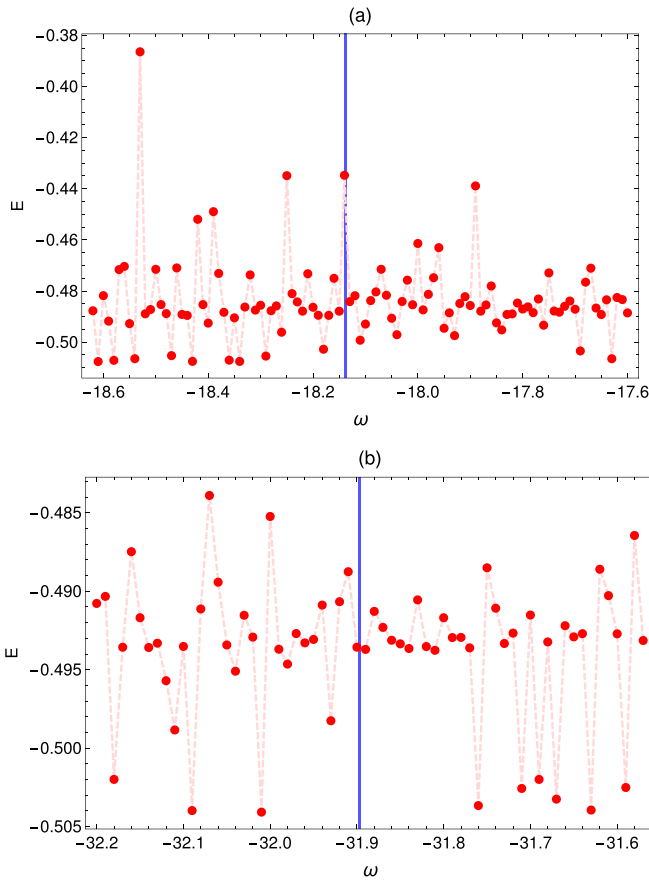


FIG. 5. Energy E as a function of the energy shift parameter ω calculated by taking the optimized ansatz that minimizes $\Omega(\psi^{\text{ansatz}}, \omega)$ for each value of ω . The results are for the system sizes $N = L \times L$ with (a) $L = 6$, (b) $L = 8$, and periodic boundary conditions. Here, $J_2/J_1 = 0.5$ and J_1 is set to $J_1 = 1$. Blue lines denote the ground-state energies obtained by exact diagonalization and the cluster update algorithm for tensor product states [35].

estimates of Ω and E for small system sizes. This is clearly visible in our results since we obtain much better energy estimates for $L = 8$ than for $L = 6$ as will be clarified further in this section. However, we obtain quite good estimates for energy gaps in both cases since the statistical error equivalently affects calculation of all energy states and cancels in the energy gap calculation.

We calculate the energy of the first excited energy state as energy of the optimized ansatz at the value of ω within $E_0 \leq \omega \leq E_1$ range where the optimized value of $\Omega(\psi, \omega)$ is minimal. For $L = 6$, the ground-state energy obtained by exact diagonalization is $E_0(L=6)/L^2 = -0.50380965$ [35,45] and we find that for $\omega \geq E_0$ value of $\Omega(\psi, \omega)$ is minimal at $\omega_1 = -17.89$. Energy of the corresponding optimized ansatz at $\omega_1 = -17.89$ is $E_1^{\text{ansatz}}/L^2 = -0.4386$. The value of E_0^{ansatz} that corresponds to Ω minimization is determined by minimizing $\Omega(\psi, \omega)$ for a range of values $\omega < E_0$ and then finding the value ω_0 such that $\Omega(\psi_0, \omega_0) \approx \Omega(\psi_1, \omega_1)$. For $L = 6$, we find that $\omega_0 = -18.42$ and $E_0^{\text{ansatz}}/L^2 = -0.45168$. The ground-state energy obtained by Ω minimization therefore significantly differs from the exact ground-state energy for the smaller system size with $L = 6$ with

