Article

Neural-Network Quantum States for Spin-1 Systems: Spin-Basis and Parameterization Effects on Compactness of Representations

Michael Y. Pei *[ID] **and Stephen R. Clark** *[ID]

H.H. Wills Physics Laboratory, University of Bristol, Bristol BS8 1TL, UK
* Correspondence: michael.yuan.pei@bristol.ac.uk (M.Y.P.); stephen.clark@bristol.ac.uk (S.R.C.)

Abstract: Neural network quantum states (NQS) have been widely applied to spin-1/2 systems, where they have proven to be highly effective. The application to systems with larger on-site dimension, such as spin-1 or bosonic systems, has been explored less and predominantly using spin-1/2 Restricted Boltzmann Machines (RBMs) with a one-hot/unary encoding. Here, we propose a more direct generalization of RBMs for spin-1 that retains the key properties of the standard spin-1/2 RBM, specifically trivial product states representations, labeling freedom for the visible variables and gauge equivalence to the tensor network formulation. To test this new approach, we present variational Monte Carlo (VMC) calculations for the spin-1 anti-ferromagnetic Heisenberg (AFH) model and benchmark it against the one-hot/unary encoded RBM demonstrating that it achieves the same accuracy with substantially fewer variational parameters. Furthermore, we investigate how the hidden unit complexity of NQS depend on the local single-spin basis used. Exploiting the tensor network version of our RBM we construct an analytic NQS representation of the Affleck-Kennedy-Lieb-Tasaki (AKLT) state in the xyz spin-1 basis using only $M = 2N$ hidden units, compared to $M \sim O(N^2)$ required in the S^z basis. Additional VMC calculations provide strong evidence that the AKLT state in fact possesses an exact *compact* NQS representation in the xyz basis with only $M = N$ hidden units. These insights help to further unravel how to most effectively adapt the NQS framework for more complex quantum systems.

Keywords: neural-network quantum states; tensor network theory; restricted boltzmann machines

1. Introduction

The strongly-interacting quantum many-body problem is crucial to our understanding of many intriguing physical phenomena, but it is also inherently difficult to treat numerically owing to the exponential growth of the Hilbert space with system size. A commonly used approximate strategy is the variational method where a trial state, characterized by a tractable number of variational parameters, is optimized in energy. The effectiveness of this approach is highly dependent on the ansatz having an expressive form that can be systematically improved, to minimize bias, while also allowing relevant observables to be evaluated efficiently. Tensor networks have provided several examples of such ansatzes, with matrix product states (MPS) [1,2] displaying impressive accuracy in one-dimensional systems, along with projected entangled-pair states (PEPS) [1,3], tree tensor networks (TTN) [4,5], and multi-scale entanglement renormalization ansatz (MERA) [6] making two-dimensional systems accessible. Recently, artificial neural networks (ANNs) have emerged as another class of highly flexible variational ansatzes with many variants, such as restricted Boltzmann machines (RBM) [7], Deep Boltzmann Machines (DBM) [8–10], convolutional neural networks (CNN) [11–15], and feed-forward neural networks (FFNN) [16–19]. An important advantage of ANNs is that they are highly flexible and can be applied to any number of spatial dimensions, making them a powerful method for tackling the subtle physics seen in two-dimensional systems.

Although one of the simplest ANN variants, RBMs have seen widespread applications, including for open quantum systems [20–23], frustrated spin problems [24,25], quantum circuit simulation [26–28], and more. There are several reasons for their continued use. First, their simple structure allows for efficient sampling crucial for applying variational Monte Carlo (VMC) [29,30]. Second, RBMs are also a good candidate for a weakly biased ansatz, given that they are capable of exactly representing arbitrary states when their hidden unit number M scales exponentially with system size N. Third, RBMs are capable of representing states with volume-law entanglement [31], which further distinguishes them from tensor networks [32], despite their conceptual similarities [33–35]. Finally, there are also numerous classes of states with efficient exact RBM representations, including graph states [8]; spin Jastrow states, such as Laughlin states [33,36,37]; and general stabilizer states, such as the toric code [38–41], as well as more exotic hypergraph states and XS-stabilizer states [40]. Recently, we found that all but the last class listed, in fact, have RBM representations requiring $M < N$ hidden units [42], illustrating how even very modestly sized RBMs have significant representational power.

Despite their efficacy for spin-$\frac{1}{2}$ systems, the application of RBMs to systems with a local on-site dimension $d > 2$, such as spin-1 or bosonic systems, has been limited with convolutional or feedforward neural networks generally being favored [16,17,43]. The typical approach in machine learning to handle models with multinomial or categorical variables is so-called "one-hot" or "unary" encoding [44]. Rather than representing a physical degree of freedom directly with one visible unit this approach encodes the possible local physical states into a set of binary visible units. While this approach leverages the power of binary or spin-$\frac{1}{2}$ RBMs, it multiplies the number of visible units by a factor d, significantly increasing the parameter count and complexity of the optimization. As a consequence, studies utilizing unary encoding so far, for example, on the Bose Hubbard model [45,46], have been limited to small system sizes. Thus, there is need to devise more efficient RBM constructions tailored for $d > 2$ systems.

Progress has been made in this direction in recent work [47], where multivalued RBMs were applied directly to the one-dimensional spin-1 anti-ferromagnetic Heisenberg model (AFH) and substantially enhanced by incorporating a transformation to a coupled SU(2) symmetric basis. Complementary to this, here, we propose and study a direct generalization of the RBMs to spin-1 systems that retains key properties of spin-$\frac{1}{2}$ RBM with a minimal increase in variational parameters (as we will not be examining other network architectures, we will use RBM and NQS interchangeably in this paper). Specifically, the ability to describe arbitrary product states without hidden units, invariance of the parameterization to the values assigned to visible variables (labeling freedom), and equivalence to the tensor network formulation. This leads to the introduction of new quadratic bias and interaction weights in the RBM effective energy function. We demonstrate the effectiveness of the new formulation via VMC calculations for the spin-1 AFH model, where it is seen to deliver the same accuracy as unary encoding but with substantially fewer variational parameters. Additionally, we also investigate how the local single-spin basis affects the hidden unit complexity of a state by performing calculations in both the S^z and xyz spin-1 bases. For the AFH model with periodic boundary conditions, we find that the S^z is more accurate. A useful advantage of our new spin-1 RBM formulation is that it permits tensor network based analytic constructions. Focusing on the paradigmatic Affleck-Kennedy-Lieb-Tasaki (AKLT) model, we give explicit exact NQS representations in both the S^z and xyz bases. Our S^z basis NQS construction displays the expected [36] $M \sim O(N^2)$ scaling, while the simplification of the amplitude structure in the xyz basis gives an NQS construction with $M = 2N$ hidden units. By using VMC calculations, we find compelling evidence that the AKLT state, in fact, only requires $M = N$ hidden units to be represented exactly in the xyz basis.

The structure of this paper is as follows. We briefly outline the VMC method applied to the many body problem in Section 2 and discuss desirable properties for variational ansatzes. Next, in Section 3, we discuss the RBM in its spin-$\frac{1}{2}$ form and analyze its key

properties. In Section 4, we introduce a new generalization of NQS to spin-1 systems designed to mimic these properties and present VMC results for the AFH model. We then introduce analytic constructions for the AKLT model in Section 5, followed by VMC calculations in both the S^z and xyz basis. Finally, in Section 6, we conclude and discuss some open problems.

2. The Many-Body Problem and Variational Monte Carlo

2.1. Quantum Many-Body Problem

In this work, we will focus on physical systems composed of N localized spinful particles. Each particle is described by a vector of spin operators $\hat{\mathbf{S}}_j = (\hat{S}_j^x, \hat{S}_j^y, \hat{S}_j^z)$ for $j = 1, 2, \ldots, N$. Typically, the eigenstates $|S_j\rangle$ of $\hat{S}_j^z$ are used as the local spin basis, from which the S^z basis for the full system is constructed as $|S\rangle = |S_1\rangle \otimes \cdots \otimes |S_N\rangle$, where $S = (S_1, S_2, \ldots, S_N)$. Any many-body quantum state for these N spins can then be expanded in this basis as

$$|\Psi\rangle = \sum_S \Psi(S)|S\rangle, \tag{1}$$

via its complex amplitudes $\Psi(S)$. In the spin-$\frac{1}{2}$ case, we have $\hat{S}_j^\alpha = \frac{1}{2}\hat{\sigma}_j^\alpha$ (taking $\hbar = 1$ throughout) for $\alpha = \{x, y, z\}$ defined by the Pauli matrices, and $S_j \in \{+\frac{1}{2} \equiv \uparrow, -\frac{1}{2} \equiv \downarrow\}$. In the spin-1 case, we have

$$\hat{S}_j^x = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix}, \quad \hat{S}_j^y = \frac{1}{\sqrt{2}}\begin{bmatrix} 0 & -i & 0 \\ i & 0 & -i \\ 0 & i & 0 \end{bmatrix}, \quad \hat{S}_j^z = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & -1 \end{bmatrix},$$

with $S_j \in \{+1 \equiv \Uparrow, 0, -1 \equiv \Downarrow\}$.

A key challenge in many-body physics is to find the ground state $|\Psi_0\rangle$ of a system governed by an interacting Hamiltonian $\hat{H}$. In the context of spin systems, $\hat{H}$ will include terms that are products of the spin operators $\hat{\mathbf{S}}_j$ over two or more spins. Given that the expectation value of any observable $\hat{A}$ for a general (unnormalized) state $|\Psi\rangle$ is

$$\langle A \rangle_\Psi = \frac{\langle \Psi | \hat{A} | \Psi \rangle}{\langle \Psi | \Psi \rangle}, \tag{2}$$

the variational approach reformulates the eigenvalue problem $\hat{H}|\Psi_0\rangle = E_0|\Psi_0\rangle$ as the minimization of the energy $E = \langle \hat{H} \rangle_\Psi$ over the exponentially many amplitudes $\Psi(S)$. Performing this task exactly is only feasible for small systems $N \sim O(10)$ [48].

2.2. Variational Monte Carlo Method

A way to circumvent the "curse of dimensionality" is to instead to restrict the optimization over a specialized class of states $|\Psi_p\rangle$, dependent on parameters p whose number n_{params} scales polynomially with N. The variational principle $E_0 \leq E_{p_0} = \min_p \langle \hat{H} \rangle_{\Psi_p}$ then provides a route to finding the best approximation $|\Psi_{p_0}\rangle$ within the ansatz for $|\Psi_0\rangle$. The flexibility and utility of variational ansatz are greatly enhanced if instead of computing expectation values $\langle A \rangle_{\Psi_p}$ exactly we evaluate them approximately using Monte Carlo sampling. This approach is called *variational Monte Carlo* (VMC) and is described in detail in Appendix A. A key feature of VMC is that only ratios of amplitudes for an ansatz $\Psi_p(S)/\Psi_p(S')$ between different spin states $|S\rangle$ and $|S'\rangle$ are needed within the algorithm allowing us to ignore the normalization of quantum states in this work. To be efficient, thus, we require that these amplitude ratios for our ansatz can be evaluated with an effort scaling polynomially with N.

All numerical calculations presented in this paper employ the powerful stochastic reconfiguration method [30,49] for optimizing the parameters p. While ansatzes with a larger number of parameters may describe more varied amplitude structure, in principle allowing greater accuracy, the effort to optimize them can scale as $O(n_{\text{params}}^3)$. It is, therefore,

crucial that the functional form of any candidate ansatz $\left|\Psi_p\right\rangle$ is judicious in its use of variational parameters to avoid excessive redundancies.

3. Neural-Network Quantum States in Spin-1/2

In this section, we review NQS in terms of RBMs, discussing their form from both an energy function and tensor network perspective. In doing so, we highlight key properties of NQS that are desirable for expressiveness and the ability to represent important families of states.

3.1. Restricted Boltzmann Machine Approach

Restricted Boltzmann machines consist of two sets of units, N *visible* units representing the physical system, and M *hidden* units to be marginalized out. The units are characterized by a Boltzmann-like combined probability distribution $p(\boldsymbol{v}, \boldsymbol{h}) = \exp(-\mathcal{E}_\lambda)$ with an effective "energy" function

$$\mathcal{E}_\lambda(\boldsymbol{v}, \boldsymbol{h}) = -\sum_{j=1}^{N} a_j v_j - \sum_{i=1}^{M} b_i h_i - \sum_{i=1}^{M}\sum_{j=1}^{N} w_{ij} h_i v_j, \tag{3}$$

where $\boldsymbol{\lambda} = \{\boldsymbol{a}, \boldsymbol{b}, \boldsymbol{w}\}$ is the set of $MN + M + N$ complex parameters consisting of N visible biases $\boldsymbol{a} = (a_1, \cdots, a_N)$, M hidden biases $\boldsymbol{b} = (b_1, \cdots, b_M)$, and $M \times N$ hidden-visible interaction weights $\boldsymbol{w} = [w_{11}, \cdots, w_{1N}, \cdots, w_{MN}]$. While $p(\boldsymbol{v}, \boldsymbol{h})$ is Boltzmann once the hidden units are traced out, we are left with a more complex marginal distribution for the visible units, from which the NQS amplitudes are derived [7]:

$$\Psi_\lambda(\boldsymbol{S} \mapsto \boldsymbol{v}) = \sum_{\boldsymbol{h}} \exp[-\mathcal{E}_\lambda(\boldsymbol{v}, \boldsymbol{h})]. \tag{4}$$

The RBM architecture is shown diagrammatically in Figure 1.
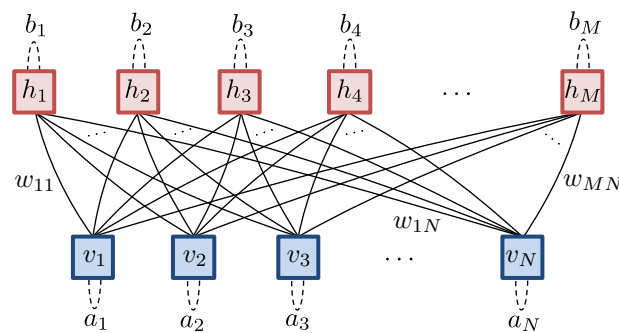

Figure 1. The bipartite graph of an RBM depicting interaction weights $\boldsymbol{w}$ as edges shown as solid arcs between hidden and visible units. For completeness, the biases $\boldsymbol{a}$ and $\boldsymbol{b}$ on each unit are depicted here as self-loop edges with dotted arcs to distinguish them from the interactions. The diagram here has been duplicated from ref. [42] and is presented here again for clarity.

