



HAL
open science

Efficiency of neural quantum states in light of the quantum geometric tensor

Sidhartha Dash, Filippo Vicentini, Michel Ferrero, Antoine Georges

► **To cite this version:**

Sidhartha Dash, Filippo Vicentini, Michel Ferrero, Antoine Georges. Efficiency of neural quantum states in light of the quantum geometric tensor. 2024. hal-04438272

HAL Id: hal-04438272

<https://hal.science/hal-04438272v1>

Preprint submitted on 6 Feb 2024

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

Efficiency of neural quantum states in light of the quantum geometric tensor

Sidhartha Dash,¹ Filippo Vicentini,^{2,1} Michel Ferrero,^{2,1} and Antoine Georges^{1,3,2,4}

¹Collège de France, Université PSL, 11 place Marcelin Berthelot, 75005 Paris, France

²CPHT, CNRS, Ecole Polytechnique, IP Paris, F-91128 Palaiseau, France.

³Center for Computational Quantum Physics, Flatiron Institute, New York, New York, 10010, USA.

⁴DQMP, Université de Genève, 24 quai Ernest Ansermet, CH-1211 Genève, Switzerland.

(Dated: February 5, 2024)

Neural quantum state (NQS) ansätze have shown promise in variational Monte Carlo algorithms by their theoretical capability of representing any quantum state. However, the reason behind the practical improvement in their performance with an increase in the number of parameters is not fully understood. In this work, we systematically study the efficiency of restricted Boltzmann Machines (RBMs) to represent the ground states in different phases of the spin-1 bilinear-biquadratic model, as the hidden layer density α increases. We train our ansatz by minimizing two different loss functions: 1) energy, and 2) infidelity of the NQS ansatz w.r.t. that of the exact ground state. We observe that the accuracy of our ansatz saturates with α in both cases. We demonstrate that this can be explained by looking at the spectrum of the quantum geometric tensor (QGT). We find that the rank of the QGT saturates beyond a certain α , and we emphasize that it corresponds to the *dimension of the relevant manifold* for an optimized NQS. This provides a useful diagnostics for the practical representation power of an NQS ansatz.

I. INTRODUCTION

Recent years have seen an immense growth in the use of machine learning (ML) methods in the field of quantum many-body physics. Central to this intersection are Neural Quantum States (NQSs), which are currently revolutionizing Variational Monte Carlo (VMC) approaches and related applications [1–8]. Their success relies on the *expressivity* of Neural-Networks (NNs), which have the theoretical capacity to represent any state, with a large enough number of parameters. This is formalized by so called (i) *universal representation theorems*, asserting that the approximation error inherent in a neural network (and by extension, in an NQS) can approach zero asymptotically as one increases the network’s width or depth, contingent on locating the global minimum of the loss function [9–12]. Additionally, it is well-understood in the standard machine learning context that (ii) increasing the number of parameters beyond the *over-parameterization* limit leads to a smoother loss landscape and faster convergence [13, 14].

Building on those principles, numerous variational studies have increased the width or depth of NNs in order to check convergence in calculations where the ground state is found by energy minimization [1, 15–17]. The most striking example is given by Fig. 2 of ref. [17] where convergence to *numerical precision* is achieved by gradually increasing the number of parameters of a CNN-ResNet from $\sim 10^4$ to $\sim 2 \times 10^5$. This approach is quite general and is also employed when simulating the real-time dynamics [18], and the steady-state in open quantum systems [5, 19].

However, exceptions to this rule do emerge in practical calculations, i.e. instances in which an NQS fails to become more accurate as the number of parameters increases. We will call this situation a *practical breakdown of universal representability*. We stress that this is not

in opposition to the representation theorems, which are theoretical concepts valid only when the global minimum can be found, in the limit of a *sufficiently large number of parameters*. This *practical breakdown of universal representability*, given a specific minimization algorithm, might happen either because the global minima cannot be found with a reasonable computational budget, or because the additional parameters cannot be used effectively even when the minimization algorithm has converged.

Moreover, it is unclear what the additional parameters encode once the optimization has converged. Several results in standard ML tasks [20–22], and Variational Quantum Algorithms [23] document the existence of redundant directions in the parameter space, suggesting that the encodings are *locally* highly degenerate. This is in clear opposition to tensor networks and matrix product states in particular, where increasing the bond dimension is linked with the increase in the maximal entanglement entropy of the state [24, 25]. Instead, while it has been shown that even simple NQSs can encode states with arbitrary entanglement [26–28], we do not know what states NQSs with a finite number of parameters cannot encode. The role of parameters in an NQS and their relationship to the overall accuracy of a calculation is therefore unclear. And while in some cases it is possible to invoke some theorems as an explanation for the effectiveness of increasing the number of parameters, we still do not understand what happens in presence of a *practical breakdown of universal representability*.

Nevertheless, it is possible to quantify the role of parameters in an NQS by looking at the Quantum Geometric Tensor (QGT), which is a special case of the Fisher information matrix (FIM), defined in the context of variational wavefunctions [29, 30]. A few studies have used the QGT to define an *effective dimension* of a variational ansatz, and to identify redundant directions in the parameter space [31, 32]. This can be used in the case of a *prac-*

tical breakdown of universal representability to identify whether it is caused by the inefficiency of the NQS to use the newly added parameters.

In this work, we investigate such a scenario in a simple setting: the search for the ground state of a one-dimensional quantum spin model. We choose the spin-1 bilinear-biquadratic (BLBQ) model, which contains a diverse set of phases including the gapped Haldane phase, and a gapless extended critical phase. Our VMC results do not show a systematic improvement in the converged infidelities/energies upon increasing α , for various phases of the BLBQ model. This contradicts the usual expectation that the accuracy of the ansatz should improve as we progress towards the limit where universal representation theorems hold. With this observation, we perform a conceptually simpler task of optimizing the infidelity of the same NQS ansatz w.r.t. the true ground state, as the loss function, to directly investigate the representation ability of the ansätze while increasing the density of the NN. We use an NQS ansatz given by a modified restricted Boltzmann machine (see supplementary material VI) for spin-1 systems [15]. The optimization of the infidelity is done by computing the infidelity and its gradient exactly to exclude biases induced from sampling in the Monte Carlo procedure. We demonstrate that the rank and spectrum of the QGT provide very useful insights into the number of *relevant* parameters that are used in practice to represent the wave-function, and how the network makes use of these parameters as its size is increased.

II. THE BILINEAR BIQUADRATIC SPIN-1 CHAIN

Model. The spin-1 bilinear-biquadratic (BLBQ) model in one dimension is defined by the Hamiltonian:

$$H = \sum_i J [\mathbf{S}_i \cdot \mathbf{S}_{i+1} + \tan(\theta) (\mathbf{S}_i \cdot \mathbf{S}_{i+1})^2]. \quad (1)$$

The model is parametrized by an angular variable θ , and $\mathbf{S}_i = (S_{ix}, S_{iy}, S_{iz})$ is the spin operator acting on the local spin-1 Hilbert space at site i . The above Hamiltonian has a gapped Haldane phase [33, 34] for $-\pi/4 < \theta < \pi/4$, an extended critical phase [35–39] for $\pi/4 \leq \theta < \pi/2$, a ferromagnetic phase for $\pi/2 \leq \theta < 5\pi/4$, and a dimerized phase for $-\pi/4 < \theta < -3\pi/4$ [40, 41] (Fig. 1). We focus on the Haldane phase and the extended critical phase in this work. The Haldane phase describes the isotropic Heisenberg antiferromagnet at $\theta = 0$. It is characterized by a hidden topological order, which is the strongest for the *Affleck-Kennedy-Lieb-Tasaki* (AKLT) state [42–44] at $\theta = \arctan(1/3)$. The AKLT state is a valence-bond state, which can be represented by two spin-1/2 particles at each site, forming singlets with the spins of the neighboring sites (see supplementary material sec. VIA). The AKLT state has an exact matrix product state (MPS) representation with bond dimension 2 [45]. It is also the state that has the lowest bipartite entanglement in comparison to other values of θ [46]. The Haldane phase is gapped and has

a ground state with an exponentially decaying antiferromagnetic spin-spin correlation in the thermodynamic limit. Up to the AKLT point, the correlation function behaves as $\langle S_0^\alpha \cdot S_j^\alpha \rangle \approx (-1)^j \exp(-j/\xi) / \sqrt{j}$ with a correlation length $\xi = 1/\ln(3)$ for the AKLT state. For $\arctan(1/3) < \theta < \pi/4$, the modulation wave-vector shifts away from $k = \pi$ to reach $k = 2\pi/3$ at $\theta = \pi/4$. The Haldane gap closes at $\theta = \pi/4$, marking a critical point, known as the *Uimin-Lai-Sutherland* (ULS) point [37, 38, 47], across a *Berezinskii-Kosterlitz-Thouless* (BKT) transition [39, 48]. At this point, the model has an exact $SU(3)$ symmetry and is integrable. Its low-energy physics is described by the Wess-Zumino-Witten ($SU(3)_{k=1}$) conformal field theory [39, 43, 48]. The model remains gapless in the thermodynamic limit throughout the region $\pi/4 \leq \theta < \pi/2$. In this phase, the dominant correlations are of quadrupolar nature, with wavevector $k = \pm 2\pi/3$ and decay as a power law with exponent $\eta = 4/3$ [35, 39].

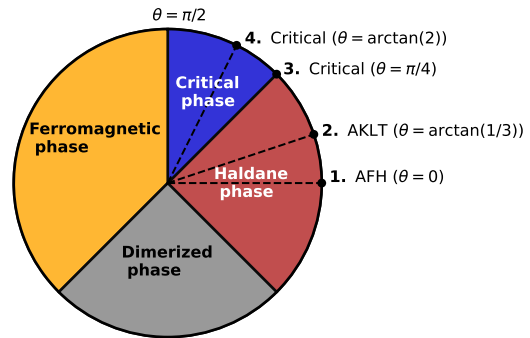


FIG. 1. Phase diagram of the spin-1 bilinear-biquadratic (BLBQ) model. We perform infidelity (Eq. (3)) minimization, and energy (Eq. (2)) minimization with an NQS ansatz given by a modified RBM for spin-1 models [15] (described in Methods V) at the four marked points in the phase diagram.