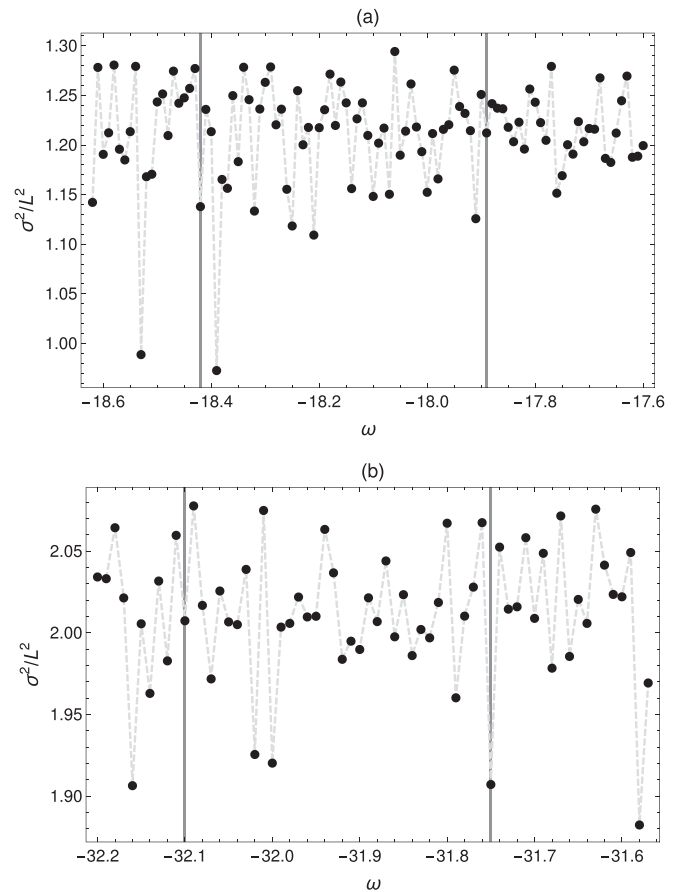


FIG. 6. Variance σ^2 as a function of the energy shift parameter ω for $J_2/J_1 = 0.5$ ($J_1 = 1$) and the system sizes $N = L \times L$ with (a) $L = 6$, (b) $L = 8$, and periodic boundary conditions. Gray lines denote values of ω for which the energy eigenvalues for the ground and first excited states are calculated.

$(E_0^{\text{ED}} - E_0^{\text{ansatz}})/E_0^{\text{ED}} \approx 0.1$. However, the energy gap $\Delta = E_1^{\text{ansatz}} - E_0^{\text{ansatz}} \approx 0.471$ is very close to the value obtained by exact diagonalization and GFMCSR technique [41,45].

For $L = 8$, influence of the statistical error in the stochastic algorithm is much smaller, and therefore better estimates for E_0 and E_1 are obtained from minimization of Ω . For the first excited-energy state, the minimum of Ω is found at $\omega = -31.75$ and corresponding energy estimate is $E_1^{\text{ansatz}}/L^2 = -0.488457$. The value of ω_0 for which $\Omega(\psi_0, \omega_0) \approx \Omega(\psi_1, \omega_1)$ is $\omega_0 = -32.1$ for $L = 8$ and the ground-state energy estimate is $E_0^{\text{ansatz}}/L^2 = -0.493472$. This value differs only by $\approx 1\%$ from the value obtained by the cluster update algorithm for TPSs [35] ($E_0^{\text{TPS}}/L^2 = -0.4984(2)$, $(E_0^{\text{ansatz}} - E_0^{\text{TPS}})/E_0^{\text{TPS}} \approx 0.00993$). For the energy gap, we obtain $\Delta = E_1^{\text{ansatz}} - E_0^{\text{ansatz}} \approx 0.321$ ($\Delta/L^2 \approx 0.005$), which is in agreement with previously obtained results calculated with the GFMCSR technique [41] and VMC combined with the quantum number projection [45].

We also note that variance (Fig. 6) of the optimized ansatz wave function obtained by minimization of Ω is quite large. This could perhaps be corrected by choosing a more complex ansatz wave function that would better mimic correlations built into exact eigenvalues, which is one of the directions for future research.

