Exact representations of many-body interactions with restricted-Boltzmann-machine neural networks

Ermal Rrapaj [1,2,*] and Alessandro Roggero [3,†]

[1]*Department of Physics, University of California, Berkeley, California 94720, USA*
[2]*School of Physics and Astronomy, University of Minnesota, Minneapolis, Minnesota 55455, USA*
[3]*Institute for Nuclear Theory, University of Washington, Seattle, Washington 98195, USA*

Restricted Boltzmann machines (RBMs) are simple statistical models defined on a bipartite graph which have been successfully used in studying more complicated many-body systems, both classical and quantum. In this work, we exploit the representation power of RBMs to provide an exact decomposition of many-body contact interactions into one-body operators coupled to discrete auxiliary fields. This construction generalizes the well known Hirsch's transform used for the Hubbard model to more complicated theories such as pionless effective field theory in nuclear physics, which we analyze in detail. We also discuss possible applications of our mapping for quantum annealing applications and conclude with some implications for RBM parameter optimization through machine learning.

I. INTRODUCTION

In recent decades, statistical methods based on artificial neural networks (ANNs) have been proven extremely valuable in studying the physical world. For a long time, experimental high energy physics has been at the forefront of these applications [1–3] and today ANN-based methods are of fundamental importance in analyzing particle accelerator data [4–7]. Thanks to the growing availability of large scale computational resources, these types of approaches have started to play an important role also in many-body theory more generally with applications as diverse as detecting phase transitions in simulations [8,9], preparing accurate variational states for lattice systems [10], accelerating sampling in Monte Carlo based simulations [11–13], constructing efficient energy density functionals [14], and performing fast approximate quantum state tomography [15,16].

A particularly interesting class is the generative models commonly used in unsupervised learning whose aim is to automatically discover underlying patterns in the data that are being analyzed (see, e.g., [17,18] for an introduction). The main advantage of this class of methods is the possibility of finding a compact and possibly accurate description of the supplied data without using any predetermined labeling. On one side, this allows the automatic discovery of appropriate labels and on the other simplifies dramatically the process of acquiring useful data points to train the model on. Once the model has been trained it can be used, for instance, to generate new data points from the reconstructed density distribution [18].

Our focus in this paper is a particularly simple ANN model composed by only two layers of neurons with a very sparse connectivity: the restricted Boltzmann machine (RBM) [19,20]. We will describe this architecture in some detail in Sec. II, but for now we want to anticipate that an RBM can be used as a universal approximator for arbitrary probability distributions (see, e.g., [21,22]). This fact, together with its simplicity, is one of the main reasons for its widespread use in many-body physics applications [9–11]. We note in passing that this construction is also the basic building block of deep learning models including the deep Boltzmann machine (DBM) and deep belief networks (DBNs) [23–25] where the number of layers is increased to provide more flexibility to the representation power of the model. For instance, three-layer DBMs have been successfully used in exact representations of ground state wave functions [26].

In this work we will use an RBM model to find simpler but exact representations for the many-body partition function

$$\exp(-\beta\hat{H}) \propto \mathrm{Tr_s}[\exp(-F_{\mathrm{rbm}}(\{\hat{\rho}_i\}, \mathbf{s}))], \tag{1}$$

in terms of the free energy F_{rbm} of an RBM where the visible layer is composed by quantum operators $\{\hat{\rho}_i\}$ and the hidden layer is made by a vector of classical auxiliary fields $\mathbf{s}$ which we marginalize over. Exact representations of this form are of fundamental importance for quantum Monte Carlo (QMC) calculations of many-body systems [27,28] which approximate ground state expectation values as

$$\langle 0|\hat{O}|0\rangle = \lim_{\beta\to\infty}\frac{1}{Z(\beta)}\mathrm{Tr}[\exp(-\beta\hat{H})\hat{O}], \tag{2}$$

where we introduced the partition function $Z(\beta) = \mathrm{Tr}[e^{-\beta\hat{H}}]$. More specifically, given a many-body Hamiltonian $\hat{H}$ written as a linear combination of Hermitian operators $\hat{H} = \sum_k \hat{H}_k$ and some initial state $|\Psi\rangle$, the basic computational step

*ermalrrapaj@gmail.com
†roggero@uw.edu

needed for QMC simulations is the map

$$|\Psi\rangle \longrightarrow |\Psi_k\rangle = \exp(-\beta\hat{H}_k)|\Psi\rangle, \tag{3}$$

for each one of the $\hat{H}_k$ terms composing the Hamiltonian operator $\hat{H}$. Efficient schemes to simplify the evolution operators $\exp(-\beta\hat{H}_k)$ using a map like Eq. (1) are key ingredients of QMC methods. Notable examples include auxiliary field methods such as the Hubbard-Stratonovich [29,30] and the Hirsch's transform [31] or the more recent approach proposed by Körber, Berkowitz, and Luu in [32].

In this work we focus in particular to two classes of Hamiltonians of fundamental importance in nuclear physics:

(i) Realistic local interactions [33,34] represented by interactions of the form

$$\hat{H}_V = \sum_i V_i(\vec{R})w_i(\vec{\sigma}, \vec{\tau}), \tag{4}$$

with $V_i(\vec{R})$ scalar coefficients dependent on the nucleon coordinates $\vec{R}$ and $w_i(\vec{\sigma}, \vec{\tau})$ a functional of spin operators $\vec{\sigma}$ and isospin operators τ. Note that it is always possible to choose the operators w_i in Eq. (4) to be involutive: $w_i(\vec{\sigma}, \vec{\tau})^2 = \mathbb{1}$.

