

Towards interpretability of neural quantum states

Fabian Döschl¹* and Annabelle Bohrdt²¹*Ludwig-Maximilians-University Munich, Theresienstr. 37, D-80333 Munich, Germany*
²*and Munich Center for Quantum Science and Technology, Schellingstr. 4, D-80799 Munich, Germany* (Received 24 October 2025; revised 15 June 2026; accepted 13 August 2026; published 13 October 2026)

Neural quantum states (NQS) have emerged as a powerful variational ansatz for representing quantum many-body wave functions. Their internal mechanisms, however, remain poorly understood. We investigate the role of correlations in NQS-like quantum state representation by employing a correlation-based interpretable neural network architecture and then proving our observations using Boolean function theory. The correlator neural network demonstrates that, even for simple product states, correlations up to the systems size in the chosen computational basis are required to represent a quantum state faithfully. We explain these observations using Fourier expansion, which reveals the correlator basis as the effective basis of the internal NQS structure, the resulting necessity for high-order correlations as supported by an entanglement bound that scales with the correlation order, the consequences of linear dependencies in constrained Hilbert spaces for correlation requirements, and the connections between spin basis rotations and the correlator basis. Furthermore, we analyze how neural networks achieve high correlation orders by increasing the magnitude of the network weights, which can be compensated by increasing the network depth. Lastly, we discuss how activation functions, network architectures, and the choice of reference basis influence correlation requirements. Our results provide insights and a better understanding of the internal structure and requirements of NQS, enabling a more systematic use of NQS in future research.

DOI: [10.1103/2cfz-dm5f](https://doi.org/10.1103/2cfz-dm5f)

I. INTRODUCTION

Two-dimensional quantum systems pose one of the most interesting challenges for current quantum many-body physics, as they are central to our understanding of high-temperature superconductivity, effects in strongly correlated spin systems, and topological phases. However, studying these systems with analytical or numerical methods is notoriously difficult.

State-of-the-art methods, such as tensor network methods (TN) or quantum Monte Carlo (QMC) approaches, are limited in their capability to find and represent quantum states in two-dimensional systems [1–4]. Recently, Carleo and Troyer proposed a variational ansatz that uses the expressive power of neural networks to represent quantum states [5]. These so-called neural quantum states (NQS) have shown remarkably good performance for many quantum systems [6–10] and have proven capable of capturing quantum states with a sign problem (limitation of QMC) [11–15] or volume-law entanglement (limitation of TN) [16–18]. However, unlike TN and QMC, NQS operate as a black-box method, making their controllability and a general understanding of how they work and when they fail elusive. In particular, it is unclear how the

network's internal structure is related to the correlations in the target quantum state.

Since the introduction of neural quantum states, there has been significant interest in understanding their representational power [16,17,19,20], optimization properties [6,21–25], and practical limitations [26]. Despite the difficulty in analyzing and interpreting these nonlinear variational functions, several studies have investigated the properties of neural quantum states analytically. A central result, extending beyond the study of NQS, is the proof that neural networks, given a potentially exponential number of parameters, can approximate any continuous function [27–29]. Although this universality also applies to the NQS-like representation of arbitrary quantum state wave functions, it does not guarantee parameter efficiency. Nevertheless, in practice, we know from many numerical [30–35] and analytical [16,19,36–39] examples that there exists, in many cases, an efficient representation of physically relevant quantum states.

However, despite the promising properties of neural quantum states, it is often extremely challenging to fine-tune a network architecture for a specific quantum system such that it yields the best results. A major hurdle for the optimization of NQS architectures is the fact that we often lack a clear understanding of the internal mechanisms of neural networks. A possibility to achieve a better understanding of the working principle is to draw connections between the neural network and the underlying correlations in the system. A step in this direction was taken in the article [40], where weighted correlations (dominant terms in the Hamiltonian) were added to a restricted Boltzmann machine to improve convergence.

*Contact author: Fa.Doeschl@physik.uni-muenchen.de

However, instead of improving a certain network architecture, we aim to reveal the required internal network structure for simulating quantum states. From the machine learning perspective, this has already been done for phase and image classification tasks, where correlation-based neural networks have been used to determine the important physical features for a successful characterization in terms of higher-order correlations [41–44]. The crucial idea for these interpretable neural network architectures is to replace the standard nonlinear activation function with a controlled expansion in higher-order correlations of the input, e.g., spin–spin correlations. Through regularization path analysis or tuning the maximal order of correlations considered as a hyperparameter, it is then possible to determine the most relevant correlation orders for a given *classification* task.

In these cases, it was enough to use low correlation orders (up to fourth order) to distinguish different phases. Yet, these insights do not directly transfer to quantum states, raising the question of which correlations are important to represent a quantum state, rather than merely characterizing it. Since representation requires the full set of wave function coefficients, it may depend on correlations far beyond those relevant for phase detection.

In this paper, we systematically investigate which correlations are relevant to represent a quantum state. Using the real-valued correlator transformer architecture [43], a polynomial ansatz with controllable expansion order, we study how truncating the expansion affects the representational power for two paradigmatic models: the two-dimensional Ising model and the toric code model. In combination with tools from Boolean function analysis, we identify conditions under which high correlation orders are unavoidable, even for simple product states, and when these requirements can be relaxed by Hilbert space truncations. This originates from the fact that the effective internal basis that the NQS uses to represent a quantum state, i.e., the basis in which the connection between neural network parameters and wave function coefficients arises naturally, is not the reference spin (or Fock) basis, but the correlator basis. This framework allows us to derive consequences for entanglement scaling, and to relate these insights to general neural quantum states by comparing their Taylor expansion to the required polynomial structure. Using the acquired knowledge, we discuss the differences between nonautoregressive and autoregressive neural networks and comment on commonly used network architectures that provide natural (dis)advantages from a Boolean function point of view.

The paper is structured as follows: In Sec. II, we introduce the NQS framework formally. Section III presents numerical results using the Correlator Quantum State (CQS), a polynomial architecture with controllable expansion order, for the 2D transverse-field Ising model and the perturbed toric code, restricted to the physical sector. Section IV reviews Boolean function analysis that provides the analytical foundation for explaining our numerical observations, which is given in Sec. V. Section VI addresses the choice of computational basis, the entanglement bounds for fixed expansion orders, and the insights for general neural quantum states. Finally, in Sec. VII, we draw conclusions about the desired NQS structure, enabling a structured optimization of its activation functions and architectures.

II. NEURAL QUANTUM STATES

For an exact representation of a quantum state, one generally needs Hilbert space dimension H_{dim} many wave function coefficients to describe the system in terms of basis states,

$$|\Psi_\lambda\rangle = \sum_{\sigma} \psi_{\text{state}}^{\sigma} |\sigma\rangle. \quad (1)$$

Without loss of generality, we will mainly focus on spin-1/2 systems in this paper. As storing and evaluating all H_{dim} coefficients is exponentially costly, one searches for a function that efficiently approximates these wave function coefficients $\psi_{\text{state}} = \{\psi_{\text{state}}^{\sigma}\}$. Neural quantum states (NQS) approach this by using a neural network to represent such a tunable function $\psi_{\theta}(\sigma)$, which maps each basis state σ to an output A_{θ}^{σ} ,

$$\psi_{\theta} : \{-1, 1\}^L \longrightarrow A_{\theta} \subseteq \mathbb{C} \quad \text{with } |A_{\theta}| \leq 2^L, \quad (2)$$

where A_{θ} is the subset of $\mathbb{C}$ that contains all wave function coefficients $\{A_{\theta}^{\sigma}\}$. For a trained model, the normalized set $\frac{A_{\theta}}{\|A_{\theta}\|_2}$ is expected to approximate the wave function of the desired state $\frac{A_{\theta}}{\|A_{\theta}\|_2} \approx \psi_{\text{state}}$. There is, however, no guarantee that there exists a nonexponentially scaling set of parameters θ such that the function $\psi_{\theta}(\sigma)$ represents the desired quantum state exactly or even approximately.

Correlator quantum states. The idea that any network architecture can be expanded using Taylor series (at least locally), makes a correlator neural network [41,43,44] that simulates the Taylor expansion up to some order n , a valuable tool to acquire information about the required network structure. In our study, we analyze the structure by changing the maximal accessible expansion (correlation) order n . We define the highest expansion order n by all $\binom{L}{n}$ correlations $\mathcal{X}_S(\vec{\sigma})$,

$$\mathcal{X}_S(\vec{\sigma}) = \prod_{i \in S} \sigma_i, \quad (3)$$

with $|S| = n$. Note that S is a set of n out of L sites i that specifies the product of spins σ_i within a spin configuration $\vec{\sigma}$.

For clarity, when discussing correlations, we refer to products over spins within one sample σ , as opposed to expectation values of correlators such as $\langle \hat{S}_i^z \hat{S}_j^z \rangle$.

For a neural quantum state, the neural network maps a basis configuration σ to a wave function coefficient $\psi_{\theta}(\vec{\sigma})$ through a combination of linear operations and applications of nonlinear activation functions. The nonlinear activation functions in principle enable contributions of arbitrarily high orders of correlation functions between the inputs (e.g., between $\hat{\sigma}_i^z$ at different lattice sites).

To numerically investigate the dependence of NQS on correlations $\mathcal{X}_S(\vec{\sigma})$, we employ the correlator transformer architecture proposed in Suresh *et al.* [43]. A detailed description is provided in Sec. S1 of the Supplemental Material [45].

The main idea of this transformer is to operate without standard nonlinear activation functions and to instead employ the nonlinearity of higher-order correlations within a snapshot. These correlations are weighted and then used to evaluate the desired system. Based on the model's performance, one can infer the importance of different correlation orders.

In the context of NQS, the correlator quantum state (CQS) wave function is defined as

$$|\psi\rangle = \sum_{\vec{\sigma}} \psi(\vec{\sigma}) |\vec{\sigma}\rangle \quad \text{with} \quad \psi(\vec{\sigma}) = W^{\text{pred}} \mathbb{X}(\vec{\sigma}) + \beta, \quad (4)$$

where $\mathbb{X}$ contains all correlation matrices $\mathbb{X} = [\bar{X}^1, \bar{X}^2, \dots, \bar{X}^i, \dots, \bar{X}^n]$. Each entry of $\bar{X}^i$ contains a weighted sum over all correlation orders j of the same parity with $j \leq i$. To simplify the interpretation, we solely focus on real-valued weights and wave function coefficients.

In the following, we analyze the general structure of CQS based on the mathematical framework of Boolean functions and investigate the underlying requirements on network architectures to represent quantum states.

III. QUANTUM STATE CHARACTERIZATION VS REPRESENTATION

It is well established that conventional phases of matter can be probed and characterized using local observables. For instance, in Ising systems, magnetization serves as a key local observable that reveals information about the underlying phase.

This concept of characterizing a quantum state based on local information can also be found in machine learning. Following the ideas of interpretability, many articles [41–44] have shown an efficient phase detection by considering a small number of low-order correlations within the system.

Similarly, one can ask a network which and how many correlations it needs to represent a quantum state. At first glance, one might assume that the highest order contributing to representing a quantum state is the same as the one required to characterize it. Although these two tasks appear related, their required structure is inherently different.

Even when performing a Taylor expansion of the neural network architecture around a point $\vec{p}$,

$$\psi_{\theta}(\vec{\sigma}) = \sum_{n=0}^{\infty} \frac{\psi_{\theta}^{(n)}(\vec{p})}{n!} (\vec{\sigma} - \vec{p})^n, \quad (5)$$

one might also expect main contributions from lower correlation orders. Here, $\psi_{\theta}(\vec{\sigma})$ can be any arbitrary network architecture with weights θ and a converging Taylor series. The $1/n!$ in the Taylor series suppresses higher-order terms by construction, as required for convergence. Therefore, one might conclude from working NQS examples that only lower-order correlations play a significant role. In the following, we will show that this is not necessarily the case. We first show numerical results obtained by using a correlator transformer architecture, and then explain the findings based on the mathematical framework of Boolean functions.