V. CONCLUSIONS

We have investigated applicability of the VMC method for direct optimization of energy eigenstates combined with appropriate ansatz for many-body wave function to study properties of complex strongly correlated and frustrated quantum systems. To demonstrate the applicability of the model, we have calculated energy gaps for the square lattice J_1 - J_2 frustrated Heisenberg model related to high- T_c superconducting materials. Our results are in good agreement with results obtained previously by other methods, particularly for larger system size where the influence of the statistical error included in the stochastic optimization scheme used in our calculations is smaller. The results therefore confirm applicability of the method to study properties of strongly correlated systems beyond the ground-state properties. This allows further insight into quantum phases and phase transitions in various correlated models.

Directions for our future research are more detailed calculations of the system properties for several values of the parameter J_2/J_1 close to quantum critical point, which is argued to be a deconfined quantum critical point [35]. We also plan to perform further calculations with several different, more complex ansatz states that could approximate correlations built into exact eigenstates better than the ansatz used in the present calculation (for example, RBM wave function with increased number of hidden units).

ACKNOWLEDGMENTS

We thank Hrvoje Buljan, Robert Pezer, Osor Barišić, and Ivan Balog for very useful suggestions and discussions. This work was supported by the QuantiXLie Centre of Excellence, a project cofinanced by the Croatian Government and European Union through the European Regional Development Fund, the Competitiveness and Cohesion Operational Programme (Grant No. KK.01.1.1.01.0004).

