Editors' Suggestion

# Interpreting convolutional neural networks' low-dimensional approximation to quantum spin systems

Yilong Ju [1,*] Shah Saad Alam [2,3,*] Jonathan Minoff [2] Fabio Anselmi [4] Han Pu [2,†] and Ankit Patel [1,5,6,‡]

[1]Department of Computer Science, Rice University, Houston, Texas 77005, USA
[2]Department of Physics and Astronomy, and Smalley-Curl Institute, Rice University, Houston, Texas 77005, USA
[3]JILA, University of Colorado, Boulder, Colorado 80309, USA
[4]Department of Mathematics, Informatics and Geoscience, University of Trieste, Trieste, Italy
[5]Center for Neuroscience and Artificial Intelligence, Baylor College of Medicine, Houston, Texas 77005, USA
[6]Department of Electrical and Computer Engineering, Rice University, Houston, Texas 77005, USA

(Received 3 September 2024; accepted 20 November 2024; published 23 January 2025)

Convolutional neural networks (CNNs) have been employed along with variational Monte Carlo methods for finding the ground state of quantum many-body spin systems with great success. However, it remains uncertain how CNNs, with a model complexity that scales at most linearly with the number of particles, solve the "curse of dimensionality" and efficiently represent wavefunctions in exponentially large Hilbert spaces. In this work, we use methodologies from information theory, group theory and machine learning, to elucidate how CNN captures relevant physics of quantum systems. We connect CNNs to a class of restricted maximum entropy (MaxEnt) and entangled plaquette correlator product state (EP-CPS) models that approximate symmetry constrained classical correlations between subsystems. For the final part of the puzzle, inspired by similar analyses for matrix product states and tensor networks, we show that the CNNs rely on the spectrum of each subsystem's entanglement Hamiltonians as captured by the size of the convolutional filter. All put together, these allow CNNs to simulate exponential quantum wave functions using a model that scales at most linear in system size as well as provide clues into when CNNs might fail to simulate Hamiltonians. We incorporate our insights into a new training algorithm and demonstrate its improved efficiency, accuracy, and robustness. Finally, we use regression analysis to show how the CNNs solutions can be used to identify salient physical features of the system that are the most relevant to an efficient approximation. Our integrated approach can be extended to similarly analyzing other neural network architectures and quantum spin systems.

DOI: 10.1103/PhysRevResearch.7.013094

## I. INTRODUCTION

A central concern in the study of a quantum many-body system is to understand how macroscopic properties emerge from microscopic interparticle interactions. However, this is in general an extremely difficult question to answer due largely to the fact that the dimension of the quantum Hilbert space grows exponentially as the number of constituent particles increases. Ingenious numerical techniques have been developed to study certain classes of many-body systems. In recent years, convolutional neural networks (CNNs), augmented with quantum Monte Carlo methods, have arisen as a powerful class of variational ansatzes for numerically solving quantum spin systems with many particles [1–6]. CNNs have often provided rapid and quite accurate numerical approximations comparable to the traditional algorithms that exist in quantum physics. As a result, there has been a flurry of research to improve the performance of these models and to apply them to broader classes of quantum spin systems with different physical constraints. However, compared to more commonly used NN tools such as the restricted Boltzmann machine (RBM) [7,8], the nature and form of the approximation used by CNNs remain unclear, thus preventing the interpretation of solutions and their domain of applicability, not to mention the extraction of useful physical insights about the quantum systems under study.

In this paper, we take a crucial step in filling this gap. Specifically, we aim to give new insights into how even a simple, one-hidden-layer CNN provides a solution to a quantum spin problem. We show how physical features, such as symmetries of the quantum spin system, naturally manifest themselves in the final trained network and during the optimization dynamics. We analyze the constraints these symmetries place on the variational parameters, and we use these insights to construct a more efficient, accurate, and robust training algorithm for CNNs. To further understand why the CNN is so adept at approximating the system using linearly many parameters, we interpret the convolutional operation in

*These authors contributed equally to this work.
†Contact author: hpu@rice.edu
‡Contact author: abp4@rice.edu

*Published by the American Physical Society under the terms of the Creative Commons Attribution 4.0 International license. Further distribution of this work must maintain attribution to the author(s) and the published article's title, journal citation, and DOI.*