Numerical observations. In the following, we apply the real-valued correlator transformer to learn the ground state of the Ising model and the perturbed toric code. To probe the representational requirements of CQS for these systems, we vary the highest accessible correlation order n . The CQS, denoted C_n , with maximal order n then contains all possible correlations up to and including order n , which we informally

write as

$$C_n = \sum_{i=0}^n \sum_{S \subseteq [L]} \mathcal{X}_S(\vec{\sigma}) \delta_{|S|,i}, \quad (6)$$

where $[L]$ denotes the set of all possible combinations of sites in the system. The resulting convergence of the energy is shown in Fig. 1(a) for the Ising model and in Fig. 1(b) for the perturbed toric code, respectively.

First, we show that our observations contradict the naive expectation; then we review tools from Boolean function analysis (Sec. IV) that allow us to explain our results in Sec. V.

Ising model. We start by applying the correlator transformer NQS (CQS) to the Ising model,

$$\hat{H}_{\text{Ising}} = \sum_{(i,j)} \hat{\sigma}_i^z \hat{\sigma}_j^z, \quad (7)$$

where σ^{α} are the standard Pauli operators. We define our system in the σ^z basis. This Hamiltonian has two-body interactions, and the emergent ground-state phases are fully characterized by first- and second-order correlations. Even with machine learning, these correlations are sufficient to characterize the phases [43,44].

As shown in Fig. 1(a), which plots the energy convergence of different CQS C_n , the correlator transformer architecture requires access to all correlation orders to obtain the best results. Therefore, a quantum many-body wave function does not generally fall off at high correlation orders. Note that poor performance of our CQS for lower correlation orders cannot be ruled out a priori, as the ansatz may be unable to represent the state or to find the optimal parameters. To exclude convergence or expressivity issues of the CQS, we also computed the exact lowest energy for the respective maximal correlation order C_n . The minimal energies for a given expansion order are displayed as horizontal lines in Fig. 1(a). The derivation of the Hamiltonian in the correlator basis is given in Sec. IV A.

While $C_1 = \mathcal{X}_{\emptyset} + \mathcal{X}_{|S|=1}$ = bias + order1 completely fails to converge to the antiferromagnetic (AFM) ground state C_8 , which includes correlations up to eighth order, achieves better results. However, the relative error remains at approximately $\epsilon \approx 10\%$. A further increase in the order of correlations significantly improves the results. With C_{11} , the relative error is reduced to approximately $\epsilon \approx \mathcal{O}(10^{-2})$. However, the best performance, $\epsilon \approx \mathcal{O}(10^{-3})$, is achieved only when all correlation orders are included ($C_L = C_{16}$). From the naive perspective, this is unexpected, as even C_8 and C_{11} , which contain all correlations up to orders 8 and 11, respectively, still fail to fully represent the antiferromagnetic ground state.

We emphasize the importance of this finding: The AFM state is a product state—the simplest form of a quantum state. While it is trivial to encode this state as a Matrix Product State with bond dimension 1, CQS in the $|\vec{\sigma}_z\rangle$ basis requires access to all L correlation orders in the system. Note that this does not imply that product states are fundamentally difficult for CQS to represent; rather, it shows that the required polynomial expansion order can scale linearly with system size, even for simple states.

In Sec. V, we explain our findings based on a mathematical framework. For instance, it can be shown analytically that

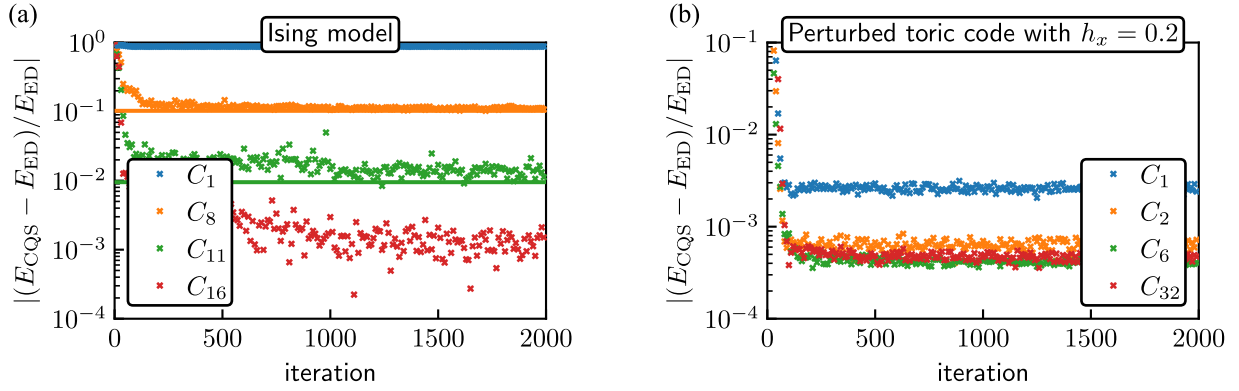


FIG. 1. Convergence of the correlator transformer (square lattice of linear size $L = 4$) applied to (a) the Ising model and (b) the perturbed toric code. C_n defines a network that uses all correlations up to order n . The horizontal lines shown in (a) are the (exact) lowest possible energies for the respective maximal correlation order. The exact energies in the correlator basis were calculated using Eq. (17).

the 16th-order correlation is (for the AFM ground state) a significant contribution to the wave function.

Toric code model. We now turn to the perturbed toric code model, which is particularly interesting for our study as it exhibits topological order arising from four-body interactions. The corresponding Hamiltonian is defined by

$$\hat{H}_{\text{TC}} = - \sum_j \prod_{i \in +_j} \hat{\sigma}_i^x - \sum_j \prod_{i \in \square_j} \hat{\sigma}_i^z + h_x \sum_i \hat{\sigma}_i^x. \quad (8)$$

Again, σ^α are the usual Pauli operators. We define the spins in the system as the connecting links between sites, such that the total Hilbert space dimension is given by $2^{2(L_x \times L_y)}$. For the CQS simulation, we choose to constrain the Hilbert space to the physical sector where Gauss's law,

$$G_i = \prod_{j \in +_i} \hat{\sigma}_j^x \stackrel{!}{=} 1 \quad (9)$$

is always fulfilled, thus reducing the effective Hilbert space to $2^{(L_x \times L_y)+1}$ [46,47]. This is achieved by choosing the σ^x basis as the computational basis, which allows us to restrict the Monte Carlo updates to states from the physical sector.

The perturbed toric code is a Hamiltonian with a topological phase, which makes it, in principle, difficult to characterize the relevant information for a given phase or wave function. For phase detection tasks, it was shown that fourth-order correlations are sufficient to achieve a successful classification [43]. Similar to the Ising model, one can show that a CQS requires all correlations in the system if one works in the full Hilbert space (see Sec. V). However, in the considered case with a Hilbert space truncation, it is at first unclear which and how many correlations are required for a full description.

Note that if we enforce Gauss's law, the limiting case of the pure toric code model ($h_x = 0$) has a trivial solution, which is given by a constant $\psi_{\text{res}}^{h_x=0}(\vec{\sigma}) = C_0$.

In Fig. 1(b), we show the convergence of the correlator transformer model for the perturbed toric code with $h_x = 0.2$. One immediately notices that the perturbed toric code, despite the topologically ordered phase, is significantly easier to learn than the AFM ground state. All evaluated correlator transformers achieve a reasonable error of $\epsilon \lesssim \mathcal{O}(10^{-3})$. This can be attributed to the fact that the system at $h_x = 0.2$

remains in the topologically ordered phase, so that only minor deviations from the unperturbed ground state $\psi_{\text{res}}^{h_x=0}(\vec{\sigma}) = C_0$ are expected.

While C_1 , which contains first-order correlations and a bias, already achieves a reasonable error of $\epsilon \approx \mathcal{O}(10^{-3})$, the results can still be improved by providing access to higher-order correlations. However, when using C_6 , an optimization limit is already reached, where we do not observe any further improvement when increasing the number of accessible correlation orders. Even when access to all correlations is allowed in C_{32} , we do not observe a major difference in the convergence or the energy errors.

Although a good performance of the CQS for the slightly perturbed toric code with enforced Gauss's law is generally expected, the stagnation observed for $C_{\geq 6}$ is surprising, given that the number of correlations accessible to the C_{32} CQS is several orders higher than those available to the C_6 CQS [compare with Eq. (16)]. In Sec. V, we explain the reduced number of required correlations using the decision tree formalism described in Sec. IV B. In particular, we show explicitly how the linear dependencies induced by Gauss's law reduce the correlation orders needed to represent the ground state.

Remark 1. The best energies that we obtained in the training are limited by the representational power of our network architecture. However, the correlator transformer was built to yield interpretability at the cost of reduced complexity. Together with the limited number of parameters, which we set to $\#\text{params} \approx \mathcal{O}(2 \times 10^4)$ for all CQS simulations, we do not expect a perfect convergence.

IV. ANALYSIS OF FUNCTIONS WITH BINARY INPUTS

As we numerically observed in the previous section, neural quantum states cannot behave the same as in typical machine learning tasks, where mostly low correlation orders are relevant [41,43,44]. In the following, we review the Boolean function formalism that reveals the correlator basis as the internal, input-independent basis of NQS-like methods (Sec. IV A). Thereafter, we discuss the decision tree formalism that provides a constructive framework for systems with truncated Hilbert spaces and thereby explains the potential relaxation of correlation order requirements (Sec. IV B). Finally, we

provide a more familiar picture of why a broad correlation order spectrum is needed, by looking at a set of equations (Sec. IV C).

A. Boolean function analysis

While neural quantum states are often perceived as a purely “black box” technique, we show that a much better understanding of the working principle can be achieved when studying the underlying structure. To do so, we follow the mathematical framework of Boolean functions [48]. Note that there is an extension of this framework to higher local Hilbert space dimensions and to one-hot encoding. We comment on these generalizations in Sec. S2 of the Supplemental Material [45].