(ii) Nuclear potentials derived in low energy effective theories expressed in terms of two- and three-body contact interaction [35] of the form

$$\hat{H}_V = \sum_{i,j} v_{ij}\hat{\rho}_i\hat{\rho}_j w_{ij}(\vec{\sigma}, \vec{\tau})$$
$$+ \sum_{i,j,k} v_{ijk}\hat{\rho}_i\hat{\rho}_j\hat{\rho}_k w_{ijk}(\vec{\sigma}, \vec{\tau}). \tag{5}$$

In this expression v_{ij} and v_{ijk} are scalar coefficients, $\hat{\rho}_i$ is a fermionic density operator [see Eq. (11) below for a more formal definition], and $w_{ij}(\vec{\sigma}, \vec{\tau})$ and $w_{ijk}(\vec{\sigma}, \vec{\tau})$ are idempotent spin and isospin operators as in Eq. (4). We note that these interactions arise naturally also in the low energy description of condensed matter systems (see, e.g., [36,37]).

Thanks to this representation it is sufficient to find an exact mapping in Eq. (1) for idempotent operators only, and this will be the focus of our present work. Note that it is always possible to express any operator on a finite Hilbert space as a linear combination of involutive operators by using (the tensor product of) Pauli operators as an operator basis.

In the rest of the paper, we provide an introduction to the RBM and present its application to represent many-body forces in nuclear physics. In Sec. II we proceed to relate the free energy $F_{\rm rbm}$ of this architecture to the physical partition functions produced by many-body interactions, starting with the familiar case of a two-body potential term. We then present a generalization to the case of three-body forces in Sec. II B. Further details on the construction for general terms is provided in Appendix A for completeness. One of the advantages of our approach is that it can be easily generalized from binary auxiliary fields to generic categorical classical variables which take values on a larger set $\{0, 1, \ldots, \mathcal{K} - 1\}$; in Sec. II C we discuss the advantages that this added flexibility can provide.

We then proceed in Sec. III to show how this RBM mapping could be used to improve the representation power of quantum annealers based on the transverse Ising model

Hamiltonian. In Sec. IV we use the exact mapping obtained before to analyze numerically the performance of various optimization protocols for the RBM parameters in reaching the known optimum for the simple case of a 2D Ising model. Note that this optimization step is not directly needed to use our results but instead addresses the question of feasibility of machine learning through RBMs and is of independent interest.

We conclude in Sec. V with a summary and possible implications of our results.

II. RESTRICTED BOLTZMANN MACHINES AS HAMILTONIANS WITH AUXILIARY FIELDS

The introduction of auxiliary fields as a mean of simplifying the interaction term of the Hamiltonian is common practice in many areas of theoretical physics, and is particularly popular in designing quantum Monte Carlo algorithms in both condensed matter and nuclear physics [27,28,38–43]. In these applications, a system of interacting particles is mapped into a free theory coupled to a background fluctuating auxiliary field and the final simulations is usually performed after integrating out the physical fields.

One famous instance of this class of mappings is the Hubbard-Stratonovich transformation [29,30] which exploits the Gaussian-integral relation

$$\exp\left(\frac{\tau}{2}\hat{O}^2\right) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} dh\, e^{-h^2/2 - \sqrt{\tau}h\hat{O}}, \tag{6}$$

with h a real auxiliary field, to provide a simpler representation for the evolution operator on the left-hand side. This transformation is used extensively to express the evolution under two-body interactions as a superposition of evolutions under one-body interactions parameterized by the value of the auxiliary field and is commonly used in auxiliary field Monte Carlo techniques [28,41–43]. This transformation can also be used for higher-order operators when they can be expressed as squares; a common example is the three-neutron force (see, e.g., [44]). When the interaction cannot be written as a perfect square one can attempt a recursive application of the transformation in Eq. (6) but this is usually accompanied by a drastically reduced efficiency (for an example of this applied to the isospin-dependent spin-orbit force, see [45]).

Another popular family of transformations that achieve a similar simplification can be obtained by considering auxiliary fields which can only take a discrete number of values. The prototypical example of this is the famous Hirsch transformation for the Hubbard model [31] which we will discuss in some detail in the next subsection (see also [46,47] for similar constructions). The main purpose of our work is to show how these discrete mappings can be described and generalized using the language of restricted Boltzmann machines (RBMs) [19,20]. As we briefly described in the introduction and show pictorially in Fig. 1, the RBM is a statistical model organized in two interconnected layers: a *visible layer* composed by a vector of N_v visible units **v** (depicted as blue dots in Fig. 1) which represent the dynamical variables whose statistics we want to model, and a *hidden layer* composed by a vector of N_h hidden units **h** (depicted as red dots in Fig. 1) which represent auxiliary variables used to create correlations among the
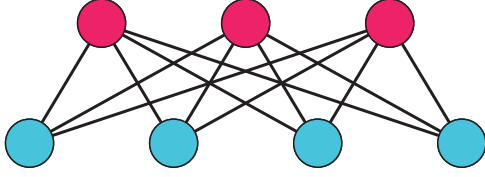
FIG. 1. Restricted Boltzmann machine; visible layer below (in blue) and hidden layer above (in red).

visible units. The RBM is an energy based statistical model, meaning that the probability distribution associated with the network can be conveniently expressed as a partition function

$$P_{\text{rbm}}(\mathbf{v}, \mathbf{h}) = \mathcal{N} \exp\left[-F_{\text{rbm}}(\mathbf{v}, \mathbf{h})\right] \quad (7)$$

with $\mathcal{N}$ a normalization constant and free energy

$$F_{\text{rbm}}(\mathbf{v}, \mathbf{h}) = \mathbf{B} \cdot \mathbf{v} + \mathbf{C} \cdot \mathbf{h} + \sum_{i=1}^{N_v} \sum_{j=1}^{N_h} W_{ij} v_i h_j. \quad (8)$$

This free energy is parametrized by the two *bias* vectors $\mathbf{B}$ and $\mathbf{C}$, which independently shift the probability density of the variables in the two layers, and the *weight* matrix W which couples the two layers. In Fig. 1 the latter is represented by the black connections. The reason for this choice as a network structure is linked to the conditional independence of the variables on the two layers which allows for an efficient sampling of the conditional probabilities

$$p_v(\mathbf{v}) = P_{\text{rbm}}(\mathbf{v}|\mathbf{h}) \quad \text{and} \quad p_h(\mathbf{h}) = P_{\text{rbm}}(\mathbf{h}|\mathbf{v}), \quad (9)$$

which can then be used to draw efficiently samples from the wanted marginal probability distribution

$$P_{\text{rbm}}^v(\mathbf{v}) = \text{Tr}_{\mathbf{h}}[P_{\text{rbm}}(\mathbf{v}, \mathbf{h})], \quad (10)$$

using block Gibbs sampling. In the expression for the marginal probability P_{rbm}^v above, we used $\text{Tr}_{\mathbf{h}}$ to denote a summation over all the possible values of the hidden variables. As we mentioned briefly in the introduction, despite its simplicity the RBM is a universal approximator [21,22], in the sense that by adding a sufficiently large number of hidden variables one can accurately approximate any marginal probability distribution P_{rbm}^v of the visible variables.