III. RESULTS

In this work, we search for the NQS approximations of the ground states in the Haldane and the critical phases of the BLBQ model by minimizing two loss functions: 1) the energy, using the VMC algorithm,

$$E = \langle H \rangle = \sum_n p_\theta(n) \frac{\langle n | H | \psi_\theta \rangle}{\langle n | \psi_\theta \rangle}, \quad (2)$$

and 2) the infidelity with respect to the exact ground state $|\Omega\rangle$ (see Methods V for more details),

$$I = 1 - \frac{\langle \psi_\theta | \Omega \rangle \langle \Omega | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle \langle \Omega | \Omega \rangle}. \quad (3)$$

Here $|\psi_\theta\rangle = \sum_n \psi_\theta(n) |n\rangle$ is the NQS and $p_\theta(n) = |\psi_\theta(n)|^2 / \langle \psi_\theta | \psi_\theta \rangle$ is the probability distribution over states $|n\rangle$ in the computational basis. The degree to which the NQS ansatz represents the true ground state

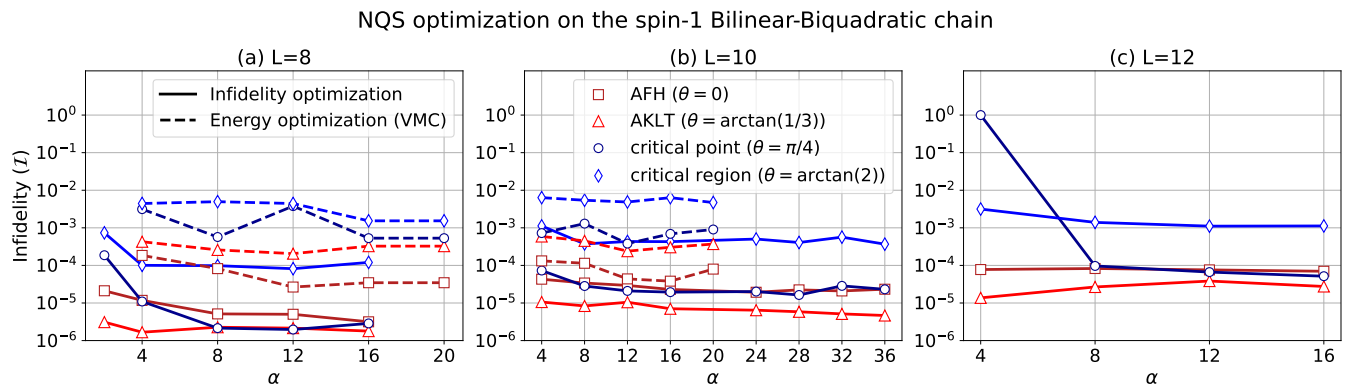


FIG. 2. This figure shows the infidelities (I) after the convergence of an exact infidelity optimization procedure, and a Variational Monte Carlo (VMC) procedure (see Methods V), as a function of α , for the four points in the phase diagram (see Fig. 1) of the spin-1 Bilinear Biquadratic (BLBQ) chain of length (a) $L = 8$, (b) $L = 10$, and (c) $L = 12$ with open boundary conditions. The NQS is a modified RBM for spin-1 systems [15] (see Methods V). The optimization of the NQS for a given model, for a given size L and a given α , is done starting from few (10 for the infidelity optimization procedure, and at least 25 for the VMC procedure) different initializations of the parameters of the NQS, and we choose the one with the smallest infidelity on convergence.

is given by the fidelity $F = 1 - I$. To quantify this for the VMC solutions, we show the infidelities (Eq. (3)) of the NQSs after the energy minimizations (as dashed lines) in Fig 2(a,b), for different hidden layer densities α , and lengths $L = 8, 10$. We find that the infidelities of the NQS do not improve with α beyond a certain value for the VMC results. To investigate this saturation of the infidelity further, we also minimize the infidelity (Eq. (3)) with the same NQS ansatz (see Methods V), while increasing α . For the rest of the paper, we focus on the results of the infidelity minimization procedure. We plot the converged infidelities (as solid lines) in Fig. 2, for different sizes of the BLBQ chain.

We observe that the converged infidelities, for the different phases, improve until a certain value of α and saturate thereafter, with some oscillations, for different sizes of the chain. This coincides with our observations for the case of energy minimization. Further, note that the NQS for the AKLT point has the minimum infidelity for most of the cases. This is likely due to the fact that the AKLT state has the lowest entanglement (has an exact MPS representation with bond dimension 2), among all other phases. However, note that in the VMC solution (dashed lines in Fig. 2(a,b)), the NQS for the AKLT point is not the most accurate [49]. This could be due to the fact that the AKLT Hamiltonian has a degenerate ground state (in case of an open boundary condition; see supplementary material sec. VIA), which makes the energy minimization problematic.

As we want to study the representation power of the NQS ansätze, we perform the infidelity minimization exactly in order to exclude biases and artefacts originating from the Monte Carlo sampling [50] in the optimization procedure and in the accuracy of the NQS. In other words, we compute all quantities during the infidelity minimization procedure by a full summation over the relevant subspace of the Hilbert space.

Additionally, we check that the infidelity minimization has converged well by inspecting the local infidelity landscape through the Hessian for $L = 8$ (see supplementary material VIC). Our solutions lie in a deep valley (many large positive eigenvalues), which has a few infinitesimally small downward slopes (few very small negative eigenvalues) as commonly seen in the ML literature [51, 52].

In order to understand the saturation of the accuracy of the NQS ansatz with an increase in α , we look at the quantum geometric tensor (QGT) [29, 30]. As we shall see in the next paragraph, the QGT gives a measure of the *relevant* directions in the parameter space [31, 32].

Quantum Geometric Tensor (QGT) The QGT provides a metric ds^2 over the space of variational quantum states. Starting from the Fubini-Study distance between two such states associated with parameters θ and ϕ :

$$\text{FS}(\psi_\theta, \psi_\phi) = \arccos \left(\sqrt{\frac{\langle \psi_\theta | \psi_\phi \rangle \langle \psi_\phi | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle \langle \psi_\phi | \psi_\phi \rangle}} \right). \quad (4)$$

One obtains, for an infinitesimal variation $\phi = \theta + d\theta$, up to second order in $d\theta$:

$$ds^2 = \text{FS}(\psi_\theta, \psi_{\theta+\delta\theta})^2 = \sum_{\alpha, \beta} G_{\alpha\beta} d\theta_\alpha^* d\theta_\beta, \quad (5)$$

In this expression, the QGT $G_{\alpha\beta}$ reads [53]:

$$G_{\alpha\beta} = \left\langle \frac{\partial \psi_\theta}{\partial \theta_\alpha} \middle| \frac{\partial \psi_\theta}{\partial \theta_\beta} \right\rangle - \left\langle \frac{\partial \psi_\theta}{\partial \theta_\alpha} \middle| \psi_\theta \right\rangle \left\langle \psi_\theta \middle| \frac{\partial \psi_\theta}{\partial \theta_\beta} \right\rangle \quad (6)$$

Let us note that the infidelity between two states differing by $d\theta$ is given by $I \approx \text{FS}^2$. Hence, the QGT appears as a metric in the second order expansion of the infidelity between two infinitesimally close variational quantum states. The QGT is hermitian, and its eigenvalues quantify how much the FS distance changes when we move

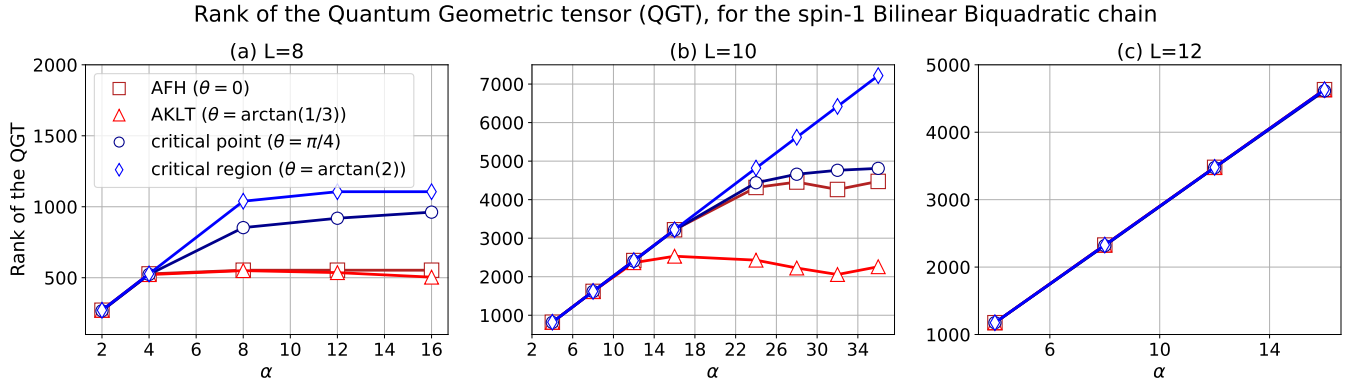


FIG. 3. This figure shows the rank of the quantum geometric tensor (QGT) after convergence of the infidelity minimization. The rank is computed as the number of eigenvalues of the QGT which are greater than 10^{-16} . (a), (b), (c) show the QGT ranks, after 120K infidelity minimization steps, for the AFH, AKLT, and the two critical phases for the BLBQ chain (with open boundary condition), with lengths $L = 8, 10, 12$ respectively.