The analysis of real-valued Boolean functions $\psi : \{-1, 1\}^L \rightarrow \mathbb{R}$ is based on studying their Fourier expansion. The Fourier expansion can be readily derived by introducing an indicator polynomial,

$$1_{\vec{a}}(\vec{\sigma}) = \frac{(1 + \alpha_1 \sigma_1)}{2} \dots \frac{(1 + \alpha_L \sigma_L)}{2} = \begin{cases} 1, & \vec{\sigma} = \vec{a}, \\ 0, & \vec{\sigma} \neq \vec{a}. \end{cases} \quad (10)$$

This indicator function is a multilinear polynomial that includes correlations up to L th order. Consequently, any wave function of a finite-size spin system has a polynomial representation of the form:

$$\psi(\vec{\sigma}) = \sum_{\vec{a} \in \{-1, 1\}^L} \psi(\vec{a}) 1_{\vec{a}}(\vec{\sigma}). \quad (11)$$

Rewriting this polynomial representation in terms of all possible (overlapping) subsets S of $[L] = [\sigma_1, \sigma_2, \dots, \sigma_L]$ yields the Fourier representation of $\psi(\vec{\sigma})$,

$$\psi(\vec{\sigma}) = \sum_{S \subseteq [L]} \langle \vec{\sigma} | \mathcal{X}_S \rangle \langle \mathcal{X}_S | \psi \rangle = \sum_{S \subseteq [L]} \bar{\psi}(S) \mathcal{X}_S(\vec{\sigma}), \quad (12)$$

where the corresponding Fourier coefficients are defined as

$$\bar{\psi}(S) = \langle \mathcal{X}_S | \psi \rangle = \frac{1}{2^L} \sum_{\vec{\sigma} \in \{-1, 1\}^L} \psi(\vec{\sigma}) \mathcal{X}_S(\vec{\sigma}), \quad (13)$$

and the monomial $\mathcal{X}_S(\vec{\sigma})$ corresponding to S is given by

$$\mathcal{X}_S(\vec{\sigma}) = \langle \vec{\sigma} | \mathcal{X}_S \rangle = \prod_{i \in S} \sigma_i, \quad (14)$$

with $\mathcal{X}_\emptyset(\vec{\sigma}) = 1$. This construction yields a total 2^L monomials that form an orthonormal basis for the $H_{\text{dim}} = 2^L$ dimensional Hilbert space [48],

$$\langle \mathcal{X}_S | \mathcal{X}_T \rangle = \frac{1}{2^L} \sum_{\vec{\sigma} \in \{-1, 1\}^L} \mathcal{X}_S(\vec{\sigma}) \mathcal{X}_T(\vec{\sigma}) = \begin{cases} 1, & S = T, \\ 0, & S \neq T. \end{cases} \quad (15)$$

Note that the k th correlation order (all S of the same degree: $|S| = k$) contributes $\frac{L!}{k!(L-k)!}$ monomials to the basis. Thus, a CQS C_n with correlations up to n th order has access to $N(C_n)$ basis vectors,

$$N(C_n) = \sum_{k=0}^n \frac{L!}{k!(L-k)!} = \sum_{k=0}^n \binom{L}{k} \stackrel{n=L}{=} 2^L. \quad (16)$$

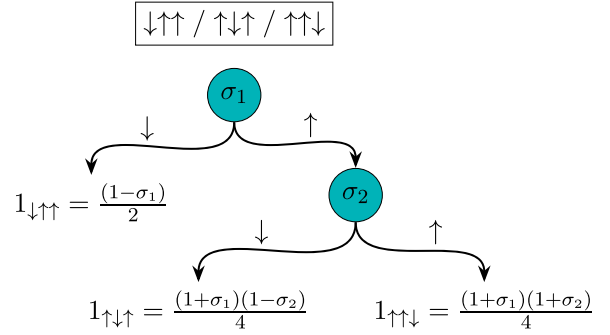


FIG. 2. Decision tree for a three spin model with a fixed magnetization of $\langle \sum_i \hat{\sigma}_i^z \rangle = 1$.

Consequently, ground states (GS) of Hamiltonians

$$\hat{H} = \sum_{S, S' \subseteq [L]} \left(\frac{1}{2^L} \sum_{\vec{\sigma}, \vec{\sigma}'} \langle \mathcal{X}_S | \vec{\sigma} \rangle h_{\vec{\sigma}, \vec{\sigma}'} \langle \vec{\sigma}' | \mathcal{X}_{S'} \rangle \right) |\mathcal{X}_S\rangle \langle \mathcal{X}_{S'}|, \quad (17)$$

with relevant contributions $h_{R, S'}$ in higher-order correlations $|R| > n$, such that $\bar{\psi}_{\text{GS}}(R) \neq 0$, cannot be fully represented by the CQS C_n for $n < |R| \leq L$. Here, $h_{\vec{\sigma}, \vec{\sigma}'}$ is the usual matrix element $h_{\vec{\sigma}, \vec{\sigma}'} = \frac{1}{2^L} \langle \vec{\sigma} | \hat{H} | \vec{\sigma}' \rangle$. Note that because of the normalization of $\langle \mathcal{X}_S | \mathcal{X}_T \rangle = \delta_{S, T}$, we used the convention $\langle \vec{\sigma} | \vec{\sigma}' \rangle = 2^L \delta_{\vec{\sigma}, \vec{\sigma}'}$ and $\mathbb{1} = \frac{1}{2^L} \sum_{\vec{\sigma} \in \{-1, 1\}^L} |\vec{\sigma}\rangle \langle \vec{\sigma}|$.

Remark 2. We want to stress the importance of this formalism for neural quantum states: Eq. (12) decouples the influence of the network weights, included in $\bar{\psi}(S)$, from the input configuration, given in $\mathcal{X}_S(\vec{\sigma})$. In other words, the machine learning algorithm approximates the quantum state in the correlator basis $|\mathcal{X}_S\rangle$ and rotates it into the desired spin basis,

$$|\psi\rangle = \sum_{S \subseteq [L]} \bar{\psi}(S) |\mathcal{X}_S\rangle \Rightarrow \sum_{\vec{\sigma} \in \{-1, 1\}^L} \sum_{S \subseteq [L]} \bar{\psi}(S) \langle \vec{\sigma} | \mathcal{X}_S \rangle |\vec{\sigma}\rangle. \quad (18)$$

Therefore, we can think of the correlator basis $|\mathcal{X}_S\rangle$ as the internal basis that neural quantum states use for quantum state representation. Consequently, operations that are simple in the spin basis, such as setting wave function coefficients $\psi(\vec{\sigma})$ to zero for selected configurations, can translate into constraints on $\bar{\psi}(S)$ that may require cancellations among multiple potentially nontrivial coefficients. This does not rule out efficient neural network representations, but shows that such efficiency cannot, in general, be inferred from the simplicity or sparsity in the spin basis for an arbitrary NQS ansatz.

B. Decision tree

Although the Fourier expansion described above provides a good starting point for the analysis, it becomes challenging to understand the behavior of NQS in systems with artificially reduced Hilbert spaces, such as fixed magnetization, Gauss's law, or particle conservation. To address this, one can introduce a decision tree representation (see Fig. 2). Similar to autoregressive NQS models [49, 50], this approach constructs

computation paths through the lattice and sequentially checks for allowed (disallowed) local spin configurations. Following this, the restricted wave function ψ_{res} can be written as

$$\psi_{\text{res}}(\vec{\sigma}) = \sum_{\vec{p} \in T} \psi(\vec{p}) 1_{\vec{p}}^{\text{res}}(\vec{\sigma}), \quad (19)$$

where $\vec{p}$ denotes a valid path in the decision tree T . Again, $1_{\vec{p}}^{\text{res}}(\vec{\sigma})$ is an indicator polynomial built from $\frac{(1+p_i\sigma_i)}{2}$ elements. Once a branch in the tree becomes trivial (e.g., all particles are distributed), this polynomial ends.

In the context of the Fourier (or Taylor) expansion, the restriction of the Hilbert space reduces the maximal correlation order in the Fourier representations to the maximal length of a branch $\max |\vec{p}|$ (compare with Fig. 2). Importantly, because of the truncation of the Hilbert space $\mathcal{H}_{\text{res}}$, many of the monomials $\mathcal{X}_S$ become linearly dependent, which implies that the Fourier basis loses (at least partially) its orthogonality: $\max |\vec{p}| \geq \log_2 \dim(\mathcal{H}_{\text{space}}^{\text{res}})$.

So far, we have considered only one single decision tree. However, decision trees are not unique. Therefore, a neural network can take the form of a sum over all possible decision trees $\mathbb{T} = [T_1, T_2, \dots]$ for each configuration and select the set of paths with minimal depth to the desired state,

$$\psi_{\text{res}}(\vec{\sigma}) = \sum_{\vec{\alpha} \in \mathcal{H}_{\text{res}}} \psi(\vec{\alpha}) \sum_{\vec{p} \in \mathbb{T}} c_{\vec{\alpha}, \vec{p}} 1_{\vec{p}}^{\text{res}}(\vec{\sigma}), \quad (20)$$

where $c_{\vec{\alpha}, \vec{p}}$ has to fulfill

$$\sum_{\vec{p} \in \mathbb{T}} c_{\vec{\alpha}, \vec{p}} 1_{\vec{p}}^{\text{res}}(\vec{\sigma}) = \begin{cases} 1, & \vec{\sigma} = \vec{\alpha}, \\ 0, & \vec{\sigma} \neq \vec{\alpha}. \end{cases} \quad (21)$$

For the three-spin example with a fixed magnetization of $\langle \sum_i \sigma_z^i \rangle = 1$, shown in Fig. 2, this is equivalent to choosing

$$1_{\uparrow\uparrow\uparrow} = \frac{(1 - \sigma_1)}{2}, \quad 1_{\uparrow\uparrow\downarrow} = \frac{(1 - \sigma_2)}{2}, \quad 1_{\uparrow\downarrow\downarrow} = \frac{(1 - \sigma_3)}{2}. \quad (22)$$

Note that σ_3 is constrained by $\sigma_3 = 1 - (\sigma_1 + \sigma_2)$.

Importantly, not every truncation of the Hilbert space or symmetry of the wave function reduces the maximal required expansion order. For example, using the spin-flip symmetry

$$\psi_{\text{eff}}(\sigma) = \frac{1}{2}[\psi(\sigma) + \psi(-\sigma)], \quad (23)$$

removes all odd correlation orders, but leaves the even ones unchanged. Therefore, a reduced number of independent amplitudes does not by itself imply a reduced maximal correlation order. The decision tree formalism is a constructive framework that allows a better understanding of how linear dependencies from restricted Hilbert spaces influence the correlation requirements of neural networks. The minimal required correlation order for an exact representation of a quantum state is given by the maximal branch depth of the optimal set of decision trees, which can be significantly smaller than the system size.

Remark 3. In classical machine learning tasks, such as image recognition, the training data constitutes only a small fraction of the full configuration space. Consequently, linear dependencies arise artificially because of sampling sparsity,

which artificially reduces the correlation requirements of the respective tasks.

C. Intuitive approach to Fourier expansion

In the Secs. IV A and IV B, we showed formally that every function $\psi : \{-1, 1\}^L \rightarrow \mathbb{R}$ can be written as a Fourier expansion [48]. Here, we provide a more intuitive picture of what this means for neural quantum states.

Writing out the Fourier expansion of a wave function $\psi(\vec{\sigma})$ explicitly reveals its structure,

$$\psi(\vec{\sigma}) = W + \sum_i A_i \sigma_i + \sum_{j < i} B_{ij} \sigma_i \sigma_j + \dots + F \sigma_1 \sigma_2 \sigma_3 \dots \sigma_L, \quad (24)$$

where each variable ($W, A_i, B_{ij}, \dots, F$) corresponds to a specific Fourier coefficient $\psi(S)$. Notice the similarity to a Taylor expansion. Writing the wave function coefficient $\psi(\vec{\sigma})$ for each basis state yields a set of 2^L equations,

state	$\psi(\sigma)$
$ \uparrow\uparrow\uparrow\dots\uparrow\rangle$	$A_1 + A_2\dots + B_{12} + B_{13}\dots + F\dots + W$
$ \downarrow\uparrow\uparrow\dots\uparrow\rangle$	$A_1 - A_2\dots - B_{12} - B_{13}\dots - F\dots + W$
$\vdots$	$\vdots$
$ \uparrow\downarrow\downarrow\dots\downarrow\rangle$	$A_1 - A_2\dots - B_{12} - B_{13}\dots \mp F\dots + W$
$ \downarrow\downarrow\downarrow\dots\downarrow\rangle$	$-A_1 - A_2\dots + B_{12} + B_{13}\dots \pm F\dots + W$

(25)

The signs in front of the coefficients reflect the products of spins in that state. Importantly, the parity of the prefactor of F , $\sigma_1 \sigma_2 \sigma_3 \dots \sigma_L = \pm 1$, changes with the number of lattice sites.