By promoting the units on which the RBM is defined to operators acting on some Hilbert space, one can also define a quantum RBM (see, e.g., [48]) allowing statistical inference on quantum states. We will defer a more in-depth discussion about this model and its possible implementation using quantum annealers to Sec. III.

The main contribution of our work is the use of the RBM network structure to define a hybrid quantum-classical architecture, where the visible unit is composed by quantum operators and the hidden units are classical discrete variables, to generalize the auxiliary-field decomposition of interactions to arbitrary many-body forces. To the best of our knowledge the closest construction to our architecture is the recently proposed map between generalized Ising models and deep Boltzmann machines presented in [49]. Our method overcomes some of the difficulties encountered in that proposal by reducing the depth of the required network (and thus simplify-

ing sampling) and removing any restriction on the numerical value of the coupling constants.

To proceed further and present the model in more details, we will now focus on the description of physical systems containing N_f species[1] of interacting fermions and discretized over N modes. General contact interactions can be expressed as powers of the density operator $\hat{\rho}_a(k)$ for mode k and species a as

$$\hat{\rho}_a(k) = c_a^\dagger(k) c_a(k), \quad \{c_a^\dagger(k), c_b(q)\} = \delta_{k,q} \delta_{a,b}, \quad (11)$$

with $c_a^\dagger(k)$ and $c_a(k)$ fermionic creation and annihilation operators and $\{\cdot, \cdot\}$ the anticommutator. For instance 2- and 3-body contact interactions can be written as

$$\hat{V}_2 = \hat{\rho}_a(k) \hat{\rho}_b(k), \quad \hat{V}_3 = \hat{\rho}_a(k) \hat{\rho}_b(k) \hat{\rho}_c(k). \quad (12)$$

Note that, due to the Pauli exclusion principle, identical fermions cannot occupy the same quantum state within the system. Thus, N_f species in a system allows for many-body forces of up to N_f for a given mode. In order to simplify the notation we will use the multi-index $\mu = (a, k)$, taking values from 1 to $M = N_f N$, and use the shorthand $\hat{\rho}_\mu$ to indicate the density operator $\hat{\rho}_a(k)$.

Using these density operators as our visible units, we can now write the free energy of our hybrid classical-quantum model as

$$F_{\text{rbm}}(\hat{\boldsymbol{\rho}}, \mathbf{h}) = \mathbf{B} \cdot \hat{\boldsymbol{\rho}} + \mathbf{C} \cdot \mathbf{h} + \sum_{\mu=1}^{M} \sum_{j=1}^{N_h} W_{ij} \hat{\rho}_\mu h_j. \quad (13)$$

In order to simplify the exposition we will consider for the moment the special case where the hidden units are binary variables $h_j = \{0, 1\}$, and generalize the construction to more general categorical variables in Sec. II C.

Before describing in detail the special case of 2- and 3-body forces, we want now to show how, by carefully choosing the parameters that define the free energy Eq. (13), we can obtain all the possible contact interactions up to the maximum order M. By tracing out the hidden layer from the total probability distribution $P_{\text{rbm}}(\hat{\boldsymbol{\rho}}, \mathbf{h})$ in Eq. (7) we obtain the following effective Hamiltonian for the visible layer:

$$H_{\text{rbm}}(\hat{\boldsymbol{\rho}}) = -\ln\left\{\text{Tr}_{\mathbf{h}} \exp\left[-F_{\text{RBM}}(\hat{\boldsymbol{\rho}}, \mathbf{h})\right]\right\}$$

$$= \mathbf{B} \cdot \hat{\boldsymbol{\rho}} + \sum_{j=1}^{N_h} \ln\left(\sum_{h_j=0}^{1} e^{(C_j + \sum_\mu^M W_{\mu j} \hat{\rho}_\mu) h_j}\right), \quad (14)$$

where we have implicitly added a (irrelevant) constant energy shift in order to cancel the normalization constant. It is now convenient to express this more compactly as follows:

$$H_{\text{rbm}} = \mathbf{B} \cdot \hat{\boldsymbol{\rho}} + \sum_{j=1}^{N_h} K_j^{(2)}\left(\sum_\mu^M W_{\mu j} \hat{\rho}_\mu\right), \quad (15)$$

---

[1]For instance $N_f = 2$ for neutrons due to the 2 possible spin projections.

where in the last line we have defined, similarly to [50], the cumulant generating function

$$K_j^{(2)}(t) = \ln\left(\sum_{h=0}^{1} e^{(C_j+t)h}\right) = \ln(1 + e^{C_j+t}), \quad (16)$$

where the superscript (2) indicates this definition is relevant to binary hidden units only (we will generalize this in Sec. II C). By performing a Taylor expansion we obtain

$$K_j^{(2)}\left(\sum_{\mu=1}^{M} W_{\mu j}\hat{\rho}_\mu\right) = \sum_{n=1}^{\infty} \frac{\kappa_{jn}^{(2)}}{n!}\left(\sum_{\mu=1}^{M} W_{\mu j}\hat{\rho}_\mu\right)^n$$

$$= \sum_{n=1}^{\infty} \frac{\kappa_{jn}^{(2)}}{n!} \sum_{k_1+\cdots+k_M=n} \binom{n}{k_1, \ldots, k_M}$$

$$\times \prod_{\mu=1}^{M} (W_{\mu j}\hat{\rho}_\mu)^{k_\mu}, \quad (17)$$

where in the second line we used the multinomial expansion and we defined the (binary) cumulants $\kappa_n^{(2)}$ as

$$\kappa_{jn}^{(2)} = \frac{d^n}{dt^n} K_j^{(2)}(t)\Big|_{t=0}. \quad (18)$$

The key observation now is noticing that, due to idempotency of the density operators, we have $\hat{\rho}_\mu^n = \hat{\rho}_\mu \; \forall n > 0$. The effective interaction coupling strengths of all the interaction terms are then given by the appropriate sums of cumulants, and can be in principle controlled by appropriately choosing the RBM parameters **B**, **C**, and W. In [50] the authors derived explicit expressions for up to 3-body interactions in this way.