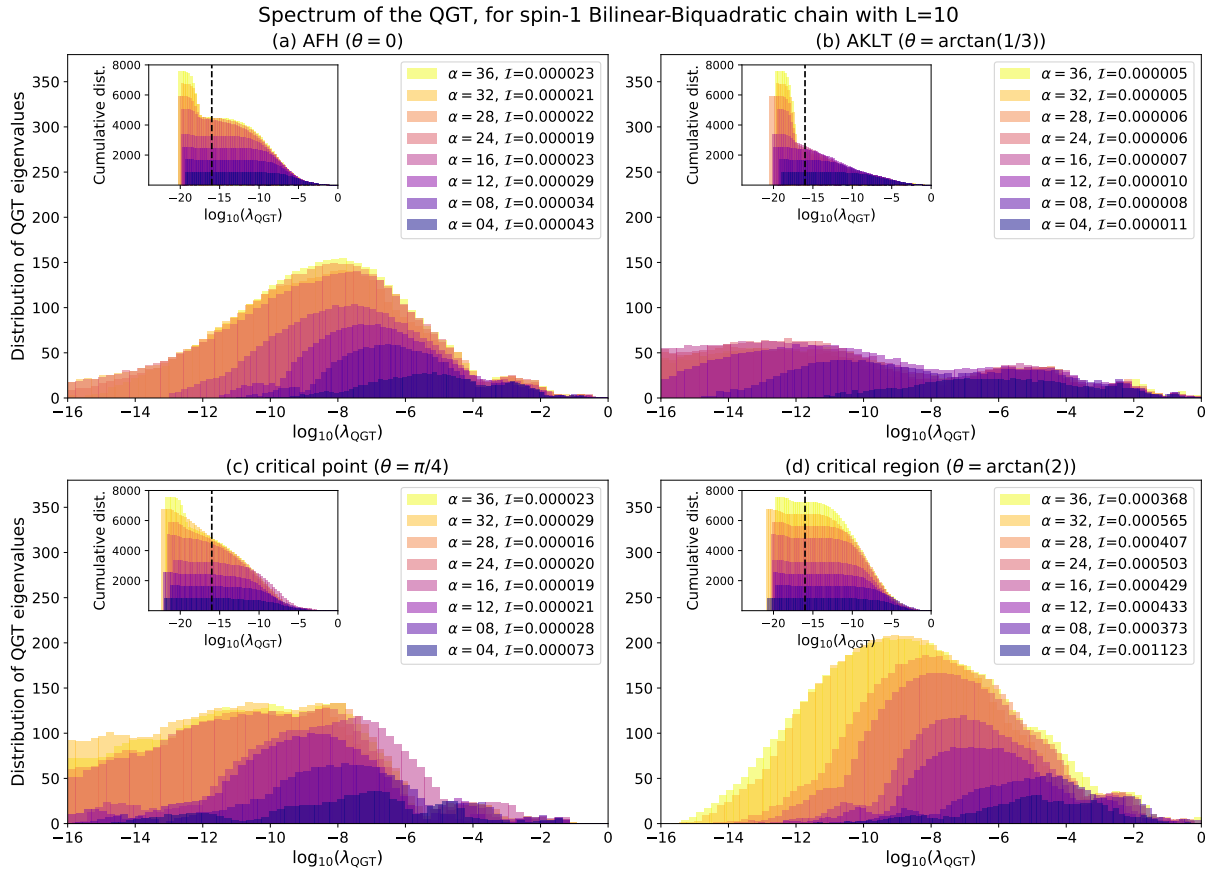


FIG. 4. This figure shows the spectra of the quantum geometric tensor (QGT) after convergence of the infidelity minimization. The plots show the distribution of the normalized eigenvalues (w.r.t. the maximum eigenvalue) of the QGT, in \log_{10} scale, after 120K iterations of infidelity minimization on the spin-1 BLBQ chain of length $L = 10$ for four points in the BLBQ phase diagram (see Fig. 1). Normalized eigenvalues with magnitudes greater than 10^{-16} (cutoff to compute the rank) are shown in the plots. The insets show the cumulative distribution of the eigenvalues (no. of eigenvalues greater than the value on the x axis) for each case. The dashed line in the inset marks the cutoff 10^{-16} .

infinitesimally along the corresponding eigenvectors. In the vicinity of a local minimum of the infidelity w.r.t. the true ground state (Eq. (3)), the eigenvectors of the QGT with non-zero eigenvalues are the ones that change the NQS ansatz away from that at the minimum. In contrast, the eigenvectors with zero eigenvalues are the *redundant directions* that do not change the ansatz from that at the minimum. Therefore, at convergence, the rank of the QGT gives the number of *relevant* parameters that the variational ansatz uses to represent a given quantum state.

Note that when the wavefunction amplitudes are real and positive, the QGT (Eq. (6)) takes the form

$$G_{\alpha\beta} = \frac{1}{4} \left\langle \frac{\partial \log p_\theta(n)}{\partial \theta_\alpha} \frac{\partial \log p_\theta(n)}{\partial \theta_\beta} \right\rangle - \frac{1}{4} \left\langle \frac{\partial \log p_\theta(n)}{\partial \theta_\alpha} \right\rangle \left\langle \frac{\partial \log p_\theta(n)}{\partial \theta_\beta} \right\rangle = \frac{1}{4} \mathcal{F}_{\alpha\beta}, \quad (7)$$

where the averages are taken w.r.t. the Born probability distribution $p_\theta(n)$ over states $|n\rangle$, and \mathcal{F} is the Fisher information matrix (FIM) associated with the probability distribution $p_\theta(n)$. The FIM is obtained as a metric from the second order expansion of the Kullback–Leibler divergence, which is a measure of the proximity between two probability distributions. In our case, the wavefunction amplitudes are complex, and thus the QGT can be seen as a generalization of the FIM to complex probability amplitudes.

Rank of the QGT. We compute the QGT at convergence in the infidelity minimization procedure for each case, and plot the rank as a function of α in Fig. 3. The rank of the QGT saturates with α for chain lengths $L = 8$ and 10 (see Fig. 3(a,b)), suggesting that the optimized NQS is unable to leverage the new parameters to improve the approximation for the ground state, as we increase α beyond a certain value. This is consistent with the apparent saturation of the infidelities (Fig. 2) as we increase α . Note that the limiting value of the rank is higher for phases within the critical region than in the Haldane phases, suggesting that the optimized NQS for the critical phases has a larger number of *relevant* parameters. Specifically, we observe that the limiting rank is the lowest for the AKLT state and highest for the critical region at $\tan\theta = 2$. This implies that the NQS for the AKLT point utilizes the fewest parameters, reflecting the fact that the wave-function at the AKLT state is inherently the simplest in the phase diagram. We also remark that the rank of the QGT for a particular phase, at fixed α , increases as we increase the size of the chain. It should be noted that the rank of the QGT for the BLBQ chain with $L = 12$ (Fig. 3(c)) has not yet saturated with α for the four phases, suggesting that the NQS is still adding *relevant* parameters with an increase in α .

Spectrum of the QGT. Although the rank counts the number of relevant directions, it does not describe how the local metric of the FS distance evolves with an increase in α . This can be characterized by the distribution of the eigenvalues of the QGT. We plot the spectrum of the QGT,

for each solution with $L = 10$, as a histogram in Fig. 4, where the eigenvalues of the QGT are normalized w.r.t. the maximum eigenvalue. The bulk of the distribution of the QGT eigenvalues eventually converges to a limiting distribution while we increase α . As α increases, the additional eigenvalues accumulate around numerical zeros ($< 10^{-16}$). The latter can be seen from the cumulative distribution functions in the respective insets of Fig. 4, which gives the number of eigenvectors with eigenvalues greater than a particular value. This indeed confirms that the NQS ansatz no longer utilizes the newly added directions to represent the ground states, as suggested by the saturation of the rank (in Fig. 3(b)). Note that the QGT spectrum is strikingly different for the AKLT phase, where the limiting distribution is multimodal with the shortest maximum height and displays the fastest convergence with α (around $\alpha = 16$). This is consistent with the fact that the limiting rank is the lowest for the AKLT phase, and reaches a plateau earlier than other phases (see Fig. 3(b)). We also see similar phenomena in the spectrum, and the rank of the QGT in case of the energy minimization (VMC), as seen in Figs. 8, 9 in the supplemental material sec. VI E.

IV. DISCUSSION

In summary, we have performed a large number of numerical experiments to find ground states in various phases of the spin-1 BLBQ chain using a 1-layer NQS ansatz (Eq. (8)). We minimize both energy (Eq. (2)) and infidelity (Eq. (3)). Increasing the hidden layer density α of the network, we observe a crossover between two regimes: at low densities, increasing α leads to higher accuracy, but eventually we observe a threshold above which increasing α does not lead to systematic improvements. This holds true for both energy minimization (VMC) and infidelity minimization procedures to optimize the NQS for representing various ground states of the spin-1 BLBQ model (Eq. (1)). To address the dependency from random initial conditions, we perform optimizations with many initial states for the NQS, and choose the solution with the lowest infidelity (Eq. (3)), for each case.

We investigated how the optimized NQS leverages the increasing number of variational parameters by means of the spectrum of the QGT. Our key finding is that the dimension of the *relevant* manifold for the optimized NQS, given by the rank of the QGT at convergence, saturates around a given value for independent (energy or infidelity) optimizations at different values of α for the ansatz. This is consistent with the fact that the accuracy of the NQS ansatz ceases to improve on increasing the width of the ansatz, since the increase in parameters mostly contributes to an increase in locally redundant directions in the parameter space, in the vicinity of the converged solution. This asserts a limitation in the practical representation ability of our NQS ansatz as the width of the single hidden layer increases, and poses as an appar-

ent hurdle for the NQS to reach the regime of universal representability.

Our analysis of the QGT provides a general diagnosis of the performance of an NQS ansatz, especially when it fails to systematically improve the approximation for the ground state on increasing the number of parameters of the ansatz, i.e. in situations of a *practical breakdown of universal representability*. The presence of a *practical breakdown of universal representability* is consistent with established findings in the ML literature which highlight that a substantial number of directions in the parameter space exhibit redundancy [20, 21]. This is in contrast to the density matrix renormalization group algorithms, with tensor network ansätze, where each variational parameter contributes to the entanglement entropy of the variational ansatz [25].

Furthermore, the onset of such a *practical breakdown of universal representability* might depend on the architecture of the NN ansatz. For instance, this might occur earlier (w.r.t. the parameter count) in shallow NN ansätze than in deeper NNs, given that deeper NNs poses a higher representation ability [54]. In future studies, it would be interesting to investigate this phenomenon for deep multilayer perceptrons and other NN architectures. It would also be interesting to study this with optimization algorithms based on the Newton methods, which promise to avoid saddle points efficiently [51, 55], but can only be implemented for smaller NNs. Besides, a compelling application of our analysis lies in characterizing different methods of encoding the input for a given NQS architecture. This involves an inspection of the rank of the QGT at initialization for different encodings, and choosing the encoding with the highest rank of the QGT, which would ensure that the chosen ansatz is the easiest to optimize.

To summarize the performance of our NQS ansatz across the four phases of the BLBQ spin chain considered in this work, we note that the NQS represents states in the gapped Haldane phase with a lower limiting dimension of the *relevant* manifold in parameter space than in the critical gapless phase. The NQS approximation for the AKLT state requires the least number of *relevant* parameters, owing to the simple yet non-trivial entanglement structure of the AKLT state. The NQS approximation for the critical phase at $\tan\theta = 2$ is the worst in both energy and infidelity minimization results, and has the largest *relevant* manifold in the parameter space. It is interesting to note, in contrast, that the ULS critical point at $\theta = \pi/4$ is more accurately represented by our NQS ansatz than the critical region at $\tan\theta = 2$. This may be due to the fact that the Hamiltonian at the ULS point has an enhanced SU(3) symmetry [35, 48].

In conclusion, we believe that the emergence of a *practical breakdown of universal representability* shows that practical calculations with NQS-based ansätze can differ considerably from the asymptotic regime addressed by universal representation theorems. We have shown that the spectrum of the QGT can diagnose the fact that, in some cases additional parameters are completely ig-

nored by the variational optimization procedure. The QGT shares the eigenvalues with the Neural Tangent Kernel [17], which has a dimension equal to the number of samples, and is hence easier to diagonalize when the number of parameters is large. Therefore, the information about the QGT spectrum can be accessed at a reasonable computational cost even for large networks, and could be integrated into an advanced optimization or regularization procedure in order to ensure the efficient usage of the network parameters.

ACKNOWLEDGEMENTS

We are grateful to Giuseppe Carleo, Juan Carrasquilla, Michele Casula, Stephen Clark, Anna Dawid, Matija Medvidović, Frédéric Mila, Schuyler Moss, Javier Robledo Moreno, Christopher Roth, André-Marie Tremblay and Agnes Valenti for useful discussions. S.D. would like to thank Olivier Simard for a careful reading of the manuscript. F.V. acknowledges support by the French Agence Nationale de la Recherche through the NDQM project, grant ANR-23-CE30-0018. This work was granted access to the HPC resources of TGCC and IDRIS under the allocation A0150510609 attributed by GENCI (Grand Equipement National de Calcul Intensif). We also acknowledge the use of computing resources at the Flatiron Institute, a division of the Simons Foundation.