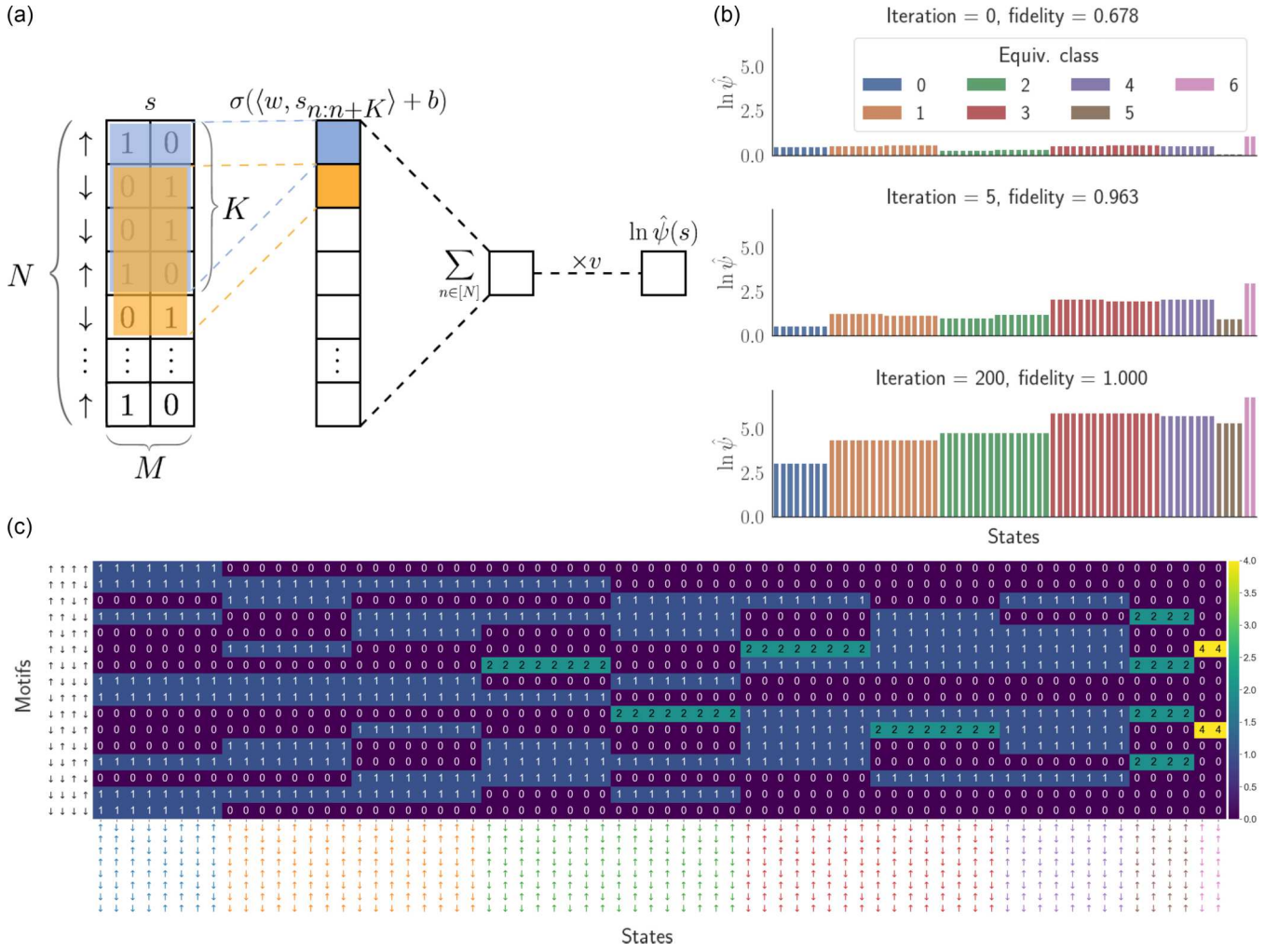


FIG. 1. (a) CNN architecture for a system of $N$ sites with $M = 2$ internal states, one convolutional layer with kernel size $K = 4$, a ReLU activation function $\sigma(\cdot)$, and a cyclic padding. (b) Learned $\ln\psi(s)$ and fidelity at different iterations. (c) Motif count matrix in the case $N = 8$, $M = 2$, and $K = 4$. All $M^K = 2^4 = 16$ motifs are labeled for each row and all $\binom{N}{N/M} = \binom{8}{4} = 70$ states for each column.

terms of its ability to capture the degrees of quantum entanglement using a CNN ansatz. We derive a mapping of the CNN to other statistical and physical ansatzes such as maximum entropy (MaxEnt) and correlator product states (CPS) [9–12], providing an interpretation of a CNN architecture in the context of quantum many-body physics. Interestingly, a multivariate regression analysis reveals which physical features are the most relevant to the low-dimensional learned solution and which ones the CNN captures correctly. Finally, we discuss how our approach and new insights can be used to design efficient approximations of complicated quantum spin systems.

## II. CONNECTING CNN TRAINING TO SYMMETRY LEARNING

We focus on the one-dimensional Sutherland model with periodic boundary conditions [13] and Hamiltonian $H = \sum_{n=1}^{N} P_{n,n+1}$, where $P_{n,n+1}$ is the operator exchanging the particles at positions $n$ and $n+1$, and the $N$ particles are evenly distributed among $M$ different species. For $M = 2$, this system reduces to the antiferromagnetic spin-1/2 Heisenberg model. We choose this Hamiltonian for two reasons. First, it is simple enough that we can benchmark the CNN's solution by comparing its energy to the exact value given by the Bethe ansatz [13]. Second, it is complex enough that the exact solution consists of $O(M^N)$ unique numbers, while we choose a CNN that has at most $O(N)$ variational parameters to work with. In order to succeed, CNNs must find a way to represent an approximation to the exact solution efficiently, and we seek to understand the nature of this approximation.