Typically, visible v_j and hidden h_i unit variables are taken as two-valued. The pair of unique values $\{\mu, \nu\}$ taken by units within the energy function Equation (3) can be freely chosen, and they need not coincide with the eigenvalues $S_j \in \{+\frac{1}{2}, -\frac{1}{2}\}$ of the local operator $\hat{S}^z$ chosen to define the physical basis. In Equation (4), we emphasize this *labeling freedom* by explicitly introducing a mapping between physical and visible configuration $\boldsymbol{S} \mapsto \boldsymbol{v}$. Commonly, with NQS, an implicit choice $\boldsymbol{v} = \boldsymbol{S}$ is made, nullifying the need for this distinction, but it will prove to be useful here. Canonical choices v_j are number-like $\{0, 1\}$ or Ising-like $\{+1, -1\}$ values. We will adopt Ising-like visible variables, in which case it follows that an arbitrary visible unit taking values

$$S_j = \left\{ \begin{array}{ccc} \uparrow & \mapsto & \mu \\ \downarrow & \mapsto & \nu \end{array} \right\} = \tilde{v}_j \tag{5}$$

is generated by a shift and rescaling of $v_j \in \{+1, -1\}$ as $\tilde{v}_j = \frac{1}{2}(\mu + \nu) + \frac{1}{2}(\mu - \nu)v_j$. Since this is a linear transformation, it is entirely accommodated by redefining the energy function $\mathcal{E}_\lambda(v, h)$ parameters λ without changing its functional form. We shall see shortly that labeling freedom is a crucial ingredient for generalizing RBMs to spin-1 system using higher-dimensional visible units in Section 4. Tracing out Ising-like hidden variables results in the amplitude expansion [7]

$$\Psi_\lambda(\boldsymbol{S} \mapsto \boldsymbol{v}) = \prod_{j=1}^{N} e^{a_j v_j} \prod_{i=1}^{M} 2\cosh\left(b_i + \sum_{j=1}^{N} w_{ij} v_j\right), \tag{6}$$

that is commonly used and numerically convenient.

A key property of spin-$\frac{1}{2}$ RBMs is that they can represent any generic product state of the form

$$\left(c_+^{(1)} | \uparrow\rangle + c_-^{(1)} | \downarrow\rangle\right) \otimes \cdots \otimes \left(c_+^{(N)} | \uparrow\rangle + c_-^{(N)} | \downarrow\rangle\right), \tag{7}$$

without hidden units by simply setting their visible biases to

$$a_j = \frac{\ln\left(c_+^{(j)}\right) - \ln\left(c_-^{(j)}\right)}{\mu - \nu}. \tag{8}$$

Hidden units are, thus, necessary to describe entangled states; however, unlike tensor networks, there is no direct quantitative relation between them. Indeed, Ref. [31] reports that, for random NQS, adding hidden units actually decreases the amount of entanglement in a state. Nevertheless, increasing the number M of hidden units increases the expressiveness of the NQS, allowing more complex correlations within $\Psi_\lambda(\boldsymbol{S} \mapsto \boldsymbol{v})$ to be encoded. Formally, the NQS ansatz can represent any arbitrary state exactly in the limit of $M \to 2^N$ [50]. More usefully, there are important classes of states that possess highly accurate approximate or even exact NQS representations with an efficient scaling $M \sim \text{poly}(N)$ [33,36,38–41]. A particularly tractable NQS hidden unit complexity is typified by the following:

Definition 1 (Compact NQS). *States with an exact NQS representation where $M \leq N$ will be denoted as compact.*

As we demonstrated in ref. [42], important classes of state, including Jastrow, graph, and stabilizer states, all have compact NQS representations.
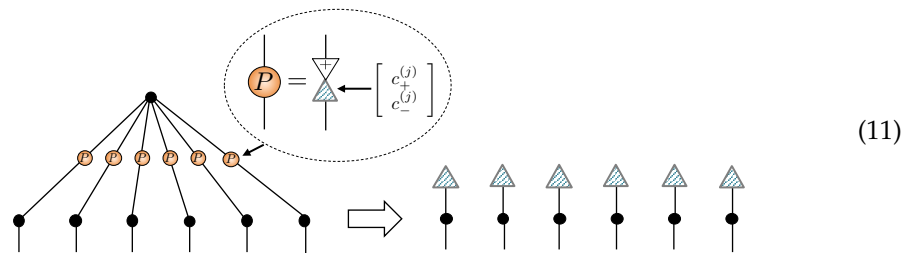
3.2. Tensor Network Approach

An alternative formulation of NQS views them as a tensor network. From this perspective, the amplitudes $\Psi(\boldsymbol{S})$ are recast as elements of an order-N tensor $\Psi_{S_1 S_2 \cdots S_N}$, and this structureless tensor can be decomposed into a set of lower order tensors contracted together [1,2]. Here, we will make repeated use of tensor network diagrams that form an important analytical tool. Here, generic tensors of any order are represented as shapes, most often a circle ∘ color shaded to guide the eye, with protruding legs for each index they possess. Contraction of tensors is then represented by the joining of legs via graphical equations like this:

$$\tag{9}$$

depicting $C_{abxz} = \sum_\alpha A_{ab\alpha} B_{x\alpha z}$. When there is a symbol inside a shape, it represents tensors with fixed elements or some specific structure. For example, the order-1 tensors with $\uparrow$ or $\downarrow$ symbols are the representation of the spin-$\frac{1}{2}$ S^z basis states $|\uparrow\rangle = \binom{1}{0}$ and $|\downarrow\rangle = \binom{0}{1}$, while

the tensor shown in Figure 2a represents $|+\rangle = |\uparrow\rangle + |\downarrow\rangle$. A special tensor we will use frequently is the COPY tensor [33,51], denoted by a dot $\bullet$, which is the multidimensional generalization of the 2×2 identity matrix. Its order-3 variant is shown in Figure 2b. It possesses a number of important properties. First, contracting any COPY tensor leg with $|+\rangle$ reduces the order of the COPY tensor by 1, as shown in Figure 2c. Second, contracting any COPY tensor leg with an S^z basis state factorizes it by duplicating the basis vector across all legs, as shown in Figure 2d, motivating its name. This property makes COPY tensors a useful glue for constructing sampleable tensor networks. An order-N COPY tensor can itself be viewed as a representation of a GHZ state, shown in Figure 2e. Finally, Figure 2f shows a generic diagonal matrix attached to a leg of a COPY tensor can be commuted across to any of the COPY dot's other legs.

Figure 2. Useful spin-$\frac{1}{2}$ tensor network diagrams. (**a**) The $|+\rangle$ state. (**b**) An order-3 COPY tensor. (**c**) Termination of a COPY tensor leg with the $|+\rangle$ state. (**d**) Copying of basis state inputs by an order-4 COPY tensor. (**e**) Expansion of an order-N COPY tensor into a GHZ state. (**f**) The commutativity of diagonal order-2 tensors across a COPY tensor [42].

With these concepts in place, the bipartite graph structure of RBM graph readily translates to a tensor network:

$$(10)$$

in which the vertices representing the visible and hidden units are replaced by COPY tensors, and order-2 tensors, or 2×2 coupling matrices $C^{(ij)}$, are introduced on each edge of the graph. Based on Figure 2e, thus, we can view each hidden unit in an NQS as contributing within an amplitude-wise product a locally deformed GHZ state. The NQS tensor network readily displays the key properties of the spin-$\frac{1}{2}$ RBM. They can trivially

represent product states in Equation (7) by using one hidden unit with rank-1 coupling matrices P containing the coefficients $\{c_+^{(j)}, c_-^{(j)}\}$ that subsequently factorizes out as

$$(11)$$

Moreover, since the elements of coupling matrices $C^{(ij)}$ have no functional dependence on any visible variables, they manifestly display labeling freedom. Crucially, while the NQS tensor network appears to have $4MN$ variational parameters, it exhibits significant gauge freedom due to the ability to reshuffle diagonal matrices across COPY tensors, as shown in Figure 2g. Consequently, most of the elements in the coupling matrices can be extracted and combined at the COPY tensors, reducing the number of independent parameters to $N + M + NM$, identical to the RBM formulation [42]. This equivalence allows either the NQS tensor network or RBM approach to be used interchangeably, and any direct generalization of RBMs to spin-1 should retain this feature.

4. Generalization to Spin-1 Systems

A tentative definition of a spin-1 RBM can be made by simply using the three-valued visible variables v_j with a natural Ising-like physical to visible mapping,

$$S_j = \left\{ \begin{array}{ccc} \Uparrow & \mapsto & +1 \\ 0 & \mapsto & 0 \\ \Downarrow & \mapsto & -1 \end{array} \right\} = v_j, \tag{12}$$

in the standard linear energy function Equation (3). By retaining two-valued Ising-like hidden units $h_i \in \{+1, -1\}$, the amplitudes $\Psi_\lambda(S \mapsto v) = \sum_h \exp[\mathcal{E}_\lambda(v, h)]$ continue to be given by Equation (6), but now admitting three-valued visible variables.

This spin-1 generalization of the RBM lacks the key properties of the spin-$\frac{1}{2}$ variant. First, $\Psi_\lambda(S \mapsto v)$ cannot easily describe a generic product spin state

$$\left(c_+^{(1)} | \Uparrow \rangle + c_0^{(1)} | 0 \rangle + c_-^{(1)} | \Downarrow \rangle \right) \otimes \cdots \otimes \left(c_+^{(N)} | \Uparrow \rangle + c_0^{(N)} | 0 \rangle + c_-^{(N)} | \Downarrow \rangle \right), \tag{13}$$

with coefficients $c_\alpha^{(j)}$, since the visible bias term $\exp(a_j v_j)$ cannot discriminate each of the local spin-1 states. This points to issues of expressiveness since we are using the same set of $MN + M + N$ parameters $\{a, b, w\}$ to describe a bigger state space. Second, the energy function Equation (3) does not possess labeling freedom for the visible units values. An arbitrary physical to visible mapping,

$$S_j = \left\{ \begin{array}{ccc} \Uparrow & \mapsto & \mu \\ 0 & \mapsto & \nu \\ \Downarrow & \mapsto & \omega \end{array} \right\} = \tilde{v}_j, \tag{14}$$

is generated from an Ising-like variable $v_j = \{+1, 0, -1\}$ via the quadratic transformation $\tilde{v}_j = \nu + \frac{1}{2}(\mu - \omega)v_j + \frac{1}{2}(\mu + \omega - 2\nu)v_j^2$. Consequently, transforming visible variables in this way cannot be accommodated in $\mathcal{E}_\lambda(v, h)$ by changing only the parameters $\{a, b, w\}$. One way to avoid these issues is to use "one-hot" or "unary" encoding which has been successfully applied to both spin-1 [40] and bosonic systems [45,46].

4.1. Unary Encoding Approach

Unary encoding applies the spin-$\frac{1}{2}$ RBM formalism of Equation (4) to a larger state space of each spin-1 by mapping the physical system into one comprising a larger number of spin-$\frac{1}{2}$ particles. Specifically, the local S^z basis of each spin-1 is mapped on to three spin-$\frac{1}{2}$'s as

$$|\Uparrow\rangle = |\uparrow\uparrow\downarrow\rangle, \quad |0\rangle = |\uparrow\downarrow\uparrow\rangle, \quad |\Downarrow\rangle = |\downarrow\uparrow\uparrow\rangle. \tag{15}$$

Physical states of the spin-1 system are now contained in the unary encoded subspace where only a single-excitation occurs within any three-site unit cell which facilitates the efficient projection of the representation. This naturally generalizes to a d-dimensional local state space.

By using this mapping, a spin-$\frac{1}{2}$ RBM can be applied inheriting its labeling freedom. Owing to the single-excitation projection, product states in Equation (13) are readily described by the visible biases $a_{j\mathrm{a}}, a_{j\mathrm{b}}$ and $a_{j\mathrm{c}}$ associated with the three spin-$\frac{1}{2}$'s {a,b,c} encoding a given spin-1. In general, for a d-dimensional local state space, unary encoding accounts for the enlarged state space by increasing the number of variational parameters to $n_{\mathrm{params}} = M + dN + dNM$ compared to the naive spin-1 RBM. However, there are some obvious deficiencies of this approach. First, unary encoding appears to unnecessarily enlarge the parameter count, as evidenced by the fact that it increases it even for $d = 2$. This will significantly increase the computational cost. Second, splitting the interaction weights w across a unary cell makes the interpretation of any individual hidden unit's contribution to the physical state rather opaque.

4.2. Defining a Spin-1 RBM and Tensor Network

These deficiencies show that a more general RBM energy function is needed that is sensitive to the three-valued nature of the visible units through the inclusion of terms involving the square of visible variables (a similar approach is likely to have been used in Ref. [47] already, although it was not explicitly stated). This motivates the following definition:

Definition 2 (spin-1 NQS). *We introduce a direct spin-1 RBM as the ansatz with amplitudes $\Psi_{\Lambda}(S \mapsto v) = \sum_h \exp[\mathcal{E}_{\Lambda}(v, h)]$ via the energy function*

$$\mathcal{E}_{\Lambda}(v, h) = \sum_{j=1}^{N} a_j v_j + \sum_{j=1}^{N} A_j v_j^2 + \sum_{i=1}^{M} b_i h_i + \sum_{i=1}^{M}\sum_{j=1}^{N} w_{ij} h_i v_j + \sum_{i=1}^{M}\sum_{j=1}^{N} W_{ij} h_i v_j^2, \tag{16}$$

defined by the parameters $\Lambda = \{a, b, w, W, A\}$.