V. METHODS

We use a modified version of the RBM, adapted for spin-1 systems [15], as the wavefunction ansatz (NQS) $\psi_\theta(\sigma) = \sum_h \exp[\mathcal{E}(\sigma, h)]$, where

$$\begin{aligned} \mathcal{E}(\sigma, h) = & \sum_{i=1}^L a_i \sigma_i + \sum_{i=1}^L A_i \sigma_i^2 + \sum_{i=1}^M \sum_{j=1}^L w_{ij} h_i \sigma_j \\ & + \sum_{i=1}^M \sum_{j=1}^L W_{ij} h_i h_j + \sum_{i=1}^M h_i b_i \end{aligned} \quad (8)$$

is the energy function of the spin-1 RBM. Here $\theta = \{\{a_i\}, \{A_i\}, \{b_i\}, \{w_{ij}\}, \{W_{ij}\}\}$ is the set of all $2L + M + 2ML$ complex parameters of the spin-1 RBM, L is the number of sites in the spin chain (no. of units in the visible layer), $M = \alpha L$ is the number of units in the hidden layer, and $\sigma = \{\sigma_i\}$ denotes a spin-configuration on the lattice. Note from Eq. (8) that the NQS ansatz given by the spin-1 RBM is holomorphic, i.e. $\psi_\theta(\sigma)$ depends only on θ , and not on θ^* .

We use this NQS ansatz (Eq. (8)) to represent the ground states of the BLBQ model in different parameter regimes. We use the infidelity measured w.r.t. the exact ground state as the loss function, for optimizing our NQS ansatz for most of our results. In addition, we also perform variational Monte Carlo, where we optimize the expectation value of the Hamiltonian, for a few cases.

Infidelity Optimization

For the infidelity optimization, we use the loss function given by the infidelity of the NQS w.r.t. the exact ground state,

$$\mathcal{L} = I = 1 - \frac{\langle \psi_\theta | \Omega \rangle \langle \Omega | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle \langle \Omega | \Omega \rangle}, \quad (9)$$

where $|\Omega\rangle$ is the exact ground state, and $\theta \in \mathbb{C}^N$. The loss function is minimized by methods based on the *stochastic gradient descent* algorithm, in which we use the conjugate gradient of the loss function,

$$\begin{aligned} F = \frac{\partial \mathcal{L}}{\partial \theta^*} = & - \frac{\langle \Omega | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle \langle \Omega | \Omega \rangle} \sum_n \langle n | \Omega \rangle \frac{\partial \psi_\theta^*(n)}{\partial \theta^*} \\ & + \frac{|\langle \psi_\theta | \Omega \rangle|^2}{\langle \psi_\theta | \psi_\theta \rangle^2 \langle \Omega | \Omega \rangle} \sum_n \psi_\theta(n) \frac{\partial \psi_\theta^*(n)}{\partial \theta^*}. \end{aligned} \quad (10)$$

We compute the infidelity and the gradient exactly, by summing over the complete basis of the total $S_z = 0$ sector of the Hilbert space. We use a combination of the *ADAM* [56] and *YOGI* [57] optimizers for our gradient descent optimization. More details of the optimization procedure are given in the supplementary material sec. VI B.

Variational Monte Carlo (VMC)

In the variational Monte Carlo (VMC) procedure, the loss function is taken to be the expectation value of the Hamiltonian,

$$\mathcal{L} = \langle H \rangle = \sum_n p_\theta(n) \frac{\langle n | H | \psi_\theta \rangle}{\langle n | \psi_\theta \rangle}. \quad (11)$$

Here $|n\rangle$ represents the basis states, and $p_\theta(n) = |\psi_\theta(n)|^2 / \langle \psi_\theta | \psi_\theta \rangle$. We approximate the above sum by a Monte Carlo sampling procedure, where we generate basis configurations as Markov chains by the Metropolis-Hastings algorithm. The Metropolis-Hastings algorithm proposes a new configuration n' , starting from n , and accepts it with probability

$$P_{\text{acc}}(n \rightarrow n') = \min \left\{ 1, \frac{|\psi_\theta(n')|^2 g(n|n')}{|\psi_\theta(n)|^2 g(n'|n)} \right\}, \quad (12)$$

where $g(n|n')$ is the conditional probability of proposing the configuration n , given n' .

The conjugate gradient of the loss function Eq. (11) is given by

$$\begin{aligned} F = \frac{\partial \mathcal{L}}{\partial \theta^*} = & \sum_n p_\theta(n) \left[\frac{\partial \log \psi_\theta^*(n)}{\partial \theta^*} \frac{\langle n | H | \psi_\theta \rangle}{\psi_\theta(n)} \right. \\ & \left. + \langle H \rangle \frac{\partial \log \psi_\theta^*(n)}{\partial \theta^*} \right], \end{aligned} \quad (13)$$

for a holomorphic wavefunction ansatz. The gradient in the above equation is also estimated approximately by the

Monte Carlo procedure described above. We use stochastic gradient descent (SGD), in combination with stochastic reconfiguration (SR) [58, 59] for the energy minimization. With SR, the parameters of the ansatz are updated as

$$\theta_\mu \rightarrow \theta_\mu - \eta \sum_\nu [\mathbf{G} + \epsilon \mathbb{1}]_{\mu\nu}^{-1} F_\nu, \quad (14)$$

where \mathbf{G} is the quantum geometric tensor (QGT), η is the learning rate, and ϵ is a small regularization constant, which is taken to be 10^{-3} . More details of the optimization procedure are given in the supplementary material sec. VI B.

The energy minimization (VMC) with SGD+SR method, and the infidelity minimization (with *ADAM* and *YOGI* optimizers from the optax library [60]) were implemented with the software library *NetKet* [61–64].

VI. SUPPLEMENTARY MATERIAL

A. The AKLT state

The AKLT state, also known as a valence-bond solid [42], is the ground state of the spin-1 BLBQ chain for $\theta = \arctan(1/3)$:

$$H_{\text{AKLT}} = \sum_{i=1}^{N-1} J \left[\mathbf{S}_i \cdot \mathbf{S}_{i+1} + \frac{1}{3} (\mathbf{S}_i \cdot \mathbf{S}_{i+1})^2 \right], \quad (15)$$

where $\mathbf{S}_i = (S_{ix}, S_{iy}, S_{iz})$ is the spin-1 operator acting on site i . The AKLT state is exactly known and can be represented by expressing the spin-1 particle by two auxiliary spin-1/2 particles. The spin-1 computational basis can be represented in terms of the triplet states formed with two spin-1/2 particles:

$$|+\rangle = \psi_{11} = |\uparrow\uparrow\rangle, \quad (16)$$

$$|0\rangle = \psi_{12} = \frac{1}{\sqrt{2}} (|\uparrow\downarrow\rangle + |\downarrow\uparrow\rangle) = \psi_{21}, \quad (17)$$

$$|-\rangle = \psi_{22} = |\downarrow\downarrow\rangle, \quad (18)$$

where $\{|+\rangle, |0\rangle, |-\rangle\}$ is the basis for the spin-1 system, and $\{|\uparrow\rangle, |\downarrow\rangle\}$ is the spin-1/2 basis. Then, the AKLT state is given by

$$|\Psi_{\text{AKLT}}(\alpha, \beta)\rangle = 2^{-(N-1)/2} \psi_{\alpha\beta 1} \epsilon^{\beta_1\alpha_2} \psi_{\alpha_2\beta_2} \epsilon^{\beta_2\alpha_3} \dots \psi_{\alpha_i\beta_i} \epsilon^{\beta_i\alpha_{i+1}} \\ \times \dots \psi_{\alpha_{N-1}\beta_{N-1}} \epsilon^{\beta_{N-1}\alpha_N} \psi_{\alpha_N\beta}, \quad (19)$$

where ϵ is the Levi-Civita tensor of rank 2, and repeated indices imply a summation. Note that the AKLT state Eq. (19) is written for the case of an open boundary condition, and has two free spin-1/2 variables α, β which correspond to the two outermost spin-1/2s on the chain. This leads to a four-fold degeneracy of the AKLT state. In the total $S_z = 0$ sector, the AKLT state is two-fold degenerate. It is interesting to note in the AKLT state that two adjacent spin-1 variables are never aligned ferromagnetically (both +s or both -s). This can be realized from Eq. (19), given that the Levi-Civita tensor only has off-diagonal terms. A broader consequence of this is that a + (-) can only be followed by a 0 or a - (+), i.e. a state of the form $|+000000000+\rangle$ is not allowed. Whereas a typical state could take the form $|00+0-+000-0+00-0\rangle$, which essentially has a Néel order when we remove all 0s. This is a consequence of a non-local order, characterized by the *string order parameter*

$$O_{ij} = \langle S_{iz} e^{i\pi \sum_{i < k < j} S_{kz}} S_{jz} \rangle, \quad (20)$$

in the Haldane phase, which attains the maximum value at the AKLT point.

Furthermore, the AKLT Hamiltonian Eq. (15) can be written as a sum of projection operators into the spin-2 subspace for every two neighboring sites, $\hat{P}_{S=2}(i, i+1)$.

The projection operator can be written by inspecting the eigenvalues of the operator $\hat{X} = (\mathbf{S}_i + \mathbf{S}_{i+1})^2$, $S(S+1) = 0, 2, 6$ for $S = 0, S = 1, S = 2$ subspaces respectively. Note that the eigenspaces of the operator \hat{X} , are shared with that of the projection operators. Therefore, we can write

$$\hat{P}_{S=2}(i, i+1) = \frac{1}{24} \hat{X}(\hat{X} - 2) \\ = \frac{1}{2} \mathbf{S}_i \cdot \mathbf{S}_{i+1} + \frac{1}{6} (\mathbf{S}_i \cdot \mathbf{S}_{i+1})^2 + \frac{1}{3}, \quad (21)$$

where we have used the fact that $\hat{X} = \mathbf{S}_i^2 + \mathbf{S}_{i+1}^2 + 2\mathbf{S}_i \cdot \mathbf{S}_{i+1} = 4 + 2\mathbf{S}_i \cdot \mathbf{S}_{i+1}$. As a consequence of Eq. (21), we can write the AKLT Hamiltonian Eq. (15) as

$$H_{\text{AKLT}} = 2J \sum_{i=1}^{N-1} \hat{P}_{S=2}(i, i+1) - \frac{2J}{3}(N-1) \quad (22)$$

An interesting observation from Eq. (22) is that the AKLT state has no two adjacent spin-1s living in the $S = 2$ sector, so as to have the minimum energy. In the picture of the auxiliary spin-1/2 particles, this leads to the formation of singlet states ($S = 0$) between two spin-1/2 particles in the adjacent sites.