In this work we use instead a different approach which allows a simpler determination of the induced coupling strengths in the visible layer by solving a linear system of equations obtained by working explicitly in the eigenbasis of the density operators. As we will see more explicitly in the next section, this approach is similar to the one used by Hirsch to find his discrete decomposition for the Hubbard interaction [31].

In the next two sections we describe in some detail our approach for the special cases of 2- and 3-body interactions and explain in Appendix A how to construct the mapping for the general case. We comment on possible extension to non-idempotent operators in Appendix C but, as we commented in the introduction, this extension is not strictly needed.

A. Two-body interactions

To set the stage we consider now the familiar case of a two-body contact interaction, and show how this can be generated by coupling a single binary auxiliary variable (hidden unit) $h \in \{0, 1\}$ to a pair of density operators (the visible units) as schematically depicted in Fig. 2. In this case the free energy of the RBM can be written as

$$F_{\text{rbm}}^{(2)}(\hat{\rho}_1, \hat{\rho}_2, h) = Ch + h\sum_{\mu=1}^{2} W_\mu\hat{\rho}_\mu, \quad (19)$$

where we neglected possible biases $\mathbf{B} = (B_1, B_2)$ on the visible layer since they only contribute to the one-body inter-
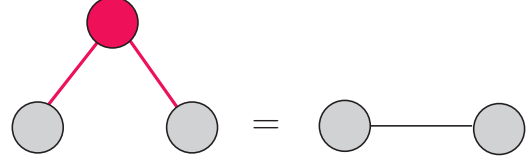

FIG. 2. Two-body interaction from the RBM mapping.

action. If needed, these biases can be used to further control the interactions generated by Eq. (19). The induced Hamiltonian in the visible layer takes the form

$$H_{\text{rbm}}^{(2)}(\hat{\rho}_1, \hat{\rho}_2) = -\ln\left(\sum_{h=0,1} e^{-F_{\text{rbm}}^{(2)}(\hat{\rho}_1,\hat{\rho}_2,h)}\right)$$

$$= A^{(2)}\hat{\rho}_1\hat{\rho}_2 + A_1^{(1)}\hat{\rho}_1 + A_2^{(1)}\hat{\rho}_2, \quad (20)$$

where a direct computation (see also Appendix A 1 for additional details) leads to the relations

$$A_\mu^{(1)} = -\ln\left(\frac{e^{-(C+W_\mu)}+1}{e^{-C}+1}\right),$$

$$A^{(2)} = -\ln\left(\frac{e^{-(C+\sum_{\mu=1}^{2}W_\mu)}+1}{e^{-C}+1}\right) - \sum_{\mu=1}^{2} A_\mu^{(1)}, \quad (21)$$

between the 3 parameters of the RBM and the 3 coupling constants of the interactions in Eq. (20). These equations can then be inverted to determine the RBM parameters needed to produce the wanted two-body term; further details on these derivations are provided in Appendix A 1.

We note at this point that the discrete Hubbard-Stratonovich transformations from Hirsch [31] are special cases of Eq. (21) for a particular choice of RBM parameters. For instance, the transformation useful for repulsive interactions (i.e., $A^{(2)} > 0$) was obtained in [31] by coupling a classical spin $\sigma = \pm 1$ to the spin density $\hat{\rho}_{\text{sp}} = \hat{\rho}_\uparrow - \hat{\rho}_\downarrow$; the resulting partition function is

$$Z_{\text{sp}} = \frac{1}{2}\sum_{\sigma=\pm} \exp(2a\sigma\hat{\rho}_{\text{sp}}) = \cosh(2a\hat{\rho}_{\text{sp}}) \quad (22)$$

where the coupling constant a is given by

$$\tanh(a)^2 = \tanh\left(\frac{A^{(2)}}{4}\right), \quad (23)$$

and Z_{sp} contains additional one-body terms corresponding to $A_1^{(1)} = A_2^{(1)} = -A^{(2)}/2$. Using the RBM model of Eq. (19) a similar transformation, with different additional one-body terms $A_1^{(1)} - 2a = A_2^{(1)} + 2a = -A^{(2)}/2$, can be obtained choosing $C = 0$ and $W_1 = -W_2 = 4a$.

Note that there is a continuous set of RBM parameters which will lead to the same two-body coupling $A^{(2)}$ and different induced one-body terms in Eq. (21). These possibly unwanted one-body contributions to the visible layer Hamiltonian can then be removed by adding the appropriate bias terms **B** in Eq. (19).

Before moving on to the results for the case of three-body forces, we want to show how the RBM mapping is not limited to contact interactions but can be immediately applied to situations where the operators associated with the visible

layer are either different idempotent operators (for instance projection operators) or, equivalently, involutive operators for which $\hat{O}^2 = \mathbb{1}$. To the latter class belong for instance the Pauli operators (and more generally any element of the Pauli group formed by arbitrary tensor products of Pauli matrices) which are routinely used to model the spin-isospin structure of nuclear interactions. For instance, consider the case of the following repulsive interaction acting on a pair of spins

$$U(\vec{\sigma}_1 \cdot \vec{\sigma}_2) = U \sum_{d=1}^{3} \sigma_d \otimes \sigma_d. \tag{24}$$

The propagator associated with it can now be written using 3 binary auxiliary variables,

$$e^{-U\alpha_t(\vec{\sigma}_1 \cdot \vec{\sigma}_2)} = \frac{e^{-3U\alpha_t}}{8} \prod_{d=1}^{3} \sum_{h_d=0}^{1} e^{a(2h_d-1)(\sigma_d \otimes \mathbb{1} - \mathbb{1} \otimes \sigma_d)}. \tag{25}$$

In the expression above α_t is the imaginary-time step size used in the calculation, $\mathbb{1}$ is the identity matrix, and a is given by Eq. (23) with $A^{(2)} = 4U\alpha_t$.