The task we aim to achieve with neural quantum states is to solve these 2^L linearly independent equations, which generally requires all 2^L Fourier coefficients $A_i, B_{ij}, \dots$, so we cannot arbitrarily assume that some coefficients are negligible. Note that setting coefficients of higher-order correlations to zero is equivalent to overdetermining the system, which can, depending on the quantum state, prevent an exact solution and make approximations difficult, if not impossible.

From this perspective, the Fourier representation provides a natural language to understand the requirements for neural quantum states. The network must be able to generate the set of correlations needed to represent the wave function. Whether this requires all orders or only a truncated subset depends on both the quantum state and the choice of computational basis, as we discuss in later sections.

V. EXPLAINING OBSERVATIONS BASED ON MATHEMATICAL FRAMEWORK

Based on the mathematical framework introduced in Secs. IV A and IV B, we now analyze the numerical observations made in Sec. III. We begin with the Ising model and thereafter turn to the perturbed toric code.

Ising model. In our numerical experiments, we found that all correlation orders are required to obtain a good ground state representation for a classical Ising model. At first glance, this may seem counterintuitive considering that the characterization of an antiferromagnetic state can be achieved by local measurements. However, with the framework discussed

in Sec. IV A, we can now prove that the highest-order Fourier coefficient $\bar{\psi}(L)$ is a relevant contribution to the ground state.

Note that we restrict this explanation to positive wave function coefficients $\psi(\sigma) \geq 0$ with an even number of spins. Depending on the parity of the number of lattice sites and the sign structure of the ground state, the highest contributing correlation order can shift to $L - 1$ such that $\bar{\psi}(L - 1) \geq \bar{\psi}(S)$.

Starting from Eq. (13), we see that the number of down spins in the subset S of the contributing basis states determines the magnitude of the Fourier coefficient. For an antiferromagnetic (AFM) state we obtain

$$\begin{aligned}\bar{\psi}_{\text{AFM}}(S) &= \frac{1}{2^L} \sum_{\vec{\sigma} \in \{-1,1\}^L} \psi_{\text{AFM}}(\vec{\sigma}) \prod_{i \in S} \sigma_i \\ &= \frac{1}{2^L} \left[\psi_A \prod_{i \in S} \sigma_i^A + \psi_B \prod_{i \in S} \sigma_i^B \right].\end{aligned}\quad (26)$$

Here, we used A and B for the two possible AFM states. In the considered 4×4 Ising model, the AFM ground states contain an even number of down spins, implying that the two corresponding highest-order monomials are $\mathcal{X}_L(\text{even } \downarrow) = 1$. Plugging this into Eq. (26), we obtain

$$\bar{\psi}(L) = \frac{1}{2^L} (\psi_A + \psi_B) \geq \bar{\psi}(S), \quad (27)$$

for any $S \subseteq [L]$, which makes the highest correlation order (for positive wave function coefficients) a significant contribution to the wave function. Therefore, the correlator transformer that was used for the Ising model in Sec. III with C_8 and C_{11} cannot represent the full ground state, although it has access to approximately 60% and 96% of all correlations [compare with Eq. (16)], respectively.

Note that the general form of the Fourier coefficients for an AFM state can be derived using Eq. (26),

$$\bar{\psi}(S) = \begin{cases} \pm \frac{1}{2^L} [\psi_A + \psi_B] & \text{for } |S| = \text{even}, \\ \pm \frac{1}{2^L} [\psi_A - \psi_B] & \text{otherwise,} \end{cases} \quad (28)$$

which shows that higher-order correlations are as relevant as lower correlation orders of the same parity. The large number of nonzero coefficients $\bar{\psi}(S)$ arises because the network must set most wave function coefficients $\psi_{\text{AFM}}(\vec{\sigma})$ to zero.

Remark 4. This analysis can be readily extended to other quantum states constructed from basis states $|\vec{\sigma}\rangle$ with an even number of down spins. The toric code ground state in the full Hilbert space is an example of a more complex quantum state that shows the same behavior $\bar{\psi}_{\text{TC}}(L) \geq \bar{\psi}_{\text{TC}}(S)$.

Toric code model. The numerical simulation of the perturbed toric code model with enforced Gauss's law [compare with Fig. 1(b)] revealed that only a limited number of correlation orders played a significant role in obtaining the best possible ground-state energy.

The Hilbert space of the toric code model restricted to the physical sector has dimension $2^{L_x \times L_y + 1}$ [46,47], which means that the system of equations is underdetermined (compare with Sec. IV C). Following the spanning tree formalism (see Sec. S3 of the Supplemental Material [45]; see also Refs. [47,51,52]), an orthogonal basis can be constructed from $L_x \times L_y + 1$ linearly independent spins. These spins uniquely determine the remaining $L_x \times L_y - 1$ spins, fully specifying

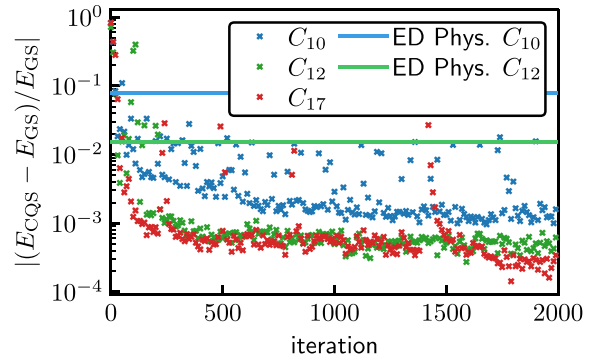


FIG. 3. Convergence of the CQS (access to all sites) for the perturbed toric code (4×4) model with enforced Gauss's law ($h_x = 1$ and star and plaquette term set to zero). We compare the CQS to exact diagonalization results (horizontal lines) in the correlation basis with correlations up to order 10 (blue) and 12 (green).

the state within the physical subspace. Therefore, using a decision tree (as described in Sec. IV B) that includes precisely these $L_x \times L_y + 1$ spins is sufficient for an exact representation of the physical sector. However, this construction requires capturing correlations up to order $L_x \times L_y + 1$. Despite this, the results shown in Fig. 1(b) indicate that lower-order correlations already suffice for an accurate approximation of the ground state $\psi_{\text{res}}^{h_x=0.2}$.

One might argue that higher-order correlations are not essential in this regime, especially since $\psi_{\text{res}}^{h_x=0.2}$ remains close to the uniform $h_x = 0$ ground state $\psi_{\text{res}}^{h_x=0.0} = C_0$. To investigate this matter, we now consider the large h_x limit. In the previous analysis, it was shown that representing the ferromagnetic ground state $|++\dots+\rangle$ in the σ^x basis (large h_x limit) requires all $(L_x \times L_y + 1)$ correlation orders in the reduced Hilbert space.

Figure 3 shows the convergence of the CQS (with access to all sites) for various maximal correlation orders in the large h_x limit. Furthermore, we compare the CQS energy to exact results obtained by diagonalizing the Hamiltonian in the orthogonal correlator basis (up to some correlation order) of the $L_x \times L_y + 1$ linearly independent spins. Similarly to the behavior observed in Fig. 1(b), we notice that the results, beyond a certain correlation order, do not improve further. All tested CQS obtain a minimal relative error of $\epsilon \approx \mathcal{O}(10^{-3})$. Furthermore, considering the same maximal correlation order, the CQS surpasses the exact results by several orders of magnitude. Here, the sole advantage of the CQS is the additional information about the remaining $L_x \times L_y - 1$ spins. Consequently, we can conclude that the information about the predetermined $L_x \times L_y - 1$ spins has to contain higher-order correlations of the $L_x \times L_y + 1$ linearly independent spins so that accurate results can be obtained even when not considering all $L_x \times L_y + 1$ correlation orders.

This is equivalent to saying that the spanning tree and therefore the decision tree in this toric code model is not unique. While we chose a specific decision tree for the exact diagonalization with a maximum correlation order of $L_x \times L_y + 1$, we allowed the CQS to choose the shortest and most efficient path.

Remark 5. We emphasize that the exact diagonalization (compare with Fig. 3) only uses states from the physical sector and is, therefore, still feasible for 32 spins.

VI. GENERAL CONSEQUENCES OF THE INTERNAL NQS BASIS

The Boolean function framework, which identifies the correlator basis as the internal NQS basis (Sec. IV), leads to several conclusions about the correlation properties required to represent a given quantum state. First, we establish the equivalence of the correlation transformation and the spin basis rotation, showing how the choice of computational basis influences the correlation structure of the target state (Sec. VIA). Second, we derive entanglement bounds for CQS, demonstrating that a fixed expansion order can at most capture logarithmic entanglement scaling, whereas states with area- or volume-law entanglement require a CQS with correspondingly quasi-area- or volume-scaling expansion order (Sec. VIB). Finally, we use these insights to draw conclusions for general NQS by comparing their Taylor expansion with the polynomial description of a quantum state given by the Boolean function framework. This comparison reveals a bias in the exploration of Hilbert space and clarifies how the required correlation order depends on architectural details of the network, as discussed in Sec. VIC.

Note that in Sec. S4 of the Supplemental Material [45], we address the relation between $\psi(\sigma)$, $p(\sigma) = |\psi(\sigma)|^2$, and $\log \psi(\sigma)$. Their Fourier coefficients are related through convolutions of $\tilde{\psi}(S)$.

A. Choosing the computational basis

In the research of neural quantum states, the problem of how to choose the most suitable reference basis for a given quantum state remains unsolved. However, with the framework introduced above, we can intuitively draw conclusions about the desired NQS structure, which provides insights about potential difficulties for the quantum state representation.

The internal NQS structure can be analyzed by rotating the variational wave function in the correlator basis,

$$\psi(\vec{\sigma}) = \sum_{S \subseteq [L]} \tilde{\psi}(S) \mathcal{X}_S(\vec{\sigma}). \quad (29)$$

Importantly, the coefficients $\tilde{\psi}(S)$ depend solely on the variational parameters and not on the input configuration and are therefore a direct consequence of the network's parameters and structure (see Sec. IVA). To determine whether the network must learn a relatively simple or complex structure of coefficients $\tilde{\psi}(S)$, it is necessary to identify known patterns or similarities.