We start with a basic CNN parameterized by $(v, w, b) \in \mathbf{R} \times \mathbf{R}^{K\times M} \times \mathbf{R}$, which contains a single convolutional layer with one filter of kernel size $K$, followed by a fully connected layer [see Fig. 1(a)]:

$$\ln\psi^{\mathrm{CNN}}(s) = v\sum_{i=1}^{N}\sigma(w\cdot s_{i:i+K-1} + b),\ \forall s \in \mathcal{S}_{N,M}, \qquad (1)$$

where $s$ is a one-hot encoded input spin configuration (see 1(a) and $s_{i:i+K-1}$ is the substring of $s$ of length $K$ starting at index $i$. We call these substrings $K$ *motifs* or simply motifs. $\psi(s)$ is the wave function at $s$ and $\sigma$ is the ReLU nonlinearity. Since the Sutherland model does not allow for changes in total

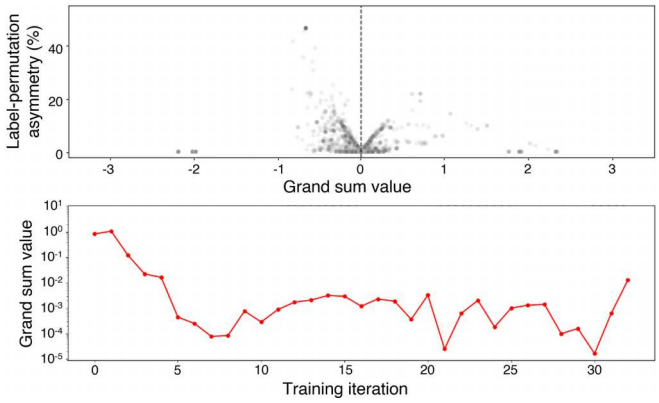


FIG. 2. (Top) Relative difference between $\ln \psi^{\mathrm{CNN}}(s)$ and $\ln \psi^{\mathrm{CNN}}(Ls)$, where $s = \uparrow\uparrow\uparrow\downarrow\downarrow\uparrow\downarrow\downarrow$, tracked throughout the CNN training. Multiple runs with different hyperparameters are included. (Bottom) The grand sum value at different training iterations. It starts with a relative large value $\approx 1$, and slowly stabilizes around a smaller number $\approx 1 \times 10^{-3}$. For both panels, $N = 8$.

magnetization, we have restricted our input spin configurations $s$ to have zero net magnetization. We also prove that a sufficiently large kernel size ($K \geqslant \lfloor N/3 \rfloor$) is necessary to assign unique wave-function values to each basis element, but not necessary for a good approximation. For more details, see Ref. [14].

Thus CNNs glean information indirectly through $K$ motifs. Global symmetries of the Hamiltonian can be learned by the CNN only through the statistics of observed occurrences of the $K$ motifs, which are visualized via a *motif count matrix* shown in Fig. 1(c) (the matrix concatenates the count vector of all motifs for all possible states [14]). As we will see later, motifs are the key to understanding why a low-dimensional approximation to the ground state exists and why the CNN is particularly suited for this task.

Visualizing the states and motifs also informs us how the symmetries of the problem appear within CNNs. By comparing the learning process of the CNN wavefunction shown in Fig. 1(b) with the input states shown in Fig. 1(c), we can see that a pattern emerges: the states that have similar $\ln \psi(s)$ are the ones that are connected to each other by a combination of symmetry operations of the Hamiltonian: translations, reflections around any point, and permutations of the spin labels. We call each set of connected $s$ an *equivalence class* [14]. Essentially, the CNN efficiently captures the symmetry constraints of the wavefunction after training. In particular, we can show that the symmetries constrain the CNN parameters via Theorem 1 (see proof in Ref. [14]).

*Theorem 1.* For systems with $M = 2$, if the grand sum condition holds, defined as $\mathrm{grandsum}(w) + 2b = 0$, then the CNN wave function has the label-permutation symmetry, i.e., $\ln \psi^{\mathrm{CNN}}(s) = \ln \psi^{\mathrm{CNN}}(Ls)$.

where $\mathrm{grandsum}(w)$ is the sum of all elements in a matrix $w$ and $Ls$ is the action of permuting the spin labels in $s$. We also use the term *grand sum value* to refer to $\mathrm{grandsum}(w) + 2b$. In Fig. 2 (top), we empirically verify that the grand sum value is indeed a sufficient condition for the label-permutation symmetry, as when the grand sum value is 0, the difference between $\ln \psi^{\mathrm{CNN}}(s)$ and $\ln \psi^{\mathrm{CNN}}(Ls)$ is also 0. Here we use $s = \uparrow\uparrow\uparrow\downarrow\downarrow\uparrow\downarrow\downarrow$ as an example. Figure 2 (bottom) confirms that CNN slowly learns to satisfy this condition by updating its parameters. Thus, if we could enforce the grand sum condition, the Hamiltonian symmetry would be even more respected.