As we have seen, the complete generality of the mapping Eq. (19) allows its application to simplify a large number of pair-interacting many-body theories.

B. Three-body interactions

The presence of three-body contact interactions is of fundamental importance in low energy effective theories of nuclear physics and some bosonic cold gases in order to avoid the Thomas collapse of finite clusters [51] and ensure renormalizability of the Hamiltonian [37,52,53]. Due to the inability to use the standard Hubbard-Stratonovich or Hirsch's transformations, the inclusion of these 3-body contact interactions in auxiliary field calculations is not straightforward. Here we will briefly review the approach used in lattice effective field theory (EFT) calculations [35] as a relevant example. In order to integrate out the fermionic degrees of freedom and carry out the auxiliary field calculation one needs to decouple the interacting part of the partition function

$$Z_{\text{int}} = \exp\left(-\frac{U\alpha_t}{2}\sum_{a,b}\hat{\rho}_a\hat{\rho}_b - \frac{V\alpha_t}{6}\sum_{a,b,c}\hat{\rho}_a\hat{\rho}_b\hat{\rho}_c\right), \tag{26}$$

which contains both a two-body and a three-body interaction, in terms of one-body fermionic operators only. Similarly to the previous section, α_t is the size of the imaginary-time step.

The strategy proposed in [54] is to represent Z_{int} using a single continuous auxiliary field

$$Z_{\text{int}} = \int_{-\infty}^{\infty} dh P(h) \exp\left(h \sum_a \hat{\rho}_a\right), \tag{27}$$

for an appropriately chosen probability distribution function $P(h)$ of the auxiliary field. In [54] it was shown how one can use the solution of the truncated Hamburger moment problem to find the wanted distribution. In order for this representation to exist one needs the interaction couplings in Eq. (26) to satisfy

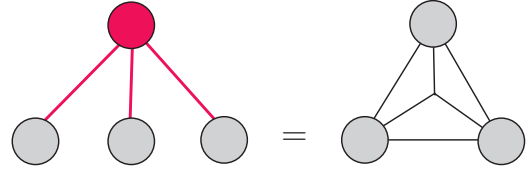$$V^2 < -2\alpha_t U^3, \tag{28}$$

FIG. 3. Three-body interaction from the RBM mapping.

which then prevents this from working for repulsive two-body couplings $U \geqslant 0$. In the more common case of an attractive two-body interaction, this mapping can be exploited only for a weak enough three-body coupling V and large enough imaginary-time step α_t. Notably, for other classes of QMC methods based on auxiliary fields such as auxiliary field diffusion Monte Carlo (AFDMC; see, e.g., [28]) this is not possible in general due to the inability to introduce explicitly an energy cutoff as is done in a lattice formulation. We note in passing that an alternative auxiliary field representation of the partition function of Eq. (26) can be obtained using the DBM mapping introduced in [49]. Also in this case the mapping can be performed only for weak enough three-body interaction, $|V| < -3U$, and attractive two-body coupling, $U < 0$.

Thanks to the flexibility of the RBM mapping scheme we can instead treat the two- and three-body interactions separately using different hidden variables. In the rest of this section we focus then directly on a mapping that reduces the three-body term by using a single binary unit coupled to three density operators as depicted in Fig. 3. The resulting free energy of the RBM is similar to Eq. (19) and reads

$$F_{\text{rbm}}^{(3)}(\hat{\rho}_1, \hat{\rho}_2, \hat{\rho}_3, h) = Ch + h \sum_{\mu=1}^{3} W_\mu \hat{\rho}_\mu, \tag{29}$$

while the target Hamiltonian can be written as

$$H_{\text{rbm}}^{(3)}(\hat{\boldsymbol{\rho}}) = A^{(3)}\hat{\rho}_1\hat{\rho}_2\hat{\rho}_3 + \sum_{\mu<\nu}^{3} A_{\mu\nu}^{(2)}\hat{\rho}_\mu\hat{\rho}_\nu + \sum_{\mu=1}^{3} A_\mu^{(1)}\hat{\rho}_\mu. \tag{30}$$

Since we are only interested in producing a target three-body term $A^{(3)}$ while removing the unwanted interactions using lower order mappings, it is convenient to simplify this to

$$H_{\text{rbm}}^{(3)}(\hat{\boldsymbol{\rho}}) = A^{(3)}\hat{\rho}_1\hat{\rho}_2\hat{\rho}_3 + A^{(2)}\sum_{\mu<\nu}^{3}\hat{\rho}_\mu\hat{\rho}_\nu + A^{(1)}\sum_{\mu=1}^{3}\hat{\rho}_\mu, \tag{31}$$

with mode-independent one- and two-body couplings. In the same way we set $W_\mu = W$ in Eq. (29). As we did for the two-body interactions above, a direct calculation leads to the following relations between the physical couplings and the RBM parameters (see also Appendix A for the general case):

$$A^{(1)} = -\ln\left(\frac{e^{-(C+W)}+1}{e^{-C}+1}\right),$$

$$A^{(2)} = -\ln\left(\frac{e^{-(C+2W)}+1}{e^{-C}+1}\right) - 2A^{(1)},$$

$$A^{(3)} = -\ln\left(\frac{e^{-(C+3W)}+1}{e^{-C}+1}\right) - 3A^{(2)} - 3A^{(1)}. \tag{32}$$

As before, in order to find the needed RBM parameters these relations need to be inverted and, depending on the sign of the physical three-body coupling, one needs to make appropriate choices for the range of the RBM parameters. For instance, if we are interested in an attractive interaction, $A^{(3)} < 0$, we can set $C = -2W$, while in the case of a repulsive interaction, $A^{(3)} > 0$, we can take $C = -W$. In both cases we obtain the same magnitude for the three-body coupling, $|A^{(3)}| = \ln[\cosh^4(\frac{W}{2})\operatorname{sech}(W)]$.