To do so, we revisit Eq. (17). In this equation we show the rotation of the Hamiltonian from the spin basis $|\sigma\rangle$ to the correlation basis $|\mathcal{X}_S\rangle$,

$$\hat{H} = \underbrace{\frac{1}{2^L} \sum_{\vec{\sigma}, S} \langle \mathcal{X}_S | \vec{\sigma} \rangle}_{\mathcal{X}} \underbrace{\hat{H} \frac{1}{2^L} \sum_{\vec{\sigma}', S'} \langle \vec{\sigma}' | \mathcal{X}'_S \rangle}_{\mathcal{X}^\dagger} \langle \mathcal{X}'_S |. \quad (30)$$

Although this rotation matrix $\mathcal{X}$ with matrix elements $\mathcal{X}_{S, \vec{\sigma}} = \langle \mathcal{X}_S | \vec{\sigma} \rangle$ can seem complicated, it has the simple and well-known form of Hadamard matrices for spin-1/2 systems [48],

$$\mathcal{X} = H_1 \otimes H_2 \otimes \dots \otimes H_L$$

$$\text{with } H_i = \begin{pmatrix} \mathcal{X}_{S_i=\{\emptyset\}, \uparrow} = 1 & \mathcal{X}_{S_i=\{\emptyset\}, \downarrow} = 1 \\ \mathcal{X}_{S_i=\{i\}, \uparrow} = 1 & \mathcal{X}_{S_i=\{i\}, \downarrow} = -1 \end{pmatrix}. \quad (31)$$

The same basis transformation is used to rotate from $|\vec{\sigma} = \vec{\sigma}_{z/x}\rangle$ to $|\vec{\sigma}_{x/z}\rangle$,

$$U = \sum_{\vec{\sigma}_x, \vec{\sigma}_z} u_{\vec{\sigma}_x, \vec{\sigma}_z} |\vec{\sigma}_x\rangle \langle \vec{\sigma}_z| = H_1^\sigma \otimes H_2^\sigma \otimes \dots \otimes H_L^\sigma$$

$$\text{with } H_i^\sigma = \begin{pmatrix} u_{+, \uparrow} = 1 & u_{+, \downarrow} = 1 \\ u_{-, \uparrow} = 1 & u_{-, \downarrow} = -1 \end{pmatrix}. \quad (32)$$

Consequently, the matrix elements $h_{n,m}^{\mathcal{X}, z/x}$ of a Hamiltonian in the correlator basis $|\mathcal{X}_S\rangle$ are the same as the elements $h_{n,m}^{\sigma_{x/z}}$ in the $|\vec{\sigma}_{x/z}\rangle$ basis,

$$\hat{H} = \sum_{n,m=1}^L h_{n,m}^{\mathcal{X}, z/x} |n_{\mathcal{X}}\rangle \langle m_{\mathcal{X}}|$$

$$= \sum_{n,m=1}^L h_{n,m}^{\sigma_{x/z}} |n_{\sigma_{x/z}}\rangle \langle m_{\sigma_{x/z}}|, \quad (33)$$

with $h_{n,m}^{\mathcal{X}, z/x} = h_{n,m}^{\sigma_{x/z}}$. Here, $n, m \leq 2^L$ are numbers that correspond to the n th basis vector in the given basis. For example, in the $|\sigma_z\rangle$ basis, the number $n = 1$ ($n = 2^L$) corresponds to the basis state,

$$|n_{\sigma_z} = 1\rangle \equiv |\uparrow \uparrow \dots \uparrow\rangle = \bigotimes_{i=1}^L \begin{pmatrix} 1 \times |\uparrow_i\rangle \\ 0 \times |\downarrow_i\rangle \end{pmatrix} \quad \text{and}$$

$$|n_{\sigma_z} = 2^L\rangle \equiv |\downarrow \downarrow \dots \downarrow\rangle = \bigotimes_{i=1}^L \begin{pmatrix} 0 \times |\uparrow_i\rangle \\ 1 \times |\downarrow_i\rangle \end{pmatrix} \quad (34)$$

and in the correlator basis $|\mathcal{X}_S\rangle$ we define it as

$$|n_{\mathcal{X}} = 1\rangle \equiv |\mathcal{X}_{\emptyset}\rangle = \bigotimes_{i=1}^L \begin{pmatrix} 1 \times |\mathcal{X}_{S_i=\{\emptyset\}}\rangle \\ 0 \times |\mathcal{X}_{S_i=\{i\}}\rangle \end{pmatrix} \quad \text{and}$$

$$|n_{\mathcal{X}} = 2^L\rangle \equiv |\mathcal{X}_L\rangle = \bigotimes_{i=1}^L \begin{pmatrix} 0 \times |\mathcal{X}_{S_i=\{\emptyset\}}\rangle \\ 1 \times |\mathcal{X}_{S_i=\{i\}}\rangle \end{pmatrix}. \quad (35)$$

Here, we want to emphasize that the sites included in S correspond to minus spins in the σ^x basis,

$$|\mathcal{X}_S^z\rangle = \frac{1}{2^L} \sum_{\vec{\sigma}^z} \prod_{i \in S} \sigma_i |\vec{\sigma}^z\rangle$$

$$= \frac{1}{2^L} \bigotimes_{i \notin S} \underbrace{\sum_{\sigma_i^z} |\sigma_i^z\rangle}_{\propto |+\rangle} \bigotimes_{j \in S} \underbrace{\left(\sum_{\sigma_j^z} \sigma_j |\sigma_j^z\rangle \right)}_{\propto |-\rangle}. \quad (36)$$

Note that we purposely added the index z/x to clarify that the matrix element $h_{n,m}^{\mathcal{X}, z/x}$ contains products of single spins in the z/x basis and therefore dependent on the computational basis

(σ_x or σ_z), while the corresponding elements $h_{n,m}^{\sigma_{x/z}}$ are the usual matrix elements in the respective x/z basis.

The consequence of having the same matrix elements $h_{n,m}^{\mathcal{X},z/x} = h_{n,m}^{\sigma_{x/z}}$ is that the coefficients for the ground-state wave function are the same in the correlator basis $\psi^{\mathcal{X},z/x}(n)$ and the corresponding spin basis $\psi^{\sigma_{x/z}}(n)$,

$$\psi^{\mathcal{X},z/x}(n) = \psi^{\sigma_{x/z}}(n), \quad (37)$$

with

$$|\Psi\rangle = \sum_{n=1}^L \psi^{\mathcal{X},z/x}(n) |n_{\mathcal{X}}\rangle = \sum_{n=1}^L \psi^{\sigma_{x/z}}(n) |n_{\sigma_{x/z}}\rangle, \quad (38)$$

which allows us to predict the difficulty for NQS simulations. Note that we use $\psi^{\mathcal{X},z/x}(n)$ instead of $\bar{\psi}(S)$ at the n^{th} basis vector of $|\mathcal{X}_S\rangle$ for better comparability.

What does this mean and how can we profit from this? In almost all cases, we intuitively pick the basis in which the wave function is peaked the most. For example, the natural choice for an $|\uparrow\uparrow\ldots\uparrow\rangle$ state would be the $|\sigma_z\rangle$ basis, since this provides an easy understanding of what we want to represent. However, as we showed in Sec. V, this requires surprisingly high computational effort (all correlation orders require the same coefficient), since it is difficult to align the weights such that most of the wave function coefficients are tuned to zero.

The analogy of correlations in the $|\vec{\sigma}_z\rangle$ basis to the wave function in the $|\sigma_x\rangle$ basis would have shown the same result. The product state $|\uparrow\uparrow\ldots\uparrow\rangle$ in the $|\vec{\sigma}_x\rangle$ basis requires all wave function coefficients to be the same: $\psi^{\sigma_x}(n) = \psi^{\mathcal{X},z}(n) = C$. While the wave function in the $|\vec{\sigma}_z\rangle$ basis is peaked, the wave function in the correlator basis $|\mathcal{X}_S\rangle$ is widely spread.

If one uses the $|\vec{\sigma}_x\rangle$ basis instead, which is counterintuitive at first glance, the wave function in the correlator basis $|\mathcal{X}_S\rangle$ matches the wave function in the $|\vec{\sigma}_z\rangle$ basis and is therefore strongly peaked: $\psi^{\sigma_x}(n) = \psi^{\mathcal{X},x}(n) = \delta_{1,n}$. In a neural network, this requires just one single bias, with all other weights set to zero.

In summary, even a rough understanding of the ground state wave function in the $|\vec{\sigma}_z\rangle$ and $|\vec{\sigma}_x\rangle$ bases can provide a useful indication of which basis may lead to a wave function that is more or less peaked. Counterintuitively, choosing the reference basis where the true ground state wave function is less peaked will yield a more peaked wave function in the corresponding correlator basis. This can prove advantageous, especially when predominantly low correlation orders need to be captured by the NQS (compare with Sec. VIC). However, one must keep in mind that the sign structure also changes with the basis, which can complicate the optimization [53,54].

Note that by applying unitary transformations, often in the form of Pauli operators $U_{z/x} = \prod_{i \in A} \sigma_i^{z/x}$, we can change the sign structure $\psi_{\text{NQS}}(\sigma) U_z |\sigma_z\rangle \rightarrow (-1)^{N_A^\sigma} \psi_{\text{NQS}}(\sigma) |\sigma\rangle$ (N_A^σ is the number of down spins on the sublattice A) and the order of the basis states $\psi_{\text{NQS}}(\sigma) U_x |\sigma_z\rangle \rightarrow \psi_{\text{NQS}}(\sigma) |-\sigma_z^A\rangle \otimes |\sigma_z^A\rangle$. In many cases, this can help to construct a basis that reduces the requirements of the NQS wave function.

Remark 6. Nonlocal basis transformations, such as a transformation to the Gaussian basis [55] or the momentum basis

[56,57] aim to trade a decrease in the complexity of representing a quantum state for increased computational costs.

B. Entanglement bound from the expansion order

A typical measure of the complexity of a quantum state is the entanglement entropy $S(\rho_A)$ of a subregion A. Based on the entanglement entropy, one usually defines three complexity scalings for quantum states: States with low entanglement entropy scaling (constant or logarithmic), states whose entanglement entropy scales with the area of the subsystem, and states with volume-law scaling. In the following, we provide an upper bound on the maximal entanglement entropy an NQS can capture, based on its maximal correlation order n . The entanglement entropy $S(\rho_A)$ of a subregion A is defined as

$$S(\rho_A) = -\text{Tr}(\rho_A \ln(\rho_A)), \quad (39)$$

with $\rho_A = \text{Tr}_B |\psi\rangle\langle\psi|$. Rewriting ρ_A in the correlator basis provides intuitive insight into the complexity of representing a given quantum state

$$\begin{aligned} (\rho_A)_{a,a'} &= \sum_B \psi(a, B) \psi^*(a', B) \\ &= \sum_B \bar{\psi}(S_a, S_B) \bar{\psi}^*(S_{a'}, S_B) \\ &= (MM^\dagger)_{a,a'}. \end{aligned} \quad (40)$$

The rank of matrix M determines the maximal entanglement entropy captured by a network with correlation order n ,

$$\begin{aligned} S(n, \rho_A) &\leq \ln(\text{rank}(M)) \\ &\stackrel{X=\min(A,B)}{\leq} S_{\max}(n, X) = \ln \left(\sum_{k=0}^n \binom{X}{k} \right) \\ &\stackrel{n/X \leq 1/2}{\leq} S_{\text{bound}}(n, X) = XH\left(\frac{n}{X}\right), \end{aligned} \quad (41)$$

where $H(\frac{n}{X})$ is the binary entropy function [58,59], valid for $n/X \leq 1/2$,

$$S_{\text{bound}}(n, X) = -n \ln \left(\frac{n}{X} \right) - (X - n) \ln \left(1 - \frac{n}{X} \right). \quad (42)$$

This upper bound reveals that the required minimal correlation order depends sensitively on the entanglement scaling of the desired quantum state. In the following, we analyze three distinct regimes (see Fig. 4):

(1) For a fixed expansion order n , the maximal entanglement entropy scales logarithmically $S \propto \ln X$.

(2) To represent states with area-law entanglement $S \propto \sqrt{X}$, the correlation order must scale at least as $n \propto \sqrt{X}/\ln(X)$.

(3) For volume-law entangled states $S \propto X$, the correlation order must scale at least as $n \propto X$.

The derivation of these scalings is shown in Sec. S5 of the Supplemental Material [45]. Consequently, representing generic many-body states requires a correlation order that increases with system size, with only weakly entangled states remaining accessible at fixed expansion order.

Remark 7. A large correlation order does not imply entanglement but can be understood as an entanglement capacity.

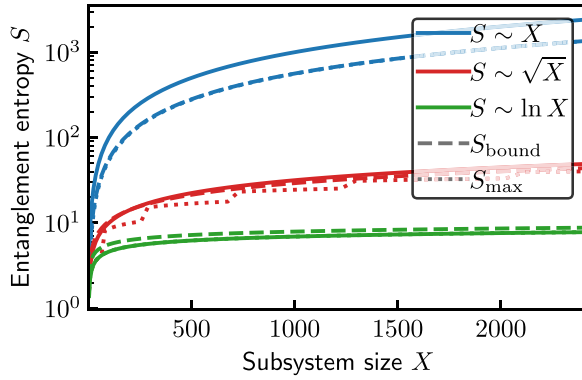


FIG. 4. Entanglement scaling for different neural networks with maximal correlation order scaling with subsystem size $n = \frac{1}{4}X$ (blue), with the area $n = \sqrt{X}/\ln(X)$ (red), and kept constant $n = 1$ (green).