Additionally, the AKLT state can be expressed exactly as an MPS with the lowest non-trivial bond dimension $\chi = 2$ [45]:

$$|\Psi_{\text{AKLT}}(\alpha, \beta)\rangle = \sum_{\{\sigma\}} A_{\alpha}^{(\sigma_1)i_1} A_{i_1}^{(\sigma_2)i_2} \dots A_{i_{N-1}}^{(\sigma_N)\beta} |\sigma_1, \sigma_2, \dots, \sigma_N\rangle, \quad (23)$$

where repeated indices imply a summation, $|\sigma_i\rangle$ denotes the local computational basis for the spin-1 particle at site i , with $\sigma_i = +, 0, -$, and

$$\mathbf{A}^{(+)} = \begin{bmatrix} 0 & \sqrt{\frac{2}{3}} \\ 0 & 0 \end{bmatrix}, \quad \mathbf{A}^{(0)} = \begin{bmatrix} -\frac{1}{\sqrt{3}} & 0 \\ 0 & \frac{1}{\sqrt{3}} \end{bmatrix}, \quad \mathbf{A}^{(-)} = \begin{bmatrix} 0 & 0 \\ -\sqrt{\frac{2}{3}} & 0 \end{bmatrix}. \quad (24)$$

α, β in Eq. (23) are the free indices, corresponding to the two auxiliary spin-1/2s at the boundary for the case of an open-boundary condition. The bond dimension of the required MPS to represent the AKLT state is just one unit higher than that for the product states (which can be represented as MPSs with bond dimension $\chi = 1$), making it the simplest entangled quantum state.

B. Details of the energy (VMC) and infidelity minimization

The evolution of the energy in the variational Monte Carlo (VMC) procedure, for various phases of the spin-1 bilinear-biquadratic (BLBQ) chain of length $L = 8$, is shown in Fig. 5(a)-(e). The details of the computations are given in table I. In Fig. 5(f), we plot the relative error in the energy of the NQS (w.r.t. the true ground state energy), after convergence of the VMC procedure, as a function of α .

We observe that the relative error in the energy (calculated exactly in Fig. 5(f)) saturates beyond a certain value of α for the four phases of the BLBQ model that we study in this paper. This corroborates with the infidelities of the optimized NQSs w.r.t. the exact ground states (dashed lines in Fig. 2(a)), for the VMC optimization, which also cease to improve beyond a certain α . Furthermore as we discussed in the main text, this apparent saturation in the accuracy of the NQS with the number of parameters of the ansatz also reappears when we perform an infidelity minimization exactly with the same ansatz.

L	Model	Symmetry of NQS	MC samples	Iterations	Initial states
8	AFH ($\theta = 0$)	global spin-flip	1000	5×10^4 (3×10^4 for $\alpha = 8$)	25 (100 for $\alpha = 8, 12$)
	AKLT ($\theta = \arctan(1/3)$)	global spin-flip	1000	5×10^4 (3×10^4 for $\alpha = 8$)	25 (100 for $\alpha = 8, 12$)
	critical ($\theta = \pi/4$)	NA	1500	5×10^4	25
	critical ($\theta = \arctan(2)$)	NA	1500	5×10^4	25
10	AFH ($\theta = 0$)	global spin-flip	1250	6×10^4 (5×10^4 for $\alpha = 12$)	25 (70 for $\alpha = 12$)
	AKLT ($\theta = \arctan(1/3)$)	global spin-flip	1250	6×10^4 (5×10^4 for $\alpha = 12$)	25 (100 for $\alpha = 12$)
	critical ($\theta = \pi/4$)	NA	1900	6×10^4	25 (100 for $\alpha = 12$)
	critical ($\theta = \arctan(2)$)	NA	1900	6×10^4	25 (94 for $\alpha = 12$)

TABLE I. Details of the VMC optimization for the spin-1 BLBQ chain (with open boundary condition) of length $L = 8, 10$. The NQS ansatz is given by the spin-1 RBM Eq. (8). The same values are used for the simulations with all values of α ($\alpha = 4, 8, 12, 16, 20$ for $L = 8$, and $\alpha = 4, 8, 12, 16, 20$ for $L = 10$) except when mentioned. All computations were done with the symmetry constraint $\sum_{i=1}^L S_{iz} = 0$. The learning rate for all optimizations was 1.5×10^{-3} .

We show the infidelity optimization curves in Fig. 6, for various phases of the spin-1 BLBQ chain with length $L = 10$. The details of the optimization are given in table II.

We impose a global spin-flip symmetry on the NQS ansätze for the AFH ($\theta = 0$) and the AKLT ($\theta = \arctan(1/3)$) states as prescribed in ref. [65]:

$$\psi_\theta(\sigma) = \frac{1}{2} [\psi_\theta(\sigma) + \psi_\theta(-\sigma)], \quad (25)$$

for both infidelity and energy minimization procedures. Here σ denotes a configuration of the spin-1s on the lattice.

L	Model	Symmetry of NQS	Iterations	Initial states
8, 10, 12	AFH ($\theta = 0$)	global spin-flip	1.2×10^5	10
	AKLT ($\theta = \arctan(1/3)$)	global spin-flip	1.2×10^5	10
	critical ($\theta = \pi/4$)	NA	1.2×10^5	10
	critical ($\theta = \arctan(2)$)	NA	1.2×10^5	10

TABLE II. Details of the infidelity minimization procedure for the spin-1 BLBQ chain (with open boundary condition) with lengths $L = 8, 10$, and 12 . The NQS ansatz is given by the spin-1 RBM Eq. (8). The values shown in the table are used for the simulations with all values of α ($\alpha = 2, 4, 8, 12, 16$ for $L = 8$, $\alpha = 4, 8, 12, 16, 24, 28, 32, 36$ for $L = 10$, and $\alpha = 4, 8, 12, 16$ for $L = 12$). Note that all computations were done by a full summation over the subspace of the Hilbert space defined by the symmetry constraint $\sum_{i=1}^L S_{iz} = 0$. During the infidelity minimization, first 3000 steps were performed with the ADAM algorithm [56] with a learning rate 5×10^{-4} , and the remaining minimization steps were performed with the YOGI algorithm [57] with a learning rate 3×10^{-4} .

C. Hessian of the infidelity loss function and its relation to QGT

The loss function in Eq. (9) (or Eq. (11)) is a scalar real valued function from \mathbb{C}^N to \mathbb{R} , $\mathcal{L} : \mathbb{C}^N \rightarrow \mathbb{R}$. It is straightforward to see that the loss function depends on both θ and θ^* , and hence is non-holomorphic. For convenience while working with complex derivatives of non-holomorphic functions, we use the following notations [66]:

$$\theta_c = \begin{bmatrix} \theta \\ \theta^* \end{bmatrix}, \quad \frac{\partial}{\partial \theta_c} = \left[\frac{\partial}{\partial \theta} \quad \frac{\partial}{\partial \theta^*} \right]; \quad \theta, \theta^* \in \mathbb{C}^N. \quad (26)$$

Then, we can write the complex Hessian of the loss function, following ref. [66], as

$$\mathbb{H} = \left(\frac{\partial}{\partial \theta_c} \right)^\dagger \frac{\partial \mathcal{L}}{\partial \theta_c} \quad (27)$$

$$= \begin{bmatrix} \frac{\partial^2 \mathcal{L}}{\partial \theta^* \partial \theta} & \frac{\partial^2 \mathcal{L}}{\partial \theta^* \partial \theta^*} \\ \frac{\partial^2 \mathcal{L}}{\partial \theta \partial \theta} & \frac{\partial^2 \mathcal{L}}{\partial \theta \partial \theta^*} \end{bmatrix}. \quad (28)$$

The generic elements of the blocks (1,1) and (1,2) in the above equation, when the loss function is given by the infidelity Eq. 9, are given by

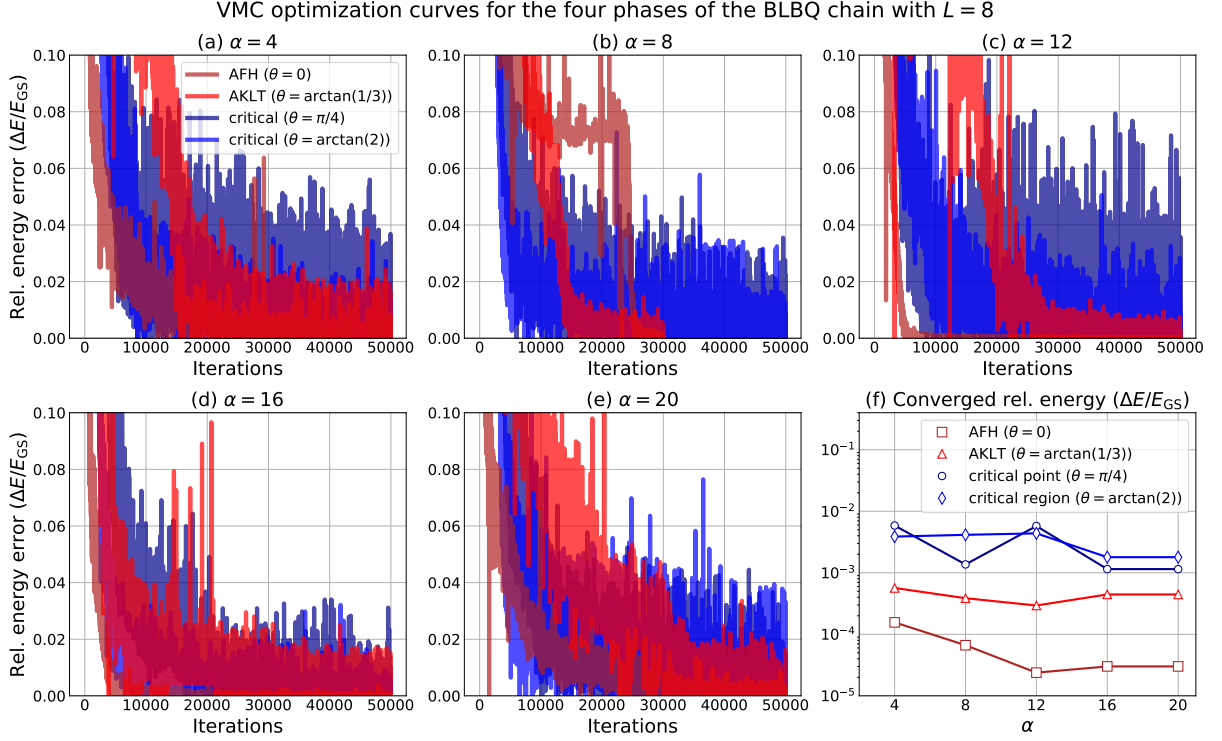


FIG. 5. This figure shows the evolution of the energy of the NQS during the VMC procedure for AFH, AKLT and two critical phases of the spin-1 BLBQ chain (with open boundary condition) with $L=8$ (a, b, c, d, e for $\alpha = 4, 8, 12, 16, 20$ respectively), and the relative errors in the converged energies with α (f). Note that the energy of the converged NQS is calculated exactly (by a full summation over the relevant subspace of the Hilbert space), to exclude biases from the Monte Carlo sampling. However, all quantities during the VMC optimization procedure are computed approximately by the Monte Carlo sampling procedure.

