Efficiency of neural-network state representations of one-dimensional quantum spin systems

Ruizhi Pan [1,*] and Charles W. Clark[1,2,†]

[1]*Joint Quantum Institute*, NIST/University of Maryland, College Park, Maryland 20742, USA
[2]*National Institute of Standards and Technology*, Gaithersburg, Maryland 20899, USA

Neural-network state representations of quantum many-body systems are attracting great attention and more rigorous quantitative analysis of their expressibility and complexity is warranted. Our analysis of the restricted Boltzmann machine (RBM) state representation of one-dimensional (1D) quantum spin systems provides new insight into their computational complexity. We define a class of long-range-fast-decay (LRFD) RBM states with quantifiable upper bounds on truncation errors and provide numerical evidence for a large class of 1D quantum systems that may be approximated by LRFD RBMs of at most polynomial complexities. These results lead us to conjecture that the ground states of a wide range of quantum systems may be exactly represented by LRFD RBMs or a variant of them, even in cases where other state representations become less efficient. At last, we provide the relations between multiple typical state manifolds. Our work proposes a paradigm for doing complexity analysis for generic long-range RBMs which naturally yields a further classification of this manifold. This paradigm and our characterization of their nonlocal structures may pave the way for understanding the natural measure of complexity for quantum many-body states described by RBMs and are generalizable for higher-dimensional systems and deep neural-network quantum states.

I. INTRODUCTION

Quantum machine learning is an emerging field that combines techniques in the disciplines of machine learning (ML) and quantum physics [1–7]. Research in this field takes three broad forms [2]: applications of classical ML techniques to quantum physical systems [4–10], quantum computing and algorithms for classical ML problems [11–14], and new ideas inspired by the intersection of the two disciplines [15,16]. In the field of learning quantum systems, there has been tremendous progress in applying ML techniques to identifying quantum phases and transitions [3,17–21], molecular modeling [22,23], quantum state tomography [24,25], and accelerating Monte Carlo simulations [26,27].

Here, we report results of an investigation of neural network quantum states in the context of quantum many-body physics [5–9,28–31], a subject of much recent interest. The core idea is to postulate an ansatz for the wave function in terms of a neural network (NN) [6], which targets a low-dimensional manifold in the exponentially large Hilbert space for state approximation [32], and apply ML algorithms to find a specific solution. The restricted Boltzmann machine (RBM)

[6–8,28,29] is a bipartite stochastic construct that combines the concepts of thermodynamic partition functions with those of classical artificial neural networks. RBMs have successfully represented a wide range of quantum states, such as low-lying eigenstates of quantum many-body-localized systems [6,33], code words of a stabilizer code [30,34,35] and chiral topological states [8,36,37].

While RBMs have demonstrated their power in numerical simulation, we have particular motivations to investigate the expressibility and complexity of the generic long-range RBMs, which are characterized by dense network architectures with full interlayer connectivity, in contrast to the so-called short-range or sparse RBMs [7,30,31]. First, the state approximators that are produced by RBM solutions returned by ML algorithms often feature a long-range form and a fast parameter decay, even when the exact RBM representations of the target states are unknown [30,31] or less efficient [8]. Increasing the number of hidden nodes captures more correlation features [6], thus increasing the approximation accuracy. The best approximators often have a form similar to that obtained by magnitude-based pruning [38,39] of a finite truncated RBM with infinitely many hidden nodes. These observations motivate the generalization of the RBM wave-function ansatz to an infinitely-many-hidden-node regime and the justification of the faithfulness of using these truncated long-range RBM approximators [40].

Another motivation for studying long-range RBMs stems from the central goal of exploring effective compressed state representations, which includes understanding the natural measure of complexity [7] and how the global physical information is encoded in that description [32]. There has been some work studying the relationship between RBMs and

other concepts about state representations, such as string-bond states [8], correlator product states [41] and tensor network states [37,42]. However, analysis of the effects of truncations through transforming RBMs into other representations may lead to redundancy and inconvenience, and it does not fully exploit the features of RBMs as architectures that naturally describe quantum states in a nonlocal manner [7,32]. Thus we choose to build a paradigm of direct analysis of the spatial complexity of long-range RBMs.

In this paper, we analyze the efficiency of long-range RBM state representation for 1D quantum spin systems. Our procedure is as follows.

(1) In Sec. II A, we generalize the RBM wave-function ansatz to an infinitely-many-hidden-node regime and define a subset of generic RBM states—the long-range-fast-decay (LRFD) RBM states, whose parameter conditions constrain the nonlocal interactions between spins (visible nodes) and virtual particles (hidden nodes).

(2) In Sec. II B, we derive an upper bound on truncation errors associated with two measures of state differences for the sequence of truncated LRFD RBM states. One measure is the l_2-norm of the state-vector difference and the other is a Hermitian-operator-based expectation-value difference.

(3) In Sec. II C, we identify the dependence of the spatial complexity for LRFD RBMs in state approximation on the decaying rates specified in the nonlocal interaction pattern.

(4) In Sec. III, we provide numerical evidence supporting a conjecture that the ground states of a wide range of 1D quantum spin systems, including some critical systems with logarithmic entanglement entropy, can be approximated by LRFD RBMs with the scaling of the spatial complexity being at most polynomial in both the system size and the inverse of approximation errors.

(5) In Sec. III, we also provide the relations between multiple typical state manifolds through which the importance of the concept of LRFD RBMs in efficiency analysis for state representation theory is manifested.

Our results offer evidence for the utility of RBMs in cases where other state parametrizations, such as matrix product states (MPSs), become less efficient. Our work actually proposes a paradigm of doing complexity analysis for general long-range RBMs, rather than limited to short-range or sparse RBMs, and naturally yields a further classification of this manifold based on the complexity scaling.

We find that the nonlocal structure of LRFD RBMs can be characterized by two conditions. These conditions are each determined by bounds associated with two degrees of freedom, defined within a framework of *levels* that is depicted in Fig. 1(a). One of the two degrees of freedom is a single-level decaying factor resembling localized orbitals and encoding information about correlations between spins (Sec. II D). The second is a level-decay factor, which has a significant influence on the complexity of the RBMs (Sec. II C).

This paradigm and our characterization of the nonlocal structures may promote the understanding of the natural measure of complexities for quantum many-body states described by RBMs and may be generalizable to higher-dimensional systems and to deep neural-network quantum states.
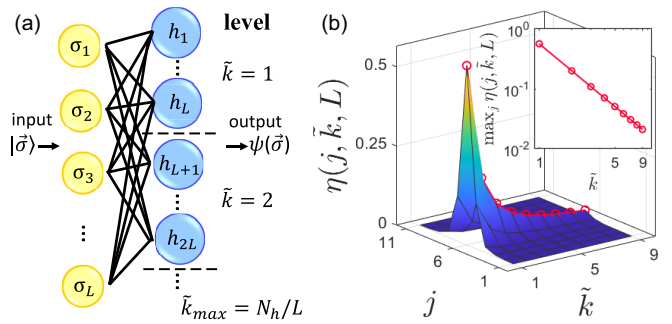

FIG. 1. (a) Network structure of RBMs as a wave-function ansatz for 1D quantum spin systems. Long-range RBMs usually implies full connectivity between the visible and hidden layer. The hidden layer is divided into N_h/L levels, each containing L hidden nodes. (b) Importance measure $\eta(j, \tilde{k}, L)$ for a LRFD RBM with translational symmetry. The RBM is constructed as Eqs. (24)–(26) show, where $\lambda(\tilde{k}) = \tilde{k}^{-\alpha_P}$, $\mu(r) = \frac{1}{2}\delta_Q r^{-\alpha_Q}$ for $r \neq 0$, $\mu(0) = \delta_Q = 0.5$, $\alpha_P = 0.75$, $\alpha_Q = 1.5$, $c_w = 1 + i$, $c_b = 0$, $a_0 = 0$, and $L = 11$. The inset plots the decay of the maximum of $\eta(j, \tilde{k}, L)$ among all j at each level with increasing $\tilde{k}$ on a log-log scale. The linearity of the curve reveals a power-law decaying of the "ridge" (red circles) of the 3D structure.

II. THE RESTRICTED BOLTZMANN MACHINE AS A WAVE-FUNCTION ANSATZ

We use the RBM as a wave-function ansatz for 1D quantum many-body spin-1/2 systems [5–7]. The RBM usually works as the building block for understanding and training deeper networks because of its relatively simple structure for inference and its power in parametric modeling as a universal approximator for discrete distribution [43]. As basic constructs of deep NNs, the RBMs have two *layers*. The first layer (a *visible* layer) represents a spin configuration $\vec{\sigma}$ in the usual way. Here, the vector $\vec{\sigma} = (\sigma_1, \ldots, \sigma_L)$ represents a system of L spins with $\sigma_j = \pm 1$ for $j = 1, \ldots, L$. The second layer is a *hidden* layer. It is composed of N_h nodes, denoted by a vector $\vec{h} = (h_1, \ldots, h_{N_h})$ with $h_k = \pm 1$ for $k = 1, \ldots, N_h$. The h_k's are introduced as auxiliary particles in the probability model; they play roles similar to those of virtual particles in the valence-bond picture for MPSs [32,44].

Given a specific spin configuration $\vec{\sigma}$, the RBM outputs the corresponding wave-function amplitude

$$\psi(\vec{\sigma}) = 2^{-N_h} \sum_{\{\vec{h}:h_k=\pm 1\}} \exp\left(\sum_{j=1}^{L} a_j \sigma_j + \sum_{k=1}^{N_h} b_k h_k \right.$$
$$\left. + \sum_{1 \leqslant j \leqslant L, 1 \leqslant k \leqslant N_h, j,k \in \mathbb{N}} W_{j,k} \sigma_j h_k \right) \quad (1)$$

$$= \prod_{j=1}^{L} e^{a_j \sigma_j} \prod_{k=1}^{N_h} \cosh\left(b_k + \sum_{j=1}^{L} \sigma_j W_{j,k} \right). \quad (2)$$

Here, a_j and b_k are the bias parameters for the jth spin and kth hidden node, respectively, $W_{j,k}$ is a weight parameter describing the interlayer interaction between the jth spin and kth hidden node, and $\mathbb{N}$ denotes the set of all natural numbers.

The a_j, b_k and $W_{j,k}$ are complex numbers. All such amplitudes defined on the computational basis yield a quantum state vector $|\Psi\rangle = \sum_{\vec{\sigma}} \psi(\vec{\sigma})|\vec{\sigma}\rangle$, where the summation is over all 2^L spin configurations. It is remarkable that we adopt the RBM form with a factor of 2^{-N_h}. This choice allows us to use infinitely many hidden nodes h_k as long as b_k and $W_{j,k}$ decay sufficiently fast to ensure the convergence of $\psi(\vec{\sigma})$ as $N_h \to \infty$ for fixed system sizes L. In other words, it ensures that adding hidden nodes with associated parameters (b_k and $W_{j,k}$) being zero will not change the value of the wave function. This choice will facilitate the asymptotic analysis as shown below.

As mentioned in Sec. I, the RBMs solved by relevant ML algorithms to approximate target states often feature a long-range form and a fast parameter decay. As more hidden nodes are added to the network, the RBM can capture more correlation features [6], thus leading to higher accuracy in approximation. The parameter decay is manifested by the decay of weight parameters $W_{j,k}$ with an increasing index separation $|j - k|$ as well as the decay of b_k with increasing k.

In this work, we assume N_h to be an integer multiple of L which will facilitate the scaling analysis, especially for translationally invariant systems. When N_h is not an integer multiple of L, we can simply fill the last fragment with hidden nodes associated with zero-value parameters without influencing the wave-function values. We divide the hidden layer into multiple *levels*, each of which contains L hidden nodes [Fig. 1(a)]. Thus there are totally N_h/L levels while the ratio N_h/L is called the hidden-unit density in some references [6]. We will show that hidden nodes at the same level can capture the same correlation feature by performing an algorithm to reorder all hidden nodes for general RBMs. This point will be further clarified when we use the RBM form with translational symmetry to represent the ground states of 1D translationally invariant quantum systems as shown below [6].

One example of the quantum states that can be exactly represented by short-range RBMs [30,31] is the 1D symmetry-protected topological (SPT) cluster state. The Hamiltonian of the SPT cluster system is defined on a 1D L-site lattice with periodic boundary conditions as $\hat{H}_{\text{cluster}} = -\sum_{j=1}^{L} \hat{\sigma}_{j-1}^z \hat{\sigma}_j^x \hat{\sigma}_{j+1}^z$, where $\hat{\sigma}^x$ and $\hat{\sigma}^z$ are Pauli matrices. A conventional r_0-range RBM is defined as an RBM satisfying $W_{j,k} = 0$ for any $|j - k| > r_0$. A short-range RBM usually refers to an r_0-range RBM with r_0 being a small constant independent of the system size L. It was shown in Refs. [30,31] that the ground state of $\hat{H}_{\text{cluster}}$ can be exactly represented by a 1-range RBM with L hidden nodes defined as

$$a_j = 0 \quad (\text{for any } j \in \{1, 2, \ldots, L\}),$$
$$b_k = i\pi/4, \quad W_{k-1,k} = i\pi/2, \quad W_{k,k} = 3i\pi/4,$$
$$W_{k+1,k} = i\pi/4 \quad (\text{for any } k \in \{1, 2, \ldots, L\}),$$
$$W_{j,k} = 0 \quad (\text{for } |j - k| > 1), \tag{3}$$

by using the stabilizer nature of the system to decrease the number of equation constraints for parameters from exponential to linear in L. Using our language of levels, this RBM just has one level and its weight parameters at this single level have a support of very short length which implies its quantum entanglement satisfying an area law. Moreover,

the translational symmetry of the system is inherited by the RBM form. The parameter patterns of this RBM also have a translational symmetry, which means that its parameters for different hidden nodes can be generated by the action of a translational-symmetry transformation operator on those for a single hidden node [6].

Inspired by the extensibility of the system of equations (3) with growing system sizes and considering the need to capture more correlation features [6] and stronger quantum entanglement between subsystem blocks [31], we expect that the RBM representation of general quantum states has multiple, possibly infinitely many, levels and the length of the support of weight parameters at each level may increase from a small constant to the maximum length L. This motivates us to analyze generic long-range RBMs with properly specified nonlocal interactions between spins and hidden nodes (virtual particles).

A. Long-range-fast-decay RBMs

We now discuss aspects of the nonlocal structure of LRFD RBMs that were summarized at the end of Sec. I. This leads to specific definitions of the two conditions that were mentioned there.

We begin by generalizing the RBM wave-function ansatz to an infinitely-many-hidden-node regime. An RBM state $|\Psi^{(L,\infty)}\rangle$ with infinitely many hidden nodes and a system size L can be defined as

$$|\Psi^{(L,\infty)}\rangle = \sum_{\vec{\sigma}} \psi^{(L,\infty)}(\vec{\sigma})|\vec{\sigma}\rangle, \tag{4}$$

where

$$\psi^{(L,\infty)}(\vec{\sigma}) = \prod_{j=1}^{L} e^{a_j^{(L)}\sigma_j} \prod_{k=1}^{\infty} \cosh\left(b_k^{(L)} + \sum_{j=1}^{L} \sigma_j W_{j,k}^{(L)}\right). \tag{5}$$

Its corresponding truncated-RBM sequence is defined as $\{|\Psi^{(L,N_h)}\rangle\}$, where

$$|\Psi^{(L,N_h)}\rangle = \sum_{\vec{\sigma}} \psi^{(L,N_h)}(\vec{\sigma})|\vec{\sigma}\rangle \tag{6}$$

and

$$\psi^{(L,N_h)}(\vec{\sigma}) = \prod_{j=1}^{L} e^{a_j^{(L)}\sigma_j} \prod_{k=1}^{N_h} \cosh\left(b_k^{(L)} + \sum_{j=1}^{L} \sigma_j W_{j,k}^{(L)}\right) \tag{7}$$

is constructed by removing the hyperbolic cosine terms with $k \geqslant N_h + 1$ from $\psi^{(L,\infty)}(\vec{\sigma})$.

Then, we define a subset of generic RBM states with infinitely many hidden nodes—long-range-fast-decay (LRFD) RBM states—as the RBMs whose parameters satisfy the following two conditions.