The new contributions to a spin-1 RBM are W, an $M \times N$-dimensional matrix of quadratic interactions, and A, an N-dimensional vector of quadratic visible biases. There are now $2MN + 2N + M$ complex parameters in total. Tracing out the two-valued Ising-like hidden units gives amplitudes

$$\Psi_{\Lambda}(S \mapsto v) = \prod_{i=1}^{N} e^{a_i v_i + A_i v_i^2} \prod_{j=1}^{M} 2\cosh\left(b_j + \sum_{i=1}^{N} w_{ij} v_i + \sum_{i=1}^{N} W_{ij} v_i^2\right). \tag{17}$$

The inclusion of a quadratic visible bias now allows any product state in Equation (7) to be described without hidden units by setting

$$a_j = \log\left(\sqrt{c_+^{(j)} c_-^{(j)}}/c_0^{(j)}\right) \quad \text{and} \quad A_j = \log\left(\sqrt{c_+^{(j)}/c_-^{(j)}}\right),$$

while the quadratic interaction term ensures labeling freedom for the visible variable.

A strong justification for Equation (16) being the appropriate spin-1 generalization of RBMs is its relation to an NQS tensor network for spin-1. To handle spin-1 systems, we introduce a COPY tensor with three-dimensional legs copying the S^z basis states

$\{|\Uparrow\rangle, |0\rangle, |\Downarrow\rangle\}$. Its properties are summarized in Figure 3a–f and are straightforward generalizations of the spin-$\frac{1}{2}$ case given in Figure 2a–f and discussed in Section 3.2. The major difference is that we now distinguish three-dimensional legs with '=' lines instead of '−'. As before, the NQS tensor network follows from the conversion of the RBM graph in Figure 1, except that visible vertices are now replaced by three-dimensional COPY tensors giving:

$$ \tag{18} $$
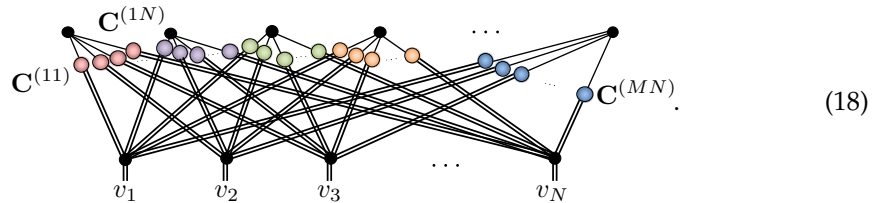
Owing to the mixed dimensionality of the COPY tensors in this network, it now requires a rectangular 2×3 coupling matrix $\mathbf{C}^{(ij)}$ between the i-th hidden and j-th visible unit. For the i-th hidden unit, its set of coupling matrices $\mathrm{Y}^{(i)} = \{\mathbf{C}^{(i1)}, \mathbf{C}^{(i2)}, \cdots, \mathbf{C}^{(iN)}\}$ can be explicitly tabulated as

$$ h_i \left\{ \begin{array}{c} +1 \\ -1 \end{array} \right. \quad \overbrace{\begin{array}{ccc} +1 & 0 & -1 \end{array}}^{v_1} \\ \begin{bmatrix} C_{++}^{(i1)} & C_{+0}^{(i1)} & C_{+-}^{(i1)} \\ C_{-+}^{(i1)} & C_{-0}^{(i1)} & C_{--}^{(i1)} \end{bmatrix} \quad \overbrace{\begin{bmatrix} C_{++}^{(i2)} & C_{+0}^{(i2)} & C_{+-}^{(i2)} \\ C_{-+}^{(i2)} & C_{-0}^{(i2)} & C_{--}^{(i2)} \end{bmatrix}}^{v_2} \quad \cdots \quad \overbrace{\begin{bmatrix} C_{++}^{(iN)} & C_{+0}^{(iN)} & C_{+-}^{(iN)} \\ C_{-+}^{(iN)} & C_{-0}^{(iN)} & C_{--}^{(iN)} \end{bmatrix}}^{v_N} $$

The amplitudes $\mathrm{Y}^{(i)}(v)$ of this hidden unit's correlator then follow by summing the product of coupling matrix elements selected by v along each row, giving

$$ \mathrm{Y}^{(i)}(v) = \sum_{h_i = \pm 1}^{1} \prod_{j=1}^{N} C_{h_j v_i}^{(ij)} = C_{+v_1}^{(i1)} C_{+v_2}^{(i2)} \cdots C_{+v_N}^{(iN)} + C_{-v_1}^{(i1)} C_{-v_2}^{(i2)} \cdots C_{-v_N}^{(iN)}. \tag{19} $$

The amplitudes of NQS tensor network are then the product of each of these hidden unit correlators

$$ \Psi_{\mathrm{NQS}}(S \to v) = \prod_{i=1}^{M} \mathrm{Y}^{(i)}(v), \tag{20} $$

which, like an RBM, can be exactly and efficiently sampled.

The spin-1 NQS tensor network appears to have $5MN$ complex variational parameters; however, again, gauge freedom allows the shuffling of diagonal matrices (of an appropriate dimension) through a COPY tensors reducing this. Specifically, its equivalence to the generalized spin-1 RBM proposed in Equation (16) is made using the coupling matrix decomposition

$$ \begin{bmatrix} C_{++}^{(ij)} & C_{+0}^{(ij)} & C_{+-}^{(ij)} \\ C_{-+}^{(ij)} & C_{-0}^{(ij)} & C_{--}^{(ij)} \end{bmatrix} = e^c \begin{bmatrix} e^{\tilde{b}_{ij}} & 0 \\ 0 & e^{-\tilde{b}_{ij}} \end{bmatrix} \begin{bmatrix} e^{w_{ij}+W_{ij}} & 1 & e^{-w_{ij}+W_{ij}} \\ e^{-w_{ij}-W_{ij}} & 1 & e^{w_{ij}-W_{ij}} \end{bmatrix} $$
$$ \times \begin{bmatrix} e^{\tilde{a}_{ij}+\tilde{A}_{ij}} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & e^{-\tilde{a}_{ij}+\tilde{A}_{ij}} \end{bmatrix}. $$

The solution for the weights w_{ij}, W_{ij} and partial biases $\tilde{a}_{ij}, \tilde{A}_{ij}, \tilde{b}_{ij}$ is outlined in Appendix B, from which the full RBM biases are formed as $a_j = \sum_{i=1}^{M} \tilde{a}_{ij}$, $A_j = \sum_{i=1}^{M} \tilde{A}_{ij}$ and $b_i = \sum_{j=1}^{N} \tilde{b}_{ij}$. The spin-1 NQS tensor network, thus, reduces to $n_{\mathrm{params}} = 2MN + 2N + M$ complex parameters.
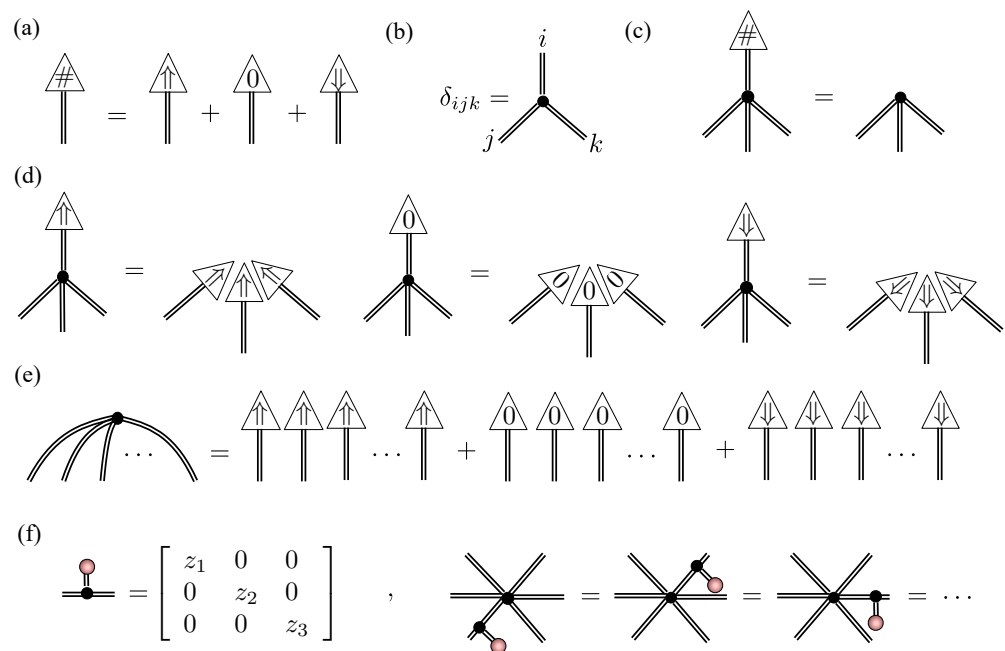
Figure 3. Useful spin-1 tensor network diagrams, directly generalizing the spin-$\frac{1}{2}$ properties outlined in Section 3.2. (**a**) The $|\#\rangle$ state. (**b**) An order-3 COPY tensor. (**c**) Termination of a COPY tensor leg with the $|\#\rangle$ state. (**d**) Copying of basis state inputs by an order-4 COPY tensor. (**e**) Expansion of an order-N COPY tensor into a spin-1 GHZ state. (**f**) The commutativity of diagonal order-2 tensors across a COPY tensor.

This correspondence between the spin-1 RBM and the spin-1 NQS tensor network highlights an advantage over unary encoding. Specifically, coupling matrices provide an intuitive tool for engineering the correlations and structures that a given hidden unit imprints on the amplitudes of an NQS. A trivial case is when all the elements of the j-th coupling matrix are 1's, denoted generically as **I**, equivalent to the hidden unit being disconnected from that visible unit. A more complex example with conditional correlations is a hidden unit with coupling matrices

$$\{\mathbf{I}, \cdots, \mathbf{I}, \overbrace{\mathbf{C}_{c\Uparrow}}^{k}, \mathbf{C}_{-0\Downarrow}, \cdots, \mathbf{C}_{-0\Downarrow}\},$$

built from

$$\mathbf{C}_{c\Uparrow} = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix} \quad \text{and} \quad \mathbf{C}_{-0\Downarrow} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \end{bmatrix}.$$

This hidden unit generates a correlator $Y(v) = \delta_{v_k,\Downarrow} + \delta_{v_k,\Uparrow}(-1)^{\mathcal{C}(v)}$, where a factor $(-1)^{\mathcal{C}(v)}$ is introduced conditional on the kth spin being in the state $|\Uparrow\rangle$, with $\mathcal{C}(v)$ being the number of $|0\rangle$ and $|\Downarrow\rangle$ states in the configuration v between spins $k+1$ and N. Similar types of hidden units will be used extensively in Section 5.1 to construct an exact representation of a state.

4.3. Projection of Unary Encoding into a Spin-1 RBM

The tensor network formalism provides further evidence that unary encoding from Section 4.1 is an over-parameterization of RBMs for spin-1 systems with $\delta n_{\mathrm{params}} = N(1+M)$ redundant parameters. Unary projection is implemented by an order-4 tensor $\mathcal{U}$ obeying:

$$(21)$$

By directly applying this projector to the spin-$\frac{1}{2}$ NQS tensor network for a unary encoded state and performing a graphical rewrite, we obtain the spin-1 NQS introduced in Equation (18). Figure 4 summarizes the crucial manipulations required.

Figure 4. (**a**) Representative examples of unary projection contractions with hidden units possessing different patterns of connectivity. (**b**) The first step in the contraction involves pulling the projection through the two-dimensional COPY tensors, leaving behind a single three-dimensional COPY tensor. (**c**) Connections to the unary spins are isolated by splitting the hidden unit COPY tensors. (**d**) The resulting blocks of the network are then contracted to obtain the 2×3 spin-1 NQS coupling matrices.

In Figure 4a, some representative examples of contractions between the projection tensor $\mathcal{U}$ and hidden units are shown. There are three steps to rewriting the network. The first step, shown in Figure 4b, essentially pulls $\mathcal{U}$ through the three unary two-dimensional COPY tensors, leaving behind a single three-dimensional COPY tensor representing the physical spin-1. Two important cases are shown in the example in Figure 4b. A hidden unit may have connections to each of the unary spin-$\frac{1}{2}$'s, where upon they get bundled up by the $\mathcal{U}$ tensor. A hidden unit may connect to only a subset of the unary spin-$\frac{1}{2}$'s, which is handled by plugging the unused legs of $\mathcal{U}$ tensor with $|+\rangle$. If a hidden unit couples to more than one of the unary spins, then the second step, shown in Figure 4c, involves splitting the hidden unit's two-dimensional COPY tensor to separate those connections. The final step is then to contract the split COPY tensor, coupling matrices and the projection $\mathcal{U}$ to form a rectangular 2×3 coupling matrix, as depicted in Figure 4d. If a hidden unit

has connections exclusively within the unary spin-$\frac{1}{2}$'s, then it becomes an entirely local visible bias contribution in the spin-1 NQS.

4.4. Change of Local Spin Basis

For tensor network representations, such as MPS or PEPS, the complexity (internal bond dimension) of a given state's description is rooted in its entanglement structure. As such changing the local basis of the spins used in a calculation has no effect on this complexity. Moreover, transforming a representation from one basis to another is accomplished by simply admixing the local tensors. Within VMC, a change of local spin basis leaves the locality and the sparsity of the Hamiltonian essentially unchanged. However, it is pivotal to the method that the amplitudes $\Psi_p(S)$ of whatever ansatz is used can be efficiently evaluated in this new basis. This is not generally true of NQS since their sampleability is intimately tied to the basis that factorizes the COPY tensors they are built from.