The two-body interaction $A^{(2)}$ can be removed by expressing it in terms of additional auxiliary variables using the identities in Eq. (21). Note that with this technique no constraint like Eq. (28) needs to be imposed on the value of the physical couplings. More details on these derivations, and their extension to the case of categorical hidden units, are provided in Appendix A 1.

We now turn to the discussion of categorical variables and provide a justification for their use in practical applications.

C. Categorical hidden variables

We proceed now to generalize the RBM mappings derived above for hidden binary variables to the more general case of categorical hidden variables which take values on a larger range $\{0, 1, 2, \ldots, \mathcal{K} - 1\}$. From the form of the general RBM free energy Eq. (13) we can expect that the effect of increasing the magnitude of the hidden variable, controlled by $\mathcal{K}$, is to correspondingly increase the energy term proportional to $\mathbf{C}$ and W. This, in turn, will increase the magnitude of the induced physical coupling at fixed RBM parameters. We expect this property to be useful for instance in the sampling process as it allows for smaller energy gaps when updating the hidden layer. Another instance when this could possibly be useful is to minimize the systematic error introduced by Trotter like decompositions of the imaginary-time propagator (see, e.g., [28]) but we leave a more detailed exploration of these possibilities to future work.

The modified expression for the induced one- and two-body coupling when coupling a pair of visible units with a single categorical hidden variable, corresponding to the generalization of Eq. (21), can be compactly written as

$$A_\mu^{(1)} = -\ln\left(\frac{e^{-\mathcal{K}(C+W_\mu)} - 1}{e^{-(C+W_\mu)} - 1}\right) + \ln\left(\frac{e^{-\mathcal{K}C} - 1}{e^{-C} - 1}\right), \quad (33)$$

for the one-body term, while for the two-body we have

$$A_\mathcal{K}^{(2)} = -\ln\left(\frac{e^{-\mathcal{K}(C+\sum_{\mu=1}^2 W_\mu)} - 1}{e^{-(C+\sum_{\mu=1}^2 W_\mu)} - 1}\right)$$
$$+ \ln\left(\frac{e^{-\mathcal{K}C} - 1}{e^{-C} - 1}\right) - \sum_{\mu=1}^2 A_\mu^{(1)}. \quad (34)$$

Note that whenever the argument of the exponentials in Eq. (33) and Eq. (34) becomes 0, one needs to take the limit continuously. To show the effect of increasing the range of the hidden variable on the induced coupling constant, we plot in Fig. 4 the magnitude of the two-body coupling as a function of $\mathcal{K}$ for two one-parameter families of RBMs:
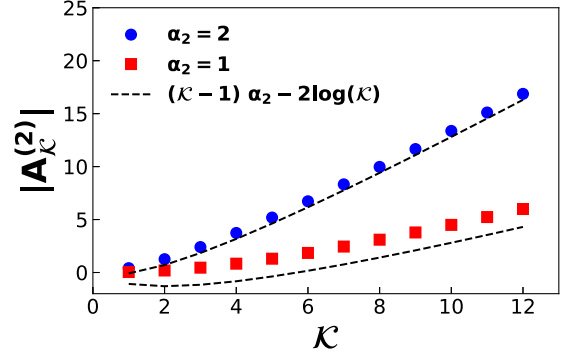

FIG. 4. Two-body coupling as a function of $\mathcal{K}$. The dashed line corresponds to the lower bound in Eq. (37).

(i) for attractive interactions $A_\mathcal{K}^{(2)} < 0$ we take

$$W_1 = W_2 = -C = \alpha_2 > 0; \quad (35)$$

(ii) for repulsive interactions $A_\mathcal{K}^{(2)} > 0$ we take

$$W_1 = -W_2 = \alpha_2 > 0, \quad C = 0. \quad (36)$$

Note that these choices were motivated only by the need to obtain all possible two-body coupling by tuning a single parameter α_2 so that we could present the results compactly in Fig. 4. This is therefore a completely general construction and it is possible that similar results could be obtained by performing different choices.

As the results show, the magnitude $|A_\mathcal{K}^{(2)}|$ increases almost linearly with $\mathcal{K}$ for a fixed choice of the RBM parameter α_2. As an easier proxy to understand the expected behavior, and especially the apparent asymptotically linear growth with $\mathcal{K}$, we also show in Fig. 4 the following lower bound,

$$\left|A_\mathcal{K}^{(2)}\right| \geqslant (\mathcal{K} - 1)\alpha_2 - 2\ln(\mathcal{K}), \quad (37)$$

valid for both parametrizations Eqs. (35) and (36). Indeed we find the approximately linear increase with $\mathcal{K}$ together with a mild logarithmic correction. A complete proof of this lower bound, and the one for three-body interactions, is provided Appendix B.

Next, we turn our attention to the more complicated case of the three-body interaction induced by a single categorical variable. The mapping between physical coupling constants and RBM parameters remains the same as Eq. (33) at the one-body level while the two- and three-body couplings become (see also Appendix A 1 for a full derivation)

$$A_{\mu\nu}^{(2)} = -\ln\left(\frac{e^{-\mathcal{K}(C+W_\mu+W_\nu)} - 1}{e^{-(C+W_\mu+W_\nu)} - 1}\right)$$
$$+ \ln\left(\frac{e^{-\mathcal{K}C} - 1}{e^{-C} - 1}\right) - A_\mu - A_\nu,$$

$$A_\mathcal{K}^{(3)} = -\ln\left(\frac{e^{-\mathcal{K}(C+W_1+W_2+W_3)} - 1}{e^{-(C+W_1+W_2+W_3)} - 1}\right)$$
$$+ \ln\left(\frac{e^{-\mathcal{K}C} - 1}{e^{-C} - 1}\right) - \sum_{\mu=1}^3 \sum_{\nu>\mu}^3 A_{\mu\nu} - \sum_{\mu=1}^3 A_\mu, \quad (38)$$