However, the width and complexity of the respective spectrum of coefficients is related to entanglement.

C. Relation to Taylor expansion

In neural quantum states, a machine learning algorithm is trained such that it can represent a quantum state. In the previous sections, we addressed the topic of representability based on the correlations that the CQS has access to and compared it to the exact polynomial description of the desired quantum state given by its Fourier expansion. In the following, we generalize this to common neural network architectures by using the complementary polynomial Taylor expansion of a neural network and apply the information we derived in the previous sections. Here, we distinguish between globally analytic NQS and NQS with only locally analytic functions before discussing general activation functions and the role of the bias.

Connection between Fourier and Taylor expansion: Globally analytic NQS. The Taylor expansion of a function $\psi(\vec{\sigma})$ around point $\vec{p}$ is defined by

$$\psi(\vec{\sigma}) = \sum_{n=0}^{\infty} \frac{\psi^{(n)}(\vec{p})}{n!} (\vec{\sigma} - \vec{p})^n. \quad (43)$$

Assuming that we work with globally analytic functions, the full NQS can be expanded around $\vec{p} = \vec{0}$ up to some significant order k ,

$$\begin{aligned} \psi(\vec{\sigma}) \approx \psi(\vec{0}) + \sum_i \frac{\partial \psi(\vec{p})}{\partial p_i} \Big|_{\vec{p}=\vec{0}} \sigma_i + \dots \\ + \frac{1}{k!} \sum_{i_1, \dots, i_k} \frac{\partial^k \psi(\vec{p})}{\partial p_{i_1} \dots \partial p_{i_k}} \Big|_{\vec{p}=\vec{0}} \sigma_{i_1} \dots \sigma_{i_k}, \end{aligned} \quad (44)$$

where the relevance of the remainder $R_k(\vec{\sigma}) = \sum_{n=k+1}^{\infty} \frac{\psi^{(n)}(\vec{p})}{n!} |_{\vec{p}=\vec{0}} \vec{\sigma}^n$ is suppressed by the factorial. As both the Taylor and Fourier expansions use the same polynomial structure, we can argue that the highest relevant Taylor expansion order defines the highest accessible correlation order.

In the previous sections, we showed that high correlation orders remain relevant: (a) peaked wave functions require broad support of the correlation weights, such that even the

highest correlation order $\bar{\psi}(L)$ can be a significant contribution: $\bar{\psi}(L) \geq \bar{\psi}(S)$ (Sec. V) and (b) depending on the entanglement scaling of the desired state, there exists a minimal relevant expansion order that depends, for most quantum states, on the system size (Sec. VIB).

Consequently, NQS-like quantum state representation requires high, nonvanishing Taylor expansion orders (often of the order of the system size) to be able to represent a quantum state faithfully. In practice, this mismatch between factorial suppression and requiring high expansion orders necessitates:

(1) the network to increase the magnitude of the variational parameters to amplify the contributions from higher-order correlations,

(2) the use of deeper neural networks that are more efficient at generating high-order correlations with small weights.

We want to highlight that this conclusion immediately leads to a significant distinction between increasing the width and increasing the depth of a neural network: While wider neural networks still have to rely on increasing the magnitude of the weights to generate higher correlation orders, a deeper neural network achieves this naturally through the consecutive use of activation functions.

Furthermore, we emphasize that requiring large network weights potentially comes with a drawback: stochastic reconfiguration, which is based on linearized imaginary-time evolution, is designed for small steps in both the imaginary-time evolution and the parameter manifold. Therefore, large weights can slow down the accurate optimization.

As an example, consider representing an $|\uparrow\uparrow\dots\uparrow\rangle$ state ($|\sigma_z\rangle$ basis) with a simple single-layer feed-forward neural network and a cosh activation function. An exact representation of this simple product state would require infinitely large network parameters W_i ,

$$\psi(\vec{\sigma}) = \cosh(W_1\sigma_1 + W_2\sigma_2 + \dots + W_L\sigma_L). \quad (45)$$

A more intuitive way to think about this is that the network needs to maximize the distance between $\psi(\uparrow\uparrow\dots\uparrow)$ and all other wave function coefficients, which can only be achieved in the large W_i limit.

In combination with Sec. VIA, the Taylor expansion perspective also suggests an intuitive picture of how common NQS initializations explore the Hilbert space. For architectures initialized with small weights, the initial state is often close to the uniform superposition $|++\dots+\rangle$ state, as their Taylor expansion is dominated by low-order correlations. Small-step optimization techniques, like stochastic reconfiguration, then gradually deform the wave function in Hilbert space. During the optimization, higher-order terms in the Taylor expansion become increasingly relevant. Through the Hadamard relation, this process has a second equivalent interpretation in the σ^x basis: as shown in Eq. (36) the correlation order $|S|$ corresponds to a set of σ^x basis states with $|S|$ minus spins. Therefore, the hierarchy of the correlation order given by the Taylor expansion implies a bias whereby NQS naturally explore basis states with few minus spins first and potentially increase the number of minus spins during training.

Locally analytic NQS. Next, we comment on NQS with activation functions that are only locally analytic. These functions have a finite convergence radius and, thus, do not

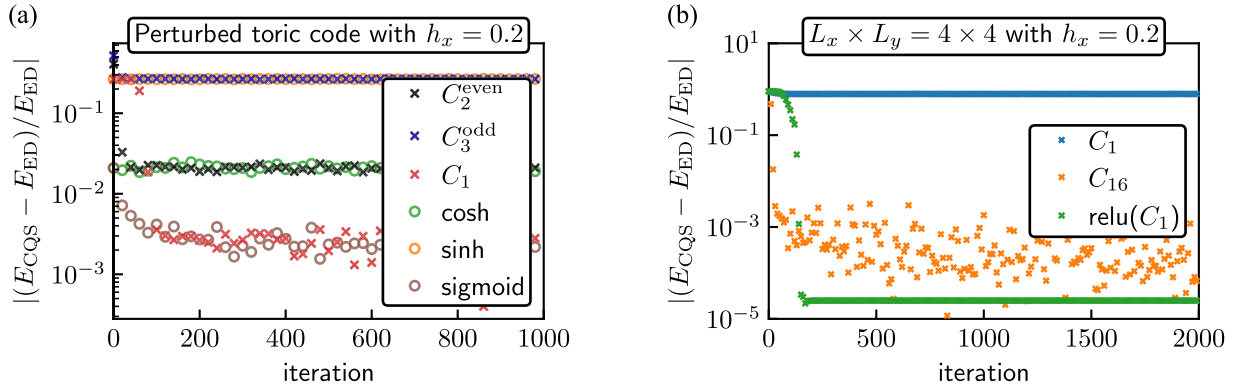


FIG. 5. Convergence of NQS models applied to (a) the perturbed toric code (Gauss's law enforced in the sampling) and (b) the transverse-field Ising model. In (a) we show a comparison between several CQS and a single-layer FFNN with different activation functions. The CQS resemble low-order expansions of the corresponding activation functions ($\cosh \approx C_2^{\text{even}}$, $\sinh \approx C_3^{\text{odd}}$, $\text{sigmoid} \approx C_1$). In (b), we compare the CQS (C_1 and C_{16}) to a single-layer FFNN (equivalent to C_1) with rectified linear unit activation function.

necessarily have the same polynomial behavior for all basis states. However, this can simplify the optimization. Locally analytic functions have at least two regimes with distinct behavior. This allows us to separate or group basis states by shifting them in different regimes. A prominent example of locally analytic functions is the rectified linear unit (ReLU),

$$\text{ReLU}(x) = \begin{cases} x & \text{for } x > 0, \\ 0 & \text{for } x \leq 0. \end{cases} \quad (46)$$

As can be readily seen, the ReLU will not increase the order of correlations in the system; nonetheless, it is a very efficient activation function. Instead of building up correlations, it has the ability to group states by tuning the bias, which allows us to set probabilities of multiple basis states to zero. The system can then use all the other network parameters to represent the wave function coefficients of the remaining states. An example for a transverse-field Ising model ($\hat{H}_{\text{TFIM}} = \hat{H}_{\text{Ising}} - h_x \sum_i \hat{\sigma}_i^x$) on an $L \times L = 4 \times 4$ lattice is shown in Fig. 5(b). Here, the feed-forward neural network achieves good results by tuning the bias such that the ReLU activation sets the wave function coefficients of most basis states to zero. Specifically, the wave function has nonzero coefficients for only 17 basis states [out of $\mathcal{O}(10^5)$]. This is not a coincidence, since the NQS has access to precisely 17 correlations [one bias ($\mathcal{X}_0 = 0$ th order) and $L \times L = 16$ first-order correlations, compare with Eq. (16)]. In the context of a system of equations, introduced in Sec. IV C, this implies that, when learning the perturbed AFM state, the network can faithfully solve the set of equations for at most 17 basis states. Similarly, other locally analytic functions like tanh or sigmoid have the ability to group basis states, such that the optimization can potentially be simplified.

Even and odd orders: The role of the bias. As shown in Sec. IV A, any wave function of a finite-size spin system can be described as a combination of even and odd orders of correlations: $\psi(\sigma) = Y_{\text{even}}(\sigma) + Y_{\text{odd}}(\sigma)$. While even orders are spin symmetric ($Y_{\text{even}}(\sigma) = Y_{\text{even}}(-\sigma)$), odd orders are antisymmetric ($Y_{\text{odd}}(\sigma) = -Y_{\text{odd}}(-\sigma)$). Therefore, the quantum states with an even (odd) parity (i.e., a ground state of the Ising model), can be represented by solely using even (odd) correlations. On the other hand, systems that are not

spin symmetric $|\psi(\sigma)| \neq |\psi(-\sigma)|$ require both even and odd orders.

While this is not surprising, one has to keep in mind that many frequently used activation functions do have a fixed parity,

$$\begin{aligned} \sinh(x) &= -\sinh(-x), \\ \cosh(x) &= +\cosh(-x), \\ \tanh(x) &= -\tanh(-x), \\ \arctan(x) &= -\arctan(-x), \\ &\vdots \end{aligned}$$

$$\frac{1}{2} - \text{sigmoid}(x) = -\left(\frac{1}{2} - \text{sigmoid}(-x)\right). \quad (47)$$

Therefore, network architectures that maintain the parity of these activation functions cannot represent quantum states with a different parity. Note that the Taylor expansion of functions with even (odd) parity exclusively contains correlations of even (odd) order. To illustrate this, we apply different network architectures to the perturbed toric code model introduced in Sec. III, where Gauss's law is enforced in the sampling. In Fig. 5(a), we show the convergence of a single-layer feed-forward NQS with different activation functions (cosh, sinh, and sigmoid) and compare them to the correlator transformer with access to either purely even (C_2^{even}), odd (C_3^{odd}) or 0th + odd-order (C_1) correlations. Although all NQS do not fully capture the ground state, the behavior of the activation functions can be understood very well by comparing them to the CQS, as those resemble their low expansion orders ($\cosh \approx C_2^{\text{even}}$, $\sinh \approx C_3^{\text{odd}}$, $\text{sigmoid} \approx C_1$). Since increasing the correlation order of just one parity does not improve the result (the activation functions contain all orders of one parity), we can attribute the inability to converge to the absence of interplay between even and odd correlations.