To understand how NQS behave, consider a representation of some state $|\Psi\rangle$, for instance, of four spin-1's sampleable in the S^z basis

$$|\Psi\rangle = \qquad\qquad\qquad . \qquad\qquad\qquad (22)$$
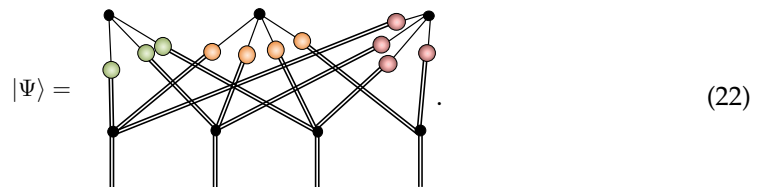
Now, suppose we transform the local S^z basis $|S\rangle$ to a new basis χ via a unitary $|\chi_S\rangle = \hat{\mathcal{B}}^\dagger|S\rangle$. Formally, from a tensor network perspective, we find the χ basis NQS representation by sandwiching $\mathbb{1} = \hat{\mathcal{B}}\hat{\mathcal{B}}^\dagger$ on each physical leg and computing the S^z basis NQS representation of $(\hat{\mathcal{B}}^\dagger)^{\otimes N}|\Psi\rangle$. However, currently, there is no known procedure for updating exactly an NQS after the application of an arbitrary single-spin unitary, even allowing for an increased number of hidden units. While we are guaranteed that an NQS representation of $\hat{\mathcal{B}}^{\otimes N}|\Psi\rangle$ exists, as illustrated here schematically

$$\qquad\qquad\qquad\qquad\qquad\qquad\qquad , \qquad (23)$$
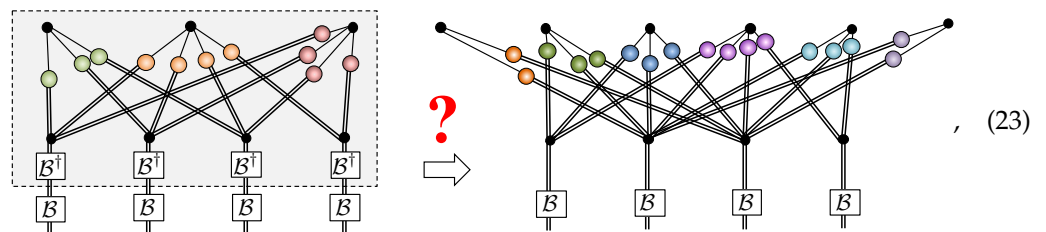
there is no guarantee it will be efficient, even if the representation of $|\Psi\rangle$ was originally. An example of such a catastrophic loss of NQS efficiency has been presented by Gao and Duan [8]. They show that, so long as the polynomial hierarchy in computational complexity theory does not collapse, a two-dimensional cluster state, which has an efficient NQS representation, has no efficient NQS representation after a specific layer of translation-invariant single-spin unitaries are applied. Thus, NQS complexity depends non-trivially on the local spin basis used.

Motivated by this, we will consider NQS calculations in two different local spin-1 bases to examine how the complexity varies, specifically, the standard S^z basis and the xyz basis defined as

$$|x\rangle = \frac{1}{\sqrt{2}}(|\Uparrow\rangle - |\Downarrow\rangle), \quad |y\rangle = \frac{i}{\sqrt{2}}(|\Uparrow\rangle + |\Downarrow\rangle), \quad |z\rangle = -|0\rangle.$$

In the *xyz* basis, the individual spin-1 operators all acquire the same off-diagonal form

$$
\hat{S}_j^x = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & i \\ 0 & -i & 0 \end{bmatrix}, \quad \hat{S}_j^y = \begin{bmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ -i & 0 & 0 \end{bmatrix}, \quad \hat{S}_j^z = \begin{bmatrix} 0 & i & 0 \\ -i & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix},
$$

as a consequence of them all contributing one eigenstate to the basis. On a practical level, using the *xyz* basis for the spin-1 RBM in Equation (17) simply requires replacing S with α and a physical-visible mapping from $\alpha \mapsto v$, such as

$$
\alpha_j = \left\{ \begin{array}{ccr} x & \mapsto & +1 \\ y & \mapsto & 0 \\ z & \mapsto & -1 \end{array} \right\} = v_j. \tag{24}
$$

Equivalently, the visible unit COPY tensors for this NQS tensor network can be considered to be rotated into this *xyz* basis.

4.5. Numerical Example—Spin-1 Anti-Ferromagnetic Heisenberg Model

To confirm the effectiveness of our spin-1 NQS, we performed VMC calculations to reach the ground state of the well-known anti-ferromagnetic Heisenberg (AFH) model in one dimension. The Hamiltonian is given by

$$
\hat{\mathcal{H}} = J \sum_i \hat{\mathbf{S}}_i \cdot \hat{\mathbf{S}}_{i+1}, \tag{25}
$$

where $J > 0$ is the magnetic interaction strength and with periodic boundary conditions $N + 1 \equiv 1$. We focus on small systems allowing direct comparison to the ground state calculated from exact diagonalization via the overlap $\mathcal{O} = |\langle \Psi_{\text{ED}} | \Psi_\Lambda \rangle|^2$. Additionally, we performed NQS optimizations in both the S^z and *xyz* bases to compare any differences in performance. This basis change alters the fixed quantum numbers of the system. In particular, in the S^z basis, the AFH model preserves the total S^z projection of the system, and its ground state lies in the $\sum_{j=1}^N S_j^z = 0$ sector of the full Hilbert space, while, in the *xyz* basis, the AFH model preserves the "parity" of the total x, y, z spin populations in the system, and, for even N, the ground state lies in the subspace of configurations, where there is an even number of each basis state. As is common for VMC calculations, we only select configurations from the relevant subspace during sampling. We perform the calculations with hidden unit numbers $M = [1, 2, \ldots, 2N]$, initializing $M = 1$ with random small complex parameters. For each successive calculation, we need the NQS with the parameters for $M - 1$ and initialize the Mth hidden unit with random small parameters, gradually increasing the size of the network in a sequential manner. To check the robustness against initialization bias of the qualitative features we have discussed, we rerun optimization sequences 5 to 10 times and present the best results here.

In Figure 5, we show how the accuracy of spin-1 NQS and unary encoding representations improve with an increasing number of hidden units M for both the S^z and *xyz* bases plotted in terms of the variational parameter count. The spin-1 NQS achieves a superior accuracy to unary encoding for a similar n_{params}. The inset of Figure 5a shows the collapse of the same S^z data plotted against M, indicating that the spin-1 NQS and unary encoding have in fact located the same solution for a given M. However, by using $\delta n_{\text{params}} = N(1 + M)$ less parameters, the spin-1 NQS is considerably more efficient to optimize, especially noting that δn_{params} scales with both system size and hidden unit number (for example, consider that the $N = 12, M = 12$ spin-1 NQS has a comparable parameter count to an $N = 12, M = 8$ unary NQS). In Figure 5b, we observe a noticeable drop in accuracy for both NQS variants in the *xyz* basis compared to the S^z basis. This suggests that the AFH ground state amplitude structure with periodic boundaries is inherently more complicated in the *xyz* basis regardless of encoding. Moreover, this confirms that hidden

unit number M of an NQS is basis-dependent quantity and cannot be used as a proxy of the entanglement.
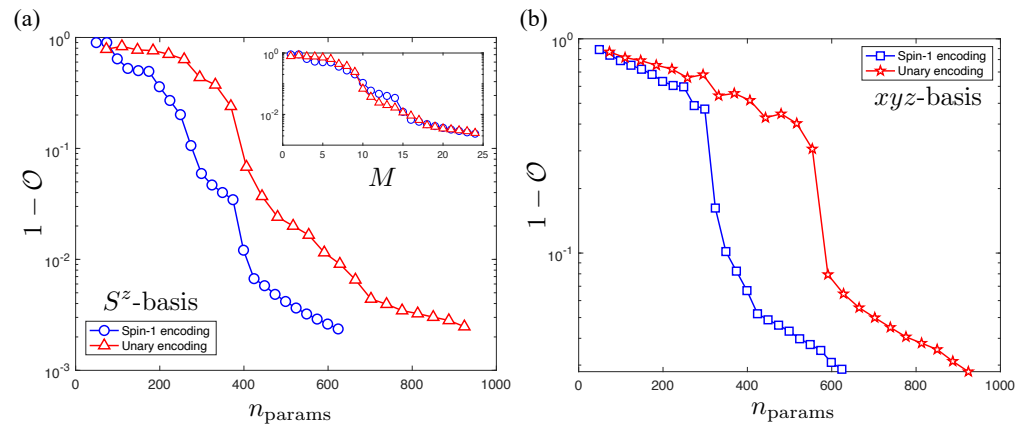

Figure 5. Plots of the infidelity of unary (red) and spin-1 NQS (blue) with the spin-1 AFH ground state with periodic boundary conditions. The calculations presented in both plots have the same number of sites $N = 12$, up to $M = 2N = 24$ hidden units. (**a**) The infidelity $1 - \mathcal{O}$ of the two NQS formulations for the S^z basis versus the RBM parameter count n_{params}. The inset shows the collapse of the same data when it is plotted in terms of the hidden unit number M. (**b**) The same calculations performed in the xyz basis.

5. Revisiting the AKLT Model

We now move on to benchmark our spin-1 NQS against the analytically solvable AKLT model [52], which is a spin-1 chain governed by a bilinear-biquadratic SU(2)-isotropic Heisenberg Hamiltonian of the form

$$\hat{H}_{\text{AKLT}} = \sum_{j=1}^{N} \left[\hat{\mathbf{S}}_j \cdot \hat{\mathbf{S}}_{j+1} + \beta (\hat{\mathbf{S}}_j \cdot \hat{\mathbf{S}}_{j+1})^2 \right] + \frac{2}{3},$$

with periodic boundary conditions $N + 1 \equiv 1$. It has special significance since it was the first solvable spin-1 chain model that exhibits the 'Haldane gap' [53]. The AKLT state $| \Psi_{\text{AKLT}} \rangle$ is the ground state of $\hat{H}_{\text{AKLT}}$ at the AKLT point $\beta = \frac{1}{3}$, and has an energy of exactly zero.

As is well-known, $| \Psi_{\text{AKLT}} \rangle$ has a special structure of correlations which are related to a valence bond solid. Specifically, each spin-1 is envisaged as being a pair of spin-$\frac{1}{2}$ particles that are correspondingly entangled in a singlet state with a partner spin-$\frac{1}{2}$ in the nearest neighboring spin-1 on the chain. The AKLT state is then the projection $\mathbb{P}$ of the local pairs of spin-$\frac{1}{2}$ particles into the triplet subspace, as depicted in Figure 6. This also leads to $| \Psi_{\text{AKLT}} \rangle$ possessing a very compact MPS representation with matrices

$$\mathbf{A}^+ = \frac{1}{\sqrt{2}} \hat{\sigma}^+, \quad \mathbf{A}^0 = -\frac{1}{2} \hat{\sigma}^z, \quad \mathbf{A}^- = -\frac{1}{\sqrt{2}} \hat{\sigma}^-,$$

where $\hat{\sigma}^\pm = \frac{1}{2}(\hat{\sigma}^x \pm i\hat{\sigma}^y)$, such that the (unnormalized) amplitudes of the ground state in the $\hat{S}^z$ basis follow as

$$\Psi_{\text{AKLT}}(\boldsymbol{S}) = \text{tr}\left(\mathbf{A}^{S_1} \mathbf{A}^{S_2} \cdots \mathbf{A}^{S_N} \right). \tag{26}$$

Since the AKLT point of $\hat{H}_{\text{AKLT}}$ lies in the gapped Haldane phase, $| \Psi_{\text{AKLT}} \rangle$ has finite-ranged magnetic correlations,

$$O_\ell^{zz} = \langle \psi_{\text{AKLT}} | \hat{S}_0^z \hat{S}_\ell^z | \psi_{\text{AKLT}} \rangle \sim e^{-\ell/\xi}, \quad \text{with} \quad \xi = \frac{1}{\ln(3)}, \tag{27}$$

yet it also has an unbroken spin rotation symmetry which is a hallmark of a symmetry protected topological order. Specifically, the string-order parameter

$$O^{\text{string}} = \lim_{N \to \infty} \lim_{\ell \to \infty} \langle \psi_{\text{AKLT}} | \hat{S}_0^z \prod_{j=1}^{\ell-1} e^{i\pi \hat{S}_j^z} \hat{S}_\ell^z | \psi_{\text{AKLT}} \rangle \sim -\frac{4}{9}, \tag{28}$$

reveals the presence of infinite-ranged anti-ferromagnetic correlations. This is evident from the structure of the MPS amplitudes. Any matrix product $\mathbf{A}^\pm \mathbf{A}^0 \cdots \mathbf{A}^0 \mathbf{A}^\pm = 0$, so any configuration containing a ferromagnetic segment, like "+ 0 0 0 +", with any number of 0's is not allowed. In contrast, allowed configurations contain only anti-ferromagnetic segments, such as "− 0 + 0 0 0 − + 0", arising from sequences, like $\mathbf{A}^\pm \mathbf{A}^0 \cdots \mathbf{A}^0 \mathbf{A}^\mp$, where every $\pm$ is partnered with a $\mp$ separated by with an arbitrary string of 0's.
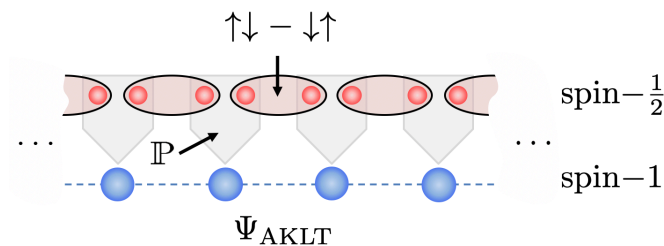
Figure 6. Valence bond solid construction of the spin-1 AKLT state Ψ_{AKLT} from the projection $\mathbb{P}$ of pairs of spin-$\frac{1}{2}$ particles shared between neighboring sites on the chain.

Despite its simple MPS representation it is surprisingly non-trivial to capture the non-commutative matrix products making up the AKLT amplitudes with an NQS. Direct conversion to an NQS from the MPS representation gives two reasons why it must contain long-ranged hidden units. First, it has been shown [38] that any short-ranged translationally invariant NQS cast into a MPS form by mapping hidden units into virtual bonds has $\mathbf{A}$ matrices that are at most rank-1. Since the matrix $\mathbf{A}^0$ in the AKLT state is rank-2, it fails this condition. Second, if we divide the chain into a sequence of three contiguous parts a, b, c, once we make b larger than the longest range of any hidden unit, so no hidden unit connects to visible units in both a and c, then the NQS amplitudes factorize as

$$\psi_{\text{sr−NQS}}(S_a, S_b, S_c) = \psi_{ab}(S_a, S_b)\psi_{bc}(S_b, S_c), \tag{29}$$

implying that S_a and S_c are uncorrelated [34]. The AKLT amplitudes $\Psi_{\text{AKLT}}(S)$ do not satisfy this property since region b can be any length of 0's, and there will always be non-zero amplitudes "+ 0 0 0 … 0 −" and "− 0 0 0 … 0 +" encoding string order correlations that are not factorizable.