-
- [1] R. Orús, *Ann. Phys.* **349**, 117 (2014).
 - [2] N. Schuch, M. M. Wolf, F. Verstraete, and J. I. Cirac, *Phys. Rev. Lett.* **100**, 040501 (2008).
 - [3] L. Wang, I. Pižorn, and F. Verstraete, *Phys. Rev. B* **83**, 134421 (2011).
 - [4] A. Sfondrini, J. Cerrillo, N. Schuch, and J. I. Cirac, *Phys. Rev. B* **81**, 214426 (2010).
 - [5] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, and J. I. Cirac, *Phys. Rev. X* **8**, 011006 (2018).
 - [6] S. R. Clark, *J. Phys. A: Math. Theor.* **51**, 135301 (2018).
 - [7] J.-P. Song and R. T. Clay, *Phys. Rev. B* **89**, 075101 (2014).
 - [8] G. Carleo and M. Troyer, *Science* **355**, 602 (2017).
 - [9] H. J. Changlani, J. M. Kinder, C. J. Umrigar, and G. K.-L. Chan, *Phys. Rev. B* **80**, 245116 (2009).
 - [10] S. Al-Assam, S. R. Clark, C. J. Foot, and D. Jaksch, *Phys. Rev. B* **84**, 205108 (2011).
 - [11] F. Mezzacapo, *Phys. Rev. B* **86**, 045115 (2012).
 - [12] F. Mezzacapo, N. Schuch, M. Boninsegni, and J. I. Cirac, *New J. Phys.* **11**, 083026 (2009).
 - [13] E. Neuscamman, H. Changlani, J. Kinder, and G. K.-L. Chan, *Phys. Rev. B* **84**, 205132 (2011).
 - [14] T. Đurić, N. Chancellor, P. J. D. Crowley, P. Di Cintio, and A. G. Green, *Phys. Rev. B* **93**, 085143 (2016).
 - [15] T. Đurić, N. Chancellor, and I. F. Herbut, *Phys. Rev. B* **89**, 165123 (2014).
 - [16] A. W. Sandvik and G. Vidal, *Phys. Rev. Lett.* **99**, 220602 (2007).
 - [17] Z.-A. Jia, B. Yi, R. Zhai, Y.-C. Wu, G.-C. Guo, and G.-P. Guo, *Adv. Quantum Technol.* **2**, 1800077 (2019).
 - [18] H.-Q. Shi, X.-Y. Sun, and D.-F. Zeng, *Commun. Theor. Phys.* **71**, 1379 (2019).
 - [19] D.-L. Deng, X. Li, and S. Das Sarma, *Phys. Rev. X* **7**, 021021 (2017).
 - [20] J. Chen, S. Cheng, H. Xie, L. Wang, and T. Xiang, *Phys. Rev. B* **97**, 085104 (2018).
 - [21] S. Das Sarma, D.-L. Deng, and L.-M. Duan, *Phys. Today* **72**(3), 48 (2019).
 - [22] K. Choo, T. Neupert, and G. Carleo, *Phys. Rev. B* **100**, 125124 (2019).
 - [23] T. Westerhout, N. Astrakhantsev, K. S. Tikhonov, M. Katsnelson, and A. A. Bagrov, *Nat. Commun.* **11**, 1593 (2020).
 - [24] S. R. White, *Phys. Rev. Lett.* **69**, 2863 (1992).
 - [25] S. Liang and H. Pang, *Phys. Rev. B* **49**, 9214 (1994).
 - [26] T. Đurić, K. Biedroń, and J. Zakrzewski, *Phys. Rev. B* **95**, 085102 (2017).
 - [27] L. Wang and A. W. Sandvik, *Phys. Rev. Lett.* **121**, 107202 (2018).
 - [28] D. M. Ceperley and B. J. Alder, *Phys. Rev. Lett.* **45**, 566 (1980).
 - [29] L. Zhao and E. Neuscamman, *J. Chem. Theory Comput.* **12**, 3436 (2016).
 - [30] E. Neuscamman, *J. Chem. Phys.* **145**, 081103 (2016).
 - [31] N. S. Blunt and E. Neuscamman, *J. Chem. Phys.* **147**, 194101 (2017).
 - [32] J. A. R. Shea and E. Neuscamman, *J. Chem. Theory Comput.* **13**, 6078 (2017).
 - [33] N. S. Blunt and E. Neuscamman, *J. Chem. Theory Comput.* **15**, 178 (2019).
 - [34] S. D. Pineda Flores, and E. Neuscamman, *J. Phys. Chem. A* **123**, 1487 (2019).
 - [35] L. Wang, Z.-C. Gu, F. Verstraete, and X.-G. Wen, *Phys. Rev. B* **94**, 075143 (2016).
 - [36] H. J. Schulz and T. A. L. Ziman, *Europhys. Lett.* **18**, 355 (1992).
 - [37] H. J. Schulz, T. A. L. Ziman, and D. Poilblanc, *J. Phys. I France* **6**, 675 (1996).
 - [38] H.-C. Jiang, H. Yao, and L. Balents, *Phys. Rev. B* **86**, 024424 (2012).
 - [39] J. Sirker, Z. Weihong, O. P. Sushkov, and J. Oitmaa, *Phys. Rev. B* **73**, 184420 (2006).
 - [40] R. Darradi, O. Derzhko, R. Zinke, J. Schulenburg, S. E. Krüger, and J. Richter, *Phys. Rev. B* **78**, 214415 (2008).
 - [41] L. Capriotti and S. Sorella, *Phys. Rev. Lett.* **84**, 3173 (2000).
 - [42] L. Capriotti, F. Becca, A. Parola, and S. Sorella, *Phys. Rev. Lett.* **87**, 097201 (2001).
 - [43] T. Li, F. Becca, W. Hu, and S. Sorella, *Phys. Rev. B* **86**, 075111 (2012).
 - [44] J. Lou and A. W. Sandvik, *Phys. Rev. B* **76**, 104432 (2007).
 - [45] S. Morita, R. Kaneko, and M. Imada, *J. Phys. Soc. Jpn.* **84**, 024720 (2015).