Although this may appear to restrict the use of certain activation functions in most models, these limitations can be easily overcome by introducing a bias in the neural network layer. The bias enables the presence of odd (even) correlation orders in an otherwise even (odd) order of the Taylor expansion, making it not merely a convenient optimization

parameter but a crucial component for an accurate quantum state representation.

VII. COMMENT ON NETWORK ARCHITECTURES

In the following, we want to use the acquired knowledge about the working principle of NQS and their representational power to comment on potential advantages and disadvantages of commonly used network architectures.

Autoregressive NQS. In research on neural quantum states, two directions have emerged, namely, nonautoregressive methods based on Metropolis–Hastings Monte Carlo, where the machine learning model outputs the full wave function at once, and autoregressive NQS that evaluate the amplitude of the full wave function as a product of local, conditional probabilities,

$$p(\vec{\sigma}) = p(\sigma_1)p(\sigma_2|\sigma_{\leq 1})\dots p(\sigma_L|\sigma_{\leq L-1}). \quad (48)$$

Thus, the machine learning model gradually acquires information about the input and has to adjust its conditional probabilities accordingly. Each conditional probability $p(\sigma_k|\sigma_{\leq k-1})$ can be written in its orthogonal correlator basis according to Eq. (12),

$$p(\sigma_k|\sigma_{\leq k-1}) = \sum_{S \subseteq [k]} \bar{p}_k(S) \mathcal{X}_S(\sigma), \quad (49)$$

where $\bar{p}_k(S)$ is defined as

$$\bar{p}_k(S) = \frac{1}{2^k} \sum_{\vec{\sigma} \in \{-1,1\}^k} p(\sigma_k|\sigma_{\leq k-1}) \mathcal{X}_S(\sigma). \quad (50)$$

Therefore, the total probability $p(\vec{\sigma})$ is by design a construction that contains all correlation functions, which is required for a good approximation of many quantum states. Thus, the relevant criterion for autoregressive NQS is the difficulty of representing the conditional probabilities $p(\sigma_k|\sigma_{\leq k-1})$, which have a worst-case complexity of $\mathcal{O}(2^k)$ (in total $\mathcal{O}(2^{\sum_{k=1}^L k})$). However, many quantum states satisfy conditional independence,

$$p(\sigma_k|\sigma_{k-1 \geq}) = p(\sigma_k|\sigma_{k-1}, \sigma_{k-2}, \dots, \sigma_{k-l}), \quad (51)$$

with small l , which reduces the complexity of the full problem to approximately $\mathcal{O}(L \times 2^l)$, such that an efficient representation can be found [39]. The term short-range conditional probabilities can be translated into our formalism in Eq. (50), as it is equivalent to setting the correlation order coefficient $\bar{p}_k(S)$ to zero if S contains any σ_j with $j \leq k-l-1$.

Note that autoregressive models have an inherent asymmetry in the architecture owing to their path-like structure [see Eq. (48)] [13], as the first site only depends on itself, while the probability on the last site depends on all sites. This will also become obvious when writing Eq. (48) in terms of $\bar{p}_k(S)$ and $\mathcal{X}_S(\sigma)$. The more short-range a conditional probability is [compare with Eq. (51)], the more symmetric the equation becomes. Furthermore, the path-like structure predefines the required decision tree in truncated Hilbert spaces, allowing for only one possible polynomial solution (compare with Sec. IV B).

Nonautoregressive NQS. Nonautoregressive NQS model the full wave function directly, without factorizing it into a

sequence of conditional probabilities. Therefore, this requires network architectures to be capable of generating high correlation orders. This is reflected in the development of the current research direction. Early approaches often relied on Restricted Boltzmann machines, because of their analytical accessibility [16,21]. However, the lack of complexity results in an inefficient way to generate large coefficients for high correlation orders. This limitation has unwittingly led to the adoption of architectures that can naturally generate high correlation orders with large coefficients.

Transformer neural quantum states are a prominent example [8,23,60–62]: the multiplicative terms in the attention mechanism naturally produce large coefficients for higher correlation orders,

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V, \quad (52)$$

where Q, K, V are all linear in the input, which itself can be a sum over multiple correlation orders.

Similarly, determinant-based architectures have become the predominant architecture for fermionic [7,63,64] systems. To illustrate that determinant wave functions generate large coefficients for high correlation orders efficiently, we discuss neural backflow-based determinant constructions for m fermions on L sites [2,38],

$$\Psi(\vec{n}) = \det \begin{vmatrix} \phi_{k_1 1}^{\text{bf}}(n_1, \dots, n_L) & \dots & \phi_{k_1 m}^{\text{bf}}(n_1, \dots, n_L) \\ \vdots & \ddots & \vdots \\ \phi_{k_m 1}^{\text{bf}}(n_1, \dots, n_n) & \dots & \phi_{k_m m}^{\text{bf}}(n_1, \dots, n_L) \end{vmatrix}, \quad (53)$$

where $\phi_{k_j j}^{\text{bf}}(n_1, \dots, n_n)$ are backflow-enhanced elements. k_i denotes the index of the i th occupied site and j represents the j th fermion. In a backflow-based determinant, one introduces the potentially deep neural network $f_{k_i j}^\theta(n_1, \dots, n_n)$ that uses Fock space configurations as inputs at an orbital level,

$$\phi_{k_j j}^{\text{bf}}(n_1, \dots, n_n) = \phi_{k_j j} + f_{k_j j}^\theta(n_1, \dots, n_n). \quad (54)$$

Therefore, $\phi_{k_j j}^{\text{bf}}(n_1, \dots, n_n)$ itself can be a function with higher-order correlations. Importantly, however, the determinant structure acts as a further nonlinear activation function: by involving products of system-size number of orbitals, it automatically generates large coefficients for higher-order correlations,

$$\Psi(\vec{n}) = \sum_{i_1, \dots, i_m} \epsilon_{i_1, \dots, i_m} \phi_{k_1 i_1}^{\text{bf}} \dots \phi_{k_m i_m}^{\text{bf}}. \quad (55)$$

A natural consequence can be that the specific choice of neural network architecture is less critical [65].

Since both network architectures are designed to naturally generate large coefficients for high correlation orders, they appear—from the correlator’s perspective—to be highly suitable for NQS simulations. This has been confirmed in numerous numerical benchmarks, where they outperformed other architectures [14,34,66–68].

VIII. CONCLUSIONS

In this work, we explored the importance of correlations in neural quantum states (NQS), in particular, how they determine a network’s ability to faithfully represent quantum states.

By using correlation-based neural networks, we were able to numerically demonstrate that even simple quantum states, such as product states, can require all possible correlation orders for an accurate representation, whereas other, more complex quantum states can be represented using a simple correlation structure in a restricted Hilbert space.

Using tools from Boolean function theory, we provided a framework to better understand the requirements for NQS-like quantum state representation. This framework reveals that the internal NQS structure effectively represents a quantum state in the correlator basis and, consequently, requires a basis transformation from the computational basis, which can lead to nontrivial correlation order structures.

Furthermore, we observed a significant relaxation of correlation requirements in the case of truncated Hilbert spaces. Restricted Hilbert spaces result in linear dependencies, which allow CQS to store information about high correlation orders in significantly lower correlation orders. Additionally, we showed connections between spin basis rotations and the correlator basis, which can provide valuable insights into the desired NQS structure and potential problems for quantum state representation.

By using Boolean function theory, we derived an entanglement bound for a given maximal correlation order, which reveals that a fixed expansion order produces logarithmic entanglement scaling, while quantum states with area-law or volume-law entanglement require a quasi-area- or volume-scaling expansion order, respectively. This has direct consequences on the neural network, as the highest relevant Taylor expansion order defines the highest accessible correlation order. To achieve these large expansion orders, neural networks need to have large variational parameters unless they are deep enough for the consecutive use of activation functions to generate high orders naturally.

Finally, we investigated how activation functions and neural network architectures influence the expressiveness of the neural quantum state ansatz. These design choices affect the ability to group certain basis states or to generate large coefficients for high-order terms, which strongly influences the correlations that can be captured and, consequently, the network's representational power.

Overall, our results offer new insights into the internal structure of NQS and allow a deeper understanding of their requirements. By identifying the role of correlations in neural quantum states and clarifying how these correlations emerge from the network architecture, we provide a foundation for systematic tailoring and optimization of NQS for a desired quantum system.

Note added. Recently, we became aware of a related work by Schurov *et al.* [54], studying the complexity of sign structures using Boolean Fourier analysis.

ACKNOWLEDGMENTS

We thank F. Pauw, T. Vovk, C. Roth, and D. Aliverti-Piuri for fruitful discussions. This research was supported by LMUexcellent, funded by the Federal Ministry of Education and Research (BMBF) and the Free State of Bavaria under the Excellence Strategy of the Federal Government and the Länder and by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy (EXC-2111 – 390814868).

DATA AVAILABILITY

The code and data used to produce the results presented in this paper are available in the following GitHub repository [69].

-
- [1] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Quantum Monte Carlo simulations of solids, *Rev. Mod. Phys.* **73**, 33 (2001).
 - [2] F. Becca and S. Sorella, *Quantum Monte Carlo Approaches for Correlated Systems* (Cambridge University Press, Cambridge, 2017).
 - [3] G. Pan and Z. Y. Meng, The sign problem in quantum Monte Carlo simulations, in *Encyclopedia of Condensed Matter Physics* (Elsevier, Amsterdam, 2024), pp. 879–893.
 - [4] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, *Ann. Phys.* **326**, 96 (2011).
 - [5] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* **355**, 602 (2017).
 - [6] A. Chen and M. Heyl, Empowering deep neural quantum states through efficient optimization, *Nat. Phys.* **20**, 1476 (2024).
 - [7] A. Chen, Z.-Q. Wan, A. Sengupta, A. Georges, and C. Roth, Neural network-augmented Pfaffian wave-functions for scalable simulations of interacting fermions, *Proc. Natl. Acad. Sci. USA* **123**, e2535288123 (2026).
 - [8] L. L. Viteritti, R. Rende, and F. Becca, Transformer variational wave functions for frustrated quantum spin systems, *Phys. Rev. Lett.* **130**, 236401 (2023).
 - [9] H. Lange, A. Van de Walle, A. Abedinnia, and A. Bohrdt, From architectures to applications: A review of neural quantum states, *Quantum Sci. Technol.* **9**, 040501 (2024).
 - [10] M. Schmitt and M. Heyl, Simulating dynamics of correlated matter with neural quantum states, *arXiv:2506.03124*.
 - [11] C. Giuliani, F. Vicentini, R. Rossi, and G. Carleo, Learning ground states of gapped quantum Hamiltonians with kernel methods, *Quantum* **7**, 1096 (2023).
 - [12] H. Lange, F. Döschl, J. Carrasquilla, and A. Bohrdt, Neural network approach to quasiparticle dispersions in doped antiferromagnets, *Commun. Phys.* **7**, 187 (2024).
 - [13] F. Döschl, F. A. Palm, H. Lange, F. Grusdt, and A. Bohrdt, Neural network quantum states for the interacting Hofstadter model with higher local occupations and long-range interactions, *Phys. Rev. B* **111**, 045408 (2025).
 - [14] Z. Denis, A. Sinibaldi, and G. Carleo, Comment on “Can neural quantum states learn volume-law ground states?” *Phys. Rev. Lett.* **134**, 079701 (2025).