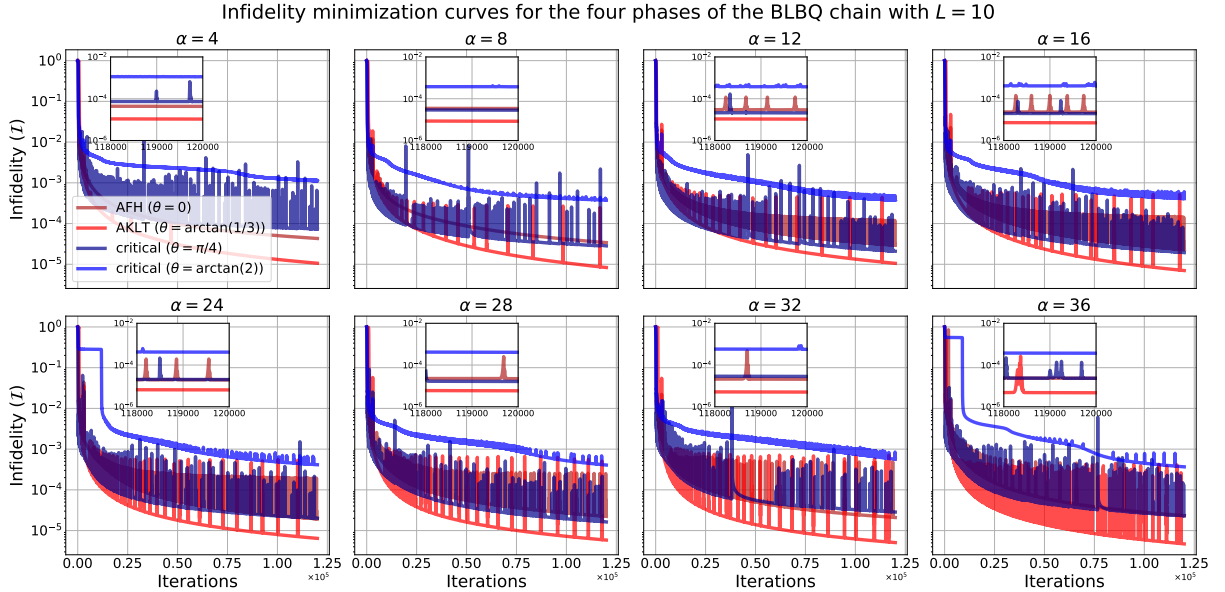


FIG. 6. This figure shows the evolution of the infidelity of the NQS w.r.t. the true ground state during the infidelity minimization procedure for the AFH, AKLT and two critical phases of the spin-1 BLBQ chain (with open boundary condition) with $L=10$. Each subplot shows the optimization curves for a different density (α) of the spin-1 RBM (Eq. (8)), and the corresponding insets show the optimization curves for the last 2000 iterations.

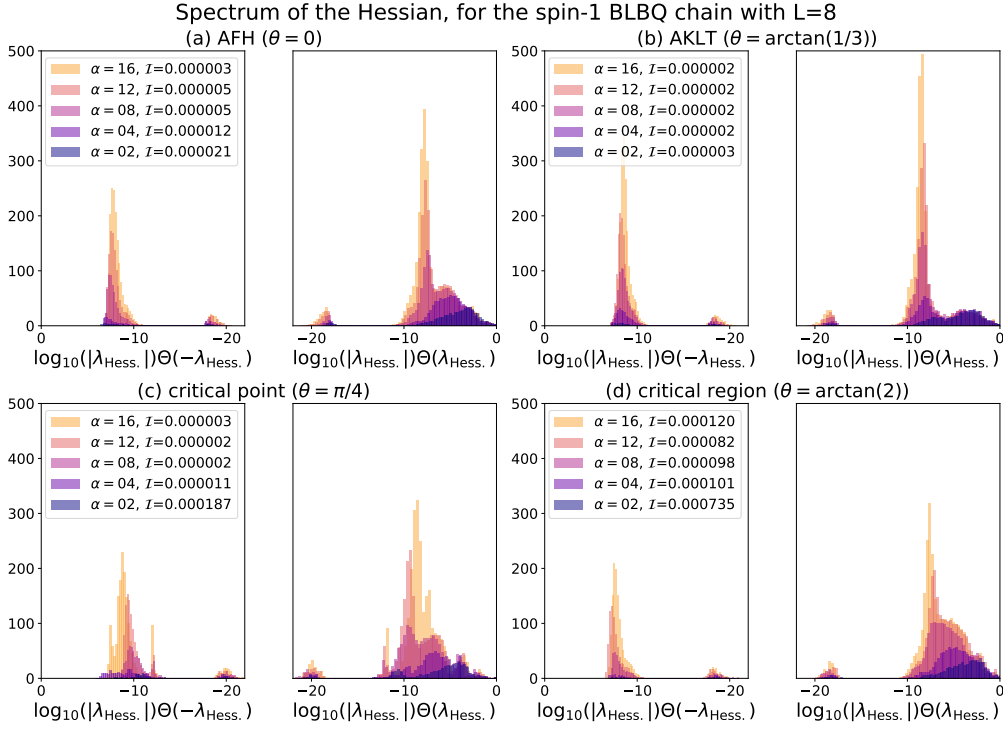


FIG. 7. This figure shows the distribution of the normalized eigenvalues (w.r.t. the maximum eigenvalue) of the Hessian, in \log_{10} scale, after 1.2×10^5 iterations of infidelity minimization on the spin-1 BLBQ chain (with open boundary condition) with length $L = 8$ for the four points (a, b, c, d) in the BLBQ phase diagram (see Fig. 1). The left panel of each subplot shows the distribution of (normalized magnitudes) of the negative eigenvalues, and the right panel of each subplot shows the distribution of the positive eigenvalues of the Hessian. In each subplot, the x-axis of the right panel increases from left to right, and that of the left panels increases from right to left.

$$\begin{aligned} \frac{\partial^2 \mathcal{L}}{\partial \theta_i^* \partial \theta_j} = & - \frac{\langle \frac{\partial \psi_\theta}{\partial \theta_i} | \Omega \rangle \langle \Omega | \frac{\partial \psi_\theta}{\partial \theta_j} \rangle}{\langle \psi_\theta | \psi_\theta \rangle \langle \Omega | \Omega \rangle} + \frac{\langle \psi_\theta | \Omega \rangle \langle \Omega | \frac{\partial \psi_\theta}{\partial \theta_j} \rangle}{\langle \psi_\theta | \psi_\theta \rangle^2 \langle \Omega | \Omega \rangle} \langle \frac{\partial \psi_\theta}{\partial \theta_i} | \psi_\theta \rangle + \frac{\langle \frac{\partial \psi_\theta}{\partial \theta_i} | \Omega \rangle \langle \Omega | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle^2 \langle \Omega | \Omega \rangle} \langle \psi_\theta | \frac{\partial \psi_\theta}{\partial \theta_j} \rangle \\ & + \frac{|\langle \psi_\theta | \Omega \rangle|^2}{\langle \psi_\theta | \psi_\theta \rangle^2 \langle \Omega | \Omega \rangle} \langle \frac{\partial \psi_\theta}{\partial \theta_i} | \frac{\partial \psi_\theta}{\partial \theta_j} \rangle - 2 \frac{|\langle \psi_\theta | \Omega \rangle|^2}{\langle \psi_\theta | \psi_\theta \rangle^3 \langle \Omega | \Omega \rangle} \langle \psi_\theta | \frac{\partial \psi_\theta}{\partial \theta_j} \rangle \langle \frac{\partial \psi_\theta}{\partial \theta_i} | \psi_\theta \rangle, \end{aligned} \quad (29)$$

$$\begin{aligned} \frac{\partial^2 \mathcal{L}}{\partial \theta_i^* \partial \theta_j^*} = & - \frac{\langle \frac{\partial^2 \psi_\theta}{\partial \theta_i \partial \theta_j} | \Omega \rangle \langle \Omega | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle \langle \Omega | \Omega \rangle} + \frac{\langle \frac{\partial \psi_\theta}{\partial \theta_j} | \Omega \rangle \langle \Omega | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle^2 \langle \Omega | \Omega \rangle} \langle \frac{\partial \psi_\theta}{\partial \theta_i} | \psi_\theta \rangle + \frac{\langle \frac{\partial \psi_\theta}{\partial \theta_i} | \Omega \rangle \langle \Omega | \psi_\theta \rangle}{\langle \psi_\theta | \psi_\theta \rangle \langle \Omega | \Omega \rangle} \langle \frac{\partial \psi_\theta}{\partial \theta_j} | \psi_\theta \rangle \\ & + \frac{|\langle \psi_\theta | \Omega \rangle|^2}{\langle \psi_\theta | \psi_\theta \rangle^2 \langle \Omega | \Omega \rangle} \langle \frac{\partial^2 \psi_\theta}{\partial \theta_i \partial \theta_j} | \psi_\theta \rangle - 2 \frac{|\langle \psi_\theta | \Omega \rangle|^2}{\langle \psi_\theta | \psi_\theta \rangle^3 \langle \Omega | \Omega \rangle} \langle \frac{\partial \psi_\theta}{\partial \theta_j} | \psi_\theta \rangle \langle \frac{\partial \psi_\theta}{\partial \theta_i} | \psi_\theta \rangle \end{aligned} \quad (30)$$

respectively. The blocks (1,1) and (2,2) are complex conjugates of each other, and are both Hermitian. The blocks (1,2) and (2,1) are Hermitian conjugates (as well as complex conjugates) of each other. As a result, the complex Hessian matrix Eq. (28) is Hermitian.

num of the infidelity landscape, i.e. when $|\Omega\rangle = |\psi_\theta\rangle$,

$$\frac{\partial^2 \mathcal{L}}{\partial \theta_i^* \partial \theta_j} = \frac{\langle \frac{\partial \psi_\theta}{\partial \theta_i} | \frac{\partial \psi_\theta}{\partial \theta_j} \rangle}{\langle \psi_\theta | \psi_\theta \rangle} - \frac{\langle \frac{\partial \psi_\theta}{\partial \theta_i} | \psi_\theta \rangle \langle \psi_\theta | \frac{\partial \psi_\theta}{\partial \theta_j} \rangle}{\langle \psi_\theta | \psi_\theta \rangle^2} = \mathbf{G}_{ij},$$

which is the quantum geometric tensor (QGT), and

$$\frac{\partial^2 \mathcal{L}}{\partial \theta_i^* \partial \theta_j^*} = \frac{\partial^2 \mathcal{L}}{\partial \theta_i \partial \theta_j} = 0.$$

It is interesting to note that when we are at the mini-

In this case, the Hessian becomes

$$\mathbb{H}|_{|\Omega\rangle=|\psi_\theta\rangle} = \begin{bmatrix} \mathbf{G} & \mathbf{0} \\ \mathbf{0} & \mathbf{G}^* \end{bmatrix}, \quad (31)$$

Hence, at the minimum of the infidelity landscape, the eigenvalues of the Hessian (of the infidelity loss function) are the same as that of the quantum geometric tensor (QGT), but with a degeneracy 2.