and similarly to above the induced physical coupling grows approximately linearly with $\mathcal{K}$. This is shown in Fig. 5 where
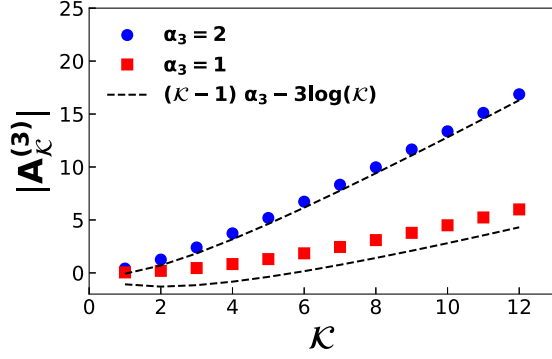
FIG. 5. Three-body coupling as a function of $\mathcal{K}$. The dashed line corresponds to the lower bound in Eq. (39).

we plot both the actual magnitude $|A_{\mathcal{K}}^{(3)}|$ and the lower bound

$$\left|A_{\mathcal{K}}^{(3)}\right| \geqslant (\mathcal{K} - 1)\alpha_3 - 3\ln(\mathcal{K}) \tag{39}$$

for two values of the RBM parameter, $\alpha_3 = 1, 2$. In the three body case we have also chosen a simple one-parameter family of RBMs given by the following choices:

(i) for attractive interactions $A_{\mathcal{K}}^{(3)} < 0$ we take

$$W_1 = W_2 = W_3 = \alpha_3 > 0, \quad C = -2\alpha_3; \tag{40}$$

(ii) for repulsive interactions $A_{\mathcal{K}}^{(3)} > 0$ we take

$$W_1 = W_2 = W_3 = \alpha_3 > 0, \quad C = -\alpha_3. \tag{41}$$

As for the case of the two-body term in Fig. 5, we see that the lower bound is relatively tight. A complete derivation of Eq. (39) is also provided in Appendix B.

Similarly to the results described in the previous two sections [see Eq. (25)], this more general construction can be easily extended to many-body operators composed by involutive operators with minimal changes.

Before concluding this section, we want to also point out that it is relatively straightforward to use the expansion of the cumulant generating function to get the effective Hamiltonian for the visible layer as proposed in [50] and explained in detail at the beginning of Sec. II. Since this can be of more general interest, we provide here the necessary generalization to the cumulant generating function of Eq. (16) which for categorical variables becomes

$$K_j^{(\mathcal{K})}(t) = \ln \left\{ \sum_{h=1}^{\mathcal{K}-1} \exp\left[-h(C_j + t)\right] \right\}$$
$$= \ln \left\{ \frac{1 - \exp\left[-\mathcal{K}(C_j + t)\right]}{1 - \exp\left[-(C_j + t)\right]} \right\}. \tag{42}$$

After this change the polynomial expansion from Eq. (17) remains essentially equivalent, with the only difference that the value of the cumulants $\kappa_{jn}^{(\mathcal{K})}$ will now need to be extracted by taking appropriate derivatives of $K_j^{(\mathcal{K})}(t)$.

III. MANY-BODY FORCE IMPLEMENTATION ON QUANTUM ANNEALERS

Quantum annealing is an algorithm for finding approximate solutions to nonconvex optimization problems using a quantum generalization of classical simulated annealing [55–57]. This approach has been applied in the past to hard optimization problems ranging from computing Ramsey numbers [58] to finding low energy configurations of proteins [59]. Used as an analog simulator, a quantum annealer was also recently used to study phase transitions in lattice models [60].

The underlying idea of quantum annealing based optimization is to first encode the solution of the problem at hand in the ground state of a k-local Hamiltonian acting on spins. This Hamiltonian is then simulated by a physical system that is cooled down to reach a low energy state which provides a good approximate solution to the optimization problem. In order to decrease the time required to find a solution of this approach, the physical system is initialized as the ground state of an auxiliary Hamiltonian H_A and then adiabatically evolved to the ground state of the problem Hamiltonian H_P. If this is done slowly enough to avoid nonadiabatic transitions, the final state will be the ground state of H_P corresponding to the optimal solution [61]. In some situations fast nonadiabatic transitions can also be used to accelerate the convergence to a close approximation to the final ground state [62,63].

Here we will focus more specifically on the class of quantum annealers that uses a tunable transverse field Ising model to implement this idea. An example of this is the device manufactured by D-Wave Systems [64]. The time-dependent spin Hamiltonian is given by

$$\mathcal{H}(\tau) = B_x(\tau)H_A + B_z(\tau)H_P,$$
$$H_A = -\sum_i \sigma_i^x,$$
$$H_P = \sum_{ij} J_{ij}\sigma_i^z\sigma_j^z + \sum_i h_i\sigma_i^z, \tag{43}$$

where $\sigma_i^{x,z}$ are Pauli matrices that operate on spin or qubit i. Note that the optimization problem is encoded in the diagonal Hamiltonian H_P by tuning appropriately both the pairwise couplings J_{ij} as well as the on-site fields h_i. The auxiliary Hamiltonian H_A is chosen for two reasons: its ground state is a complete superposition of all the basis states and it is easy to prepare by applying a physical transverse field. The time evolution is then performed by modifying the applied magnetic fields B_x and B_z starting from $B_x(0) \ll B_z(0)$ and evolving to $B_x(1) \gg B_z(1)$. Here $\tau = t/t_a$, where t is the physical time and t_a is the annealing time: the time that it takes to perform the transition from the Hamiltonian H_A into the problem Hamiltonian H_P.