- [46] L. Capriotti, *Int. J. Mod. Phys. B* **15**, 1799 (2001).
- [47] P. W. Anderson, *Science* **235**, 1196 (1987).
- [48] P. A. Lee, N. Nagaosa, and X. G. Wen, *Rev. Mod. Phys.* **78**, 17 (2006).
- [49] V. Kalmeyer and R. B. Laughlin, *Phys. Rev. Lett.* **59**, 2095 (1987).
- [50] X.-G. Wen, F. Wilczek, and A. Zee, *Phys. Rev. B* **39**, 11413 (1989).
- [51] T. Đurić and A. Lazarides, *Phys. Rev. B* **86**, 115135 (2012).
- [52] N. Read and S. Sachdev, *Phys. Rev. Lett.* **66**, 1773 (1991).
- [53] R. Moessner and S. L. Sondhi, *Phys. Rev. Lett.* **86**, 1881 (2001).
- [54] X.-G. Wen, *Phys. Rev. B* **44**, 2664 (1991).
- [55] H. Yao and S. A. Kivelson, *Phys. Rev. Lett.* **108**, 247206 (2012).
- [56] W.-J. Hu, F. Becca, A. Parola, and S. Sorella, *Phys. Rev. B* **88**, 060402(R) (2013).
- [57] F. Figueirido, A. Karlhede, S. Kivelson, S. Sondhi, M. Rocek, and D. S. Rokhsar, *Phys. Rev. B* **41**, 4619 (1990).
- [58] J. Richter and J. Schulenburg, *Eur. Phys. J. B* **73**, 117 (2010).
- [59] K. S. D. Beach, *Phys. Rev. B* **79**, 224431 (2009).
- [60] M. Mambrini, A. Läuchli, D. Poilblanc, and F. Mila, *Phys. Rev. B* **74**, 144422 (2006).
- [61] C.-P. Chou and H.-Y. Chen, *Phys. Rev. B* **90**, 041106(R) (2014).
- [62] P. Merchant, B. Normand, K. W. Krämer, M. Boehm, D. F. McMorrow, and Ch. Rüegg, *Nat. Phys.* **10**, 373 (2014).
- [63] S. Sachdev and B. Keimer, *Phys. Today* **64**(2), 29 (2011).
- [64] T. Senthil, A. Vishwanath, L. Balents, S. Sachdev, and M. P. A. Fisher, *Science* **303**, 1490 (2004).
- [65] S. Wenzel, L. Bogacz, and W. Janke, *Phys. Rev. Lett.* **101**, 127202 (2008).
- [66] S. Sachdev, The quantum phases of matter, in *Proceedings of the 25th Solvay Conference on Physics: Theory of the Quantum World*, edited by D. Gross, M. Henneaux, and A. Sevrin (World Scientific, Singapore, 2011).
- [67] B. Sutherland, *Phys. Rev. B* **37**, 3786(R) (1988).
- [68] K. S. D. Beach and A. W. Sandvik, *Nucl. Phys. B* **750**, 142 (2006).
- [69] S. Liang, B. Douçot, and P. W. Anderson, *Phys. Rev. Lett.* **61**, 365 (1988).
- [70] A. W. Sandvik and H. G. Evertz, *Phys. Rev. B* **82**, 024407 (2010).
- [71] N. Read and B. Chakraborty, *Phys. Rev. B* **40**, 7133 (1989).
- [72] Y. Tang, A. W. Sandvik, and C. L. Henley, *Phys. Rev. B* **84**, 174427 (2011).
- [73] H. Robbins and S. Monro, *Ann. Math. Stat.* **22**, 400 (1951).
- [74] J. C. Spall, in *Wiley Encyclopedia of Electrical and Electronics Engineering*, edited by J. G. Webster (Wiley, New York, 1999), Vol. 20.
- [75] W. Marshall, *Proc. R. Soc. London, Ser. A* **232**, 48 (1955).
- [76] U. Schollwöck, *Phys. Rev. B* **58**, 8194 (1998).
- [77] J. Richter, N. B. Ivanov, and K. Retzlaff, *Europhys. Lett.* **25**, 545 (1994).
- [78] A. Voigt, J. Richter, and N. B. Ivanov, *Phys. A (Amsterdam, Neth.)* **245**, 269 (1997).
- [79] C. Peterson and E. Hartman, *Neural Networks* **2**, 475 (1989).
- [80] T. K. Leen and J. E. Moody, *Phys. Rev. E* **56**, 1262 (1997).
- [81] S. Kirkpatrick, C. D. Gelatt, and M. P. Vecchi, *Science* **220**, 671 (1983).