It has been previously found that the AKLT state in the S^z basis requires an NQS with $M \sim O(N^2)$ long-ranged hidden units [36], and this was borne out in numerical calculations for small systems. In Appendix C, we explicitly construct an NQS for the S^z basis AKLT amplitudes using $M = 2N^2 + N\lfloor \frac{1}{2}(N-1) \rfloor + 1$ hidden units, many of which are extensive over the system. The $O(N^2)$ scaling can be readily understood as a consequence of having hidden units that each eliminate disallowed configurations, such as "$\pm$ 0 0 0 $\pm$", and impose the sign for allowed configurations, such as "$\pm$ 0 0 0 $\mp$", for all N separations and N translations over the system. This is rather less efficient than the compact spin-$\frac{1}{2}$ NQS found for Jastrow, graph and stabilizer states in ref. [42]. The AKLT state can be expressed with $O(N)$ hidden units but at the expense of needing a 2-layer DBM network [38] that cannot in general be exactly sampled, complicating its use numerically. However, as we saw for the AFH model numerical results, the hidden unit complexity is basis dependent. Surprisingly, we will show next that an exact $M \sim O(N)$ spin-1 NQS representation of the AKLT state is obtained in the *xyz* basis.

5.1. Exact Spin-1 NQS for AKLT State in the xyz Basis

The AKLT state provides an instructive example of how a single spin basis change can significantly alter the amplitude structure. Transforming the MPS representation into the *xyz* basis yields matrices

$$\mathbf{B}^x = \frac{1}{\sqrt{2}}(\mathbf{A}^+ - \mathbf{A}^{-1}) = \frac{1}{2}\hat{\sigma}^x, \quad \mathbf{B}^y = \frac{-i}{\sqrt{2}}(\mathbf{A}^{+1} + \mathbf{A}^{-1}) = \frac{1}{2}\hat{\sigma}^y, \quad \mathbf{B}^z = -\mathbf{A}^0 = \frac{1}{2}\hat{\sigma}^z,$$

and, thus, renders the amplitudes into products of Pauli matrices

$$\Psi_{\text{AKLT}}(\boldsymbol{\alpha}) = \text{tr}(\mathbf{B}^{\alpha_1}\mathbf{B}^{\alpha_2}\cdots\mathbf{B}^{\alpha_N}),$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_N)$ with $\alpha_j = \{x, y, z\}$ label the *xyz* basis. As expected, there is no change in the complexity/internal dimension of the MPS representation.

The structure of the amplitudes $\Psi_{\text{AKLT}}(\boldsymbol{\alpha})$ in the *xyz* basis is significantly simpler than $\Psi_{\text{AKLT}}(\boldsymbol{S})$ in the S^z basis. Amplitudes are now evaluated by tracking the anticommutations of Pauli matrices required to make the matrices of each the type form a contiguous sequence, e.g., $xx \cdots xyy \cdots yzz \cdots z$, and then reducing the product repeatedly via $(\hat{\sigma}^\alpha)^2 = \mathbb{1}$. The resulting matrix trace is non-zero only when the overall product is $\propto \mathbb{1}$, and so all non-zero amplitudes have an equal magnitude. Depending on whether N is even or odd, this condition requires that there is either an even or odd number of x, y and z's in any configuration string, respectively. Using this, we arrive at the following result:

Theorem 1 (AKLT state *xyz* NQS). *The AKLT state in the xyz spin basis has an exact spin-1 NQS representation requiring $M = 2N$ hidden units.*

Proof. We establish this result using a direct and intuitive construction for $\Psi_{\text{AKLT}}(\boldsymbol{\alpha})$ in which hidden units are devised to implement the nodal structure and sign structure of this state. The rules governing the amplitudes are as follows:

1. To implement the parity constraint on the number of x, y and z's in any configuration string $\boldsymbol{\alpha}$, we introduce the following 2×3 coupling matrices:

$$\mathbf{C}_{xy} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & -1 & 1 \end{bmatrix}, \quad \mathbf{C}_{yz} = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \end{bmatrix}.$$

    By defining two hidden units from these matrices as $Y_{xy} = \{\mathbf{C}_{xy}, \mathbf{C}_{xy}, \cdots, \mathbf{C}_{xy}\}$ and $Y_{yz} = \{\mathbf{C}_{yz}, \mathbf{C}_{yz}, \cdots, \mathbf{C}_{yz}\}$, we arrive at the product filter

$$Y_{xy}(\boldsymbol{\alpha})Y_{yz}(\boldsymbol{\alpha}) = \left(1 + (-1)^{\#\,x's\,+\,\#\,y's}\right)\left(1 + (-1)^{\#\,y's\,+\,\#\,z's}\right),$$

    in which the hidden units cancel out any strings $\boldsymbol{\alpha}$ that have odd numbers of both x's and y's, and y's and z's, respectively. Together, these hidden units completely establish for any N the nodal structure of the AKLT state amplitudes in this basis.

2. To reproduce the sign structure arising from anticommuting Pauli matrices into a contiguous sequence, we require two types of hidden units. The first type of hidden unit uses a conditional coupling matrix for the local state $|x\rangle$

$$\mathbf{C}_x = \begin{bmatrix} 0 & 1 & 1 \\ 1 & 0 & 0 \end{bmatrix},$$

    along with $\mathbf{C}_{yz}$ to define a hidden unit of the form

$$Y_{x(k)} = \{\mathbf{C}_{yz}, \cdots, \mathbf{C}_{yz}, \overbrace{\mathbf{C}_x}^{k}, \mathbf{I}, \cdots, \mathbf{I}\},$$

where $\mathbf{C}_x$ appears in the kth position in the sequence. The action of $Y_{x(k)}$ is to induce on a configuration a factor $(-1)^{\#y's \,+\, \#z's}$ between site $k-1$ and the left boundary, conditional on site k being in state $|x\rangle$. This is the sign that would occur if a σ^x matrix was anticommuted to this boundary through the corresponding product of Pauli matrices. The second type of hidden unit uses two further coupling matrices

$$\mathbf{C}_y = \begin{bmatrix} 1 & 1 & 1 \\ 1 & -1 & 1 \end{bmatrix}, \quad \mathbf{C}_z = \begin{bmatrix} 1 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix},$$

defining a hidden unit of the form

$$Y_{z(k)} = \{\mathbf{I}, \cdots, \mathbf{I}, \overbrace{\mathbf{C}_z}^{k}, \mathbf{C}_y, \cdots, \mathbf{C}_y\},$$

where $\mathbf{C}_z$ appears in the kth position in the sequence. The action of $Y_{z(k)}$ is to induce on a configuration a factor $(-1)^{\#y's}$ between site $k+1$ and the right boundary, conditional on site k being in state $|z\rangle$. This is the sign that would occur if a $\hat{\sigma}^z$ matrix was anticommuted to the right boundary through the corresponding product of Pauli matrices, assuming that any $\hat{\sigma}^x$'s have already been anticommuted to the left boundary. To capture all locations k for both types, thus, requires $2(N-1)$ hidden units which entirely establish the sign structure of the AKLT state amplitudes in this basis.

This gives a total of $M = 2N$ hidden units. $\square$

The resulting amplitude-wise product decomposition of the AKLT into hidden unit correlators is depicted Figure 7 for $N = 6$.
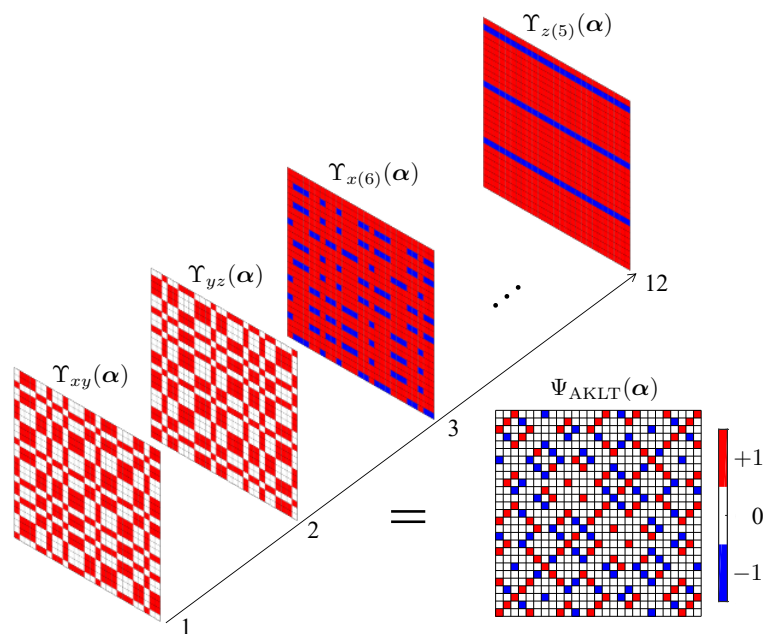
Figure 7. The 3^6 (unnormalized) amplitudes of the $N = 6$ AKLT state in the xyz basis $\Psi_{\mathrm{AKLT}}(\boldsymbol{\alpha})$ are displayed here as a $3^3 \times 3^3$ matrix with the color of an element designating which of the values $+1, 0, -1$ the amplitude has. The bottom left corner element corresponds to the amplitude of $|xxxxxx\rangle$, while the top right corner corresponds to $|zzzzzz\rangle$. The amplitudes $\Psi_{\mathrm{AKLT}}(\boldsymbol{\alpha})$ are reconstructed exactly by the product of $M = 2N = 12$ hidden unit filters given in the main text shown. The first two filters $Y_{xy}(\boldsymbol{\alpha})$ and $Y_{yz}(\boldsymbol{\alpha})$ establish the nodal structure, while the other ten filters $Y_{x(k)}(\boldsymbol{\alpha})$ and $Y_{z(k)}(\boldsymbol{\alpha})$ imprint the sign structure.

### 5.2. Analytic Example—AKLT Unary Stabilizer State

An explicit NQS construction for the AKLT state in the xyz basis has been given before in Ref. [40] using unary encoded cells of $\{a, b, c\}$ spin-$\frac{1}{2}$'s. Their construction involves initializing the b spin-$\frac{1}{2}$'s in state $|+\rangle$ while entangling the a and c spin-$\frac{1}{2}$'s between adjacent unary cells in the state $|\uparrow\rangle|\uparrow\rangle + |\downarrow\rangle|\downarrow\rangle$. As a tensor network, this is represented as

$$\text{(30)}$$
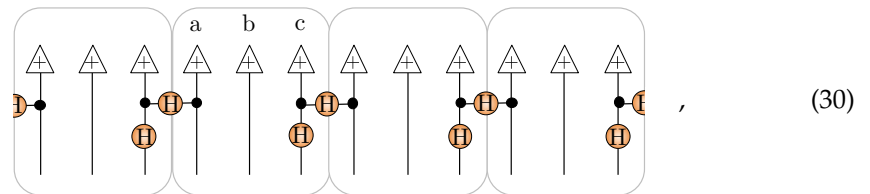
where each box is a unary cell, and H denotes the Hadamard matrix. Each unary cell then has the following unitary applied to it

$$\text{(31)}$$

where S is the phase gate [40] (we have switched the controlled-NOT gates in the circuit given in Figure 5b of ref. [40] into controlled-Z gates here to expose the graph state equivalence). Putting these pieces together, the unary encoded spin-$\frac{1}{2}$ state $\left|\Psi_{\text{unary}}\right\rangle$ is a stabilizer state constructed by the circuit

$$\text{(32)}$$
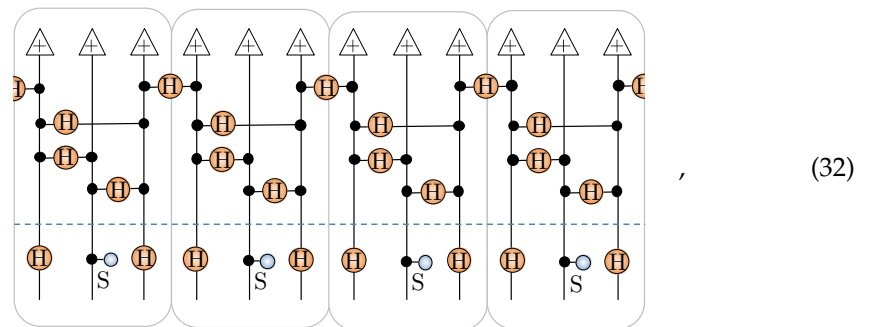
where the top part (above the dashed line) generates a graph state, and the bottom part applies local Clifford gates. Once unary projection to the spin-1 is applied, $\left|\Psi_{\text{unary}}\right\rangle$ generates the AKLT state [40].