D. Spectra of the Hessian for infidelity minimization on the BLBQ chain with $L = 8$

We plot the spectra of the Hessian (Eq. (31)) as histograms in Fig. 7, at the end of the infidelity minimization for the four phases of the BLBQ chain (Fig. 1) with length $L = 8$. Eigenvalues are normalized w.r.t. the maximum eigenvalue. The positive and negative eigenvalues are shown separately in the subplots of Fig. 7 for clarity on the nature of the landscape around the solution.

We observe that the positive eigenvalues dominate around the converged NQSs, for all the phases that we studied. The magnitudes of the negative eigenvalues are smaller than the largest positive eigenvalue at most by a factor $\sim 10^{-6}$. This suggests that we have converged reasonably in a valley with steep positive curvatures along most directions (for all values of $\alpha = 2, 4, 8, 12, 16$), and only a few almost flat directions (with very small negative curvatures).

E. Spectra of the QGT for VMC computations on the BLBQ chain with $L = 8$

We plot the spectra of the QGT (Eq. (6)), as histograms in Fig. 8 at the end of the energy minimization (VMC) procedure for the four phases of the BLBQ chain (Fig. 1) with length $L = 8$, along with the cumulative distributions in the respective insets. We normalize the eigenvalues of the QGT w.r.t. the maximum eigenvalue for each case.

We observe that the distributions eventually converge to a limiting one as we increase α . As we further keep increasing α after the distributions converge, we only add eigenvalues which are essentially numerical zeros ($< 10^{-16}$), signifying redundant directions in the parameter space. This reiterates our observation from the QGT at the end of the infidelity minimization procedure (Fig. 4), that the NQS stops using the newly added parameters to represent the ground states more accurately. Furthermore, the distribution of the QGT converges at the smallest α for the AKLT state, as we also saw from the infidelity minimization results.

We also plot the rank of the QGT in Fig. 9 for the energy minimization results on the BLBQ chain with $L = 8$, and $L = 10$. The rank of the QGT for all phases at $L = 8$, saturates as expected from the spectra in Fig. 8, confirming the presence of redundant directions in the parameter space.

Therefore, we conclude that the NQS ansatz given by the spin-1 RBM (Eq. (8)), does not efficiently use all of its parameters as we increase the width of the network, both in cases of energy and infidelity minimization.

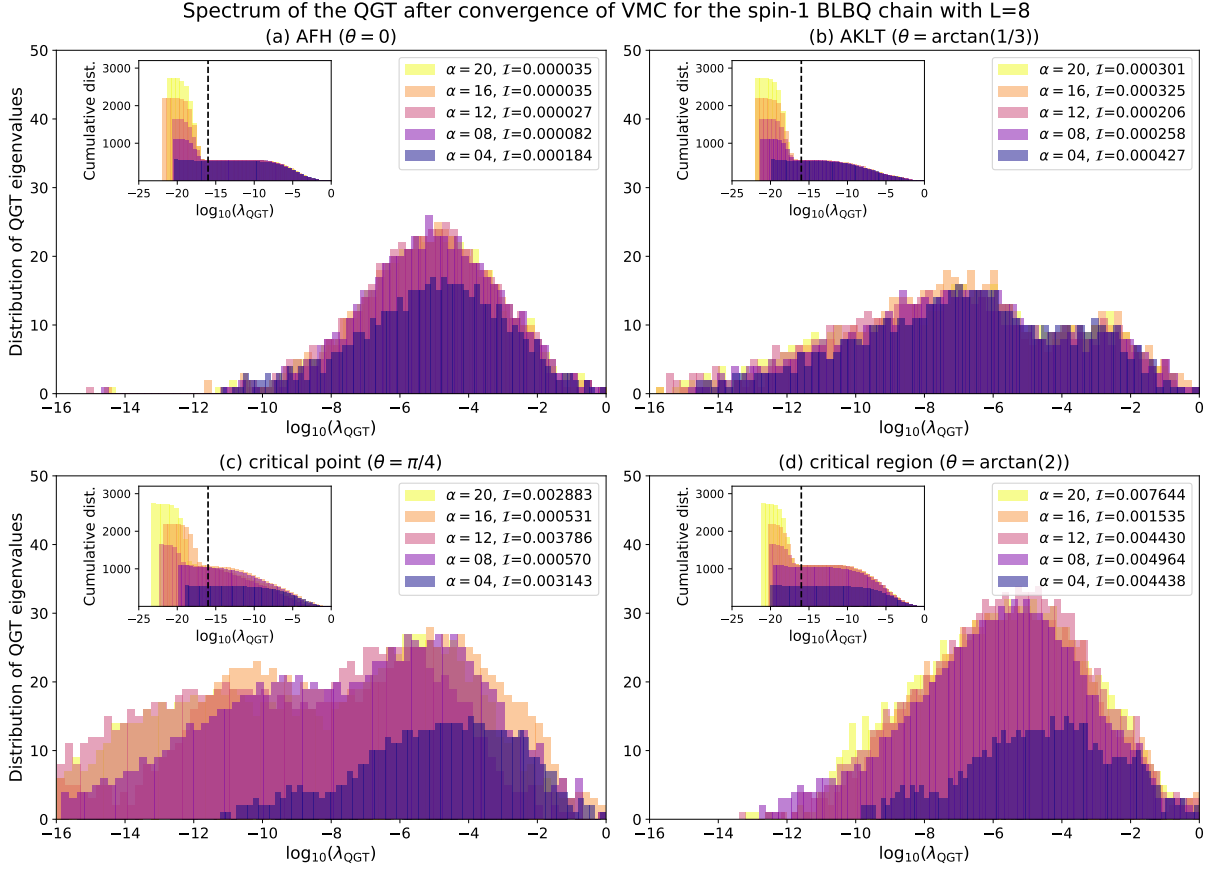


FIG. 8. This figure shows the distribution of the normalized eigenvalues (w.r.t. the maximum eigenvalue) of the QGT, in \log_{10} scale, after the convergence of the VMC procedure on the spin-1 BLBQ chain (with open boundary condition) with length $L = 8$ for the (a) AFH, (b) AKLT, and (c,d) two critical phases in the BLBQ phase diagram (see Fig. 1). Computations were done starting with at least 25 different initializations of the NQS ansatz, and the solution with the best infidelity was chosen. The plots show normalized eigenvalues with magnitudes greater than 10^{-16} , which is taken as a cutoff. The eigenvalues below this cutoff are considered numerical zeros. Note that while the VMC procedure involved MC sampling for computing the energies, gradients, and the QGT to implement the stochastic reconfiguration method, we compute the QGT at the end of the optimization exactly by a full summation over the relevant subspace of the Hilbert space. The insets show the cumulative distribution of the eigenvalues (no. of eigenvalues greater than the value on the x axis) for each case. The dashed line in the inset marks the cutoff 10^{-16} .

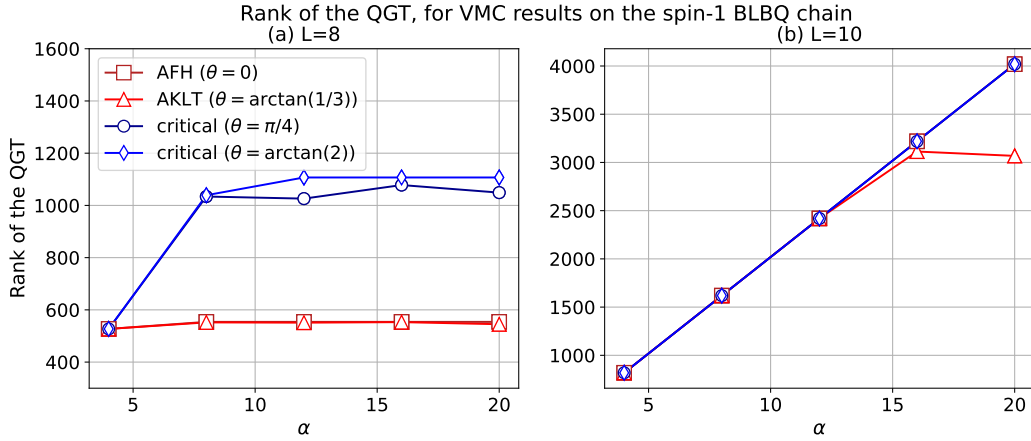


FIG. 9. This figure shows the rank of the QGT after the convergence of the VMC procedure on the spin-1 BLBQ chain (with open boundary condition) with lengths (a) $L = 8$, and (b) $L = 10$ for the four phases in the BLBQ phase diagram (see Fig. 1). The rank is computed by using a cutoff of 10^{-16} in the eigenvalues of the QGT.