Although this condition does not directly guarantee the reflection symmetry, for strings in certain equivalence classes [e.g., the orange and red classes in Fig. 1(c)], we note that applying a reflection is equivalent to first permuting the spin labels followed by a proper translation. Then, since the output of our CNN is shift-invariant by construction, imposing the grand sum condition ensures that the strings in these equivalence classes will also have reflection symmetry. Interestingly, as we will show later, enforcing the grand sum condition improves the CNN performance.

## III. CNN AS A CLASSICAL MAXIMUM ENTROPY ANSATZ

Given that the CNN is a fundamentally motif-based ansatz and is agnostic to everything else, it is then natural to compare the CNN to a classical maximum entropy (MaxEnt) ansatz constrained by the expectation values of the motif counting operators $m_{s'}(s)$, defined as the number of times that a motif $s'$ appears in a string $s$. By exploiting the positive definiteness of the ground state wave function of our model, we define the classical probability distribution $P(s) = |\psi(s)|^2$ without sacrificing any information about the wave function. Thus our quantum variational problem becomes a classical information theory problem of maximizing entropy with the constraint that the $K$-particle motif expectation values (MEVs) defined as $\langle m_{s'} \rangle = \sum_s P(s) m_{s'}(s)$ should match those of the ground state wavefunction, i.e., $\langle m_{s'} \rangle = \langle m_{s'} \rangle_{GS}$, $\forall s'$.

This results in a MaxEnt ansatz for our wave function [15]

$$\ln \psi^{\mathrm{MaxEnt}}(s) = \sum_{s'} \lambda_{s'} m_{s'}(s), \tag{2}$$

where each $\lambda_{s'}$ is a Lagrange multiplier associated with the constraint on $\langle m_{s'} \rangle$ (for full derivation, see Ref. [14]).

It is immediately apparent that our CNN ansatz [Eq. (1)] can be rewritten as a restricted classical MaxEnt ansatz constrained on $K$ marginals of contiguous substrings [Eq. (2)], by converting between a sum over motifs to a sum over positions

$$\ln \psi^{\mathrm{CNN}}(s) = \sum_{s'} \lambda_{s'}^{\mathrm{CNN}} m_{s'}(s) \tag{3}$$

with $\lambda_{s'}^{\mathrm{CNN}} = v\sigma(w \cdot s' + b)$. Both are exponential distributions, and both are agnostic to information about the full input $s$ other than the motif frequencies. The ReLU CNN is inherently a classical solution to a quantum problem. It treats the quantum wavefunction as a classical probability distribution, maximizing the *quantal entropy* [16].

A stronger connection between CNN and MaxEnt comes when we consider the effect of the Sutherland Hamiltonian's symmetries on the constraints. As derived in the next section, motifs $s', s''$ that are connected via reflection and label switching have the same $m_{s'} = m_{s''}$. We can thus group motifs into motif equivalence classes $\mathcal{M} := \bigcup_e M_e$. Motifs in the same equivalence class $M_e$ have the same MEV and so the MaxEnt constraints are then degenerate. Taking into account

TABLE I. Connections between CNN, CPS, and MaxEnt ansatzes. The equation unifying these three frameworks is given by $\ln \psi^{\text{ansatz}} = \sum_{s'} \mathcal{C}_{s'} m_{s'}(s)$.

| Method/Ansatz | CNN | CPS | MaxEnt |
|---|---|---|---|
| Field | machine learning | physics | statistics |
| Coefficients $\mathcal{C}_{s'}$ | $v\sigma(ws' + b)$ | $\ln \phi_{s'}$ | Lagrange multipliers $\lambda_{s'}$ |
| Functional Form | product and exponential | product | exponential |
| Training Goal | minimize energy | any | maximize entropy |
| Known Information | Hamiltonian | effective range of interactions | moment constraints and symmetries |
| Parameters | $v, w, b$ | $\phi$ | $\lambda$ |
| Hyperparameters | filter size $K$ | plaquette size $K$ | $K$-marginal distributions preserved |

such degeneracies, the *symmetric constraints* MaxEnt ansatz is (see Fig. S1 in Ref. [14])

$$\ln \psi^{\text{MaxEnt}}(s) = \sum_{e \in \mathcal{M}} \lambda_e m_e(s). \tag{4}$$

The CNN thus can be understood as a particular choice of the classical MaxEnt ansatz constrained over $K$ marginals where we have only $|\mathcal{M}| < M^K$ Lagrange multipliers $\lambda_e$, one for each equivalence class. In other words, we have "gauge" freedom in assigning the individual $\lambda_{s'}$ in $\lambda_e = \sum_{s' \in M_e} \lambda_{s'}$ for each equivalence class $M_e$. While the $\lambda_e$ are linearly independent, the individual $\lambda_{s'}$ are not [14]. The CNN becomes a particular gauge choice where, for label-switched pairs of motifs, one of the Lagrange multiplers is set to zero by the ReLU function. We thus have a connection between the MaxEnt parameters and the CNN's variational parameters.