Previously, in ref. [42], we showed how any stabilizer state for N spin-$\frac{1}{2}$'s can be readily converted into an NQS with $M \leq N - 1$. Here, we just summarize the basic process. The first step in this conversion is to use the local Clifford equivalence of stabilizer states to graph states to relocate all the non-diagonal Clifford gates to independent vertices of the graph. This conversion takes the simple chain-like graph state and pattern of Clifford gates from Equation (32) and gives the following for $N = 12$ spins:
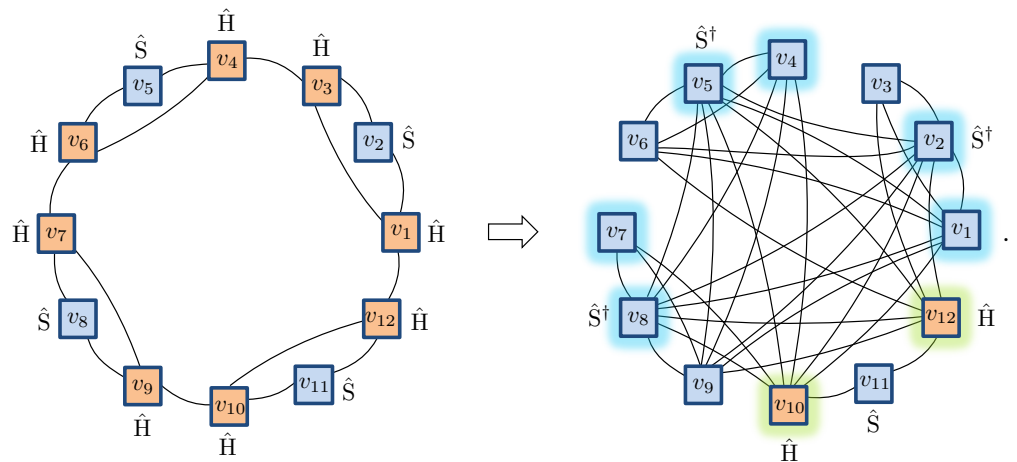
As required, the resulting graph has diagonal Clifford gates on all vertices, except for a small independent set $\mathcal{I} = \{10, 12\}$ highlighted. Notice that the three-site translational invariance of $\left| \Psi_{\text{unary}} \right\rangle$, mirrored by the initial chain graph, is still formally present in the transformed graph but is now obscured by its highly connected topology. By forming a vertex cover $\mathcal{C} = \{\mathcal{I}, 1, 2, 4, 5, 7, 8\}$, we obtain a NQS [42] with $M = 8$ hidden units (despite the more complex graph, this is the same number of hidden units required to describe the initial chain graph state as an NQS):

$$. \qquad (33) $$
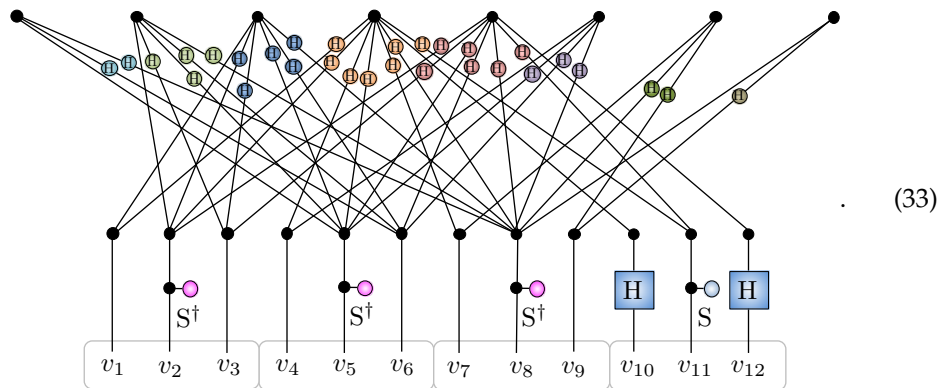
More generally, for N spin-$\frac{1}{2}$'s, this procedure generates an NQS with $M = 2N/3$ hidden units.

After applying the unary projection and contraction process directly to this spin-$\frac{1}{2}$ NQS, we obtain spin-1 NQS tensor network, whose schematic structure is shown here for $N = 4$:

$$. \qquad (34) $$
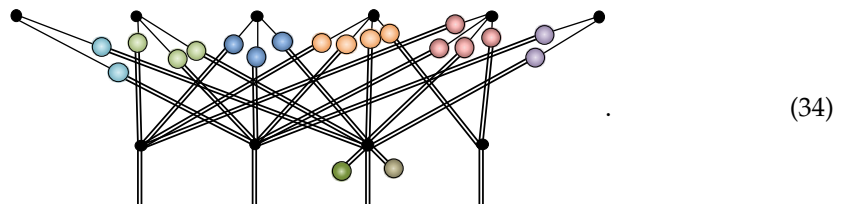
It is evident that pairs of hidden units possess coordination $4, 3$ and 2, while one pair gets projected down to a spin-1 visible bias, giving $M = 6$ overall. The same structure applies for general N with pairs of equal coordination spanning $N, N-1, \ldots, 2$, giving a spin-1 NQS for the AKLT state in the xyz basis requiring $M = 2N - 2$ hidden units. This representation is essentially identical to the one presented in Section 5.1, except that the hidden units implementing the nodal structure (the fully connected pair) also contribute to the sign structure, reducing the total hidden unit count by one pair. This raises an

interesting question of whether an even more compressed spin-1 NQS representation of the AKLT state is possible. We finish by examining this using direct numerical optimizations.

5.3. Numerical Example—AKLT in xyz and S^z Bases

Although the AKLT state is translationally invariant, the hidden units encoding the sign structure of this solution are neither individually translationally invariant nor do their translates appear. Consequently, we performed VMC calculations with increasing M using both the spin-1 NQS and unary NQS for $N = \{6, 8, 10, 12\}$ and compared against exact diagonalization. As with the AFH model (which is $\hat{H}_{AKLT}$ with $\beta = 0$), earlier in Section 4.5, we considered both the S^z and xyz basis with their corresponding nodal structure enforced by sampling. We also utilize the same sequential growth scheme as we used in the Heisenberg calculations, again confirming the robustness of our qualitative conclusions by performing reruns of the optimization sequence and presenting the best results here.

As we are performing stochastic optimization, intrinsic sampling noise will limit the accuracy to which any formally exact solution can be found. It is, therefore, crucial to quantitatively characterize when exactness may have been reached numerically. For a gapped Hamiltonian, the average energy of an approximate state E can be related to its infidelity with the ground state using [30]

$$1 - \mathcal{O} \leq \frac{E - E_0}{\delta} = \frac{\epsilon}{\delta}, \tag{35}$$

where δ is the energy gap of the Hamiltonian, and $\epsilon = E - E_0$ is the energy deviation. As the ground state energy E_0 of the AKLT Hamiltonian is zero, the energy deviation is simply the sampled energy of the state ϵ_{samp}. As shown in Figure 8a, even if the exact analytic spin-1 NQS solution is used, ϵ_{samp} fluctuates when using a finite number of samples typical of an optimization step. For all optimizations presented in this paper, we used the following hyperparameters: number of samples per optimization step $n_{samp} = 8000$, number of optimization steps $n_{step} \sim 5000$. Typically, the full wavefunction and its fidelity are calculated and checked every 1000 steps to gauge whether the solution has converged or requires further optimization. To account for fluctuations caused by a finite n_{samp} employed throughout the stochastic optimization, therefore, we use, in Equation (35), $\epsilon = \Delta\epsilon$, the standard deviation of the sampled energy. As shown in Figure 8b, $\Delta\epsilon$ vanishes as $1/\sqrt{n_{samp}}$, and we estimate an algorithmic fidelity resolution of $\mathcal{R} = \Delta\epsilon/\delta \approx 1.2 \times 10^{-5}$ for $N = 12$ sites below, in which it may be hard to discriminate an exact solution from an extremely good approximate one.

In Figure 9, we show the smooth decrease in $1 - \mathcal{O}$ against M for the S^z basis. While the $N = 6$ curve in Figure 9a drops below $\mathcal{R}$, this is not attained for larger N shown in Figure 9b–d within the M's tested. This indicates that no "exact" NQS solutions have been located. As with the AFH model, the spin-1 NQS and unary NQS achieve a similar accuracy verses M, although the former utilizes less variational parameters.
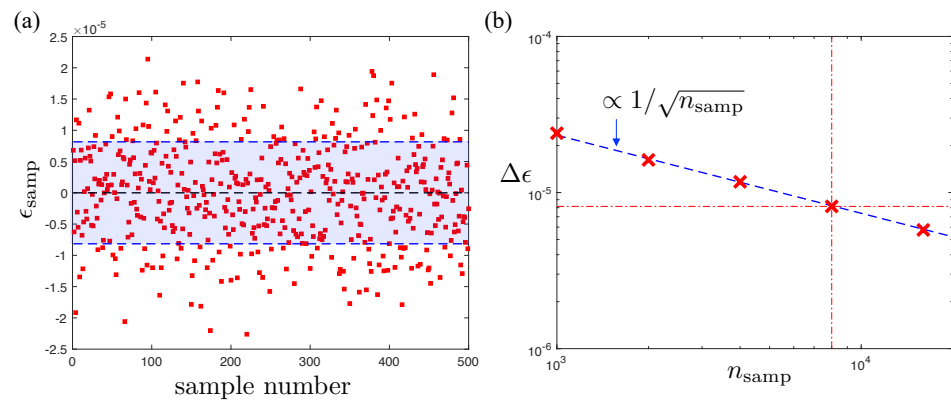
Figure 8. (**a**) The energy ϵ_{samp} (red squares) for 500 independent Monte Carlo runs of the $N = 12$ exact AKLT spin-1 NQS each consisting of $n_{\text{samp}} = 8000$ individual sampling steps separated by N individual Markov chain moves to reduce autocorrelation effects. The standard deviation $\Delta\epsilon$ for these samples around the exact zero ground state energy is shown by the blue band. (**b**) The standard deviations $\Delta\epsilon$ of energies sampled versus the number of samples n_{samp} used. Each point is also calculated from 500 independent samples, and the number of sampling steps per evaluation $n_{\text{samp}} = \{1000, 2000, 4000, 8000, 16{,}000\}$. The fluctuations in energy closely follow a $1/\sqrt{n_{\text{samp}}}$ scaling (dashed line), as expected for a Monte Carlo sampling process [30]. The red dotted lines are to guide the eye to the point for $n_{\text{samp}} = 8000$, which is representative of an optimization step for our VMC calculations.
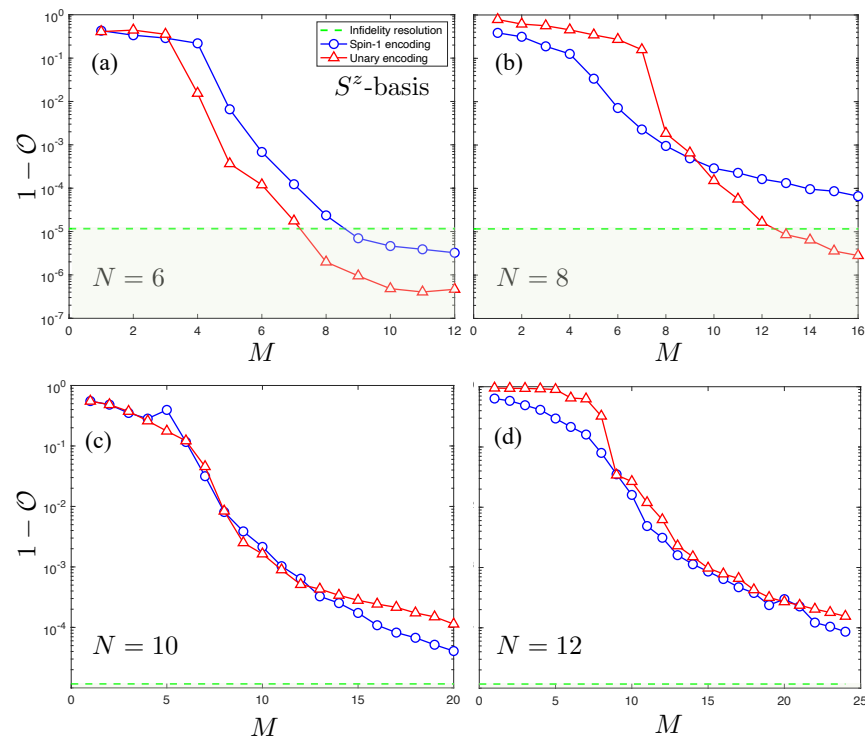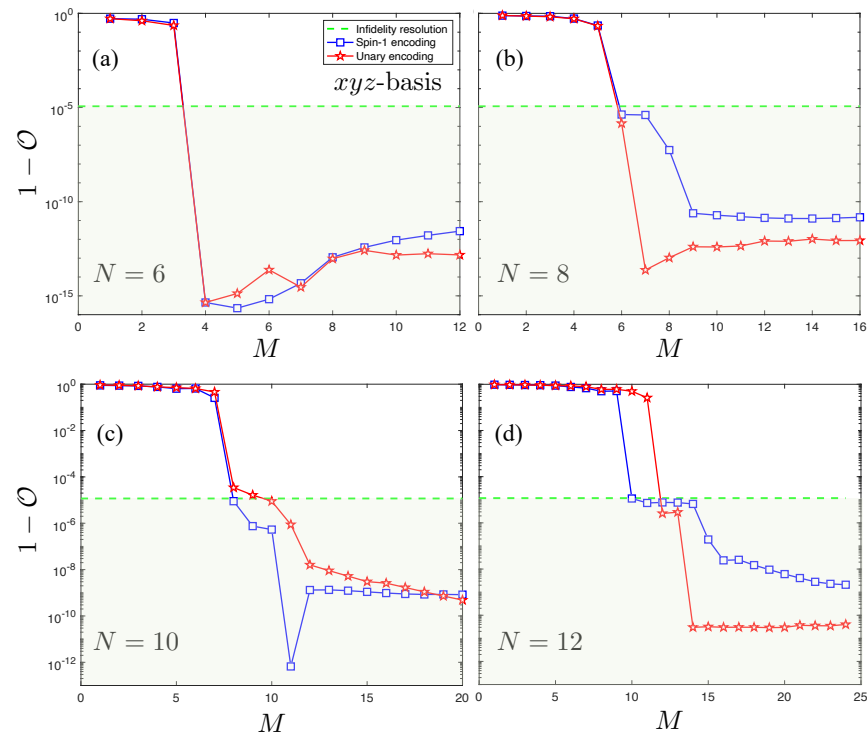

Figure 9. Plots of the infidelity of unary (red stars) and spin-1 (blue squares) NQS with the AKLT state in the S^z basis. The infidelity bound $\mathcal{R}$ is also plotted as a dotted green line. Results for four system sizes are plotted: (**a**) $N = 6$, (**b**) $N = 8$, (**c**) $N = 10$, (**d**) $N = 12$.