Condition 1 (boundedness of $W_{j,k}$). There exists an L-independent integer $\tilde{k}_s \in \mathbb{N}$ and three nonnegative monotonically decreasing real functions $\lambda_R(\tilde{k})$, $\lambda_I(\tilde{k})$, and $\mu(r)$ such that, after a reordering of all hidden nodes, for all $k > \tilde{k}_s L$,

$$\left|\text{Re}\left(W_{j,k}^{(L)}\right)\right| \leqslant \lambda_R(\tilde{k})\mu(|j - j_c|_{\text{circ}}), \tag{8}$$

$$\left|\text{Im}\left(W_{j,k}^{(L)}\right)\right| \leqslant \lambda_I(\tilde{k})\mu(|j - j_c|_{\text{circ}}), \tag{9}$$

where $\tilde{k} \in \{1, 2, \ldots, N_h/L\}$ designates the numerical index of levels; j_c, the center spin for the kth hidden node, denotes the site index of the spin with which the interaction of the kth hidden node reaches its maximum among all $j \in \{1, 2, \ldots, L\}$; $|m|_{\mathrm{circ}} = \min\{m, L - m\}$ in accordance with the periodic boundary conditions; and $r \in \{0, 1, \ldots, (L - 1)/2\}$ denotes the distance between j and j_c assuming L is odd without influencing the validity of the following asymptotic analysis. The functions $\lambda_R(\tilde{k})$, $\lambda_I(\tilde{k})$ and $\mu(r)$ satisfy the conditions that there exist finite L-independent nonnegative constants P_0 and μ_0 such that

$$\sum_{\tilde{k}=\tilde{k}_s+1}^{\infty} \left(\lambda_R^2(\tilde{k}) + \beta_1^2 \lambda_I^2(\tilde{k})\right) = P_0 < \infty, \tag{10}$$

$$\mu(r) \leqslant \mu_0 < \infty \quad (\text{for all } r \geqslant 0), \tag{11}$$

where $\beta_1 = 3\sqrt{2 \ln 2}/\pi$ is found in the convergence proof given in Appendix A.

We provide an interpretation of each new variable as follows. After a reordering of all hidden nodes which is usually associated with sorting based on the value of $|\mathrm{Re}(W_{j,k}^{(L)})|^2 + \beta_1^2 |\mathrm{Im}(W_{j,k}^{(L)})|^2$, starting from the level $\tilde{k}_s + 1$, $\tilde{k} = \tilde{k}(k)$ and $j_c = j_c(k)$ are both functions of k and the correspondence between the pair $(\tilde{k}, j_c)$ and k is a bijective map. It means that every hidden node with $k > \tilde{k}_s L$ is associated with a unique pair and thus can be uniquely positioned in the RBM network after the reordering [Fig. 1(a)]. The hidden nodes capturing the same correlation feature are grouped at the same level so that the new indices of these hidden nodes characterized by the pair $(\tilde{k}, j_c)$ actually manifest the level of correlations. This characterization can also facilitate a symmetry manifestation for quantum states holding translational symmetry. The reordering step is to solve the problem that ML algorithms with a stochastic nature are often unable to automatically group the hidden nodes according to level stratification and their site positions usually exhibit randomness. The condition $\tilde{k}_s = 0$ implies that all hidden nodes satisfy the boundedness conditions so that the level stratification can be applied to the whole hidden layer [Fig. 1(a)].

Condition 2 (boundedness of b_k). After the same reordering of all hidden nodes that is described in Condition 1, for all $k > \tilde{k}_s L$,

$$\left|\mathrm{Re}\left(b_k^{(L)}\right)\right| \leqslant \lambda_R(\tilde{k})\mu(0), \tag{12}$$

$$\left|\mathrm{Im}\left(b_k^{(L)}\right)\right| \leqslant \lambda_I(\tilde{k})\mu(0). \tag{13}$$

The definition of LRFD RBMs should be understood from the point of view of state manifolds [32,45]. A state manifold for quantum many-body states usually refers to a subspace of the whole Hilbert space spanned by a parameterized wave-function family [45] and thus is a set containing specific types of quantum states. So the manifold of LRFD RBMs can be defined as a space spanned by all parameterized wave functions, every one of which belongs to a quantum-state sequence associated with a varying system size and satisfying the above conditions 1 and 2. One LRFD-RBM state refers to an element in this manifold. So this definition is in the same spirit as the definition of MPSs with different scaling laws [32,44].

Condition 1 gives an upper bound on the magnitude of RBM weight parameters and actually provides a description of the nonlocal interaction between spins and hidden nodes (virtual particles). It requires that $|\mathrm{Re}(W_{j,k}^{(L)})|$ and $|\mathrm{Im}(W_{j,k}^{(L)})|$ are upper bounded, respectively, by the products $\lambda_R(\tilde{k})\mu(r)$ and $\lambda_I(\tilde{k})\mu(r)$. The monotonically decreasing functions $\lambda_R(\tilde{k})$ and $\lambda_I(\tilde{k})$ can be regarded as level-decay factors, while $\mu(r)$ is a factor describing the decay due to the increase of the distance between the spin-site index (j) and the corresponding spin-site index of the center spin ($j_c(k)$) for the kth hidden node. The function $\mu(r)$ has a localization feature and resembles a single-modal localized orbital in the physics of periodic potentials, such as Wannier modes [46], which can be reflected by its monotonically decreasing with increasing r. So this description can effectively capture the parameter decays induced by both the level increase and the growth of system size, providing two degrees of freedom in characterizing the nonlocal interaction pattern. The separate treatments for the real and imaginary parts originate from their inequivalent positions in the RBM wave-function form, which is shown in Appendix A.

Condition 2 implies that the contribution of b_k-related terms can be upper bounded by the largest $W_{j,k}$-related terms at each level so that the $W_{j,k}$ weight parameters play a dominant role in the asymptotic analysis (Appendix A). Since there is often a degree of freedom in choosing the value of $\mu(0)$, condition 2 can be satisfied for a wide range of RBM states.

Conditions 1 and 2 are proposed to ensure the convergence of the state vector [Eq. (5)] and provide a clear quantification for the rate of parameter decays, on the basis of which a complexity analysis can be conducted. A rigorous proof of the convergence of the state vector when conditions 1 and 2 are satisfied is given in Appendix A. This proof is important not only because it ensures that the generalization of RBMs to an infinitely-many-hidden-node regime makes sense by defining them as the limits of some infinite sequences, but also because it introduces the key mathematical tricks and concepts that are necessary for analyzing the effects of truncations.

The core idea of the proof is that we can prove the sequence $\{\psi^{(L,nL)}(\vec{\sigma}) : n \in \mathbb{N}\}$ is a Cauchy sequence in the field of complex numbers $\mathbb{C}$ [47]. This proof is inspired by the fact that, when $b_k^{(L)}$ and $W_{j,k}^{(L)}$ decay sufficiently fast, the complex-valued ratio $\psi^{(L,(n+m)L)}(\vec{\sigma})/\psi^{(L,nL)}(\vec{\sigma})$ will quickly fall into the neighborhood of the point $z = 1$ in the complex plane as n increases. So we derive an upper and lower bound on the ratio's modulus $|\psi^{(L,(n+m)L)}(\vec{\sigma})/\psi^{(L,nL)}(\vec{\sigma})|$ which converge to 1 and an upper bound on the magnitude of its argument $|\arg(\psi^{(L,(n+m)L)}(\vec{\sigma})/\psi^{(L,nL)}(\vec{\sigma}))|$ which converges to 0 as n increases. Then we show that the corresponding magnitude sequence $\{|\psi^{(L,nL)}(\vec{\sigma})|\}$ and the argument sequence $\{\arg(\psi^{(L,nL)}(\vec{\sigma}))\}$ are Cauchy sequences in the field of real numbers $\mathbb{R}$, thus $\{\psi^{(L,nL)}(\vec{\sigma})\}$ is a Cauchy sequence in $\mathbb{C}$.

B. Effects of wave-function truncation for fixed system sizes

We derive upper bounds on truncation errors associated with two measures of state differences for the sequence of truncated LRFD RBM states. Define $\varepsilon(L, N_h)$ to be a specific type of truncation error for using $|\Psi^{(L,N_h)}\rangle$ to approximate $|\Psi^{(L,\infty)}\rangle$.

A natural measure of state differences is the square of the l_2-norm [48] of the state-vector difference, $\||\tilde{\Psi}^{(L,\infty)}\rangle - |\tilde{\Psi}^{(L,N_h)}\rangle\|_2^2$, where the tilde symbol is used to represent corresponding states after a normalization operation. It is remarkable that the RBM wave-function ansatz is not automatically normalized and an estimation of the normalization factor $\langle\Psi|\Psi\rangle$ is important and often tricky as shown in Appendix B. This measure of truncation errors is adopted in fundamental works about the faithfulness and efficiency of other wave-function ansätze, such as MPSs [40,49]. So our work is potentially a crucial step towards making a comparison between the efficiencies of RBMs and other state representations.

A second measure of state differences is a Hermitian-operator-based expectation-value difference defined as

$$
\begin{aligned}
&|\langle\hat{B}\rangle^{(L,\infty)} - \langle\hat{B}\rangle^{(L,N_h)}| \\
&= |\langle\tilde{\Psi}^{(L,\infty)}|\hat{B}|\tilde{\Psi}^{(L,\infty)}\rangle - \langle\tilde{\Psi}^{(L,N_h)}|\hat{B}|\tilde{\Psi}^{(L,N_h)}\rangle|. \quad (14)
\end{aligned}
$$

Here, $\hat{B}$ can be any Hermitian operator of the form $\hat{B} = \bigotimes_{j=1}^{L}\hat{\sigma}_j^{(m_j)}$, where $\bigotimes$ is the tensor product symbol, $m_j \in \{0, 1, 2, 3\}$, $\hat{\sigma}_j^{(0)} = I_{2\times 2}$ is a 2×2 identity matrix, and $\{\hat{\sigma}_j^{(1)}, \hat{\sigma}_j^{(2)}, \hat{\sigma}_j^{(3)}\}$ denote the Pauli matrices. We also use this measure as $\{\hat{\sigma}_j^{(m)} : m = 0, 1, 2, 3\}$ is a complete basis set for the local Hilbert space for the jth spin and a wide range of typical physical observables, such as spin correlations and total energy, correspond to Hermitian operators of such type or linear combinations of polynomially many such operators.

Then we can prove a lemma which provides upper bounds on truncation errors of the above two types for the sequence of truncated LRFD RBM states.

Lemma 3 (upper bounds on truncation errors). For LRFD RBMs satisfying conditions 1 and 2, after the same reordering of all hidden nodes described in condition 1, there exists $n_{\Theta}(L) > \tilde{k}_s$ such that, for all $N_h > n_{\Theta}(L)L$,

$$
\||\tilde{\Psi}^{(L,\infty)}\rangle - |\tilde{\Psi}^{(L,N_h)}\rangle\|_2^2 \leqslant F_1(LQ(L)P(N_h/L)), \quad (15)
$$

$$
|\langle\hat{B}\rangle^{(L,\infty)} - \langle\hat{B}\rangle^{(L,N_h)}| \leqslant F_2(LQ(L)P(N_h/L)), \quad (16)
$$

where

$$
F_1(x) = 2 - 2\exp\left[-2\left(1 + \beta_1^2\right)x\right]\cos(4\beta_2 x) \quad (17)
$$

$$
= c_1 x + \mathcal{O}(x^2) \quad (\text{as } x \to 0), \quad (18)
$$

$$
F_2(x) = \max\left\{|\exp(4x) - 1|, \left|1 - \exp\left(-4\beta_1^2 x\right)\right|\right\}
$$
$$
\qquad + \max\left\{[\exp(8x) - 2\exp(4x)\cos(8\beta_2 x) + 1]^{1/2},\right.
$$
$$
\qquad \left[\exp\left(-8\beta_1^2 x\right) - 2\exp\left(-4\beta_1^2 x\right)\cos(8\beta_2 x)\right.
$$
$$
\qquad \left.\left. + 1\right]^{1/2}\right\} \quad (19)
$$

$$
= c_2 x + \mathcal{O}(x^{3/2}) \quad (\text{as } x \to 0). \quad (20)
$$

$$
P(m) = \sum_{\tilde{k}=m+1}^{\infty}\lambda^2(\tilde{k}) \quad (m \geqslant \tilde{k}_s, m \in \mathbb{N}), \quad (21)
$$

$$
Q(L) = \left(\sum_{r=0}^{(L-1)/2}\mu(r)\right)^2, \quad (22)
$$

the relevant constants are $\beta_2 = 3\sqrt{3}/\pi$, $c_1 = 4(1 + \beta_1^2)$ and $c_2 = 4\beta_1^2 + 4(\beta_1^4 + 4\beta_2^2)^{1/2}$, $n_{\Theta}(L)$ can be estimated by inequality (B11), and we have assumed that $\lambda_R(\tilde{k}) = \lambda_I(\tilde{k}) = \lambda(\tilde{k})$ for simplicity which holds throughout the following discussion.

The proof is given in Appendix B which uses arguments similar to those described in the proof for the convergence of LRFD RBMs. Based on the intuition that the ratio $\psi^{(L,\infty)}(\vec{\sigma})/\psi^{(L,N_h)}(\vec{\sigma})$ will fastly converge to 1 with increasing N_h, we derive an upper bound $\sqrt{R_1}$ and a lower bound $\sqrt{R_2}$ on the ratio's modulus $|\psi^{(L,\infty)}(\vec{\sigma})/\psi^{(L,N_h)}(\vec{\sigma})|$ and an upper bound Θ on the magnitude of its argument $|\arg(\psi^{(L,\infty)}(\vec{\sigma})/\psi^{(L,N_h)}(\vec{\sigma}))|$. The two types of truncation errors can be upper bounded using these three variables and the two upper bounds can be finally expressed as functions $[F_1(x)$ and $F_2(x)]$ of $LQ(L)P(N_h/L)$ which decreases to zero with increasing N_h and fixed L. The idea of the proof is shown schematically in Fig. 2(c).

Based on our description of the nonlocal interactions between spins and virtual particles and using the language of levels, $P(N_h/L)$ is a summation of all level-decay factors for hidden nodes at levels starting from $\tilde{k} = N_h/L + 1$ to $\tilde{k} = \infty$, while $Q(L)$ corresponds to the localized "orbitals" at every single level and contributes a factor reflecting the pure influence of system-size growing regardless of levels. The two different types of truncation errors correspond to two different forms of the function $F(x)$, but both of them are analytic at the point $x = 0$.

We give the scaling of truncation errors in N_h as below. It can be obtained that if $Q(x) = \mathcal{O}(q(x))$ as $x \to \infty$, $P(x) = \mathcal{O}(1/p_d(x))$ as $x \to \infty$, and $F(x) = \mathcal{O}(f(x))$ as $x \to 0$, then

$$
\varepsilon(L, N_h) = \mathcal{O}\left(f\left(\frac{Lq(L)}{p_d(N_h/L)}\right)\right) \quad (\text{as } N_h \to \infty). \quad (23)
$$