## IV. CNN AS A CORRELATOR PRODUCT STATE (CPS) ANSATZ

Interestingly, another formal connection can be drawn between the CNN ansatz and the correlator product state ansatzes. The CPS ansatz has already widespread use in 1D and 2D spin systems, and its wave function is given by

$$\psi^{\text{CPS}}(s) = \prod_i \phi_{s_{i:i+K-1}}, \tag{5}$$

where $\phi_{s_{i:i+K-1}}$ is the wave function for $i$th plaquette. Our CNN then is a disguised CPS with correlator parameters $\phi_{s'}$ set to $e^{v\sigma(ws'+b)}$ and constrained to be real and positive. In this respect, the CPS ansatz can also be interpreted as a less constrained classical MaxEnt ansatz on the $K$-marginal distributions. This also reinforces that CPS ansatzes are different from purely quantum ansatzes.

The formal connections between the CNN, CPS, and MaxEnt wave functions are shown in Table I, where each ansatz is expressed as

$$\ln \psi = \sum_{s'} \mathcal{C}_{s'} m_{s'}(s). \tag{6}$$

A full discussion of the derivation and mapping is given in Ref. [14]. One prediction from our result is that the $\{\lambda_e\}$ are constrained to be the same for CNN and CPS. We confirm this by running a regression analysis of $\lambda_e^{\text{CNN}}$ using $\lambda_e^{\text{CPS}}$ as predictors on a CNN trained with $K = 3$ for an $N = 16$ system. We obtain a coefficient of $0.9770 \pm 0.002$ and $R^2 = 1.000$. Furthermore, we formally derive the slope and intercept of the learned model, which precisely match the empirical values. See Ref. [14] and Table S1 for details and more discussion regarding the case of $K = 6$.

## V. EFFICIENTLY APPROXIMATING THE MAXENT CONSTRAINTS

We now see how the CNN reduces the $O(M^N)$ problem to a solution in $O(M^K)$ space. We are still left with the mystery of how it further approximates the solution using only $O(K)$ training parameters. To investigate this, we look into where there is a lower dimensional approximation possible of the MEV constraints themselves.

Our first insight comes noting that the classical marginal probability distributions $m_{s'}$ are directly connected to quantum reduced density matrix $\rho_K = \sum_\alpha e^{-\alpha} |\alpha\rangle\langle\alpha|$ as [17]

$$\langle m_{s'} \rangle = \langle s' | \rho_k | s' \rangle = \sum_\alpha e^{-\alpha} |\langle s' | \alpha \rangle|^2. \tag{7}$$

The set $\{\alpha\}$ is known as the entanglement spectrum and has size $M^K$. However, if the underlying Hamiltonian has area law quantum entanglement (as for the Sutherland Hamiltonian, see also Fig. S2 in Ref. [14]), then the entanglement spectrum can be approximated statistically using a $O(K)$ size truncated set i.e., $\sum_{\alpha=1}^{M^K}(\cdot) \approx \sum_{\alpha}^{\alpha_c}(\cdot)$ where $\alpha_c \sim O(K)$. Thus, while the CNN is a classical MaxEnt solution to a quantum wave function, its success in simulating the MaxEnt using $O(K)$ parameters depends in part on the range of the quantum correlations.

To make this observation more explicit, we write $\rho_K = e^{-\beta H_K}/Z$, where $H_K$ is the effective entanglement Hamiltonian, $\beta$ the inverse entanglement temperature, and $Z = \text{Tr}[e^{-\beta H_K}]$ the partition function [18]. Calculating the symmetries of $H_K$ provides us the symmetry constraints on the MEVs we use to construct the motif symmetric MaxEnt in the previous section. In our case, $\beta$ and $H_K$ are obtainable using results from conformal field theory (CFT) [19], from which we can derive that the eigenvalues of $\rho_K$ decays sufficiently fast to allow the partition function to be efficiently represented by only $O(K)$ terms (see Fig. S1 in Ref. [14]), analogous to the truncation of the entanglement spectrum necessary for matrix product states [20–22]. This means that the expectation values of any observables (such as $m_{s'}(s)$) can be sufficiently approximated using only $O(K)$ values instead of $M^K$.