The analogous results in Figure 10 for the *xyz* basis display remarkable features in comparison. For each N, a sharp drop in $1 - \mathcal{O}$ by over 4 orders of magnitude is observed at $M = N - 2$ that consistently pushes the infidelity below $\mathcal{R}$. After reaching this point, $1 - \mathcal{O}$ versus M plateaus, and subsequent hidden units have negligible bearing on the accuracy of the wavefunction due to statistical fluctuations originating from the stochastic

optimization. These features are consistently produced by both the spin-1 NQS and unary NQS, aside from the largest system size $N = 12$ in Figure 10d, indicating that the increasing number of redundant variational parameters in the unary encoding is complicating the optimization. The overall behavior of $1 - \mathcal{O}$ observed in this basis is strong evidence of a numerically exact solution with $M = N$ (once the 2 hidden units implementing the nodal structure are included). This is substantially smaller than the analytic solutions introduced and is very suggestive of there being a *compact* exact spin-1 NQS representation of the AKLT state in the xyz basis.

Figure 10. Plots of the infidelity of unary (red stars) and spin-1 (blue squares) NQS with the AKLT state for the xyz basis. The infidelity resolution $\mathcal{R}$ is plotted as a green dotted line. Results for four system sizes are plotted: (**a**) $N = 6$, (**b**) $N = 8$, (**c**) $N = 10$, (**d**) $N = 12$.

6. Conclusions

We have introduced the most natural and direct generalization of RBM from spin-$\frac{1}{2}$ to spin-1. This necessitated including a quadratic visible bias and a quadratic visible-hidden interaction in the RBM energy function to ensure trivial product state representation, labeling freedom and gauge equivalence to the tensor network formulation. We demonstrated its use numerically for the spin-1 AFH model in both the S^z and xyz bases, illustrating how the choice of basis can affect the accuracy and hidden unit complexity of an NQS representation. Using our spin-1 NQS, we then re-examined how to represent the AKLT state exactly. In the S^z basis, it is known to require $M \sim O(N^2)$ hidden units, yet, by changing to the xyz basis, we construct an NQS with $M \sim O(N)$ hidden units.

Numerical VMC calculations have indicated that, by capturing the nodal structure, either implicitly within the sampling or explicitly through the inclusion of extra hidden units, the optimization can find even more efficient constructions for the sign structure. The resulting spin-1 NQS for the AKLT state in the xyz basis requires $M = N$ hidden units in total making it *compact*. This example raises the interesting possibility of improving the efficiency and accuracy of NQS calculations by including single-spin basis transformations to lower the hidden unit complexity.

Several important open questions remain about NQS representations. In particular, it would be instructive to build representations of classes of bosonic states using multinomial

RBMs. In this case, a local Fock basis is typically employed; however, our findings suggest that it could be useful to explore a local basis that breaks the particle number symmetry when describing condensates. Moreover, the elevation of visible units from binary to multinomial raises the question of whether also using multinomial hidden units can enhance the expressiveness of NQS. This has been explored in the context of binary visible units in ref. [54] in an analytical context to precisely represent certain two- and three-body interactions. The use of multinomial hidden units for numerical VMC calculations has been largely unexplored and is the subject of forthcoming work [55] for the Bose Hubbard model in two dimensions.

Appendix A. Monte Carlo Method and Sampling

Any many-body quantum state can be expressed in terms of its amplitudes on a set of configurations states in a chosen basis, but exhaustively computing the amplitudes for every possible configuration in the basis is often prohibitively expensive. Monte Carlo methods circumvent this by randomly selecting a subset of these configurations, and calculating an estimate of key observables using this selection. In this case, we require only that the (unnormalized) amplitudes $\Psi_p(S)$ of the ansatz (defined by parameters p) in a fixed basis, such as S^z, can be computed in a time with polynomial complexity in N. We then rewrite the expectation value as

$$\langle \hat{A} \rangle_{\Psi_p} = \sum_S P(S) A(S), \quad \text{where} \quad P(S) = \frac{|\Psi_p(S)|^2}{\sum_S |\Psi_p(S)|^2} \tag{A1}$$

is the probability of the spin configuration S, and

$$A(S) = \sum_{S'} \frac{\Psi_p(S')}{\Psi_p(S)} \langle S | \hat{A} | S' \rangle \tag{A2}$$

is the local estimator of $\hat{A}$. The sum over S' in Equation (A2) is restricted to only those configurations for which the matrix element $\langle S | \hat{A} | S' \rangle \neq 0$, and, so as long as $\hat{A}$ is sparse in the chosen fixed basis, then $A(S)$ can be efficiently evaluated. Typical product terms comprising short-ranged Hamiltonians $\hat{H}$ fulfill this requirement. Thus, an unbiased estimate of $\langle \hat{A} \rangle_{\Psi_p}$ is made by drawing independent samples of spin configurations S from the distribution $P(S)$. A powerful way to accomplish this is Markov-chain Monte Carlo in which a sequence of spin configurations $S_0 \to S_1 \to \cdots$ is generated using the

Metropolis-Hastings algorithm that accepts a proposed configuration S_{prop} for the nth iteration with a probability $P_n(S_{\text{prop}}) = \min[1, |\Psi_p(S_{\text{prop}})|^2/|\Psi_p(S_{n-1})|^2]$. A subset of visited configurations in the sequence, suitably separated to ensure they are decorrelated, is then used to estimate of $\langle \hat{A} \rangle_{\Psi_p}$. Crucially, both $A(S)$ and Monte Carlo sampling rely only on the ratio of ansatz amplitudes, so the intractable state normalization $\sum_S |\Psi_p(S)|^2$ is entirely avoided.

Variational Monte Carlo proceeds by evaluating the ansatz energy $E_p = \langle \hat{H} \rangle_{\Psi_p}$ and its variance, along with their gradient vectors with respect to parameters p. The parameters can then be updated by a small step along the direction of steepest descent, with this iterated until convergence to the minimum p_0 [29]. The efficiency of this minimization process is strongly problem and ansatz dependent. In challenging cases, more sophisticated approaches, such as modified stochastic optimization [57], the 'linear method' [58,59], and stochastic reconfiguration [30,49], are needed. All numerical calculations in this paper employ stochastic reconfiguration. The method involves the construction of a $n_{\text{params}} \times n_{\text{params}}$ matrix, requiring $O(n_{\text{samp}} n_{\text{params}}^2)$ operations, though with a conjugate gradient method this can be brought down to $O(n_{\text{samp}} n_{\text{params}})$ [60]. The parameter changes can then be calculated from a set of linear equations involving the matrix, which scales as $O(n_{\text{params}}^2)$ using matrix-vector product methods. As a rule of thumb, the matrix requires a number of samples $n_{\text{samp}} > 10 n_{\text{params}}$ to ensure the sampled matrix is not rank-deficient [30], giving an overall minimum scaling of $O(n_{\text{params}}^2)$ for the algorithm, though, typically, it is prudent to have a large number of samples $n_{\text{samp}} \gg n_{\text{params}}$ as the error of any sampled observable, typically scales, as $1/\sqrt{n_{\text{samp}}}$ [30], including the elements of the matrix.

Appendix B. Boltzmann Parameterization of Coupling Matrices

Every 2×3 coupling matrix within the spin-1 NQS tensor network can be parameterized in Boltzmann form using the generalized energy function introduced in Section 4.2 as

$$C_{h_i v_j}^{(ij)} = \exp\left(\tilde{c}_{ij} + w_{ij} h_i v_j + W_{ij} h_i v_j^2 + \tilde{b}_{ij} h_i + \tilde{a}_{ij} v_j + \tilde{A}_{ij} v_j^2\right), \tag{A3}$$

for $h_i \in \{+1, -1\}$ and $v_j \in \{+1, 0, -1\}$, which includes a quadratic weight W_{ij} and quadratic partial bias $\tilde{A}_{ij}$. These complex parameters are found from the coupling matrix elements as

$$\tilde{a}_{ij} = \tfrac{1}{4}\left[\log\left(C_{++}^{(ij)}\right) + \log\left(C_{-+}^{(ij)}\right) - \log\left(C_{+-}^{(ij)}\right) - \log\left(C_{--}^{(ij)}\right)\right],$$

$$\tilde{b}_{ij} = \tfrac{1}{2}\left[\log\left(C_{+0}^{(ij)}\right) - \log\left(C_{-0}^{(ij)}\right)\right],$$

$$\tilde{c}_{ij} = \tfrac{1}{2}\left[\log\left(C_{+0}^{(ij)}\right) + \log\left(C_{-0}^{(ij)}\right)\right],$$

$$w_{ij} = \tfrac{1}{4}\left[\log\left(C_{++}^{(ij)}\right) - \log\left(C_{-+}^{(ij)}\right) - \log\left(C_{+-}^{(ij)}\right) + \log\left(C_{--}^{(ij)}\right)\right],$$

$$W_{ij} = \tfrac{1}{4}\left[\log\left(C_{++}^{(ij)}\right) - \log\left(C_{-+}^{(ij)}\right) - 2\log\left(C_{+0}^{(ij)}\right) + 2\log\left(C_{-0}^{(ij)}\right) + \log\left(C_{+-}^{(ij)}\right) - \log\left(C_{--}^{(ij)}\right)\right],$$

$$\tilde{A}_{ij} = \tfrac{1}{4}\left[\log\left(C_{++}^{(ij)}\right) + \log\left(C_{-+}^{(ij)}\right) - 2\log\left(C_{+0}^{(ij)}\right) - 2\log\left(C_{-0}^{(ij)}\right) + \log\left(C_{+-}^{(ij)}\right) + \log\left(C_{--}^{(ij)}\right)\right].$$

In both cases, the presence of zero matrix elements in $\mathbf{C}^{(ij)}$, which appear frequently in our 'by hand' constructions, return diverging parameters. However, as discussed in ref. [42], this can be handled by softening the zeros via

$$C_{h_i v_j}^{(ij)} \mapsto \max\left(C_{h_i v_j}^{(ij)}, e^{-\mathcal{S}}\right), \tag{A4}$$

where typically $\mathcal{S} \approx 5\text{–}10$.

Appendix C. AKLT State NQS in S^z Basis

In this appendix, we construct a spin-1 NQS for the AKLT state amplitudes in the S^z basis given in Equation (26) by using hidden units as successive filters. The rules governing the structure of the amplitudes can be summarized as follows:

1. Zero out the amplitude for any configuration containing a substring for any $\ell = 0, 1, \ldots, N - 2$ of the form $+[0 \cdots 0]^\ell+$ or $-[0 \cdots 0]^\ell-$, where $[0 \cdots 0]^\ell$ is a string of 0's of length ℓ. To implement this rule, we introduce the following 2×3 coupling matrices:

    $$\mathbf{C}_{i\Uparrow} = \begin{bmatrix} 1 & 1 & 1 \\ i & 0 & 0 \end{bmatrix}, \quad \mathbf{C}_{i\Downarrow} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & i \end{bmatrix}, \quad \mathbf{C}_{+0} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 1 & 0 \end{bmatrix},$$

    which, in turn, single out S^z basis spin-1 states $|\Uparrow\rangle$, $|0\rangle$ and $|\Downarrow\rangle$. To cancel out in any configuration substrings $+[0 \cdots 0]^\ell+$ and $-[0 \cdots 0]^\ell-$, starting at a given site k with a given separation ℓ, we construct pairs of hidden units with

    $$\{\mathbf{I}, \cdots, \mathbf{I}, \overbrace{\mathbf{C}_{+i}}^{k}, \mathbf{C}_{+0}, \cdots, \mathbf{C}_{+0}, \overbrace{\mathbf{C}_{+i}}^{k+1+\ell}, \mathbf{I}, \cdots, \mathbf{I}\}, \quad \text{and}$$
    $$\{\mathbf{I}, \cdots, \mathbf{I}, \overbrace{\mathbf{C}_{-i}}^{k}, \mathbf{C}_{+0}, \cdots, \mathbf{C}_{+0}, \overbrace{\mathbf{C}_{-i}}^{k+1+\ell}, \mathbf{I}, \cdots, \mathbf{I}\}.$$

    To capture all separations $\ell = 0, 1, \ldots, N - 2$ starting on any site $k = 1, 2, \ldots, N$, we require $2N(N - 1)$ hidden units.

2. Apply a factor of $(-1)^\ell$ to the amplitude of a configuration for each substring of the form $+[0 \cdots 0]^\ell-$ it contains. To implement this rule, we introduce two additional coupling matrices:

    $$\mathbf{C}_{2\Uparrow} = \begin{bmatrix} 1 & 1 & 1 \\ 2 & 0 & 0 \end{bmatrix}, \quad \mathbf{C}_{-\Downarrow} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & -1 \end{bmatrix}.$$

    A factor $(-1)^\ell$ is induced on any configuration containing the substring $+[0 \cdots 0]^\ell-$, starting at a given site k with a given separation ℓ, by a hidden unit with

    $$\{\mathbf{I}, \cdots, \mathbf{I}, \overbrace{\mathbf{C}_{2\Uparrow}}^{k}, \mathbf{C}_{+0}, \cdots, \mathbf{C}_{+0}, \overbrace{\mathbf{C}_{-\Downarrow}}^{k+1+\ell}, \mathbf{I}, \cdots, \mathbf{I}\}.$$

    To capture all odd separations $\ell = 1, 3, \ldots, N - 2$ starting on any site $k = 1, 2, \ldots, N$ requires $N \lfloor \frac{1}{2}(N - 1) \rfloor$ hidden units.

3. Zero out the single excitation configurations $+[0 \cdots 0]^{N-1}$ and $-[0 \cdots 0]^{N-1}$ and all their translates. To implement this rule, we introduce another coupling matrix:

    $$\mathbf{C}_{-\Uparrow} = \begin{bmatrix} 1 & 1 & 1 \\ -1 & 0 & 0 \end{bmatrix}.$$

    Configurations $+[0 \cdots 0]^{N-1}$ and $-[0 \cdots 0]^{N-1}$ are then cancelled out by a pair of hidden units with $\{\mathbf{C}_{-\Uparrow}, \mathbf{C}_{+0}, \cdots, \mathbf{C}_{+0}\}$ and $\{\mathbf{C}_{-\Downarrow}, \mathbf{C}_{+0}, \cdots, \mathbf{C}_{+0}\}$. Capturing all the translates is achieved with $2N$ hidden units.