Our construction of LRFD RBMs and theoretical analysis of the truncation errors can be further clarified with results from numerical computations. We can construct LRFD RBMs with translation symmetry whose parameters exactly satisfy

$$
W_{j,k}^{(L)} = c_w\lambda(\tilde{k})\mu(|j - j_c|_{\text{circ}}), \quad (24)
$$

$$
b_k^{(L)} = c_b\lambda(\tilde{k})\mu(0), \quad (25)
$$

$$
a_j^{(L)} = a_0 \quad (26)
$$

for any $1 \leqslant j \leqslant L$, $1 \leqslant k \leqslant N_h$, where a_0, c_w and c_b are complex constants with $|c_b| \leqslant |c_w|$ to satisfy condition 2, $\tilde{k} = \tilde{k}(k) = \lceil k/L\rceil$, $j_c = j_c(k) = k - (\tilde{k} - 1)L$, N_h is an integer multiple of L, $\lceil x\rceil$ denotes the ceiling function, and $\tilde{k}_s = 0$ in this case. It can be shown that such an RBM form can be directly transformed into the RBM form proposed to represent ground states of 1D translationally invariant systems [6] for any finite N_h but we generalize it to an infinitely-many-hidden-node regime ($N_h \to \infty$). Since the parameters for different hidden nodes can be generated by the action of a translational-symmetry transformation operator on those for a single hidden node, we just need to focus on one representative hidden node for each level. So we propose an importance measure $\eta(j, \tilde{k}, L)$ to measure the importance of a set of edges which
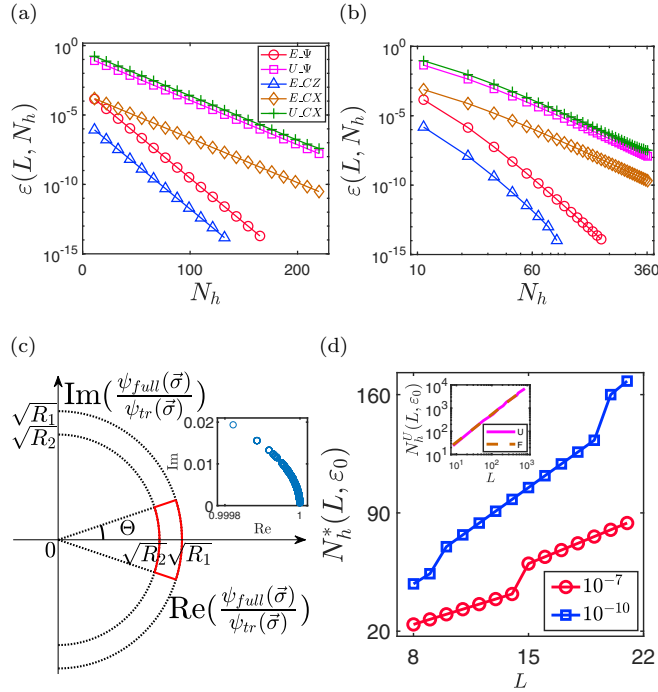
FIG. 2. [(a) and (b)] Comparison between the exact and estimated truncation errors $\varepsilon(L, N_h)$ as a function of N_h with fixed L for 1D SPT cluster states with a perturbation part. E_Ψ: exact values, first-type truncation errors; U_Ψ: upper-bound-based estimation, first-type; E_CZ: exact, second-type, $\hat{B} = \hat{\sigma}_1^z \hat{\sigma}_2^z$; E_CX: exact, second-type, $\hat{B} = \hat{\sigma}_1^x \hat{\sigma}_2^x$; U_CX: upper-bound-based estimation, second-type, $\hat{B} = \hat{\sigma}_1^x \hat{\sigma}_2^x$. The perturbation part is constructed as Eqs. (24)–(26) show. $\mu(r) = \frac{1}{2}\delta_Q r^{-\alpha_Q}$ for $r \neq 0$, $\mu(0) = \delta_Q = 0.1$, $\alpha_Q = 3$, $c_w = c_b = 1 + i$, $a_0 = 0$, and $L = 11$. (a) Exponential decay of $\lambda(\tilde{k})$ (vertical axis: logarithmic scale). $\lambda(\tilde{k}) = 0.2\delta_P^{-(\tilde{k}-1)}$ with $\delta_P = 1.5$. (b) Power-law decay of $\lambda(\tilde{k})$ (on a log-log scale). $\lambda(\tilde{k}) = \tilde{k}^{-\alpha_P}$ with $\alpha_P = 3$. (c) Schematic interpretation of variables used in proofs. The inset in (c) shows the distribution of data points of the ratio $\psi^{(L,\infty)}(\vec{\sigma})/\psi^{(L,N_h)}(\vec{\sigma})$ with $N_h = L$, which corresponds to a truncation removing hidden nodes starting from the second level. The parameter setting is the same as that in (b). The data points are all localized in the neighborhood of $z = 1$ in the complex plane enclosed by the red solid curves in (c) as we analyzed. (d) Scaling of $N_h^*(L, \varepsilon_0)$ in L for two fixed values of the first-type truncation errors ε_0 with the same parameter setting as in (b). Red circle: $\varepsilon_0 = 10^{-7}$; blue square: $\varepsilon_0 = 10^{-10}$. The inset in (d) shows the scaling of N_h estimated based on our upper bounds with $\varepsilon_0 = 10^{-3}$. Magenta solid: using exact values of upper bounds; brown dashed: using leading-order estimations. The two curves almost coincide.

is defined as

$$
\begin{aligned}
\eta(j, \tilde{k}, L) &= \left| \text{Re}\left(W^{(L)}_{j,(L+1)/2+(\tilde{k}-1)L} \right) \right|^2 \\
&\quad + \beta_1^2 \left| \text{Im}\left(W^{(L)}_{j,(L+1)/2+(\tilde{k}-1)L} \right) \right|^2
\end{aligned} \tag{27}
$$

and present it as a function of the spin-site index j and level index $\tilde{k}$. Its 3D structure can reflect the decay of both $\lambda(\tilde{k})$ and $\mu(r)$ while the center of the "orbital" at every level is localized around $j = (L + 1)/2$. So a plotting of the peak at every level as a function of the level index ($\tilde{k}$) can reflect the decay of

$\lambda(\tilde{k})$. One example of such LRFD RBM with a power-law decaying $\lambda(\tilde{k})$ is shown in Fig. 1(b).

We show the two types of truncation errors $\varepsilon(L, N_h)$ as a function of N_h with fixed L for 1D SPT cluster states with a perturbation part [Figs. 2(a) and 2(b)]. It means that the RBM is constructed as a summation of the setting defined in the system of equations (3) and a perturbation part specified as Eqs. (24)–(26) show. The numerical results for $\lambda(\tilde{k})$ with exponential and power-law decays are given. As described above, the 1D SPT cluster states can be exactly represented by a short-range (1-range) RBM [30]. Using our description, its RBM representation just has one level, and the corresponding $\lambda(\tilde{k})$ and $\mu(r)$ quickly go down to zero for $\tilde{k} > 1$ and $r > 1$. The addition of the perturbation part makes the composite RBM a LRFD RBM so that we can study the truncation errors. We give the results for both types of truncation errors and let $\hat{B}$ be the operator of spin correlations between spin 1 and 2 in z and x directions.

Our numerical experiments on the scaling of the truncation errors in N_h with fixed L are well upper bounded by our estimations given in inequalities (15) and (16), which substantiates our theoretical analysis. Those experiments also indicate that our estimations in Eq. (23) correctly capture the asymptotic properties of $\varepsilon(L, N_h)$ with varying N_h. Moreover, the fact that the curve of exact $\varepsilon(L, N_h)$ and that of our estimation associated with $\hat{B} = \hat{\sigma}_1^x \hat{\sigma}_2^x$ have exactly the same slope implies that our estimation in Eq. (23) gives an asymptotically optimal upper bound. It means that, for the second-type truncation errors [inequality (16)], there is still room to improve the constant prefactors in our estimation, but we cannot qualitatively further improve the upper bound. In comparison, there is room to both qualitatively improve the upper bound and improve the constant prefactors for the first-type truncation errors [inequality (15)].

C. Scaling of complexity

We can investigate the scaling of spatial complexity in system sizes for LRFD RBMs as the results in Secs. II A and II B still hold for varying L. We give an upper-bound estimation of the complexity of RBM representations which depends on the asymptotic behavior at $x = \infty$ of the functions $P(x)$ [Eq. (21)] and $Q(x)$ [Eq. (22)] and thus is determined by the decaying rates specified by $\lambda(\tilde{k})$ and $\mu(r)$.

Define the minimum N_h to achieve a sufficiently small approximation error ε_0 as

$$
N_h^*(L, \varepsilon_0) = \inf\{N_h : \varepsilon(L, N_h) \leqslant \varepsilon_0\}. \tag{28}
$$

Using lemma 3, the sufficient condition for $\varepsilon(L, N_h) \leqslant \varepsilon_0$ is that the corresponding upper bound on truncation errors is no larger than ε_0. So this provides one way to get an upper bound on $N_h^*(L, \varepsilon_0)$ for LRFD RBMs. It can be shown that

$$
N_h^*(L, \varepsilon_0) = \mathcal{O}\left(L\, p_d^{-1}\left(\frac{L\, q(L)}{f^{-1}(\varepsilon_0)} \right) \right) \quad (\text{as } L \to \infty), \tag{29}
$$

where $q(x)$, $p_d(x)$, and $f(x)$ are functions to specify the asymptotic behaviors of $Q(x)$, $P(x)$, and $F(x)$ as defined above and the superscript "-1" denotes the inverse of the corresponding function. This upper-bound estimation is usually asymptotically larger than, thus not influenced by, the $n_\Theta(L)L$.

TABLE I. Complexity estimations for distinct typical settings of $\mu(r)$ and $\lambda(\tilde{k})$. "−" in the $\mu(r)$ column denotes all $\mu(r)$ functions that make $Q(L)$ converge as $L \to \infty$. "−" in the $\lambda(\tilde{k})$ column denotes all $\lambda(\tilde{k})$ functions that make $P(N_h/L)$ have the asymptotic behavior of $\mathcal{O}(1/\ln(N_h/L))$ as $N_h \to \infty$. Note that $\delta_P > 1$ and $\alpha_P > 1/2$ in these settings.

Manifold	$\mu(r)$	$Q(L)$	$\lambda(\tilde{k})$	$P(N_h/L)$	$N_h^*(L, \varepsilon)$
$S_2^{(1)}$	−	Converge	$\delta_P^{-\tilde{k}}$	$\mathcal{O}(\delta_P^{-2N_h/L})$	$\mathcal{O}(L\ln(L/\varepsilon))$
$S_2^{(2)}$	−	Converge	$\tilde{k}^{-\alpha_P}$	$\mathcal{O}((N_h/L)^{1-2\alpha_P})$	$\mathcal{O}((L^{2\alpha_P}/\varepsilon)^{1/(2\alpha_P-1)})$
$S_2^{(3)}$	$r^{-1}\ (r \geqslant 1)$	$\mathcal{O}((\ln L)^2)$	$\delta_P^{-\tilde{k}}$	$\mathcal{O}(\delta_P^{-2N_h/L})$	$\mathcal{O}(L\ln(L/\varepsilon))$
$S_2^{(4)}$	$r^{-1}\ (r \geqslant 1)$	$\mathcal{O}((\ln L)^2)$	$\tilde{k}^{-\alpha_P}$	$\mathcal{O}((N_h/L)^{1-2\alpha_P})$	$\mathcal{O}((L^{2\alpha_P}(\ln L)^2/\varepsilon)^{1/(2\alpha_P-1)})$
$S_2^{(5)}$	$\to \mu_\infty > 0\ (r \to \infty)$	$\mathcal{O}(L^2)$	$\delta_P^{-\tilde{k}}$	$\mathcal{O}(\delta_P^{-2N_h/L})$	$\mathcal{O}(L\ln(L/\varepsilon))$
$S_2^{(6)}$	$\to \mu_\infty > 0\ (r \to \infty)$	$\mathcal{O}(L^2)$	$\tilde{k}^{-\alpha_P}$	$\mathcal{O}((N_h/L)^{1-2\alpha_P})$	$\mathcal{O}((L^{2\alpha_P+2}/\varepsilon)^{1/(2\alpha_P-1)})$
$S_2^{(7)}$	−	Converge	−	$\mathcal{O}(1/\ln(N_h/L))$	$\mathcal{O}(L\exp(L/\varepsilon))$

Rich information can be extracted from Eq. (29). First, the first factor L comes from our assumption that N_h is an integer multiple of the system size L and the second factor L in front of $q(L)$ is extracted using the translational symmetry of the wave function. So these two factors reflect the growing system sizes and the remaining factors reflect the distinction in complexity for different LRFD RBMs.

Second, $P(N_h/L)$ and $Q(L)$ [thus $\mu(r)$ and $\lambda(\tilde{k})$] which characterize the nonlocal structure of RBMs in our description have qualitatively different influence on the complexity. Specifically, $Q(L)$ can converge to a finite L-independent constant in the thermodynamic limit and does not influence the complexity for sufficiently localized "orbitals" in the cases where $\mu(r)$ decays sufficiently fast. With the upper boundedness condition for $\mu(r)$ [Eq. (11)], $Q(L)$ can contribute an at-most-quadratic factor to this upper bound on $N_h^*(L, \varepsilon_0)$. By contrast, the asymptotic property of $P(x)$ significantly influences the complexity and may lead to the inefficiency of RBM representations if $\lambda(\tilde{k})$ decays sufficiently slowly. That would imply that there are too many correlation features to be captured by the RBM so polynomially many parameters are not enough to fully compress the information into the RBM form. However, as long as $p_d^{-1}(x)$ has an at-most-power-law dependence on x, this upper-bound estimation will imply that the complexity is definitely at most polynomial in both system size L and $1/\varepsilon_0$ with the above two types of truncation errors. Moreover, it is also remarkable that our estimation only provides an upper bound on the complexity, so a faster-than-polynomial scaling of the bound (such as that for $S_2^{(7)}$ manifold defined later in Table I) does not necessarily imply the inefficiency of the representation. It is possible that the upper bound is not tight and the real complexity is at most polynomial in this case.

Third, the asymptotic behavior of $F(x)$ at $x = 0$ also influences the scaling of the complexity and it directly acts on ε_0. We have demonstrated that, for the two types of truncation errors described above, the corresponding $F(x)$'s [$F_1(x)$ and $F_2(x)$] are both analytic at $x = 0$. For general types of truncation errors that can be upper bounded by a function $F(LQ(L)P(N_h/L))$, $1/f^{-1}(\varepsilon_0)$ has a power-law dependence on $1/\varepsilon_0$ as long as $F(x)$ is analytic at $x = 0$ based on the Taylor series expansion of the function.

This result suggests separate effects of the factors $\mu(r)$ and $\lambda(\tilde{k})$. The scaling of entanglement entropy, which is an important measure of the complexity of quantum many-body states, is influenced by $\mu(r)$, whereas $\lambda(\tilde{k})$ significantly influences the spatial complexity of parametrization in LRFD RBM representations. The length of the support of $\mu(r)$, which determines the "range" r_0 of RBMs, directly influences the scaling of the entanglement entropy of the states between subregions but does not directly contribute a faster-than-polynomial factor to the parametrization complexity. This result possibly provides further theoretical evidence for the high efficiency of RBMs in representing states with entanglement entropy scaling faster than an area law in system sizes [31].

We apply our complexity estimation to several typical settings of $\mu(r)$ and $\lambda(\tilde{k})$ in Table I. In this work, we define $S_{l'}$ ($1 \leqslant l' \leqslant 5$, $l' \in \mathbb{N}$) to be state manifolds associated with distinct RBM-related or other constraints with l' as a constraint index for clarifying their hierarchical relationships in Sec. IV. We define S_1 and S_2 to be the manifolds spanned by all short-range and LRFD RBMs, respectively. Several typical manifolds as subsets of S_2 defined by specifying distinct $\mu(r)$ and $\lambda(\tilde{k})$ settings are denoted by $S_2^{(j')}$ ($1 \leqslant j' \leqslant 7$, $j' \in \mathbb{N}$) with j' as a subset-of-manifold index, among which $S_2^{(j')}$ with $1 \leqslant j' \leqslant 6$ all correspond to spatial complexities at most polynomial in L based on our estimations.

We also apply this complexity analysis to RBMs constructed as the 1D SPT cluster states with a perturbation part. Our numerical results on the scaling of $N_h^*(L, \varepsilon_0)$ in L with fixed ε_0 [Fig. 2(d)] for small system sizes are consistent with our theoretical analysis summarized in Table I. The piecewise linearity of $N_h^*(L, \varepsilon_0)$ as a function of L with a slope growing very slowly implies that the scaling is perhaps just slightly faster than linear, consistent with our estimation based on parameter settings. The piecewise linearity is due to our assumption that N_h is an integer multiple of L. So it applies a ceiling operation to the ratio N_h/L which will not change when L varies within a small range. The inset in Fig. 2(d) shows that $N_h^U(L, \varepsilon_0)$ serve as upper bounds on $N_h^*(L, \varepsilon_0)$ as in our analysis. The $N_h^U(L, \varepsilon_0)$ are obtained by using the exact values of the right-hand side of inequality (15) and its leading-order estimations. These are almost the same and both have a power-law scaling in L as indicated by Eq. (29), supporting the validity of our complexity analysis. Moreover, it is remarkable that the above scaling analysis still

qualitatively holds for general cases where truncation points N_h's are not necessarily integer multiples of L. This stems from the fact that it is usually the long-range decaying behavior of RBM parameters among levels, which is not influenced by fragmentation within a single level, that dominates the complexity scaling in the context of wide-range varying N_h we are concerned with.

D. Spin-correlation information

In this section, we analyze what information about the physical properties of the quantum states can be extracted from the LRFD RBM form using our description of the non-local structure. Here, we focus on a small-parameter regime in which all $|a_j|$, $|b_k|$ and $|W_{j,k}|$ are no larger than ε_1, and $\varepsilon_1 \ll 1/L$, $\varepsilon_1 \ll 1/N_h$. We do not explicitly write the superscript "$^{(L)}$" for RBM parameters and assume that the RBM just has a finite number (N_h) of hidden nodes in this section.

Based on the proof given in Appendix C, we find that the unnormalized correlation in the z direction between spins with a distance of r for a LRFD RBM with translational symmetry is

$$C_{\text{unnorm}}^z(r) = \langle \Psi^{(L,N_h)} | \hat{\sigma}_1^z \hat{\sigma}_{1+r}^z | \Psi^{(L,N_h)} \rangle \tag{30}$$

$$= 2(\text{Re}(WW^T))_{1,1+r} + 4\,\text{Re}(a_1)\,\text{Re}(a_{1+r})$$

$$+ \mathcal{O}(\varepsilon_1^3) \quad (\text{as } \varepsilon_1 \to 0). \tag{31}$$

Note that the above result is the r-related part of the spin correlation, while the real value of the correlation is $C_{\text{unnorm}}^z(r)$ divided by an r-independent normalization factor $\langle \Psi^{(L,N_h)} | \Psi^{(L,N_h)} \rangle$. So for RBMs constructed as Eqs. (24)–(26) show with $a_0 = 0$ and $c_w \in \mathbb{R}$ for simplicity,

$$C_{\text{unnorm}}^z(r) \approx 2|c_w|^2 \sum_{\tilde{k}=1}^{N_h/L} |\lambda(\tilde{k})|^2 \sum_{j_c=1}^{L} \mu(|1 - j_c|_{\text{circ}})$$

$$\times \mu(|1 + r - j_c|_{\text{circ}}). \tag{32}$$

So the $\mu(r)$-related factor as shown above describes the decaying rate of spin correlations in the z direction as a function of the distance r, while the $\lambda(\tilde{k})$-related factors independent of r do not influence the decaying rate if we only consider the leading-order terms in Eq. (31).