TABLE II. Regression Results for MEVs vs Physical Features of Interest. $R^2 = 0.786$, number of observations = 64, condition number = 42.4 ($*$ indicates significance at the 99.9% level).

| Variable | Intercept | $d_{\text{Neel}}$ | $n_{\text{like}}$ | $d_{\text{Neel}} \cdot n_{\text{like}}$ |
|---|---|---|---|---|
| Coeff. (std. error) | 8.95*(0.60) | −2.97*(0.35) | −3.60*(0.30) | 1.25*(0.13) |

We now explore if characteristic physical features correlate strongly with the MEVs, and whether a further lower-dimensional description of the MEVs exists. We perform this analysis using two physically-informed features based on the Hamiltonian: (1) $n_{\text{like}}(s')$, the number of pairs of adjacent like spins in motif $s'$, and (2) $d_{\text{Neel}}(s')$, the edit distance from each motif $s'$ to the closest Neel motif.

Both these metrics are rooted in the physical features of the system. (1) $n_{\text{like}}(m)$: the Heisenberg model eigenspectrum can be thought of as one that favors having unlike pairs in basis states and penalizes having like pairs. The Neel states with the highest $\ln \psi(s)$ are the states with the least number of like pairs, whereas the ferromagnetic states with the lowest $\ln \psi(s)$ have the highest number of like pairs. (2) $d_{\text{Neel}}(s')$: the Sutherland Hamiltonian can be thought of as generating swaps of adjacent spins at position $n$, and all basis states of the Sutherland system can be generated using iterative local swaps from the Neel states.

We focus on the $N = 60$ system and examine the CNN with a kernel size $K = 6$ with the best training hyperparameters, algorithm, and random seed, which gives $\delta_E = -2.27\%$, where $\delta_E = (\hat{E}_0 - E_0)/(E_1 - E_0)$ is the error between the predicted and actual ground state energy relative to the energy gap between the first excited state and the ground state. In Table II, we show our best regression model. The model has a high $R^2 = 0.786$, and shows that larger $n_{\text{like}}(s')$ and $d_{\text{Neel}}(s')$ lead to lower MEVs. This effect saturates since the coefficient for the interaction term is positive but relatively small. This regression analysis reveals that MEVs, the MaxEnt constraints which are crucial to the CNN ansatz, have a much simpler dependence on the physics of the system than anticipated and that an accurate low-dimensional approximation of the CNN exists.

## VI. EXPLOITING UNDERSTANDING TO IMPROVE CNN PERFORMANCE

In this section, we show how our novel understanding of the CNN ansatz can be exploited to design a better-performing architecture. Theorem 1 motivates us to directly impose symmetry into the CNN by enforcing the grand sum condition. We show that in this way, we can improve the accuracy, robustness to initial conditions, and training speed.

Based on the grand sum condition, we propose two symmetry-forcing algorithms: SYMFORCE-INIT, which enforces the grand sum condition only at initialization, and SYMFORCE-TRAJ, which enforces the grand sum condition throughout the entire learning trajectory (i.e., after each parameter update at every iteration, see Alg. S2 and Alg. S3 for details). Both are computationally efficient, simple to implement, and compatible with any training scheme since calculating the grand sum is just summing over $2K + 1$ parameters (typically $K \ll N$). Note that there are multiple ways to implement this constraint. Our algorithm sets $v \leftarrow v$, $w \leftarrow w - (\text{grandsum}(w) + 2b)/(2K)$ and $b \leftarrow b$ after updating $(v, w, b)$ at each iteration. In addition, we can prove that (see theorem S2) once the CNN has learned the symmetries, the update of $\psi(s)$ equals the update of $\psi(gs)$ for any state $s$ and transformations $g$ of interest. This means once the symmetry is learned, it will be retained throughout the rest of the training. Thus we can expect SYMFORCE-INIT to have a similar performance to SYMFORCE-TRAJ.

We adopt the variational Monte Carlo (VMC) learning framework from Ref. [23] and modify the CNN implemented in the paper to a shallow CNN. We focus on very large SU(2) systems, where $N \in \{60, 240\}$ and $M = 2$. We use both the shallow and deep CNNs as baselines, labeled as "Original" and "Deep ($L$ layers)," respectively. See Ref. [14] for more hyperparameter settings and tuning details.

We monitor the number of iterations until convergence $T_{\text{cv}}$ and the error between the predicted and actual ground state energy $\Delta_E = \hat{E}_0 - E_0$ for each experiment setting, averaging over five random initializations. $T_{\text{cv}}$ is defined as the first iteration when the relative change of the rolling average of $\hat{E}_0$ compared to that at the previous five iterations is smaller than 0.01%. Since we train the models for a maximum of 500 iterations, $T_{\text{cv}}$ is set to 500 if this criterion is never met.