- [1] G. Carleo and M. Troyer, Solving the quantum many-body problem with artificial neural networks, *Science* **355**, 602 (2017).
- [2] Y. Nomura, A. S. Darmawan, Y. Yamaji, and M. Imada, Restricted boltzmann machine learning for solving strongly correlated quantum systems, *Physical Review B* **96**, 205152 (2017).
- [3] K. Choo, T. Neupert, and G. Carleo, Two-dimensional frustrated j_1 - j_2 model studied with neural network quantum states, *Physical Review B* **100**, 125124 (2019).
- [4] M. Hibat-Allah, M. Ganahl, L. E. Hayward, R. G. Melko, and J. Carrasquilla, Recurrent neural network wave functions, *Physical Review Research* **2**, 023358 (2020).
- [5] F. Vicentini, A. Biella, N. Regnault, and C. Ciuti, Variational neural-network ansatz for steady states in open quantum systems, *Phys. Rev. Lett.* **122**, 250503 (2019).
- [6] J. Robledo Moreno, G. Carleo, A. Georges, and J. Stokes, Fermionic wave functions from neural-network constrained hidden states, *Proceedings of the National Academy of Sciences* **119**, e2122059119 (2022).
- [7] A. Lovato, C. Adams, G. Carleo, and N. Rocco, Hidden-nucleons neural-network quantum states for the nuclear many-body problem, *Physical Review Research* **4**, 043178 (2022).
- [8] J. Kim, G. Pescia, B. Fore, J. Nys, G. Carleo, S. Gandolfi, M. Hjorth-Jensen, and A. Lovato, Neural-network quantum states for ultra-cold fermi gases, *arXiv preprint arXiv:2305.08831* (2023).
- [9] G. Cybenko, Approximation by superpositions of a sigmoidal function, *Mathematics of control, signals and systems* **2**, 303 (1989).
- [10] M. Leshno, V. Y. Lin, A. Pinkus, and S. Schocken, Multilayer feedforward networks with a nonpolynomial activation function can approximate any function, *Neural Networks* **6**, 861 (1993).
- [11] Z. Lu, H. Pu, F. Wang, Z. Hu, and L. Wang, The expressive power of neural networks: A view from the width, *Advances in neural information processing systems* **30** (2017).
- [12] K. Hornik, M. Stinchcombe, and H. White, Multilayer feedforward networks are universal approximators, *Neural networks* **2**, 359 (1989).
- [13] Y. Li and Y. Liang, Learning overparameterized neural networks via stochastic gradient descent on structured data, *Advances in neural information processing systems* **31** (2018).
- [14] Z. Allen-Zhu, Y. Li, and Z. Song, A convergence theory for deep learning via over-parameterization, in *International conference on machine learning* (PMLR, 2019) pp. 242–252.
- [15] M. Y. Pei and S. R. Clark, Neural-network quantum states for spin-1 systems: spin-basis and parameterization effects on compactness of representations, *Entropy* **23**, 879 (2021).
- [16] L. L. Viteritti, R. Rende, and F. Becca, Transformer variational wave functions for frustrated quantum spin systems, *Physical Review Letters* **130**, 236401 (2023).
- [17] A. Chen and M. Heyl, Efficient optimization of deep neural quantum states toward machine precision, *arXiv preprint arXiv:2302.01941* (2023).
- [18] M. Schmitt and M. Heyl, Quantum many-body dynamics in two dimensions with artificial neural networks, *Phys. Rev. Lett.* **125**, 100503 (2020).
- [19] F. Vicentini, R. Rossi, and G. Carleo, Positive-definite parametrization of mixed quantum states with deep neural networks, *arXiv preprint arXiv:2206.13488* (2022).
- [20] K. Fukumizu, A regularity condition of the information matrix of a multilayer perceptron network, *Neural networks* **9**, 871 (1996).
- [21] R. Karakida, S. Akaho, and S.-i. Amari, Universal statistics of fisher information in deep neural networks: Mean field approach, in *The 22nd International Conference on Artificial Intelligence and Statistics* (PMLR, 2019) pp. 1032–1041.
- [22] J. Xu, D. J. Hsu, and A. Maleki, Benefits of overparameterization with em, in *Advances in Neural Information Processing Systems*, Vol. 31, edited by S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett (Curran Associates, Inc., 2018).
- [23] M. Larocca, N. Ju, D. García-Martín, P. J. Coles, and M. Cerezo, Theory of overparameterization in quantum neural networks, *Nature Computational Science* **3**, 542–551 (2023).
- [24] M. B. Hastings, An area law for one-dimensional quantum systems, *Journal of statistical mechanics: theory and experiment* **2007**, P08024 (2007).
- [25] L. Tagliacozzo, T. R. de Oliveira, S. Iblisdir, and J. I. Latorre, Scaling of entanglement support for matrix product states, *Phys. Rev. B* **78**, 024410 (2008).
- [26] D.-L. Deng, X. Li, and S. D. Sarma, Quantum entanglement in neural network states, *Physical Review X* **7**, 021021 (2017).
- [27] Y. Levine, O. Sharir, N. Cohen, and A. Shashua, Quantum entanglement in deep learning architectures, *Physical review letters* **122**, 065301 (2019).
- [28] C. Gauvin-Ndiaye, J. Tindall, J. R. Moreno, and A. Georges, Mott transition and volume law entanglement with neural quantum states, *arXiv preprint arXiv:2311.05749* (2023).
- [29] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, Quantum natural gradient, *Quantum* **4**, 269 (2020).
- [30] C.-Y. Park and M. J. Kastoryano, Geometry of learning neural quantum states, *Physical Review Research* **2**, 023232 (2020).
- [31] T. Haug, K. Bharti, and M. Kim, Capacity and quantum geometry of parametrized quantum circuits, *PRX Quantum* **2**, 040309 (2021).
- [32] M. Medvidović and D. Sels, Variational quantum dynamics of two-dimensional rotor models, *PRX Quantum* **4**, 040302 (2023).
- [33] F. D. M. Haldane, Continuum dynamics of the 1-d heisenberg antiferromagnet: Identification with the $o(3)$ nonlinear sigma model, *Physics letters a* **93**, 464 (1983).
- [34] M. Den Nijs and K. Rommelse, Preroughening transitions in crystal surfaces and valence-bond phases in quantum spin chains, *Physical Review B* **40**, 4709 (1989).
- [35] A. Läuchli, G. Schmid, and S. Trebst, Spin nematics correlations in bilinear-biquadratic $s=1$ spin chains, *Physical Review B* **74**, 144426 (2006).
- [36] G. Uimin, One-dimensional problem for $s=1$ with modified antiferromagnetic hamiltonian, *Soviet Journal of Experimental and Theoretical Physics Letters* **12**, 225 (1970).
- [37] C. Lai, Lattice gas with nearest-neighbor interaction in one dimension with arbitrary statistics, *Journal of Mathematical Physics* **15**, 1675 (1974).
- [38] B. Sutherland, Model for a multicomponent quantum sys-

- tem, *Physical Review B* **12**, 3795 (1975).
- [39] C. Itoi and M.-H. Kato, Extended massless phase and the haldane phase in a spin-1 isotropic antiferromagnetic chain, *Physical Review B* **55**, 8295 (1997).
- [40] J. Parkinson, The $s=1$ quantum spin chain with pure biquadratic exchange, *Journal of Physics C: Solid State Physics* **21**, 3793 (1988).
- [41] M. N. Barber and M. T. Batchelor, Spectrum of the biquadratic spin-1 antiferromagnetic chain, *Physical Review B* **40**, 4621 (1989).
- [42] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki, Valence bond ground states in isotropic quantum antiferromagnets, *Communications in Mathematical Physics* **115**, 477 (1988).
- [43] U. Schollwöck, T. Jolicoeur, and T. Garel, Onset of incommensurability at the valence-bond-solid point in the $s=1$ quantum spin chain, *Physical Review B* **53**, 3304 (1996).
- [44] I. Affleck, T. Kennedy, E. H. Lieb, and H. Tasaki, Rigorous results on valence-bond ground states in antiferromagnets, *Condensed Matter Physics and Exactly Soluble Models: Selecta of Elliott H. Lieb*, 249 (2004).
- [45] U. Schollwöck, The density-matrix renormalization group in the age of matrix product states, *Annals of physics* **326**, 96 (2011).
- [46] Y.-W. Dai, A.-M. Chen, X.-J. Liu, and Y.-H. Su, The commensurate and incommensurate haldane phases for the spin-1 bilinear-biquadratic model, *Chinese Physics B* (2022).
- [47] G. Uimin and J. Lett, 12, 225 (1970); ck lai, *J. Math. Phys* **15**, 1675 (1974).
- [48] M. Binder and T. Barthel, Low-energy physics of isotropic spin-1 chains in the critical and haldane phases, *Physical Review B* **102**, 014447 (2020).
- [49] We compute the infidelity Eq. (3) of the NQS for the AKLT state w.r.t. the projection of the NQS into the degenerate subspace. In other words, $|\Omega\rangle = \langle\Omega_1|\psi_\theta\rangle|\Omega_1\rangle + \langle\Omega_2|\psi_\theta\rangle|\Omega_2\rangle$, where $\{|\Omega_1\rangle, |\Omega_2\rangle\}$ span the degenerate subspace.
- [50] A. Sinibaldi, C. Giuliani, G. Carleo, and F. Vicentini, Unbiasing time-dependent variational monte carlo by projected quantum evolution, *arXiv preprint arXiv:2305.14294* (2023).
- [51] Y. N. Dauphin, R. Pascanu, C. Gulcehre, K. Cho, S. Ganguli, and Y. Bengio, Identifying and attacking the saddle point problem in high-dimensional non-convex optimization, *Advances in neural information processing systems* **27** (2014).
- [52] S. Fort and A. Scherlis, The goldilocks zone: Towards better understanding of neural network loss landscapes, in *Proceedings of the aaai conference on artificial intelligence*, Vol. 33 (2019) pp. 3574–3581.
- [53] This expression holds for the case when the variational state $|\psi_\theta\rangle$ is normalized. In the general case, the QGT takes the form $G_{ij} = \frac{\langle \frac{\partial \psi_\theta}{\partial \theta_i} | \frac{\partial \psi_\theta}{\partial \theta_j} \rangle}{\langle \psi_\theta | \psi_\theta \rangle} - \frac{\langle \frac{\partial \psi_\theta}{\partial \theta_i} | \psi_\theta \rangle \langle \psi_\theta | \frac{\partial \psi_\theta}{\partial \theta_j} \rangle}{\langle \psi_\theta | \psi_\theta \rangle^2}$.
- [54] S.-B. Lin, Generalization and expressivity for deep nets, *IEEE Transactions on Neural Networks and Learning Systems* **30**, 1392 (2018).
- [55] C. Umrigar and C. Filippi, Energy and variance optimization of many-body wave functions, *Physical review letters* **94**, 150201 (2005).
- [56] D. P. Kingma and J. Ba, Adam: A method for stochastic optimization, *arXiv preprint arXiv:1412.6980* (2014).
- [57] M. Zaheer, S. Reddi, D. Sachan, S. Kale, and S. Kumar, Adaptive methods for nonconvex optimization, *Advances in neural information processing systems* **31** (2018).
- [58] S. Sorella, Green function monte carlo with stochastic reconfiguration, *Physical review letters* **80**, 4558 (1998).
- [59] S. Sorella, Wave function optimization in the variational monte carlo method, *Physical Review B* **71**, 241103 (2005).
- [60] M. Hessel, D. Budden, F. Viola, M. Rosca, E. Sezener, and T. Hennigan, Optax: composable gradient transformation and optimisation, in *jax!*, 2020, URL <http://github.com/deepmind/optax> **16** (2010).
- [61] G. Carleo, K. Choo, D. Hofmann, J. E. T. Smith, T. Westerhout, F. Alet, E. J. Davis, S. Efthymiou, I. Glasser, S.-H. Lin, M. Mauri, G. Mazzola, C. B. Mendl, E. van Nieuwenburg, O. O'Reilly, H. Théveniaut, G. Torlai, F. Vicentini, and A. Wietek, Netket: A machine learning toolkit for many-body quantum systems, *SoftwareX*, 100311 (2019).
- [62] F. Vicentini, D. Hofmann, A. Szabó, D. Wu, C. Roth, C. Giuliani, G. Pescia, J. Nys, V. Vargas-Calderón, N. Astrakhantsev, and G. Carleo, NetKet 3: Machine Learning Toolbox for Many-Body Quantum Systems, *SciPost Phys. Codebases*, 7 (2022).
- [63] D. Häfner and F. Vicentini, mpi4jax: Zero-copy mpi communication of jax arrays, *Journal of Open Source Software* **6**, 3419 (2021).
- [64] A. Sinibaldi and F. Vicentini, Netket fidelity package (2023).
- [65] Y. Nomura, Helping restricted boltzmann machines with quantum-state representation by restoring symmetry, *Journal of Physics: Condensed Matter* **33**, 174003 (2021).
- [66] K. Kreutz-Delgado, The complex gradient operator and the cr-calculus, *arXiv preprint arXiv:0906.4835*; see page 14, 35 (2009).