The above result in Eq. (31) gives an interpretation of the roles of hidden nodes. The hidden nodes can be viewed as intermediate virtual particles that relate spins (physical particles) at different lattice sites. When an RBM is short-range, the term $(\text{Re}(WW^T))_{1,1+r}$ will vanish for large enough r as there is no virtual particle that can have both nonzero connectivity to two spins separated by r. Then, more intermediate hidden nodes are needed to transport such relations, which means that we need to consider higher-order terms. This is additional evidence that long-range RBMs have the potential to represent states with strong quantum correlations. It is shown in Appendix C that, even when $\mu(r) \to 0$ as $r \to 0$, we can still construct LRFD RBMs in which the spin correlations in the z direction can have long-range decayings lower bounded by $\Theta(1/r^{\alpha_Q})$ (for $\mu(r) = \Theta(1/r^{\alpha_Q})$) with $\alpha_Q > 1$, $\Theta(\ln r/r)$ (for $\mu(r) = \Theta(1/r)$), and even $\Theta(1)$ (for $\mu(r) = \Theta(1/r^{\alpha_Q})$ with $0 < \alpha_Q \leqslant \frac{1}{2}$). These three kinds of decaying rates of spin correlations are demonstrated by numerical computations
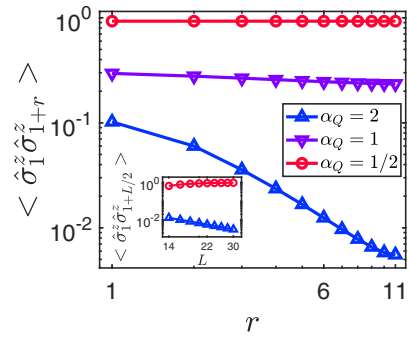

FIG. 3. Spin correlations in the z direction as a function of distance r on a log-log scale. The LRFD RBMs are constructed as Eqs. (24)–(26) show, where $\mu(r) = \frac{1}{2}\delta_Q r^{-\alpha_Q}$ for $r \neq 0$, $\mu(0) = \delta_Q = 0.2$, $\lambda(\tilde{k}) = \tilde{k}^{-\alpha_P}$, $\alpha_P = 3.5$, $c_w = 1$, $c_b = 0$, $a_0 = 0$, $L = 22$, and $N_h = 5L$. The inset shows the spin correlation $\langle \hat{\sigma}_1^z \hat{\sigma}_{1+L/2}^z \rangle$ with r being the half-chain length for varying L (on a log-log scale). It shows a convergence of $\langle \hat{\sigma}_1^z \hat{\sigma}_{1+L/2}^z \rangle$ to an L-independent constant (almost attaining the maximum value of 1) for $\alpha_Q = 1/2$ and a decay for $\alpha_Q = 2$.

(Fig. 3). The spin correlation $\langle \hat{\sigma}_1^z \hat{\sigma}_{1+r}^z \rangle$ almost saturates the maximum value of 1 for $\alpha_Q = 1/2$. In comparison, these spin correlations have different long-range decaying rates for $\alpha_Q = 1$ and $\alpha_Q = 2$ as r increases.

III. GROUND-STATE APPLICATIONS

Based on the proposal of the concept of LRFD RBMs and the theoretical analysis of their spatial complexity, it is natural to explore their applications to learning quantum states associated with specific models.

First, in Appendix D, we prove that the state with all spins pointing up in the z direction, which is the ground state of a spin-1/2 system with a single magnetic field in the z direction and has a form of the Kronecker delta function, can be approximated by LRFD RBMs with arbitrary accuracy. We find that the RBM construction is not unique for such a target state even when fixing the global phase which implies eliminating the degree of freedom associated with a global gauge transformation and we give one construction. Thus we provide one example of the utility of LRFD RBMs in state representation for arbitrarily large system sizes.

Second, we are particularly interested in the behavior of RBMs in cases where other state representations become less efficient. We numerically study the representation of the ground states of critical systems with finite sizes for which the MPS representation becomes less efficient [40,49], while MPS has achieved notable success in representing quantum many-body states with entanglement entropy satisfying an area law [32,44].

To accomplish this, we use RBMs with translational symmetry and apply the conventional quantum Monte Carlo algorithm (also a variational method) with stochastic-reconfiguration optimizations [6,50–52] to learn the ground states of two typical quantum models: the 1D transverse-field Ising model (TFIM) [Eq. (33)] and XXZ model [Eq. (34)],

described by Hamiltonians

$$\hat{H} = -\sum_{j=1}^{L} \hat{\sigma}_j^z \hat{\sigma}_{j+1}^z - B_x \sum_{j=1}^{L} \hat{\sigma}_j^x \tag{33}$$

and

$$\hat{H} = \sum_{j=1}^{L} \left(-\hat{\sigma}_j^x \hat{\sigma}_{j+1}^x - \hat{\sigma}_j^y \hat{\sigma}_{j+1}^y + J_z \hat{\sigma}_j^z \hat{\sigma}_{j+1}^z \right) \tag{34}$$

with periodic boundary conditions, respectively, where B_x denotes the strength of a transverse field and J_z denotes the strength of coupling in the z direction. We use RBMs to learn the ground state of the TFIM with $B_x = 1$ which implies that the quantum system is exactly in the phase-transition point between a ferromagnetic and a paramagnetic phase [53] and of the XXZ model with $J_z = -0.2$ which implies that the system is in a gapless disordered XY phase [54]. Both systems are critical systems with the entanglement entropy of the ground states scaling logarithmically in system sizes [53,55,56]. The ground states of these two Hamiltonians (at least for small system sizes) can be well learned by RBMs, which is demonstrated by the high accuracy in spin-correlation calculations given in Appendix E. The importance measures $\eta(j, \tilde{k}, L)$ for these two RBMs are provided in Figs. 4(a) and 4(b) as representatives and we find that the structure of $\eta(j, \tilde{k}, L)$ for the ground state of a TFIM deeply in a paramagnetic phase (far from critical points) is very similar to that shown in Fig. 4(a).

The numerical results show that the RBM representations of the two ground states of the above two critical systems have forms very similar to LRFD RBMs. The overall 3D structures for the importance measures $\eta(j, \tilde{k}, L)$ are similar to the one presented in Fig. 1(b) which corresponds to a standard LRFD RBM. The weight parameters for hidden nodes at the same level are quite localized and decay fastly as the level index $\tilde{k}$ increases and as the spin-site index j goes away from the center. Moreover, it seems that the "ridge" of $\eta(j, \tilde{k}, L)$ for varying system sizes can be upper bounded by an L-independent power-law decay curve, based on which we can extract a corresponding α_P characterizing the rate of level decay for these small-system-size wave functions. If these features still hold as L increases and approaches infinity, these states will form LRFD RBMs which belong to the set $S_2^{(2)}$ or $S_2^{(6)}$ in Table I and the corresponding $\lambda(\tilde{k})$ and $\mu(r)$ can be defined.

Moreover, the above results exhibit a feature that is also manifested in the theory of MPS representations. It has been shown that [40], though MPS becomes less efficient in representing the ground states of critical systems, the bond dimension required to achieve an approximation error ε_0 can still be upper bounded by a function scaling polynomially in the system size L. The exponent in the power-law dependence of spatial complexity of MPSs on L depends on the central charge c, which is a quantity roughly quantifying the "degrees of freedom of the theory" in conformal field theory [44]. A larger c leads to a higher exponent in that estimation which implies a higher complexity in MPS representation. While the TFIM at the above phase-transition point has $c = \frac{1}{2}$ and the XXZ model in the disordered XY phase has $c = 1$ [57], our numerical results do show a smaller fitted α_P for the
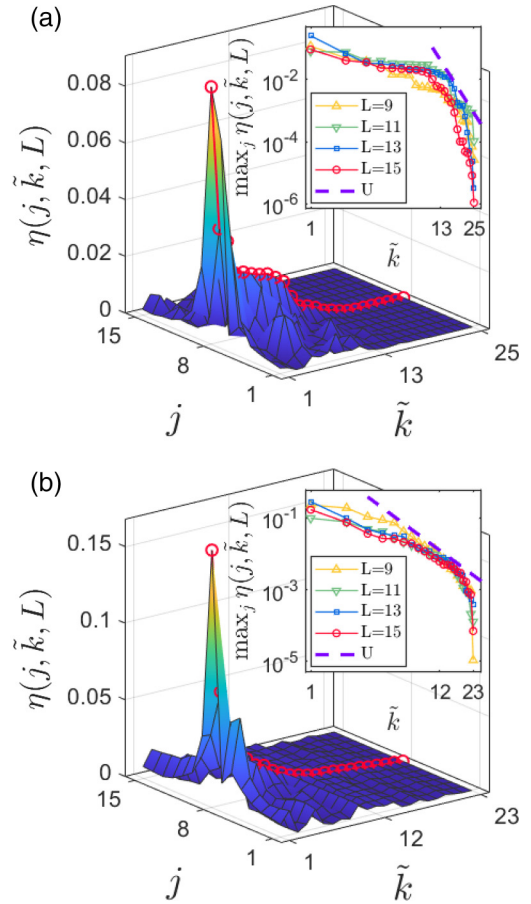

FIG. 4. Importance measure $\eta(j, \tilde{k}, L)$ for the RBMs approximating ground states of two critical systems with $L = 15$. (a) TFIM with $B_x = 1$. (b) XXZ model with $J_z = -0.2$. The insets in each subfigure show the decays of the maximum importance measure at each level as level index $\tilde{k}$ increases on a log-log scale. The system size $L = 9$, 11, 13, and 15. The purple dashed curve implies that these decaying curves can be upper bounded by a power-law decay. By numerical fitting, the corresponding α_P for the dashed lines in the insets of (a) and (b) are 2.957 and 1.232, respectively.

XXZ model, which implies that the XXZ model has more intrinsic "complexity" compared to the TFIM, thus needing more parameters to capture this complexity.

Furthermore, it should be emphasized that the importance measure $\eta(j, \tilde{k}, L)$ proposed in this work can be viewed as an information condensation and reformulation of weight parameters for generic RBMs with translational symmetry involving large numbers of (or infinitely many) parameters. It can be used for understanding how close such RBMs are to LRFD RBMs and even generally measuring their intrinsic complexities by delving into its decaying behaviors in j and $\tilde{k}$ domains, which can be supported by the above comparison of the TFIM and XXZ model.

IV. STATE MANIFOLDS AND COMPLEXITY CLASSIFICATION

Rigorously speaking, the numerical results for systems of finite sizes only provide evidence supporting that the states
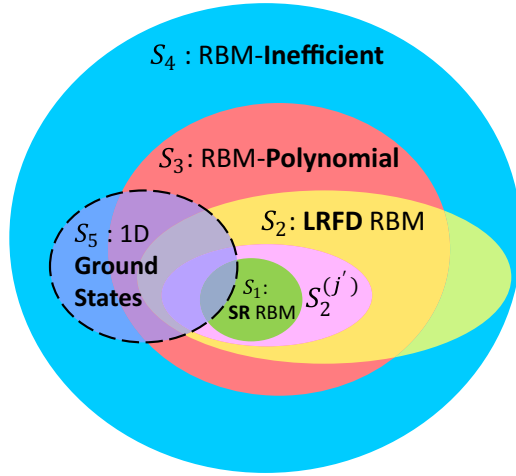
FIG. 5. Relations between multiple typical state manifolds. S_1: short-range RBMs; S_2: LRFD RBMs; $S_2^{(j')}$ (for $1 \leqslant j' \leqslant 6$, $j' \in \mathbb{N}$): LRFD RBMs with distinct parameter conditions, specified in Table I; S_3: RBMs with spatial complexities scaling at most polynomially in system sizes; S_4: RBMs with a faster-than-polynomial scaling of spatial complexities in system sizes, corresponding to inefficiency of representation; S_5: ground states of 1D quantum spin systems. The dashed boundary of S_5 means that its relations with other manifolds have not been fully determined.

may be LRFD RBMs but cannot prove it, since the properties of RBMs in the process of approaching the thermodynamic limit are not yet known. Based on the success of RBMs in numerical simulations and the fact that they can often achieve high accuracy even with a constant number of levels (at least for small system sizes), we conjecture that the ground states of a wide range of quantum systems may be exactly represented by LRFD RBMs, or a variant of them. Here, the term "variant" means a generalization of the forms specified in conditions 1 and 2 by including additional factors that can be naturally incorporated into our complexity analysis. For example, the $\lambda(\tilde{k})$ and $\mu(r)$ functional forms, which are L-independent in our definition of LRFD RBMs, can be generalized into $\lambda(\tilde{k}, L)$ and $\mu(r, L)$, respectively, while their effects can be easily evaluated using our paradigm for complexity analysis.

We summarize the relations between multiple typical state manifolds so that the significance of proposing the concept of LRFD RBMs can be better understood. A state manifold usually refers to a subspace of the whole Hilbert space spanned by a parameterized wave-function family [45] and thus it is a set containing a specific scope of quantum states. The manifolds S_1, S_2, $S_2^{(j')}$ (for $1 \leqslant j' \leqslant 6$, $j' \in \mathbb{N}$), S_3 and S_4 are defined to be the space spanned by quantum states represented by RBMs satisfying corresponding conditions as given in Fig. 5, while S_5 is defined to be the manifold spanned by all ground states of 1D quantum many-body spin systems.

The definitions of these manifolds directly implies that $S_1 \subsetneq S_2^{(j')} \subsetneq S_2$ (for $1 \leqslant j' \leqslant 6$). Our complexity analysis for LRFD RBMs (Sec. II C) gives the result that $S_2^{(j')} \subseteq (S_2 \cap S_3)$.

Previous research shows that a set of problems where RBMs appear to be powerful are related to topological states, among which the 1D SPT cluster states belong to $S_5 \cap S_1$ [30]. The Laughlin wave functions, which have the structure of Jastrow wave functions and are associated with chiral topological order, can be exactly represented by RBMs in S_3 with a quadratic scaling of N_h in L but their approximations with RBMs of a long-range form and less complexity are often used [8]. S_4 contains all other sets mentioned in Fig. 5 as RBMs without restriction on the number of hidden nodes are universal approximators for discrete distribution [43]. Numerical results seem to support that a "large fraction" of S_5 is contained in its intersection with S_2. We argue that the concept of S_2 may benefit the understanding of which fraction of S_5 falls into its intersection with S_3, thus also promoting the understanding of the complexity of quantum many-body states.

It is remarkable that our paradigm for complexity analysis and our characterization of the nonlocal structures of RBMs for 1D quantum spin systems can be generalized to higher-dimensional systems, e.g., lattices. This is done by generalizing the description of single-level "orbitals" from $\mu(r)$ to $\mu(\vec{r})$ while keeping $\lambda(\tilde{k})$ as a level-decay factor. For deep NN quantum states, we can still view each single hidden layer as a combination of multiple levels which capture different correlation features. We can calculate the truncation errors for each hidden layer associated with specific nodal functions and analyze the propagation of errors through layers.

V. SUMMARY

In this work, we define a subset of generic RBM quantum states—long-range-fast-decay (LRFD) RBM states. Using the language of levels, the nonlocal structure of LRFD RBMs is described with two functions: one of which, $\mu(r)$, captures the localization of the spatial distribution of the wave function for each single level and encodes information about spin correlations; the other, $\lambda(\tilde{k})$, is a level-decay factor capturing different correlation features and significantly influencing the complexity of the RBMs. We derive upper bounds on truncation errors, which allow us to analyze the scaling of the spatial complexity in system sizes and approximation errors for LRFD RBMs. We provide numerical results supporting that the ground states of a wide range of 1D quantum spin systems, including some critical systems, may be approximated by LRFD RBMs with an at-most-polynomial complexity. Finally, we describe the relationships between state manifolds of different computational complexity and identify hierarchies of RBM-efficient approximation.

Generalizing the RBM wave-function ansatz to an infinitely-many-hidden-node regime and proposing the concept of LRFD RBMs does not imply the use of an infinitely-large neural network for state representations. These serve to define the completeness of a set of variational states and serve as a tool for complexity analysis based on the good extensibility and analyzability of LRFD-RBM forms. This concept may promote general understanding of the intrinsic complexity of quantum many-body states.