Figure 3 shows that for both $N$ values, the proposed algorithms can indeed improve the CNN training. The top panel shows that the symmetry-forcing algorithms reduce $T_{\text{cv}}$ by roughly $1/3 \sim 1/2$ compared to the original training algorithm, achieving efficiency similar to that of deeper CNNs. Interestingly, the bottom panel shows that with our symmetry-forcing algorithms, the shallow CNNs can achieve the same level of accuracy as the deeper CNNs, even only using orders of magnitude fewer parameters, while still being robust to initial conditions and choice of hyperparameters. Without the SYMFORCE algorithms, deep networks become more sensitive to initial conditions and the choice of hyperparameters as we reduce the number of parameters. We can see this from the large error bars and the fact that only one set of hyperparameters works for CNNs with two layers when the number of parameters is $<10^2$. Also, it is more difficult to interpret deeper CNNs, since higher-level features tend to be more abstract. In addition, we observe that SYMFORCE-INIT and SYMFORCE-TRAJ have similar behavior in all aspects, which verifies theorem S2.

## VII. CONCLUSION

Our work adds to the interpretability of neural networks when employed to solve quantum systems, which joins recent investigations of the success of neural quantum states [24] and the role of quantum entanglement. Convolutional

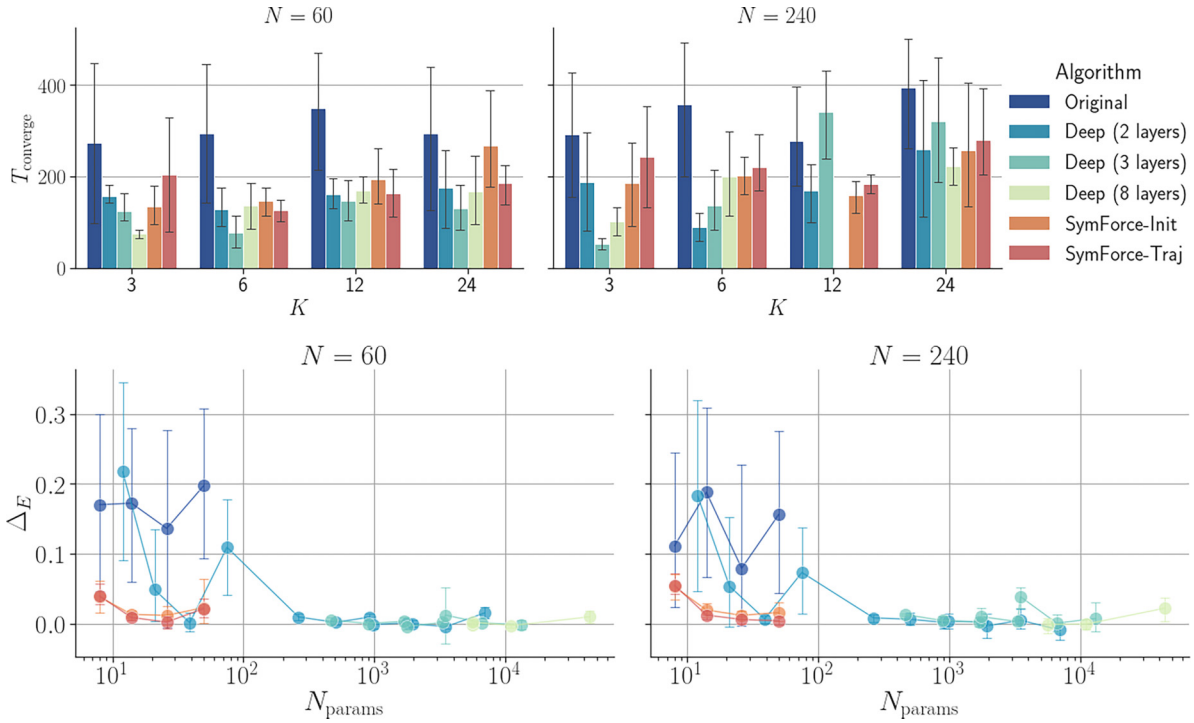


FIG. 3. (Top) The number of iterations until convergence $T_{\text{convergence}}$ vs kernel size $K$ for $N = 60$ and $N = 240$. (Bottom) The error between the predicted and actual ground state energy $\Delta_E = \hat{E}_0 - E_0$ vs the number of parameters in a CNN $N_{\text{params}}$ (in a logarithmic scale) for $N = 60$ and $N = 240$. We select the hyperparameters for each algorithm and $K \in \{3, 6, 12, 24\}$ corresponding to the minimum $\delta_E$, averaged over five random initializations. We use the same color map for the algorithms in both panels.

neural networks are essentially classical maximum entropy models that are constrained to replicate the correct marginal probability distributions of the underlying quantum wavefunction. We also show that correlator product states, a popular ansatz in quantum physics, can also be unified with CNNs and MaxEnts. CNNs' success in replicating the correct constraints using $O(K)$ parameters relies on the existence of lower-dimensional structure in the marginal probability constraints via area law entanglement and strong correlation with two physics-informed features rooted in the Hamiltonian. We develop a training algorithm that performs as well as deep CNNs by applying our understanding of the symmetries of the marginal probability distributions and how they constrain the CNN parameters. We believe that our analysis can be extended to other spin-1/2 systems and informs us of both the aptitude and limitations of CNNs when applied to quantum Hamiltonians.