- [15] D. S. Kufel, J. Kemp, D. Vu, S. M. Linsel, C. R. Laumann, and N. Y. Yao, Approximately symmetric neural networks for quantum spin liquids, *Phys. Rev. Lett.* **135**, 056702 (2025).
- [16] X. Gao and L.-M. Duan, Efficient representation of quantum many-body states with deep neural networks, *Nat. Commun.* **8**, 662 (2017).
- [17] O. Sharir, A. Shashua, and G. Carleo, Neural tensor contractions and the expressive power of deep neural quantum states, *Phys. Rev. B* **106**, 205136 (2022).
- [18] D.-L. Deng, X. Li, and S. Das Sarma, Quantum entanglement in neural network states, *Phys. Rev. X* **7**, 021021 (2017).
- [19] 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).
- [20] B. Barton, J. Carrasquilla, C. Roth, and A. Valenti, Connectivity determines the capability of sparse neural network quantum states, *Mach. Learn.: Sci. Technol.* **7**, 035025 (2026).
- [21] G. Carleo, Y. Nomura, and M. Imada, Constructing exact representations of quantum many-body systems with deep neural networks, *Nat. Commun.* **9**, 5322 (2018).
- [22] H. Kol-Namer and M. Goldstein, Neural network ground state from the neural tangent kernel perspective: The sign bias, [arXiv:2411.03980](https://arxiv.org/abs/2411.03980).
- [23] R. Rende, L. L. Viteritti, L. Bardone, F. Becca, and S. Goldt, A simple linear algebra identity to optimize large-scale neural network quantum states, *Commun. Phys.* **7**, 260 (2024).
- [24] S. Dash, L. Gravina, F. Vicentini, M. Ferrero, and A. Georges, Efficiency of neural quantum states in light of the quantum geometric tensor, *Commun. Phys.* **8**, 92 (2025).
- [25] M. S. Moss, A. Orfi, C. Roth, A. M. Sengupta, A. Georges, D. Sels, A. Dawid, and A. Valenti, Double descent: When do neural quantum states generalize? *Phys. Rev. E* **113**, 045303 (2026).
- [26] G. Passeti, D. Hofmann, P. Neiteimeier, L. Grunwald, M. A. Sentef, and D. M. Kennes, Can neural quantum states learn volume-law ground states? *Phys. Rev. Lett.* **131**, 036502 (2023).
- [27] G. Cybenko, Approximation by superpositions of a sigmoidal function, *Math. Control Signal Systems* **2**, 303 (1989).
- [28] K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, *Neural Networks* **2**, 359 (1989).
- [29] K. Hornik, M. Stinchcombe, and H. White, Universal approximation of an unknown mapping and its derivatives using multilayer feedforward networks, *Neural Networks* **3**, 551 (1990).
- [30] M. Hibat-Allah, R. G. Melko, and J. Carrasquilla, Investigating topological order using recurrent neural networks, *Phys. Rev. B* **108**, 075152 (2023).
- [31] M. Hibat-Allah, E. M. Inack, R. Wiersema, R. G. Melko, and J. Carrasquilla, Variational neural annealing, *Nat. Mach. Intell.* **3**, 952 (2021).
- [32] D. Pfau, S. Axelrod, H. Sutterud, I. von Glehn, and J. S. Spencer, Accurate computation of quantum excited states with neural networks, *Science* **385**, eadn0137 (2024).
- [33] Y. Gu, W. Li, H. Lin, B. Zhan, R. Li, Y. Huang, D. He, Y. Wu, T. Xiang, M. Qin, L. Wang, and D. Lv, Solving the Hubbard model with neural quantum states, *Nat. Commun.* **17**, 7838 (2026).
- [34] A. Chen, V. D. Naik, and M. Heyl, Convolutional transformer wave functions, *Phys. Rev. Res.* **8**, L022040 (2026).
- [35] Z. Denis and G. Carleo, Accurate neural quantum states for interacting lattice bosons, *Quantum* **9**, 1772 (2025).
- [36] D.-L. Deng, X. Li, and S. Das Sarma, Machine learning topological states, *Phys. Rev. B* **96**, 195145 (2017).
- [37] R. Kaubruegger, L. Pastori, and J. C. Budich, Chiral topological phases from artificial neural networks, *Phys. Rev. B* **97**, 195136 (2018).
- [38] S. Lu, X. Gao, and L.-M. Duan, Efficient representation of topologically ordered states with restricted Boltzmann machines, *Phys. Rev. B* **99**, 155136 (2019).
- [39] T.-H. Yang, M. Soleimanifar, T. Bergamaschi, and J. Preskill, When can classical neural networks represent quantum states? [arXiv:2410.23152](https://arxiv.org/abs/2410.23152).
- [40] A. Valenti, E. Greplova, N. H. Lindner, and S. D. Huber, Correlation-enhanced neural networks as interpretable variational quantum states, *Phys. Rev. Res.* **4**, L012010 (2022).
- [41] C. Miles, A. Bohrdt, R. Wu, C. Chiu, M. Xu, G. Ji, M. Greiner, K. Q. Weinberger, E. Demler, and E.-A. Kim, Correlator convolutional neural networks as an interpretable architecture for image-like quantum matter data, *Nat. Commun.* **12**, 3905 (2021).
- [42] K. Fischer, A. René, C. Keup, M. Layer, D. Dahmen, and M. Helias, Decomposing neural networks as mappings of correlation functions, *Phys. Rev. Res.* **4**, 043143 (2022).
- [43] A. Suresh, H. Schlömer, B. Hashemi, and A. Bohrdt, Interpretable correlator transformer for image-like quantum matter data, *Mach. Learn.: Sci. Technol.* **6**, 025006 (2025).
- [44] K. Cybiński, J. Enouen, A. Georges, and A. Dawid, Speak so a physicist can understand you! TetrisCNN for detecting phase transitions and order parameters, [arXiv:2411.02237](https://arxiv.org/abs/2411.02237).
- [45] See Supplemental Material at <http://link.aps.org/supplemental/10.1103/2cfz-dm5f> for additional details on the methods and their generalizations, the relation between the correlator basis and spin moments and its connection to the logarithmic wave function encoding, and derivation of entanglement bounds, which also contains Refs. [70–73].
- [46] D. Luo, G. Carleo, B. K. Clark, and J. Stokes, Gauge equivariant neural networks for quantum lattice gauge theories, *Phys. Rev. Lett.* **127**, 276402 (2021).
- [47] Y. Sheffer, E. Berg, and A. Stern, Preparing topological states with finite depth simultaneous commuting gates, *Phys. Rev. B* **111**, 115160 (2025).
- [48] R. O'Donnell, *Analysis of Boolean Functions* (Cambridge University Press, Cambridge, 2014).
- [49] M. Hibat-Allah, M. Ganahl, L. E. Hayward, R. G. Melko, and J. Carrasquilla, Recurrent neural network wave functions, *Phys. Rev. Res.* **2**, 023358 (2020).
- [50] M. Hibat-Allah, R. G. Melko, and J. Carrasquilla, Supplementing recurrent neural network wave functions with symmetry and annealing to improve accuracy, [arXiv:2207.14314](https://arxiv.org/abs/2207.14314).
- [51] N. Lang and H. P. Büchler, Minimal instances for toric code ground states, *Phys. Rev. A* **86**, 022336 (2012).
- [52] W. Zhao, M. Fang, and D. Su, Graph-based approach to entanglement entropy of quantum error-correcting codes, *Phys. Rev. A* **113**, 052412 (2026).
- [53] M. Bukov, M. Schmitt, and M. Dupont, Learning the ground state of a non-stoquastic quantum Hamiltonian in a rugged neural network landscape, *SciPost Phys.* **10**, 147 (2021).
- [54] I. Schurov, A. Kravchenko, M. I. Katsnelson, A. A. Bagrov, and T. Westerhout, Learning complexity of many-body quantum

- sign structures through the lens of Boolean Fourier analysis, [arXiv:2508.09870](#).
- [55] Z. Liu, D. Qian, and J. Wang, Enhancing neural-network variational Monte Carlo through basis transformation, [arXiv:2604.15888](#).
- [56] J. R. Moreno, J. Cohn, D. Sels, and M. Motta, Enhancing the expressivity of variational neural, and hardware-efficient quantum states through orbital rotations, [arXiv:2302.11588](#).
- [57] J. Augusto Sobral, M. Perle, and M. S. Scheurer, Physics-informed transformers for electronic quantum states, *Nat. Commun.* **16**, 10811 (2025).
- [58] T. M. Cover and J. A. Thomas, Kolmogorov complexity, in *Elements of Information Theory* (John Wiley & Sons, Ltd, 2005), Chap. 14, pp. 463–508.
- [59] D. Galvin, Three tutorial lectures on entropy and counting, [arXiv:1406.7872](#).
- [60] Y. Wu, C. Guo, Y. Fan, P. Zhou, and H. Shang, NNQS-transformer: An efficient and scalable neural network quantum states approach for *ab initio* quantum chemistry, in *Proceedings of the International Conference for High Performance Computing, Networking, Storage and Analysis* (Association for Computing Machinery, New York, 2023).
- [61] L. L. Viteritti, R. Rende, A. Parola, S. Goldt, and F. Becca, Transformer wave function for two dimensional frustrated magnets: Emergence of a spin-liquid phase in the Shastry-Sutherland model, *Phys. Rev. B* **111**, 134411 (2025).
- [62] H. Lange, G. Bornet, G. Emperauger, C. Chen, T. Lahaye, S. Kienle, A. Browaeys, and A. Bohrdt, Transformer neural networks and quantum simulators: A hybrid approach for simulating strongly correlated systems, *Quantum* **9**, 1675 (2025).
- [63] J. R. Moreno, G. Carleo, A. Georges, and J. Stokes, Fermionic wave functions from neural-network constrained hidden states, *Proc. Natl. Acad. Sci. USA* **119**, e2122059119 (2022).
- [64] B. J. Wurst, D. M. Kennes, and J. B. Profe, Efficiency of the hidden fermion determinant states ansatz in the light of different complexity measures, *Phys. Rev. Res.* **7**, 023316 (2025).
- [65] H. Lange, A. Böhrer, C. Roth, and A. Bohrdt, Simulating the two-dimensional t - j model at finite doping with neural quantum states, *Phys. Rev. Lett.* **135**, 136504 (2025).
- [66] D. Luo and B. K. Clark, Backflow transformations via neural networks for quantum many-body wave functions, *Phys. Rev. Lett.* **122**, 226401 (2019).
- [67] R. Rende, S. Goldt, F. Becca, and L. L. Viteritti, Fine-tuning neural network quantum states, *Phys. Rev. Res.* **6**, 043280 (2024).
- [68] K. Choo, A. Mezzacapo, and G. Carleo, Fermionic neural-network states for *ab-initio* electronic structure, *Nat. Commun.* **11**, 2368 (2020).
- [69] F. Döschl and A. Bohrdt, CoTra_NQS: Code and data supporting the results presented in this paper, GitHub repository, 2026, https://github.com/FabianDoeschl/CoTra_NQS.
- [70] G. Carleo, K. Choo, D. Hofmann, J. E. T. Smith, T. Westerhout, F. Alet, E. J. Davis, S. Efthymiou, I. Glasser, S.-H. Lin, *et al.*, NetKet: A machine learning toolkit for many-body quantum systems, *SoftwareX* **10**, 100311 (2019).
- [71] E. Waring, VII. Problems concerning interpolations, *Philos. Trans. R. Soc. London* **69**, 59 (1779).
- [72] A. M. Bayen and T. Siau, Interpolation, in *An Introduction to MATLAB® Programming and Numerical Methods for Engineers*, edited by A. M. Bayen and T. Siau (Academic Press, Boston, 2015), Chap. 14, pp. 211–223.
- [73] M. Lindsey, Multiscale interpolative construction of quantized tensor trains, [arXiv:2311.12554](#).