APPENDIX A: PROOF OF THE CONVERGENCE OF $|\Psi^{(L,\infty)}\rangle$ FOR LONG-RANGE-FAST-DECAY RBMs

In this section, we prove the convergence of $|\Psi^{(L,\infty)}\rangle$ [Eq. (5)] which satisfies conditions 1 and 2 in the definition of the long-range-fast-decay (LRFD) RBMs given in Sec. II A.

We have defined the wave-function amplitude $\psi^{(L,\infty)}(\vec{\sigma})$ for LRFD RBMs and the amplitude $\psi^{(L,N_h)}(\vec{\sigma})$ for its corresponding truncated RBMs with the first N_h hidden nodes kept.

For any fixed L and any $\vec{\sigma}$, define

$$A_n(\vec{\sigma}) = |\psi^{(L,nL)}(\vec{\sigma})| \tag{A1}$$

and

$$\phi_n(\vec{\sigma}) = \arg(\psi^{(L,nL)}(\vec{\sigma})) \tag{A2}$$

to be the modulus and the argument of $\psi^{(L,nL)}(\vec{\sigma})$, respectively, as we assume N_h to be an integer multiple of L throughout this work. Then

$$\psi^{(L,nL)}(\vec{\sigma}) = A_n(\vec{\sigma})e^{i\phi_n(\vec{\sigma})}. \tag{A3}$$

In the following part, we will prove that the sequence $\{\psi^{(L,nL)}(\vec{\sigma}) : n \in \mathbb{N}\}$ is a Cauchy sequence in $\mathbb{C}$ with d as a metric, where $d : \mathbb{C} \times \mathbb{C} \to \mathbb{R}$ is just the commonly used distance between two complex numbers defined as $d(z_1, z_2) = |z_1 - z_2|$ for $z_1, z_2 \in \mathbb{C}$.

Proof. If there exists any finite $n' \in \mathbb{N}$ such that $\psi^{(L,n'L)}(\vec{\sigma}) = 0$, then $\psi^{(L,nL)}(\vec{\sigma}) = 0$ for all $n \geqslant n'$. Then $\{\psi^{(L,nL)}(\vec{\sigma}) : n \in \mathbb{N}\}$ must be a Cauchy sequence in the metric space $(\mathbb{C}, d)$ as it converges to 0. So we will just focus on the cases where $\psi^{(L,nL)}(\vec{\sigma}) \neq 0$, which implies $A_n(\vec{\sigma}) \neq 0$, for all $n \in \mathbb{N}$ in the following part.

Define the effective angles [6] $\theta_k(\vec{\sigma}) = b_k^{(L)} + \sum_{j=1}^L \sigma_j W_{j,k}^{(L)}$. Then, for any $n > \tilde{k}_s$, $n \in \mathbb{N}$,

$$A_n(\vec{\sigma}) = \left| \prod_{j=1}^L e^{a_j^{(L)}\sigma_j} \prod_{k=1}^{nL} \cosh\left(b_k^{(L)} + \sum_{j=1}^L \sigma_j W_{j,k}^{(L)}\right) \right| \tag{A4}$$

$$= \left| \prod_{j=1}^L e^{a_j^{(L)}\sigma_j} \prod_{k=1}^{\tilde{k}_s L} \cosh(\theta_k(\vec{\sigma})) \right| \left| \prod_{k=\tilde{k}_s L+1}^{nL} \cosh(\theta_k(\vec{\sigma})) \right| \tag{A5}$$

$$= A_0 \left| \prod_{k=\tilde{k}_s L+1}^{nL} \cosh(\theta_k(\vec{\sigma})) \right|, \tag{A6}$$

where the contribution of terms with $1 \leqslant k \leqslant \tilde{k}_s L$ is summarized in an n-independent constant

$A_0 = |\prod_{j=1}^L e^{a_j^{(L)}\sigma_j} \prod_{k=1}^{\tilde{k}_s L} \cosh(\theta_k(\vec{\sigma}))|$ as the boundedness conditions (conditions 1 and 2) only apply to the range $k > \tilde{k}_s L$ and the terms in this range will be evaluated separately.

Define intermediate variables

$$u_k(\vec{\sigma}) = \mathrm{Re}\left(b_k^{(L)} + \sum_{j=1}^L \sigma_j W_{j,k}^{(L)}\right) \in \mathbb{R}, \tag{A7}$$

$$v_k(\vec{\sigma}) = \mathrm{Im}\left(b_k^{(L)} + \sum_{j=1}^L \sigma_j W_{j,k}^{(L)}\right) \in \mathbb{R}, \tag{A8}$$

$$U_k = \max_{\{\vec{\sigma}\}}\{|u_k(\vec{\sigma})|\} = \left|\mathrm{Re}\left(b_k^{(L)}\right)\right| + \sum_{j=1}^L \left|\mathrm{Re}\left(W_{j,k}^{(L)}\right)\right|, \tag{A9}$$

$$V_k = \max_{\{\vec{\sigma}\}}\{|v_k(\vec{\sigma})|\} = \left|\mathrm{Im}\left(b_k^{(L)}\right)\right| + \sum_{j=1}^L \left|\mathrm{Im}\left(W_{j,k}^{(L)}\right)\right|, \tag{A10}$$

where U_k and V_k can be viewed as universal measures of the amplitudes of the real and imaginary parts of RBM parameters associated with the kth hidden node regardless of spin configurations. Then

$$A_n(\vec{\sigma}) = A_0 \prod_{k=\tilde{k}_s L+1}^{nL} \frac{1}{2}\sqrt{e^{2u_k(\vec{\sigma})} + e^{-2u_k(\vec{\sigma})} + 2\cos(2v_k(\vec{\sigma}))}. \tag{A11}$$

So we can upper bound $A_n(\vec{\sigma})$ by

$$A_n(\vec{\sigma}) \leqslant A_0 \prod_{k=\tilde{k}_s L+1}^{nL} \cosh(u_k(\vec{\sigma})) \tag{A12}$$

$$\leqslant A_0 \prod_{k=\tilde{k}_s L+1}^{nL} \exp\left(\frac{1}{2}u_k^2(\vec{\sigma})\right) \tag{A13}$$

$$\leqslant A_0 \prod_{k=\tilde{k}_s L+1}^{nL} \exp\left(\frac{1}{2}U_k^2\right) \tag{A14}$$

$$= A_0 \exp\left(\frac{1}{2}\sum_{k=\tilde{k}_s L+1}^{nL}\left(\left|\mathrm{Re}\left(b_k^{(L)}\right)\right| + \sum_{j=1}^L \left|\mathrm{Re}\left(W_{j,k}^{(L)}\right)\right|\right)^2\right). \tag{A15}$$

Using conditions 1 and 2, we can get an upper bound on $A_n(\vec{\sigma})$ expressed in terms of $\lambda_R(\tilde{k})$ and $\mu(r)$ as

$$A_n(\vec{\sigma}) \leqslant A_0 \exp\left[\frac{1}{2}\sum_{k=\tilde{k}_s L+1}^{nL}\left(\lambda_R(\tilde{k}(k))\mu(0) + \sum_{j=1}^L \lambda_R(\tilde{k}(k))\mu(|j - j_c(k)|_{\mathrm{circ}})\right)^2\right] \tag{A16}$$

$$= A_0 \exp\left[\frac{1}{2}\sum_{k=\tilde{k}_s L+1}^{nL}(\lambda_R(\tilde{k}(k)))^2\right.$$

$$\left. \times \left(\mu(0) + \sum_{\triangle j=0,\pm 1,\ldots,\pm\frac{L-1}{2}}\mu(|\triangle j|)\right)^2\right] \quad \text{(A17)}$$

$$= A_0 \exp\left[2L\sum_{\tilde{k}=\tilde{k}_s+1}^{n}\lambda_R^2(\tilde{k})\left(\sum_{r=0}^{(L-1)/2}\mu(r)\right)^2\right]. \quad \text{(A18)}$$

Note that Eq. (10) and inequality (11) as constraints on $\lambda_R(\tilde{k})$ and $\mu(r)$ can be used to upper bound the first and second summations in the exponent of the above expression. Since $\sum_{k=\tilde{k}_s+1}^{n}\lambda_R^2(\tilde{k}) \leqslant \sum_{\tilde{k}=\tilde{k}_s+1}^{\infty}\lambda_R^2(\tilde{k}) \leqslant P_0$, we have

$$A_n(\vec{\sigma}) \leqslant M_0 := A_0 \exp(2LQ(L)P_0). \quad \text{(A19)}$$

Therefore we prove that, for any fixed $L > 0$, $A_n(\vec{\sigma})$ is bounded above by an n-independent constant M_0 for all $n > \tilde{k}_s$, which is the first step for the overall proof.

Using similar mathematical tricks with a modified range of k and $\tilde{k}$, for any $n > \tilde{k}_s$ and $m \in \mathbb{N}$, we can upper bound $|A_{n+m}(\vec{\sigma})/A_n(\vec{\sigma})|$ by

$$\left|\frac{A_{n+m}(\vec{\sigma})}{A_n(\vec{\sigma})}\right| = \prod_{k=nL+1}^{(n+m)L}|\cosh(\theta_k(\vec{\sigma}))| \quad \text{(A20)}$$

$$\leqslant \exp\left[2LQ(L)\sum_{\tilde{k}=n+1}^{n+m}\lambda_R^2(\tilde{k})\right]. \quad \text{(A21)}$$

We can also lower bound $|A_{n+m}(\vec{\sigma})/A_n(\vec{\sigma})|$ by

$$\left|\frac{A_{n+m}(\vec{\sigma})}{A_n(\vec{\sigma})}\right| = \prod_{k=nL+1}^{(n+m)L}\frac{1}{2}\sqrt{e^{2u_k(\vec{\sigma})} + e^{-2u_k(\vec{\sigma})} + 2\cos(2v_k(\vec{\sigma}))}$$

$$\text{(A22)}$$

$$\geqslant \prod_{k=nL+1}^{(n+m)L}\frac{1}{2}\sqrt{2 + 2\cos(2v_k(\vec{\sigma}))} \quad \text{(A23)}$$

$$= \prod_{k=nL+1}^{(n+m)L}|\cos(|v_k(\vec{\sigma})|)|. \quad \text{(A24)}$$

Since

$$V_k \leqslant 2\lambda_I(\tilde{k}(k))\sum_{r=0}^{(L-1)/2}\mu(r), \quad \text{(A25)}$$

and $\lambda_I(\tilde{k}(k)) \to 0$ as $k \to \infty$, we know that $V_k \to 0$ as $k \to \infty$. Define $n_I = n_I(L) = \min\{n \geqslant \tilde{k}_s, n \in \mathbb{N} : \forall k \geqslant n_I L + 1, V_k \leqslant \pi/3\}$. Using the fact that $|\cos(x)| \geqslant \exp(-\beta_1^2 x^2/2)$ with $\beta_1 = 3\sqrt{2\ln 2}/\pi$ for all $x \in [0, \pi/3]$, for $n \geqslant n_I$,

$$\left|\frac{A_{n+m}(\vec{\sigma})}{A_n(\vec{\sigma})}\right| \geqslant \prod_{k=nL+1}^{(n+m)L}|\cos(V_k)| \quad \text{(A26)}$$

$$\geqslant \exp\left(-\frac{1}{2}\beta_1^2\sum_{k=nL+1}^{(n+m)L}V_k^2\right) \quad \text{(A27)}$$

$$\geqslant \exp\left[-2\beta_1^2 LQ(L)\sum_{\tilde{k}=n+1}^{n+m}\lambda_I^2(\tilde{k})\right]. \quad \text{(A28)}$$

In fact, $\cos(x)$ converges to $\exp(-x^2/2)$ as $x \to 0$ and a constant β_1 slightly larger than 1 is introduced to ensure the holding of the inequality while capturing the leading-order asymptotic properties.

Since $P(\tilde{k}_s)$ converges to a constant $P_0/(1 + \beta_1^2)$ and all terms are nonnegative, we have $P(m') \to 0$ as $m' \to \infty$. So $\sum_{\tilde{k}=m'+1}^{\infty}\lambda_R^2(\tilde{k})$ and $\sum_{\tilde{k}=m'+1}^{\infty}\lambda_I^2(\tilde{k})$ also approach 0 as $m' \to \infty$.

Therefore, for any sufficiently small $\varepsilon > 0$, there exists $n_1 \geqslant n_I \geqslant \tilde{k}_s$ such that, for all $m' > n_1$,

$$\sum_{\tilde{k}=m'+1}^{\infty}\lambda_R^2(\tilde{k}) < \ln(1 + \varepsilon/M_0)/(2LQ(L)), \quad \text{(A29)}$$

$$\sum_{\tilde{k}=m'+1}^{\infty}\lambda_I^2(\tilde{k}) < \ln(1/(1 - \varepsilon/M_0))/(2\beta_1^2 LQ(L)). \quad \text{(A30)}$$

Then, for all $n > n_1$ and $m \in \mathbb{N}$,

$$\left|\frac{A_{n+m}(\vec{\sigma})}{A_n(\vec{\sigma})}\right| \leqslant \exp\left[2LQ(L)\sum_{\tilde{k}=n+1}^{\infty}\lambda_R^2(\tilde{k})\right] \quad \text{(A31)}$$

$$< 1 + \varepsilon/M_0, \quad \text{(A32)}$$

$$\left|\frac{A_{n+m}(\vec{\sigma})}{A_n(\vec{\sigma})}\right| \geqslant \exp\left[-2\beta_1^2 LQ(L)\sum_{\tilde{k}=n+1}^{\infty}\lambda_I^2(\tilde{k})\right] \quad \text{(A33)}$$

$$> 1 - \varepsilon/M_0, \quad \text{(A34)}$$

which implies

$$|A_{n+m}(\vec{\sigma})/A_n(\vec{\sigma}) - 1| < \varepsilon/M_0. \quad \text{(A35)}$$

Therefore combining inequalities (A19) and (A35),

$$|A_{n+m}(\vec{\sigma}) - A_n(\vec{\sigma})| = A_n(\vec{\sigma})\left|\frac{A_{n+m}(\vec{\sigma})}{A_n(\vec{\sigma})} - 1\right| \quad \text{(A36)}$$

$$< M_0\frac{\varepsilon}{M_0} = \varepsilon. \quad \text{(A37)}$$

Thus we prove that the modulus sequence $\{A_n(\vec{\sigma}) : n \in \mathbb{N}\}$ is a Cauchy sequence in $\mathbb{R}$. So $A_n(\vec{\sigma})$ must converge to some nonnegative constant $A_\infty(\vec{\sigma}) \in \mathbb{R}$ as $n \to \infty$.