## ACKNOWLEDGMENTS

H.P. acknowledges support from NSF (Grant No. PHY-2207283) and the Welch Foundation (Grant No. C-1669). S.S.A. acknowledges support from NSF (Grant No. PHY-2317149). A.B.P. and Y.J acknowledge support from NSF (Grant No. DBI-1707400) and NIH (Grant No. P42ES027725) F.A. and A.B.P. were supported by the Intelligence Advanced Research Projects Activity (IARPA) via Department of Interior/Interior Business Center (DoI/IBC) Contract No. D16PC00003. The US Government is authorized to reproduce and distribute reprints for governmental purposes notwithstanding any copyright annotation thereon. This work is also supported by the Lifelong Learning Machines (L2M) Program of the Defense Advanced Research Projects Agency (DARPA) via Contracts No. HR0011-18-2-0025 and No. R01 EY026927 to AT and by NSF NeuroNex Grant 1707400.

[1] X. Liang, S.-J. Dong, and L. He, Hybrid convolutional neural network and projected entangled pair states wave functions for quantum many-particle states, Phys. Rev. B **103**, 035138 (2021).

[2] 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).

[3] C. Roth and A. H. MacDonald, Group convolutional neural networks improve quantum state accuracy, arXiv:2104.05085.

[4] X. Liang, W.-Y. Liu, P.-Z. Lin, G.-C. Guo, Y.-S. Zhang, and L. He, Solving frustrated quantum many-particle models with convolutional neural networks, Phys. Rev. B **98**, 104426 (2018).

[5] H. Zheng, Z. Li, J. Liu, S. Strelchuk, and R. Kondor, Speeding up learning quantum states through group equivariant convolutional quantum Ansätze, PRX Quantum **4**, 020327 (2023).

[6] K. Choo, T. Neupert, and G. Carleo, Two-dimensional frustrated $J_1$-$J_2$ model studied with neural network quantum states, Phys. Rev. B **100**, 125124 (2019).

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

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

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

[10] H. J. Changlani, J. M. Kinder, C. J. Umrigar, and G. K.-L. Chan, Approximating strongly correlated wave functions with correlator product states, Phys. Rev. B **80**, 245116 (2009).

[11] 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).

[12] F. Mezzacapo, N. Schuch, M. Boninsegni, and J. I. Cirac, Ground-state properties of quantum many-body systems: Entangled-plaquette states and variational monte carlo, New J. Phys. **11**, 083026 (2009).

[13] B. Sutherland, Model for a multicomponent quantum system, Phys. Rev. B **12**, 3795 (1975).

[14] See Supplemental Materials at http://link.aps.org/supplemental/10.1103/PhysRevResearch.7.013094 for further information on training algorithm, derivation of MaxEnt ansatz with symmetries, definitions of motif count vectors and matrices, discussion on the need of nonlinear activation in CNN, and more.

[15] N. Canosa, A. Plastino, and R. Rossignoli, Ground-state wave functions and maximum entropy, Phys. Rev. A **40**, 519 (1989).

[16] N. Canosa, A. Plastino, and R. Rossignoli, Asymptotic density matrix for many-body ground states, Nucl. Phys. A **453**, 417 (1986).

[17] A consequence of this is that, for any symmetry of the entanglement Hamiltonian $H_K$, any two substrings $s'$ and $s''$ connected by the same symmetry fall into the same motif equivalence class and have $\langle m_{s'} \rangle = \langle m_{s''} \rangle$.

[18] H. Li and F. D. M. Haldane, Entanglement spectrum as a generalization of entanglement entropy: Identification of topological order in non-abelian fractional quantum hall effect states, Phys. Rev. Lett. **101**, 010504 (2008).

[19] T. Mendes-Santos, G. Giudici, M. Dalmonte, and M. A. Rajabpour, Entanglement hamiltonian of quantum critical chains and conformal field theories, Phys. Rev. B **100**, 155122 (2019).

[20] I. Peschel, M. Kaulke, and Ö. Legeza, Density-matrix spectra for integrable models, Annalen der Physik **511**, 153 (1999).

[21] D. Perez-Garcia, F. Verstraete, M. M. Wolf, and J. I. Cirac, Matrix product state representations, arXiv:quant-ph/0608197.

[22] A. Klümper, A. Schadschneider, and J. Zittartz, Matrix product ground states for one-dimensional spin-1 quantum antiferromagnets, Europhys. Lett. **24**, 293 (1993).

[23] L. Yang, Z. Leng, G. Yu, A. Patel, W.-J. Hu, and H. Pu, Deep learning-enhanced variational Monte Carlo method for quantum many-body physics, Phys. Rev. Res. **2**, 012039 (2020).

[24] Z.-A. Jia, B. Yi, R. Zhai, Y.-C. Wu, G.-C. Guo, and G.-P. Guo, Quantum neural network states: A brief review of methods and applications, Adv. Quantum Technol. **2**, 1800077 (2019).