4. For N, odd zero out the amplitude for configuration $[0 \cdots 0]^N$, or, for N, even double its amplitude. To implement this rule, we introduce one final coupling matrix:

    $$\mathbf{C}_{-0} = \begin{bmatrix} 1 & 1 & 1 \\ 0 & -1 & 0 \end{bmatrix},$$

    and defining a single hidden unit as $\{\mathbf{C}_{-0}, \mathbf{C}_{-0}, \cdots, \mathbf{C}_{-0}\}$.

5. Apply a factors of $2^{(\# +'s)} \times (-1)^{(\# 0's)} \times (-1)^{(\# -'s)}$ to the amplitudes of all configurations. This does not require any additional hidden units. Instead, we can pick any hidden unit and right multiply each of its N coupling matrices by $\mathbf{D} = \text{diag}(2, -1, -1)$. In RBM formalism, this is equivalent to setting non-zero visible unit biases.

This gives a total of $M = 2N^2 + N\lfloor\frac{1}{2}(N-1)\rfloor + 1$ hidden units. Notice that the origin of the $O(N^2)$ scaling arises from spatial translations, so if this symmetry is directly imposed within the NQS ansatz, as it often is numerically, then the AKLT state in the S^z basis instead has its number of unique hidden units scaling as $M = 2N + \lfloor\frac{1}{2}(N-1)\rfloor + 1$.

References

1. Verstraete, F.; Murg, V.; Cirac, J. Matrix product states, projected entangled pair states, and variational renormalization group methods for quantum spin systems. *Adv. Phys.* **2008**, *57*, 143–224. [CrossRef]
2. Orús, R. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Ann. Phys.* **2014**, *349*, 117–158. [CrossRef]
3. Cirac, J.I.; Sierra, G. Infinite matrix product states, Conformal Field Theory and the Haldane-Shastry model. *Phys. Rev. B* **2010**, *81*, 104431. [CrossRef]
4. Shi, Y.Y.; Duan, L.M.; Vidal, G. Classical simulation of quantum many-body systems with a tree tensor network. *Phys. Rev. A* **2006**, *74*, 022320. [CrossRef]
5. Murg, V.; Verstraete, F.; Legeza, O.; Noack, R.M. Simulating strongly correlated quantum systems with tree tensor networks. *Phys. Rev. B* **2010**, *82*, 205105. [CrossRef]
6. Evenbly, G.; Vidal, G. Quantum Criticality with the Multi-scale Entanglement Renormalization Ansatz. In *Strongly Correlated Systems*; Springer Series in Solid-State Sciences; Springer: Berlin/Heidelberg, Germang, 2013; Volume 176, pp. 99–130.
7. Carleo, G.; Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **2017**, *355*, 602606. [CrossRef] [PubMed]
8. Gao, X.; Duan, L.M. Efficient representation of quantum many-body states with deep neural networks. *Nat. Commun.* **2017**, *8*, 662. [CrossRef] [PubMed]
9. Carleo, G.; Nomura, Y.; Imada, M. Constructing exact representations of quantum many-body systems with deep neural networks. *Nat. Commun.* **2018**, *9*, 5322. [CrossRef]
10. He, H.; Zheng, Y.; Bernevig, B.A.; Sierra, G. Multi-Layer Restricted Boltzmann Machine Representation of 1D Quantum Many-Body Wave Functions. *arXiv* **2019**, arXiv:1910.13454.
11. Choo, K.; Neupert, T.; Carleo, G. Two-dimensional frustrated $J_1 - J_2$ model studied with neural network quantum states. *Phys. Rev. B* **2019**, *100*, 125124. [CrossRef]
12. Irikura, N.; Saito, H. Neural-network quantum states at finite temperature. *Phys. Rev. Res.* **2020**, *2*, 013284. [CrossRef]
13. Schmitt, M.; Heyl, M. Quantum Many-Body Dynamics in Two Dimensions with Artificial Neural Networks. *Phys. Rev. Lett.* **2020**, *125*, 100503. [CrossRef]
14. Liang, X.; Dong, S.J.; He, L. Hybrid convolutional neural network and projected entangled pair states wave functions for quantum many-particle states. *Phys. Rev. B* **2021**, *103*, 035138. [CrossRef]
15. Levine, Y.; Sharir, O.; Cohen, N.; Shashua, A. Quantum Entanglement in Deep Learning Architectures. *Phys. Rev. Lett.* **2019**, *122*, 065301. [CrossRef]
16. Saito, H.; Kato, M. Machine Learning Technique to Find Quantum Many-Body Ground States of Bosons on a Lattice. *J. Phys. Soc. Jpn.* **2018**, *87*, 014001. [CrossRef]
17. Choo, K.; Carleo, G.; Regnault, N.; Neupert, T. Symmetries and Many-Body Excitations with Neural-Network Quantum States. *Phys. Rev. Lett.* **2018**, *121*, 167204. [CrossRef]
18. Luo, D.; Clark, B.K. Backflow Transformations via Neural Networks for Quantum Many-Body Wave Functions. *Phys. Rev. Lett.* **2019**, *122*, 226401. [CrossRef] [PubMed]
19. Adams, C.; Carleo, G.; Lovato, A.; Rocco, N. Variational Monte Carlo Calculations of $\mathbf{A} \leq \mathbf{4}$ Nuclei with an Artificial Neural-Network Correlator Ansatz. *Phys. Rev. Lett.* **2021**, *127*, 022502. [CrossRef]
20. Torlai, G.; Melko, R.G. Latent Space Purification via Neural Density Operators. *Phys. Rev. Lett.* **2018**, *120*, 240503. [CrossRef] [PubMed]
21. Vicentini, F.; Biella, A.; Regnault, N.; Ciuti, C. Variational Neural-Network Ansatz for Steady States in Open Quantum Systems. *Phys. Rev. Lett.* **2019**, *122*, 250503. [CrossRef]
22. Hartmann, M.J.; Carleo, G. Neural-Network Approach to Dissipative Quantum Many-Body Dynamics. *Phys. Rev. Lett.* **2019**, *122*, 250502. [CrossRef] [PubMed]
23. Yoshioka, N.; Hamazaki, R. Constructing neural stationary states for open quantum many-body systems. *Phys. Rev. B* **2019**, *99*, 214306. [CrossRef]
24. Westerhout, T.; Astrakhantsev, N.; Tikhonov, K.S.; Katsnelson, M.I.; Bagrov, A.A. Neural Quantum States of frustrated magnets: Generalization and sign structure. *Nat. Commun.* **2020**, *11*, 1593. [CrossRef] [PubMed]

25. Nomura, Y. Helping restricted Boltzmann machines with quantum-state representation by restoring symmetry. *J. Phys. Condens. Matter* **2021**, *33*, 174003. [CrossRef]

26. Jónsson, B.; Bauer, B.; Carleo, G. Neural-network states for the classical simulation of quantum computing. *arXiv* **2018**, arXiv:1808.05232.

27. Freitas, N.; Morigi, G.; Dunjko, V. Neural network operations and Susuki–Trotter evolution of neural network states. *Int. J. Quantum Inf.* **2018**, *16*, 1840008. [CrossRef]

28. Bausch, J.; Leditzky, F. Quantum codes from neural networks. *New J. Phys.* **2020**, *22*, 023005. [CrossRef]

29. Gubernatis, J.; Kawashima, N.; Werner, P. *Quantum Monte Carlo Methods: Algorithms for Lattice Models*; Cambridge University Press: Cambridge, UK, 2016. [CrossRef]

30. Becca, F.; Sorella, S. *Quantum Monte Carlo Approaches for Correlated Systems*; Cambridge University Press: Cambridge, UK, 2017. [CrossRef]

31. Deng, D.L.; Li, X.; Das Sarma, S. Quantum Entanglement in Neural Network States. *Phys. Rev. X* **2017**, *7*, 021021. [CrossRef]

32. Eisert, J.; Cramer, M.; Plenio, M.B. Colloquium: Area laws for the entanglement entropy. *Rev. Mod. Phys.* **2010**, *82*, 277. [CrossRef]

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

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

35. Collura, M.; Dell'Anna, L.; Felser, T.; Montangero, S. On the descriptive power of Neural-Networks as constrained Tensor Networks with exponentially large bond dimension. *SciPost Phys. Core* **2021**, *4*. [CrossRef]

36. Glasser, I.; Pancotti, N.; August, M.; Rodriguez, I.D.; Cirac, J.I. Neural-Network Quantum States, String-Bond States, and Chiral Topological States. *Phys. Rev. X* **2018**, *8*, 011006. [CrossRef]

37. Kaubruegger, R.; Pastori, L.; Budich, J.C. Chiral topological phases from artificial neural networks. *Phys. Rev. B* **2018**, *97*, 195136. [CrossRef]

38. Zheng, Y.; He, H.; Regnault, N.; Bernevig, B.A. Restricted Boltzmann Machines and Matrix Product States of 1D Translational Invariant Stabilizer Codes. *Phys. Rev. B* **2019**, *99*, 155129. [CrossRef]

39. Zhang, Y.H.; Jia, Z.A.; Wu, Y.C.; Guo, G.C. An Efficient Algorithmic Way to Construct Boltzmann Machine Representations for Arbitrary Stabilizer Code. *arXiv* **2018**, arXiv:1809.08631.

40. Lu, S.; Gao, X.; Duan, L.M. Efficient representation of topologically ordered states with restricted Boltzmann machines. *Phys. Rev. B* **2019**, *99*, 155136. [CrossRef]

41. Jia, Z.A.; Zhang, Y.H.; Wu, Y.C.; Kong, L.; Guo, G.C.; Guo, G.P. Efficient Machine Learning Representations of Surface Code with Boundaries, Defects, Domain Walls and Twists. *Phys. Rev. A* **2019**, 99, 012307. [CrossRef]

42. Pei, M.Y.; Clark, S.R. Compact Neural-network Quantum State representations of Jastrow and Stabilizer states. *arXiv* **2021**, arXiv:2103.09146.

43. Saito, H. Solving the Bose–Hubbard Model with Machine Learning. *J. Phys. Soc. Jpn.* **2017**, *86*, 093001. [CrossRef]

44. Guo, C.; Berkhahn, F. Entity Embeddings of Categorical Variables. *arXiv* **2016**, arXiv:1604.06737.

45. McBrian, K.; Carleo, G.; Khatami, E. Ground state phase diagram of the one-dimensional Bose-Hubbard model from restricted Boltzmann machines. *J. Phys. Conf. Ser.* **2019**, *1290*, 012005. [CrossRef]

46. Vargas-Calderón, V.; Vinck-Posada, H.; González, F.A. Phase Diagram Reconstruction of the Bose–Hubbard Model with a Restricted Boltzmann Machine Wavefunction. *J. Phys. Soc. Jpn.* **2020**, *89*, 094002. [CrossRef]

47. Vieijra, T.; Casert, C.; Nys, J.; De Neve, W.; Haegeman, J.; Ryckebusch, J.; Verstraete, F. Restricted Boltzmann Machines for Quantum States with Non-Abelian or Anyonic Symmetries. *Phys. Rev. Lett.* **2020**, *124*, 097201. [CrossRef] [PubMed]

48. Läuchli, A.M.; Sudan, J.; Sørensen, E.S. Ground-state energy and spin gap of spin-$\frac{1}{2}$ Kagomé-Heisenberg antiferromagnetic clusters: Large-scale exact diagonalization results. *Phys. Rev. B* **2011**, *83*, 212401. [CrossRef]

49. Sorella, S. Generalized Lanczos algorithm for variational quantum Monte Carlo. *Phys. Rev. B* **2001**, *64*, 024512. [CrossRef]

50. Le Roux, N.; Bengio, Y. Representational Power of Restricted Boltzmann Machines and Deep Belief Networks. *Neural Comput.* **2008**, *20*, 1631–1649. [CrossRef] [PubMed]

51. Biamonte, J.D.; Clark, S.R.; Jaksch, D. Categorical Tensor Network States. *AIP Adv.* **2011**, *1*, 042172. [CrossRef]

52. Affleck, I.; Kennedy, T.; Lieb, E.H.; Tasaki, H. Rigorous results on valence-bond ground states in antiferromagnets. *Phys. Rev. Lett.* **1987**, *59*, 799–802. [CrossRef] [PubMed]

53. Haldane, F.D.M. Nonlinear Field Theory of Large-Spin Heisenberg Antiferromagnets: Semiclassically Quantized Solitons of the One-Dimensional Easy-Axis Néel State. *Phys. Rev. Lett.* **1983**, *50*, 1153–1156. [CrossRef]

54. Rrapaj, E.; Roggero, A. Exact representations of many-body interactions with restricted-Boltzmann-machine neural networks. *Phys. Rev. E* **2021**, *103*, 013302. [CrossRef] [PubMed]

55. Pei, M.Y.; Clark, S.R. Neural-network quantum states for bosons revisited. 2021, in preparation.

56. Clark, S.R.; Pei, M.Y. NQS Spin-1 Numerics Data. Data Repository for NQS Spin-1 Calculation Data and Plotting Scripts. Available online: https://data.bris.ac.uk/data/dataset/1ln9kyt6i86n12ehhftht27edp (accessed on 3 May 2021).

57. Lou, J.; Sandvik, A.W. Variational ground states of two-dimensional antiferromagnets in the valence bond basis. *Phys. Rev. B* **2007**, *76*, 104432. [CrossRef]

58. Nightingale, M.P.; Melik-Alaverdian, V. Optimization of Ground- and Excited-State Wave Functions and van der Waals Clusters. *Phys. Rev. Lett.* **2001**, *87*, 043401. [CrossRef] [PubMed]
59. Toulouse, J.; Umrigar, C.J. Optimization of quantum Monte Carlo wave functions by energy minimization. *J. Chem. Phys.* **2007**, *126*, 084102. [CrossRef] [PubMed]
60. Neuscamman, E.; Umrigar, C.J.; Chan, G.K.L. Optimizing large parameter sets in variational quantum Monte Carlo. *Phys. Rev. B* **2012**, *85*, 045103. [CrossRef]