Then, we analyze the argument sequence $\{\phi_n(\vec{\sigma}) : n \in \mathbb{N}\}$. For $n > \tilde{k}_s$ and $m \in \mathbb{N}$, we have

$$|\phi_{n+m}(\vec{\sigma}) - \phi_n(\vec{\sigma})| = \left|\arg\left(\frac{\psi^{(L,(n+m)L)}(\vec{\sigma})}{\psi^{(L,nL)}(\vec{\sigma})}\right)\right| \quad \text{(A38)}$$

$$= \left|\arg\left(\prod_{k=nL+1}^{(n+m)L}\cosh(u_k(\vec{\sigma}))\cos v_k(\vec{\sigma})\right.\right.$$

$$\left.\left. + i\sinh(u_k(\vec{\sigma}))\sin v_k(\vec{\sigma})\right)\right|. \quad \text{(A39)}$$

For $n \geqslant n_I > \tilde{k}_s$, $|v_k(\vec{\sigma})| \leqslant V_k \leqslant \pi/3$, so $\cosh(u_k(\vec{\sigma}))$ $\cos v_k(\vec{\sigma}) > 0$, which implies that the argument of each

term is in the range $(-\pi/2, \pi/2)$. So we have

$$|\phi_{n+m}(\vec{\sigma}) - \phi_n(\vec{\sigma})|$$

$$= \left| \sum_{k=nL+1}^{(n+m)L} \arctan(\tanh(u_k(\vec{\sigma})) \tan(v_k(\vec{\sigma}))) \right| \quad \text{(A40)}$$

$$\leqslant \sum_{k=nL+1}^{(n+m)L} \arctan(\tanh(|u_k(\vec{\sigma})|) \tan(|v_k(\vec{\sigma})|)). \quad \text{(A41)}$$

Using

$$\arctan(\tanh(x) \tan(y)) \leqslant \beta_2 \tanh(x) y \leqslant \beta_2 xy \quad \text{(A42)}$$

with $\beta_2 = 3\sqrt{3}/\pi$ for all $x \geqslant 0$ and $0 \leqslant y \leqslant \pi/3$, we can get

$$|\phi_{n+m}(\vec{\sigma}) - \phi_n(\vec{\sigma})|$$

$$\leqslant \beta_2 \sum_{k=nL+1}^{(n+m)L} |u_k(\vec{\sigma})| |v_k(\vec{\sigma})| \quad \text{(A43)}$$

$$\leqslant \beta_2 \sum_{k=nL+1}^{(n+m)L} U_k V_k \quad \text{(A44)}$$

$$= \beta_2 \sum_{k=nL+1}^{(n+m)L} \left(\left| \text{Re} \left(b_k^{(L)} \right) \right| + \sum_{j=1}^{L} \left| \text{Re} \left(W_{j,k}^{(L)} \right) \right| \right)$$

$$\times \left(\left| \text{Im} \left(b_k^{(L)} \right) \right| + \sum_{j=1}^{L} \left| \text{Im} \left(W_{j,k}^{(L)} \right) \right| \right) \quad \text{(A45)}$$

$$\leqslant \beta_2 \left[4L \sum_{\tilde{k}=n+1}^{n+m} \lambda_R(\tilde{k}) \lambda_I(\tilde{k}) \left(\sum_{r=0}^{(L-1)/2} \mu(r) \right)^2 \right] \quad \text{(A46)}$$

$$= 4\beta_2 L Q(L) \sum_{\tilde{k}=n+1}^{n+m} \lambda_R(\tilde{k}) \lambda_I(\tilde{k}). \quad \text{(A47)}$$

Again, since $\sum_{\tilde{k}=m'+1}^{\infty} \lambda_R^2(\tilde{k})$ and $\sum_{\tilde{k}=m'+1}^{\infty} \lambda_I^2(\tilde{k})$ approach 0 as $m' \to \infty$, $\sum_{\tilde{k}=m'+1}^{\infty} \lambda_R(\tilde{k}) \lambda_I(\tilde{k}) \leqslant \frac{1}{2} \sum_{\tilde{k}=m'+1}^{\infty} (\lambda_R^2(\tilde{k}) + \lambda_I^2(\tilde{k}))$ also approaches 0 as $m' \to \infty$.

Therefore, for any sufficiently small $\varepsilon > 0$, there exists $n_2 \geqslant n_I \geqslant \tilde{k}_s$ such that, for all $m' > n_2$,

$$\sum_{\tilde{k}=m'+1}^{\infty} \lambda_R(\tilde{k}) \lambda_I(\tilde{k}) < \varepsilon / (4\beta_2 L Q(L)). \quad \text{(A48)}$$

Then, for all $n > n_2$ and $m \in \mathbb{N}$,

$$|\phi_{n+m}(\vec{\sigma}) - \phi_n(\vec{\sigma})| < 4\beta_2 L Q(L) \frac{\varepsilon}{4\beta_2 L Q(L)} = \varepsilon. \quad \text{(A49)}$$

Thus we prove that the argument sequence $\{\phi_n(\vec{\sigma}) : n \in \mathbb{N}\}$ is a Cauchy sequence in $\mathbb{R}$. So $\phi_n(\vec{\sigma})$ must converge to some constant $\phi_\infty(\vec{\sigma}) \in \mathbb{R}$ as $n \to \infty$.

Combining the above two conclusions, we can get that, for any sufficiently small $\varepsilon > 0$, there exists $n_3 \geqslant n_I \geqslant \tilde{k}_s$ such that, for all $n > n_3$ and $m \in \mathbb{N}$, $|A_{n+m}(\vec{\sigma}) - A_n(\vec{\sigma})| < \varepsilon/\sqrt{2}$ and

$|\phi_{n+m}(\vec{\sigma}) - \phi_n(\vec{\sigma})| < \arccos(1 - \frac{\varepsilon^2}{4M_0^2})$. So

$$|\psi^{(L,(n+m)L)}(\vec{\sigma}) - \psi^{(L,nL)}(\vec{\sigma})|$$

$$= [(A_{n+m}(\vec{\sigma}) - A_n(\vec{\sigma}))^2 + 2A_{n+m}(\vec{\sigma}) A_n(\vec{\sigma})$$

$$\times (1 - \cos(\phi_{n+m}(\vec{\sigma}) - \phi_n(\vec{\sigma})))]^{1/2} \quad \text{(A50)}$$

$$< \left[\frac{1}{2} \varepsilon^2 + 2M_0^2 \frac{\varepsilon^2}{4M_0^2} \right]^{1/2} = \varepsilon. \quad \text{(A51)}$$

Thus we prove that the sequence $\{\psi^{(L,nL)}(\vec{\sigma}) : n \in \mathbb{N}\}$ is a Cauchy sequence in the metric space $(\mathbb{C}, d)$, and $\psi^{(L,nL)}(\vec{\sigma})$ actually converges to the constant $A_\infty(\vec{\sigma}) e^{i\phi_\infty(\vec{\sigma})} \in \mathbb{C}$ as $n \to \infty$. ∎

APPENDIX B: PROOF OF UPPER BOUNDS ON TRUNCATION ERRORS FOR LRFD RBMS

In this section, we provide a proof of the upper bounds on the two types of truncation errors for LRFD RBMs (Lemma 3) given in Sec. II B.

We give the proof for the first-type truncation errors as follows.

Proof.

$$\left\| |\tilde{\Psi}^{(L,\infty)}\rangle - |\tilde{\Psi}^{(L,N_h)}\rangle \right\|^2$$

$$= 2 - (\langle \tilde{\Psi}^{(L,\infty)} | \tilde{\Psi}^{(L,N_h)}\rangle + \text{c.c.}) \quad \text{(B1)}$$

$$= 2 - \frac{(\langle \Psi^{(L,\infty)} | \Psi^{(L,N_h)}\rangle + \text{c.c.})}{\sqrt{\langle \Psi^{(L,\infty)} | \Psi^{(L,\infty)}\rangle \langle \Psi^{(L,N_h)} | \Psi^{(L,N_h)}\rangle}} \quad \text{(B2)}$$

$$= 2 - \frac{2 \sum_{\vec{\sigma}} |\psi^{(L,N_h)}(\vec{\sigma})|^2 \text{Re} \left(\frac{\psi^{(L,\infty)}(\vec{\sigma})}{\psi^{(L,N_h)}(\vec{\sigma})} \right)}{\sqrt{\sum_{\vec{\sigma}} |\psi^{(L,\infty)}(\vec{\sigma})|^2 \sum_{\vec{\sigma}} |\psi^{(L,N_h)}(\vec{\sigma})|^2}}. \quad \text{(B3)}$$

Define

$$\Theta = \Theta(L, N_h) = \max_{\vec{\sigma}} \left\{ \left| \arg \left(\frac{\psi^{(L,\infty)}(\vec{\sigma})}{\psi^{(L,N_h)}(\vec{\sigma})} \right) \right| \right\} \quad \text{(B4)}$$

to be the maximal argument for the ratio factor $\psi^{(L,\infty)}(\vec{\sigma})/\psi^{(L,N_h)}(\vec{\sigma})$. For $N_h/L \geqslant n_I(L) > \tilde{k}_s$,

$$\left| \arg \left(\frac{\psi^{(L,\infty)}(\vec{\sigma})}{\psi^{(L,N_h)}(\vec{\sigma})} \right) \right|$$

$$\leqslant 4\beta_2 L Q(L) \sum_{\tilde{k}=N_h/L+1}^{\infty} \lambda_R(\tilde{k}) \lambda_I(\tilde{k}). \quad \text{(B5)}$$

Therefore

$$\Theta \leqslant 4\beta_2 L Q(L) \sum_{\tilde{k}=N_h/L+1}^{\infty} \lambda_R(\tilde{k}) \lambda_I(\tilde{k}). \quad \text{(B6)}$$

Define

$$R_1 = R_1(L, N_h) = \max_{\vec{\sigma}} \left\{ \left| \frac{\psi^{(L,\infty)}(\vec{\sigma})}{\psi^{(L,N_h)}(\vec{\sigma})} \right|^2 \right\} \quad \text{(B7)}$$

and

$$R_2 = R_2(L, N_h) = \min_{\vec{\sigma}} \left\{ \left| \frac{\psi^{(L,\infty)}(\vec{\sigma})}{\psi^{(L,N_h)}(\vec{\sigma})} \right|^2 \right\} \quad \text{(B8)}$$

to be the maximal and minimal moduli for the square of the ratio factor, respectively. We have

$$R_1 \leqslant \exp\left[4LQ(L)\sum_{\tilde{k}=N_h/L+1}^{\infty}\lambda_R^2(\tilde{k})\right], \quad (B9)$$

$$R_2 \geqslant \exp\left[-4\beta_1^2 LQ(L)\sum_{\tilde{k}=N_h/L+1}^{\infty}\lambda_I^2(\tilde{k})\right]. \quad (B10)$$

We know that there exists $n_\Theta(L) > n_I(L) > \tilde{k}_s$ such that, for all $N_h > n_\Theta(L)L$,

$$4\beta_2 LQ(L)\sum_{\tilde{k}=N_h/L+1}^{\infty}\lambda_R(\tilde{k})\lambda_I(\tilde{k}) \leqslant \frac{\pi}{4}, \quad (B11)$$

which implies $\Theta \leqslant \pi/4$. Consider

$$\mathrm{Re}\left(\frac{\psi^{(L,\infty)}(\vec{\sigma})}{\psi^{(L,N_h)}(\vec{\sigma})}\right) \geqslant \sqrt{R_2}\cos(\Theta), \quad (B12)$$

$$\sum_{\vec{\sigma}}|\psi^{(L,\infty)}(\vec{\sigma})|^2 \leqslant R_1\sum_{\vec{\sigma}}|\psi^{(L,N_h)}(\vec{\sigma})|^2. \quad (B13)$$

We can get

$$\||\tilde{\Psi}^{(L,\infty)}\rangle - |\tilde{\Psi}^{(L,N_h)}\rangle\|^2$$

$$\leqslant 2 - \frac{2\sum_{\vec{\sigma}}|\psi^{(L,N_h)}(\vec{\sigma})|^2\sqrt{R_2}\cos(\Theta)}{\sqrt{R_1\left(\sum_{\vec{\sigma}}|\psi^{(L,N_h)}(\vec{\sigma})|^2\right)^2}} \quad (B14)$$

$$= 2 - 2\sqrt{\frac{R_2}{R_1}}\cos(\Theta) \quad (B15)$$

$$\leqslant 2 - 2\exp\left[-2LQ(L)\sum_{\tilde{k}=N_h/L+1}^{\infty}\left(\lambda_R^2(\tilde{k}) + \beta_1^2\lambda_I^2(\tilde{k})\right)\right]$$

$$\times \cos\left[4\beta_2 LQ(L)\sum_{\tilde{k}=N_h/L+1}^{\infty}\lambda_R(\tilde{k})\lambda_I(\tilde{k})\right]. \quad (B16)$$

For simplicity, we have assumed $\lambda_R(\tilde{k}) = \lambda_I(\tilde{k}) = \lambda(\tilde{k})$ which implies that the real part and imaginary part of RBM parameters have the same decaying rate and

$$P(N_h/L) = \sum_{\tilde{k}=N_h/L+1}^{\infty}\lambda^2(\tilde{k}). \quad (B17)$$

Then

$$\||\tilde{\Psi}^{(L,\infty)}\rangle - |\tilde{\Psi}^{(L,N_h)}\rangle\|^2$$

$$\leqslant 2 - 2\exp\left[-2(1+\beta_1^2)LQ(L)P(N_h/L)\right]$$

$$\times \cos[4\beta_2 LQ(L)P(N_h/L)] \quad (B18)$$

$$= F_1(LQ(L)P(N_h/L)), \quad (B19)$$

where $F_1(x)$ is defined in Eq. (17).

The idea of the proof is shown schematically in Fig. 2(c), where $\psi_{\mathrm{full}}(\vec{\sigma})$ denotes the amplitude for the full LRFD RBM $\psi^{(L,\infty)}(\vec{\sigma})$ and $\psi_{\mathrm{tr}}(\vec{\sigma})$ denotes the amplitude for the truncated RBM $\psi^{(L,N_h)}(\vec{\sigma})$. ∎

We give the proof for the second-type truncation errors as follows.

We have defined a Hermitian operator $\hat{B}$ of the form $\hat{B} = \bigotimes_{j=1}^{L}\hat{\sigma}_j^{(m_j)}$ where $\bigotimes$ is the tensor product symbol, $m_j \in \{0, 1, 2, 3\}$, $\hat{\sigma}_j^{(0)} = I_{2\times2}$ is the identity matrix, and $\{\hat{\sigma}_j^{(1)}, \hat{\sigma}_j^{(2)}, \hat{\sigma}_j^{(3)}\}$ denote the Pauli matrices.

Proof.

$$|\langle\hat{B}\rangle^{(L,\infty)} - \langle\hat{B}\rangle^{(L,N_h)}|$$

$$= \left|\frac{\langle\Psi^{(L,\infty)}|\hat{B}|\Psi^{(L,\infty)}\rangle}{\langle\Psi^{(L,\infty)}|\Psi^{(L,\infty)}\rangle} - \frac{\langle\Psi^{(L,N_h)}|\hat{B}|\Psi^{(L,N_h)}\rangle}{\langle\Psi^{(L,N_h)}|\Psi^{(L,N_h)}\rangle}\right| \quad (B20)$$

$$\leqslant \left|\frac{\langle\Psi^{(L,\infty)}|\hat{B}|\Psi^{(L,\infty)}\rangle}{\langle\Psi^{(L,\infty)}|\Psi^{(L,\infty)}\rangle} - \frac{\langle\Psi^{(L,\infty)}|\hat{B}|\Psi^{(L,\infty)}\rangle}{\langle\Psi^{(L,N_h)}|\Psi^{(L,N_h)}\rangle}\right|$$

$$+ \left|\frac{\langle\Psi^{(L,\infty)}|\hat{B}|\Psi^{(L,\infty)}\rangle}{\langle\Psi^{(L,N_h)}|\Psi^{(L,N_h)}\rangle} - \frac{\langle\Psi^{(L,N_h)}|\hat{B}|\Psi^{(L,N_h)}\rangle}{\langle\Psi^{(L,N_h)}|\Psi^{(L,N_h)}\rangle}\right| \quad (B21)$$

$$= G_1 + G_2, \quad (B22)$$

where

$$G_1 = \left|\frac{\langle\Psi^{(L,\infty)}|\Psi^{(L,\infty)}\rangle}{\langle\Psi^{(L,N_h)}|\Psi^{(L,N_h)}\rangle} - 1\right||\langle\hat{B}\rangle^{(L,\infty)}| \quad (B23)$$

and

$$G_2 = |\langle\Psi^{(L,\infty)}|\hat{B}|\Psi^{(L,\infty)}\rangle - \langle\Psi^{(L,N_h)}|\hat{B}|\Psi^{(L,N_h)}\rangle|$$

$$\times (\langle\Psi^{(L,N_h)}|\Psi^{(L,N_h)}\rangle)^{-1} \quad (B24)$$

$$= \sum_{\vec{\sigma}_1}\sum_{\vec{\sigma}_2}\left|B_{\vec{\sigma}_1\vec{\sigma}_2}\psi^{(L,N_h)}(\vec{\sigma}_1)^*\psi^{(L,N_h)}(\vec{\sigma}_2)\right.$$

$$\times\left.\left(\frac{\psi^{(L,\infty)}(\vec{\sigma}_1)^*\psi^{(L,\infty)}(\vec{\sigma}_2)}{\psi^{(L,N_h)}(\vec{\sigma}_1)^*\psi^{(L,N_h)}(\vec{\sigma}_2)} - 1\right)\right|$$

$$\times\left(\sum_{\vec{\sigma}}|\psi^{(L,N_h)}(\vec{\sigma})|^2\right)^{-1} \quad (B25)$$

capture the contribution of the deviations in the normalization factor and the unnormalized expectation value to the approximation error, respectively, with $B_{\vec{\sigma}_1\vec{\sigma}_2} = \langle\vec{\sigma}_1|\hat{B}|\vec{\sigma}_2\rangle$.

Considering $R_2 \leqslant \sum_{\vec{\sigma}}|\psi^{(L,\infty)}(\vec{\sigma})|^2/\sum_{\vec{\sigma}}|\psi^{(L,N_h)}(\vec{\sigma})|^2 \leqslant R_1$ and $|\langle\hat{B}\rangle^{(L,\infty)}| \leqslant 1$, we can get

$$G_1 \leqslant \max\{|R_1 - 1|, |1 - R_2|\}. \quad (B26)$$

Define

$$\xi = \max_{(\vec{\sigma}_1,\vec{\sigma}_2)}\left|\frac{\psi^{(L,\infty)}(\vec{\sigma}_1)^*\psi^{(L,\infty)}(\vec{\sigma}_2)}{\psi^{(L,N_h)}(\vec{\sigma}_1)^*\psi^{(L,N_h)}(\vec{\sigma}_2)} - 1\right|. \quad (B27)$$

Then

$$G_2 \leqslant \xi\sum_{\vec{\sigma}_1}\sum_{\vec{\sigma}_2}\left|B_{\vec{\sigma}_1\vec{\sigma}_2}\right||\psi^{(L,N_h)}(\vec{\sigma}_1)^*||\psi^{(L,N_h)}(\vec{\sigma}_2)|$$

$$\times\left(\sum_{\vec{\sigma}}|\psi^{(L,N_h)}(\vec{\sigma})|^2\right)^{-1} \quad (B28)$$

$$= \xi \sum_{\vec{\sigma}_2} |\psi^{(L,N_h)}(\hat{B}(\vec{\sigma}_2))^*| |\psi^{(L,N_h)}(\vec{\sigma}_2)|$$

$$\times \left(\sum_{\vec{\sigma}} |\psi^{(L,N_h)}(\vec{\sigma})|^2 \right)^{-1}, \qquad (B29)$$

where $\hat{B}(\vec{\sigma}_2)$ is the only spin configuration that makes $B_{\vec{\sigma}_1 \vec{\sigma}_2} \neq 0$ for a specific $\vec{\sigma}_2$. Using the Cauchy-Schwarz inequality,

$$\sum_{\vec{\sigma}_2} |\psi^{(L,N_h)}(\hat{B}(\vec{\sigma}_2))^*| |\psi^{(L,N_h)}(\vec{\sigma}_2)|$$

$$\leqslant \left(\sum_{\vec{\sigma}_2} |\psi^{(L,N_h)}(\hat{B}(\vec{\sigma}_2))^*|^2 \sum_{\vec{\sigma}_2} |\psi^{(L,N_h)}(\vec{\sigma}_2)|^2 \right)^{1/2} \quad (B30)$$

$$= \sum_{\vec{\sigma}} |\psi^{(L,N_h)}(\vec{\sigma})|^2. \qquad (B31)$$

Therefore, combined with the geometric features, we can get

$$G_2 \leqslant \xi \qquad (B32)$$

$$\leqslant \max \left\{ \sqrt{(R_1 \cos(2\Theta) - 1)^2 + R_1^2 \sin^2(2\Theta)}, \right.$$

$$\sqrt{(R_2 \cos(2\Theta) - 1)^2 + R_2^2 \sin^2(2\Theta)},$$

$$|R_1 - 1|, |1 - R_2| \Big\} \qquad (B33)$$

$$= \max \left\{ \sqrt{R_1^2 - 2R_1 \cos(2\Theta) + 1}, \right.$$

$$\left. \sqrt{R_2^2 - 2R_2 \cos(2\Theta) + 1} \right\}. \qquad (B34)$$

Therefore

$$|\langle \hat{B} \rangle^{(L,\infty)} - \langle \hat{B} \rangle^{(L,N_h)}| \leqslant \max\{|R_1 - 1|, |1 - R_2|\}$$

$$+ \max \left\{ \sqrt{R_1^2 - 2R_1 \cos(2\Theta) + 1}, \right.$$

$$\left. \sqrt{R_2^2 - 2R_2 \cos(2\Theta) + 1} \right\} \qquad (B35)$$

$$\leqslant F_2(LQ(L)P(N_h/L)), \qquad (B36)$$

where $F_2(x)$ is defined in Eq. (19). ∎

APPENDIX C: PROOF OF SPIN-CORRELATION FORMULA

In this section, we provide a proof of the relevant spin-correlation (in the z direction) formulas in Sec. II D.

Proof. We assume that all RBM parameters a_j's, b_k's, and $W_{j,k}$'s are no larger than ε_1, and $\varepsilon_1 \ll 1/L$, $\varepsilon_1 \ll 1/N_h$. In this section, we will not explicitly write the superscript "(L,N_h)" and assume that the RBM just has a finite number (N_h) of hidden nodes. Then we can get the Taylor series expansion of $|\psi(\vec{\sigma})|^2$ about $a_j = b_k = W_{j,k} = 0$ in the small-parameter regime:

$$|\psi(\vec{\sigma})|^2 = \exp \left(\sum_{j=1}^{L} 2 \operatorname{Re}(a_j) \sigma_j \right) \prod_{k=1}^{N_h} |\cosh(\theta_k)|^2 \quad (C1)$$

$$= F^{(0)} + \sum_{j_1 \in \{1,2,\dots,L\}} F_{j_1}^{(1)} \sigma_{j_1} + \cdots \qquad (C2)$$

$$= \sum_{n=0}^{L} \sum_{1 \leqslant j_1 < \cdots < j_n \leqslant L} F_{j_1,\dots,j_n}^{(n)} \prod_{j=j_1}^{j_n} \sigma_j, \qquad (C3)$$

where $F_{j_1,\dots,j_n}^{(n)}$ is the coefficient associated with the spin array $\prod_{j=j_1}^{j_n} \sigma_j$ only depending on the RBM parameters and the contraction $\sigma_j^2 = 1$ has been used.

Then, the unnormalized spin correlation between spin 1 and spin $1 + r$ in the longitudinal (z) direction is

$$C_{\text{unnorm}}^z(r) = \sum_{\sigma_1 = \pm 1} \cdots \sum_{\sigma_L = \pm 1} \sigma_1 \sigma_{1+r} |\psi(\vec{\sigma})|^2 \qquad (C4)$$

$$= \sum_{\sigma_1 = \pm 1} \cdots \sum_{\sigma_L = \pm 1} \sigma_1 \sigma_{1+r}$$

$$\times \sum_{n=0}^{L} \sum_{1 \leqslant j_1 < \cdots < j_n \leqslant L} F_{j_1,\dots,j_n}^{(n)} \prod_{j=j_1}^{j_n} \sigma_j. \quad (C5)$$

By applying the Taylor series expansions to each term, we can get

$$\exp \left(\sum_{j=1}^{L} 2 \operatorname{Re}(a_j) \sigma_j \right)$$

$$= \left(1 + 2 \sum_{j=1}^{L} \operatorname{Re}(a_j)^2 \right) + 2 \sum_{j=1}^{L} \operatorname{Re}(a_j) \sigma_j$$

$$+ 2 \sum_{j_1=1}^{L} \sum_{1 \leqslant j_2 \leqslant L, j_2 \neq j_1} \operatorname{Re}(a_{j_1}) \operatorname{Re}(a_{j_2}) \sigma_{j_1} \sigma_{j_2}$$

$$+ \mathcal{O}(\varepsilon_1^3) \quad \text{(as } \varepsilon_1 \to 0), \qquad (C6)$$

$$\theta_k^2 = \left(b_k^2 + \sum_{j=1}^{L} W_{j,k}^2 \right) + 2 b_k \sum_{j=1}^{L} W_{j,k} \sigma_j$$

$$+ \sum_{j_1=1}^{L} \sum_{1 \leqslant j_2 \leqslant L, j_2 \neq j_1} W_{j_1,k} W_{j_2,k} \sigma_{j_1} \sigma_{j_2}, \qquad (C7)$$

and then

$$|\cosh(\theta_k)|^2$$

$$= 1 + \frac{1}{2} \theta_k^2 + \frac{1}{2} (\theta_k^*)^2 + \mathcal{O}(\varepsilon_1^4) \qquad (C8)$$

$$= \left[1 + \operatorname{Re} \left(b_k^2 + \sum_{j=1}^{L} W_{j,k}^2 \right) \right] + 2 \operatorname{Re} \left(b_k \sum_{j=1}^{L} W_{j,k} \right) \sigma_j$$

$$+ \operatorname{Re} \left(\sum_{j_1=1}^{L} \sum_{1 \leqslant j_2 \leqslant L, j_2 \neq j_1} W_{j_1,k} W_{j_2,k} \right) \sigma_{j_1} \sigma_{j_2}$$

$$+ \mathcal{O}(\varepsilon_1^4) \quad \text{(as } \varepsilon_1 \to 0). \qquad (C9)$$

Only the $F_{1,1+r}^{(1)}$-related term will not vanish in the summation over all $\sigma_j = \pm 1$ as its associated $\prod_{j=j_1}^{j_n} \sigma_j$ exactly

cancels $\sigma_1\sigma_{1+r}$. Therefore

$$C_{\text{unnorm}}^z(r) = F_{1,1+r}^{(1)} \tag{C10}$$

$$= 2(\text{Re}(WW^T))_{1,1+r} + 4\,\text{Re}(a_1)\,\text{Re}(a_{1+r})$$

$$+ \mathcal{O}(\varepsilon_1^3) \quad (\text{as } \varepsilon_1 \to 0). \tag{C11}$$

So for RBMs constructed as Eqs. (24)–(26) show with $a_0 = 0$ and $c_w \in \mathbb{R}$ for simplicity,

$$C_{\text{unnorm}}^z(r) \approx 2(WW^T)_{1,1+r} \tag{C12}$$

$$= 2|c_w|^2 \sum_{\tilde{k}=1}^{N_h/L} |\lambda(\tilde{k})|^2 \sum_{j_c=1}^{L} \mu(|1 - j_c|_{\text{circ}})$$

$$\times \mu(|1 + r - j_c|_{\text{circ}}). \tag{C13}$$

Considering the definition of $|\dots|_{\text{circ}}$ and assume L is even without influencing the asymptotic analysis, the r-related part in $C_{\text{unnorm}}^z(r)$ is

$$G^z(r) = \sum_{j_c=1}^{L} \mu(|1 - j_c|_{\text{circ}})\mu(|1 + r - j_c|_{\text{circ}}) \tag{C14}$$

$$= \sum_{j_c=1}^{r-1} \mu(j_c)\mu(r - j_c) + \sum_{j_c=L/2+1-r}^{L/2-1} \mu(j_c)\mu(L - r - j_c)$$

$$+ 2\sum_{j_c=r+1}^{L/2} \mu(j_c)\mu(j_c - r) + 2\mu(0)\mu(r). \tag{C15}$$

Considering the variation of each term with varying j_c in the above summation in a ring-shaped geometry corresponding to the periodic boundary conditions, we know that the contribution of the first two summations dominates $G^z(r)$ as $r \to L/2$, and the contribution of the second summation is no larger than that from the first one. So we can just focus on the contribution of the first summation, which works as a lower bound on $G^z(r)$, when analyzing the asymptotic long-range behavior of $G^z(r)$ as $L \to \infty$ and $r \to L/2$.

For $\mu(r) = \delta_Q/r$,

$$\sum_{j_c=1}^{r-1} \mu(j_c)\mu(r - j_c)$$

$$= \delta_Q^2 \sum_{j_c=1}^{r-1} \frac{1}{r}\left(\frac{1}{j_c} + \frac{1}{r - j_c}\right) = \frac{2\delta_Q^2}{r} \sum_{j_c=1}^{r-1} \frac{1}{j_c} \tag{C16}$$

$$= 2\delta_Q^2 \frac{\ln r}{r} + \mathcal{O}\left(\frac{1}{r}\right) \quad (\text{as } L \to \infty, r \to L/2). \tag{C17}$$

For $\mu(r) = \delta_Q/r^{\alpha_Q}$ with $\alpha_Q > 1$,

$$\sum_{j_c=1}^{r-1} \mu(j_c)\mu(r - j_c)$$

$$= \delta_Q^2 \sum_{j_c=1}^{r-1} \frac{1}{r^{\alpha_Q}}\left(\frac{1}{j_c} + \frac{1}{r - j_c}\right)^{\alpha_Q} \tag{C18}$$

$$\geqslant \frac{\delta_Q^2}{r^{\alpha_Q}} \sum_{j_c=1}^{r-1}\left(\frac{1}{j_c^{\alpha_Q}} + \frac{1}{(r - j_c)^{\alpha_Q}}\right) = \frac{2\delta_Q^2}{r^{\alpha_Q}} \sum_{j_c=1}^{r-1} \frac{1}{j_c^{\alpha_Q}} \tag{C19}$$

$$= 2\delta_Q^2 \zeta(\alpha_Q)\frac{1}{r^{\alpha_Q}} + \mathcal{O}\left(\frac{1}{r^{2\alpha_Q}}\right)$$

$$(\text{as } L \to \infty, r \to L/2), \tag{C20}$$

where $\zeta(x)$ denotes the Riemann zeta function [58].

For $\mu(r) = \delta_Q/r^{\alpha_Q}$ with $0 < \alpha_Q \leqslant \frac{1}{2}$,

$$G^z\left(\frac{L}{2}\right) = 2\sum_{j_c=1}^{\frac{L}{2}-1} \mu(j_c)\mu\left(\frac{L}{2} - j_c\right) + 2\mu(0)\mu\left(\frac{L}{2}\right) \tag{C21}$$

$$\geqslant 2\delta_Q^2 \sum_{j_c=1}^{L/2-1} \frac{1}{j_c^{\alpha_Q}} \frac{1}{(L/2 - j_c)^{\alpha_Q}} \tag{C22}$$

$$\geqslant 2\delta_Q^2 \sum_{j_c=1}^{L/2-1} \frac{1}{[(j_c + L/2 - j_c)/2]^{2\alpha_Q}} \tag{C23}$$

$$\geqslant 2\delta_Q^2 \frac{L/2 - 1}{L/4} \tag{C24}$$

$$= 4\delta_Q^2 + o(1) \quad (\text{as } L \to \infty, r \to L/2). \tag{C25}$$

∎

While the truncation error for a geometric series can be directly estimated, we derive an estimation of the order of the truncation error for a series with terms having a power-law-decaying form as follows. This result serves for the complexity analysis in this work. We know that [58], for any $s > 1$, $b \geqslant a \geqslant 1$, and $a, b \in \mathbb{N}$,

$$\sum_{n=a+1}^{b} \frac{1}{n^s} = \frac{a^{1-s} - b^{1-s}}{s - 1} - s\int_a^b \frac{\left(x - \lfloor x \rfloor - \frac{1}{2}\right)}{x^{s+1}}dx$$

$$+ \frac{b^{-s} - a^{-s}}{2}, \tag{C26}$$

where $\lfloor x \rfloor$ denotes the floor function. Consider

$$\left|s\int_a^b \frac{\left(x - \lfloor x \rfloor - \frac{1}{2}\right)}{x^{s+1}}dx\right| \leqslant \frac{s}{2}\int_a^b \frac{1}{x^{s+1}}dx \tag{C27}$$

$$= \frac{a^{-s} - b^{-s}}{2}. \tag{C28}$$

Therefore

$$\sum_{n=a+1}^{b} \frac{1}{n^s} = \frac{a^{1-s} - b^{1-s}}{s - 1} + \mathcal{O}(a^{-s}) \quad (\text{as } a \to \infty). \tag{C29}$$

The above results also hold for $b \to \infty$. Therefore

$$\sum_{n=a+1}^{\infty} \frac{1}{n^s} = \frac{a^{1-s}}{s - 1} + \mathcal{O}(a^{-s}) = \mathcal{O}(a^{1-s}) \quad (\text{as } a \to \infty). \tag{C30}$$

Another remarkable point is that calculations of spin correlations in the x or y direction can be conducted in a similar way, but their forms are more complicated as the basis states in RBM forms are not naturally the eigenstates of $\hat{\sigma}_j^x$ or $\hat{\sigma}_j^y$ spin operators and thus new mathematical tricks are possibly needed to reformalize them.

APPENDIX D: LRFD RBMs APPROXIMATING THE KRONECKER DELTA FUNCTION

In this section, we construct LRFD RBMs which can approximate the quantum state with all spins pointing up in the z direction, which has a form of the Kronecker delta function, with arbitrary accuracy for any finite system size. The RBMs are constructed as Eqs. (24)–(26) show with

$$\mu(r) = \mu_0 > 0 \quad \text{(for any } r \geqslant 0\text{)}, \tag{D1}$$

$$c_w = c_b = 1, \tag{D2}$$

$$a_0 = 0. \tag{D3}$$

Since all b_k and $W_{j,k}$ parameters for these RBMs are real positive numbers, the wave-function amplitude for spin configurations reaches its maximum at $\vec{\sigma}_0 = (1, 1, \ldots, 1)$ which corresponds to all spins up. We can get that, for any other spin configurations $\vec{\sigma}'$,

$$|\psi^{(L,\infty)}(\vec{\sigma}_0)/\psi^{(L,\infty)}(\vec{\sigma}')|$$
$$\geqslant |\psi^{(L,\infty)}(\vec{\sigma}_0)/\psi^{(L,\infty)}(\vec{\sigma}_1)| \tag{D4}$$
$$= \prod_{\tilde{k}=1}^{\infty} \left[\frac{\cosh((L+1)\mu_0\lambda(\tilde{k}))}{\cosh((L-1)\mu_0\lambda(\tilde{k}))} \right]^L, \tag{D5}$$

where $\vec{\sigma}_1 = (1, 1, \ldots, 1, -1)$ refers to the spin configuration obtained by flipping the last spin in $\vec{\sigma}_0$. Let $x = (L-1)\mu_0\lambda(\tilde{k})$ and $\triangle x = 2\mu_0\lambda(\tilde{k})$. By performing the Taylor series expansion of $\cosh(x + \triangle x)$ about x and comparing the leading-order terms, we can get that there exist constants $0 < \beta_3 < 1$ and $x_0 > 0$ such that, for any $0 < \triangle x < x < x_0$,

$$\frac{\cosh(x + \triangle x)}{\cosh(x)} \geqslant 1 + \tanh(x)\triangle x \geqslant e^{\beta_3 x \triangle x}. \tag{D6}$$

For any fixed L and μ_0, there exists $\tilde{k}_0 \in \mathbb{N}$ such that $(L-1)\mu_0\lambda(\tilde{k}) < x_0$ for all $\tilde{k} > \tilde{k}_0$. So it can be proven that

$$|\psi^{(L,\infty)}(\vec{\sigma}_0)/\psi^{(L,\infty)}(\vec{\sigma}')|$$
$$\geqslant \prod_{\tilde{k}=1}^{\tilde{k}_0} \left[\frac{\cosh((L+1)\mu_0\lambda(\tilde{k}))}{\cosh((L-1)\mu_0\lambda(\tilde{k}))} \right]^L$$
$$\times \prod_{\tilde{k}=\tilde{k}_0+1}^{\infty} \left[\exp\left(2\beta_3(L-1)\mu_0^2\lambda^2(\tilde{k})\right) \right]^L \tag{D7}$$
$$\geqslant \exp\left(2\beta_3 L(L-1)\mu_0^2 \sum_{\tilde{k}=\tilde{k}_0+1}^{\infty} \lambda^2(\tilde{k})\right). \tag{D8}$$

The above lower bound contains infinitely many terms, though each of which is upper bounded by an expression associated with x_0, and can reach arbitrarily high values by decreasing the decaying rate of $\lambda(\tilde{k})$. Therefore the long-range nature of these RBMs allows the shape of $\psi^{(L,\infty)}(\vec{\sigma})$ in the spin-configuration space to approach the Kronecker delta function as $\lambda(\tilde{k})$ decays more and more slowly. Our numerical results in Fig. 6 with $\lambda(\tilde{k}) = \delta_P^{-(\tilde{k}-1)}$ imply that the distribution of the square of normalized wave-function amplitudes in the spin-configuration space can approximate the Kronecker delta function with increasing accuracy as $1/\delta_P$ grows up and the
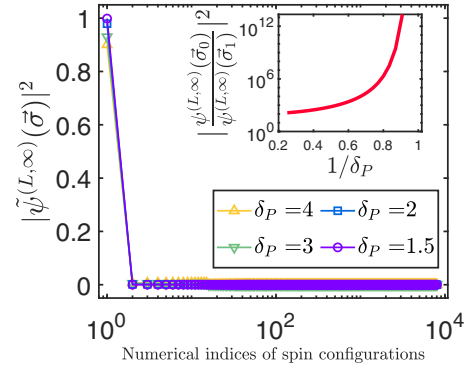

FIG. 6. Distribution of the square of normalized wave-function amplitudes in the spin-configuration space which can approximate the Kronecker delta function. The LRFD RBMs are constructed as Eqs. (D1)–(D3) show and $\mu_0 = 0.1$, $L = 13$, and $\lambda(\tilde{k}) = \delta_P^{1-\tilde{k}}$ with $\delta_P > 1$. The horizontal axis denotes the numerical indices of all 2^L spin configurations which are sorted in a monotonically decreasing order of their corresponding amplitudes. The inset shows the ratio $|\psi^{(L,\infty)}(\vec{\sigma}_0)/\psi^{(L,\infty)}(\vec{\sigma}_1)|^2$ as a function of $1/\delta_P$.

ratio $|\psi^{(L,\infty)}(\vec{\sigma}_0)/\psi^{(L,\infty)}(\vec{\sigma}_1)|^2$ as a measure of the approximation accuracy can reach arbitrarily high values as $1/\delta_P$ approaches 1, thus supporting our argument.

APPENDIX E: ERROR CURVES

In this section, we provide the approximation errors as a function of the number of hidden nodes for truncated LRFD RBMs [Fig. 7(a)] and the optimal RBMs as well as the truncation errors for the calculations of spin correlations in the x and z directions as a function of the number of levels kept in the truncated LRFD RBMs [Fig. 7(b)]. In Fig. 7(a), $\varepsilon(L_0, N_h)$ for LRFD RBMs denotes the second-type truncation errors with $\hat{B}$ being generalized into the Hamiltonian of the quantum system in the ground-state learning which usually has a form of a linear combination of polynomially many original $\hat{B}$-type operators. Then the approximation accuracy of using the optimal RBM with the number of hidden nodes not exceeding N_h is definitely better (at least no worse) than that
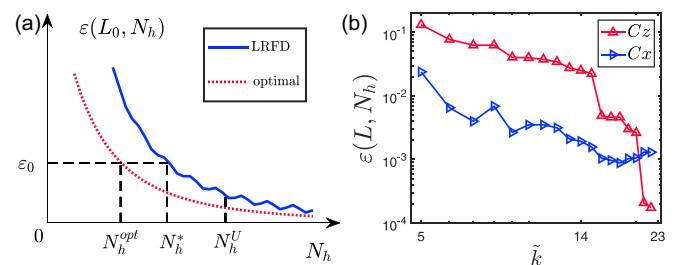

FIG. 7. (a) Schematic diagram of approximation errors as a function of the number of hidden nodes for truncated LRFD RBMs and the optimal RBMs with a fixed system size L_0. (b) Approximation errors for the calculations of spin correlations in the x (C_x) and z (C_z) directions as a function of the number of levels kept in the truncated LRFD RBMs for the ground states of the XXZ model compared with results from exact-diagonalization methods. The figure is plotted on a log-log scale.

of using an RBM which is a finite truncation of a LRFD RBM keeping N_h hidden nodes based on definitions. So $N_h^U(L_0, \varepsilon_0)$ in our complexity analysis actually also provides an upper bound on $N_h^{\text{opt}}(L_0, \varepsilon_0)$ defined as the minimum number of hidden nodes to achieve a specific approximation error ε_0 for any kinds of RBMs (not limited to LRFD RBMs), which is of great importance for pre-training computational-resource estimations in ML tasks. The truncation error for LRFD

RBMs will converge to 0 and is possibly not monotonically decreasing as N_h approaches infinity. Figure 7(b) shows that the decaying curves of the second-type truncation errors for the ground-state learning of the XXZ model with the same parameter setting as in Fig. 4 are consistent with our analysis and can be upper bounded by power-law decaying curves and reach high accuracy as the number of preserved levels increases.

---

[1] G. Carleo, I. Cirac, K. Cranmer, L. Daudet, M. Schuld, N. Tishby, L. Vogt-Maranto, and L. Zdeborová, Machine learning and the physical sciences, Rev. Mod. Phys. **91**, 045002 (2019).

[2] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature (London) **549**, 195 (2017).

[3] J. Carrasquilla and R. G. Melko, Machine learning phases of matter, Nat. Phys. **13**, 431 (2017).

[4] H.-Y. Huang, R. Kueng, G. Torlai, V. V. Albert, and J. Preskill, Provably efficient machine learning for quantum many-body problems, Science **377**, eabk3333 (2022).

[5] S. Das Sarma, D.-L. Deng, and L.-M. Duan, Machine learning meets quantum physics, Phys. Today **72**(3), 48 (2019).

[6] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, Science **355**, 602 (2017).

[7] R. G. Melko, G. Carleo, J. Carrasquilla, and J. I. Cirac, Restricted Boltzmann machines in quantum physics, Nat. Phys. **15**, 887 (2019).

[8] I. Glasser, N. Pancotti, M. August, I. D. Rodriguez, and J. I. Cirac, Neural-network quantum states, string-bond states, and chiral topological states, Phys. Rev. X **8**, 011006 (2018).

[9] X. Gao and L.-M. Duan, Efficient representation of quantum many-body states with deep neural networks, Nat. Commun. **8**, 662 (2017).

[10] G. Torlai and R. G. Melko, Learning thermodynamics with Boltzmann machines, Phys. Rev. B **94**, 165134 (2016).

[11] S. Lloyd, M. Mohseni, and P. Rebentrost, Quantum algorithms for supervised and unsupervised machine learning, arXiv:1307.0411.

[12] N. Wiebe, A. Kapoor, and K. M. Svore, Quantum algorithms for nearest-neighbor methods for supervised and unsupervised learning, Quantum Info. Comput. **15**, 316 (2015).

[13] N. Wiebe, A. Kapoor, and K. M. Svore, Quantum deep learning, Quantum Info. Comput. **16**, 541 (2016).

[14] P. Rebentrost, M. Mohseni, and S. Lloyd, Quantum support vector machine for big data classification, Phys. Rev. Lett. **113**, 130503 (2014).

[15] M. H. Amin, E. Andriyash, J. Rolfe, B. Kulchytskyy, and R. Melko, Quantum Boltzmann machine, Phys. Rev. X **8**, 021050 (2018).

[16] P. Mehta and D. J. Schwab, An exact mapping between the variational renormalization group and deep learning, arXiv:1410.3831.

[17] W. Hu, R. R. P. Singh, and R. T. Scalettar, Discovering phases, phase transitions, and crossovers through unsupervised machine learning: A critical examination, Phys. Rev. E **95**, 062122 (2017).

[18] K. Ch'ng, J. Carrasquilla, R. G. Melko, and E. Khatami, Machine learning phases of strongly correlated fermions, Phys. Rev. X **7**, 031038 (2017).

[19] L. Wang, Discovering phase transitions with unsupervised learning, Phys. Rev. B **94**, 195105 (2016).

[20] R. Pan and C. W. Clark, Implementing Majorana fermions in a cold-atom honeycomb lattice with textured pairings, Phys. Rev. A **98**, 033604 (2018).

[21] Y.-T. Hsu, X. Li, D.-L. Deng, and S. Das Sarma, Machine learning many-body localization: Search for the elusive nonergodic metal, Phys. Rev. Lett. **121**, 245701 (2018).

[22] M. Rupp, A. Tkatchenko, K.-R. Müller, and O. A. von Lilienfeld, Fast and accurate modeling of molecular atomization energies with machine learning, Phys. Rev. Lett. **108**, 058301 (2012).

[23] A. P. Bartók, S. De, C. Poelking, N. Bernstein, J. R. Kermode, G. Csányi, and M. Ceriotti, Machine learning unifies the modeling of materials and molecules, Sci. Adv. **3**, e1701816 (2017).

[24] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Neural-network quantum state tomography, Nat. Phys. **14**, 447 (2018).

[25] M. Kieferová and N. Wiebe, Tomography and generative training with quantum Boltzmann machines, Phys. Rev. A **96**, 062327 (2017).

[26] J. Liu, Y. Qi, Z. Y. Meng, and L. Fu, Self-learning Monte Carlo method, Phys. Rev. B **95**, 041101(R) (2017).

[27] L. Huang and L. Wang, Accelerated Monte Carlo simulations with restricted Boltzmann machines, Phys. Rev. B **95**, 035105 (2017).

[28] Z. Cai and J. Liu, Approximating quantum many-body wave functions using artificial neural networks, Phys. Rev. B **97**, 035116 (2018).

[29] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, Restricted Boltzmann machine learning for solving strongly correlated quantum systems, Phys. Rev. B **96**, 205152 (2017).

[30] D.-L. Deng, X. Li, and S. Das Sarma, Machine learning topological states, Phys. Rev. B **96**, 195145 (2017).

[31] D.-L. Deng, X. Li, and S. Das Sarma, Quantum entanglement in neural network states, Phys. Rev. X **7**, 021021 (2017).

[32] J. I. Cirac, D. Pérez-García, N. Schuch, and F. Verstraete, Matrix product states and projected entangled pair states: Concepts, symmetries, theorems, Rev. Mod. Phys. **93**, 045003 (2021).

[33] K. Choo, G. Carleo, N. Regnault, and T. Neupert, Symmetries and many-body excitations with neural-network quantum states, Phys. Rev. Lett. **121**, 167204 (2018).

[34] Y. Zheng, H. He, N. Regnault, and B. A. Bernevig, Restricted Boltzmann machines and matrix product states

of one-dimensional translationally invariant stabilizer codes, Phys. Rev. B **99**, 155129 (2019).

[35] H. He, Y. Zheng, B. A. Bernevig, and G. Sierra, Multi-layer restricted Boltzmann machine representation of 1D quantum many-body wave functions, arXiv:1910.13454.

[36] R. Kaubruegger, L. Pastori, and J. C. Budich, Chiral topological phases from artificial neural networks, Phys. Rev. B **97**, 195136 (2018).

[37] Y. Huang and J. E. Moore, Neural network representation of tensor network and chiral states, Phys. Rev. Lett. **127**, 170601 (2021).

[38] A. Golubeva and R. G. Melko, Pruning a restricted Boltzmann machine for quantum state reconstruction, Phys. Rev. B **105**, 125124 (2022).

[39] D. Sehayek, A. Golubeva, M. S. Albergo, B. Kulchytskyy, G. Torlai, and R. G. Melko, Learnability scaling of quantum states: Restricted Boltzmann machines, Phys. Rev. B **100**, 195125 (2019).

[40] F. Verstraete and J. I. Cirac, Matrix product states represent ground states faithfully, Phys. Rev. B **73**, 094423 (2006).

[41] S. R. Clark, Unifying neural-network quantum states and correlator product states via tensor networks, J. Phys. A: Math. Theor. **51**, 135301 (2018).

[42] J. Chen, S. Cheng, H. Xie, L. Wang, and T. Xiang, Equivalence of restricted Boltzmann machines and tensor network states, Phys. Rev. B **97**, 085104 (2018).

[43] N. Le Roux and Y. Bengio, Representational power of restricted Boltzmann machines and deep belief networks, Neural Comput. **20**, 1631 (2008).

[44] J. Eisert, M. Cramer, and M. B. Plenio, Colloquium: Area laws for the entanglement entropy, Rev. Mod. Phys. **82**, 277 (2010).

[45] Z. Huang and A. V. Balatsky, Complexity and geometry of quantum state manifolds, arXiv:1711.10471.

[46] M. Koshino, N. F. Q. Yuan, T. Koretsune, M. Ochi, K. Kuroki, and L. Fu, Maximally localized wannier orbitals and the extended Hubbard model for twisted bilayer graphene, Phys. Rev. X **8**, 031087 (2018).

[47] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 2012), Chap. 5, p. 274.

[48] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 2012), Chap. 5, p. 265.

[49] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, Matrix product state representations, Quantum Info. Comput. **7**, 401 (2007).

[50] S. Sorella, M. Casula, and D. Rocca, Weak binding between two aromatic rings: Feeling the van der Waals attraction by quantum Monte Carlo methods, J. Chem. Phys. **127**, 014105 (2007).

[51] E. Neuscamman, C. J. Umrigar, and G. K.-L. Chan, Optimizing large parameter sets in variational quantum Monte Carlo, Phys. Rev. B **85**, 045103 (2012).

[52] A. Harju, B. Barbiellini, S. Siljamäki, R. M. Nieminen, and G. Ortiz, Stochastic gradient approximation: An efficient method to optimize many-body wave functions, Phys. Rev. Lett. **79**, 1173 (1997).

[53] D. Jaschke, K. Maeda, J. D. Whalen, M. L. Wall, and L. D. Carr, Critical phenomena and Kibble–Zurek scaling in the long-range quantum Ising chain, New J. Phys. **19**, 033032 (2017).

[54] M. F. Maghrebi, Z.-X. Gong, and A. V. Gorshkov, Continuous symmetry breaking in 1D long-range interacting quantum systems, Phys. Rev. Lett. **119**, 023001 (2017).

[55] T. Koffel, M. Lewenstein, and L. Tagliacozzo, Entanglement entropy for the long-range Ising chain in a transverse field, Phys. Rev. Lett. **109**, 267203 (2012).

[56] P. Chen, Z. long Xue, I. P. McCulloch, M.-C. Chung, M. Cazalilla, and S.-K. Yip, Entanglement entropy scaling of the XXZ chain, J. Stat. Mech. (2013) P10007.

[57] F. C. Alcaraz and M. J. Martins, Conformal invariance and the Heisenberg model with arbitrary spin, Phys. Rev. Lett. **63**, 708 (1989).

[58] E. C. Titchmarsh and D. R. Heath-Brown, *The Theory of the Riemann Zeta-Function* (Oxford University Press, Oxford, 1986), Chap. 2, p. 14.