Due to the presence of two-body spin coupling in H_P, this system can be employed to approximately solve quadratic unconstrained binary optimization (QUBO) problems. The goal of this section is to show how the RBM mapping we derived in this work can be used to increase the representation power of the problem Hamiltonian H_P by implementing higher-order diagonal interactions using some of the available qubits as auxiliary spins in the hidden layer. This process of reducing many-body interactions to two-body terms has been extensively explored in the past (see, e.g., [65–67] and [68] for a recent summary). Known mappings are limited in their application to specific signs of the many-body coupling, for instance only interactions with positive cubic terms are known
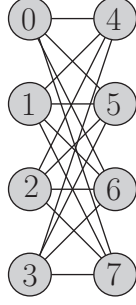
FIG. 6. Cell of Chimera graph in D-Wave systems.

to embed into the Chimera topology without using additional auxiliary qubits for the embedding [68]. Here we show that our RBM construction can be applied at the same cost for cubic terms of both signs. The technique can also be easily extended to higher-body interactions.

For illustration purposes, we will consider a simple extension to the QUBO model to contain also cubic terms implemented as effective couplings among three spins. In particular consider the following cost function,

$$E = a \sum_i^3 q_i + b \sum_{i<j}^3 q_i q_j + c\, q_1 q_2 q_3, \qquad (44)$$

with only 3 binary variables q_i. This can be converted into a generalized classical Ising model by introducing the spin variables $s_i = 2q_i - 1$, and then mapped to a diagonal quantum Hamiltonian acting on physical spins ($s_i \to \sigma_i^z$):

$$H = A \sum_i^3 \sigma_i^z + B \sum_{i<j}^3 \sigma_i^z \sigma_j^z + C \sigma_1^z \sigma_2^z \sigma_3^z,$$

$$A = 2(a+b-c), \quad B = 4(b-c), \quad C = 8c. \qquad (45)$$

Using the notation from Ref. [68], this Hamiltonian corresponds to the K_4 gadget.

Due to hardware restrictions the topology of the interactions between qubits is restricted to only a finite set. In this exploratory work we consider the Chimera graph [69] found on D-Wave devices; the scheme can however be generalized to different topologies. In the Chimera graph topology, all qubits in one partition are linked to the qubits in the other partition but not among themselves as shown in Fig. 6.

Given the coupling map in Fig. 6, we can implement the K_4 gadget using 4 ancilla qubits:

(i) qubits 1, 2, 3 represent the logical qubits;
(ii) qubit 4 mediates the pair interaction between (1,3);
(iii) qubit 5 mediates the pair interaction between (2,3);
(iv) qubit 6 mediates the pair interaction between (1,2);
(v) qubit 7 mediates the cubic interaction.

This is also depicted in Fig. 7. The full Hamiltonian takes the following form,

$$H_P = W_1 \sum_{i=1}^3 \sigma_i^z - \mathrm{sgn}(C) W_3 \sigma_7^z$$

$$+ |W_2| \big[\sigma_1^z + \mathrm{sgn}(A_3)\sigma_2^z \big] \sigma_4^z$$

$$+ |W_2| \big[\sigma_2^z + \mathrm{sgn}(A_3)\sigma_3^z \big] \sigma_5^z$$
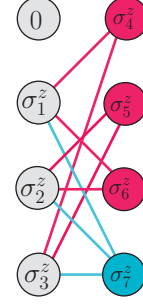

FIG. 7. Three-body interactions implemented on the Chimera topology.

$$+ |W_2| \big[\sigma_1^z + \mathrm{sgn}(A_2)\sigma_2^z \big] \sigma_6^z$$

$$+ W_3 \sum_{i=1}^3 \sigma_i^z \sigma_7^z, \qquad (46)$$

where the first line represents on-site energy shifts, the second through fourth represent the pair interactions induced by qubits 4–6, and the last one contains the pair interactions with qubit 7 responsible for the three-body term. The coupling terms can then be obtained using the RBM mapping described above (see also Appendix A). In practice we first compute the coefficient W_3 from the relation

$$|C| = \tfrac{1}{8} \ln[\cosh^4(2W_3)\,\mathrm{sech}(4W_3)], \qquad (47)$$

and the other parameters can then be found using

$$A_2 = \frac{\mathrm{sgn}(C)}{4} \ln[\cosh(4W_3)] - 2B,$$

$$W_2 = \frac{\mathrm{sgn}(A_2)}{2} \ln[\cosh(A_2)],$$

$$W_1 = -\frac{\mathrm{sgn}(C)}{8} \ln[\cosh(4W_3)] - A. \qquad (48)$$

This mapping was derived from Eqs. (A9) and (A14) by replacing 0 with -1. If we trace out the auxiliary qubits belonging to the hidden layer we recover the effective Hamiltonian Eq. (45). Furthermore, one could reduce the ancilla requirement even further by using a single ancilla to embed direct pair interactions between qubit 1 and qubits 2 and 3 while using a single ancilla to represent the missing two-body term between qubit 2 and qubit 3 (qubit 5 in the mapping above). This would have the same qubit overhead as the best method from Ref. [68] while requiring only one ferromagnetic coupling for the embedding.

Note that the RBM identities are valid for any direct product of Pauli matrices $\sigma^{x,y,z}$ as we discussed in the introduction. This means that extensions like the one presented here for the simple Hamiltonian Eq. (43) can be generalized to the interesting situation where non-stoquastic physical interactions are available.

In practical applications an effective temperature T_{eff} is usually introduced to properly describe the generated statistical distribution [70,71]. Since the effective Hamiltonian is obtained by tracing over the global partition function including the auxiliary spins, one will need to estimate T_{eff} first

(see, e.g., [72]) in order to appropriately tune the RBM parameters. We leave the exploration of these issues for future work.

Before concluding we note that the RBM mapping derived in this paper can also be used to show the equivalence between various gadgets presented in Ref. [68]. For instance, it is simple to use the RBM mapping to show that the gadget $K_6 - e$ is equivalent, upon appropriate rescaling of the interactions, to the simpler gadget K_5.

IV. ON TRAINING THE RBM FOR PHYSICAL SYSTEMS

In this final section we turn our attention to machine learning, specifically on training neural architectures for physical systems with many-body interactions. The relative entropy is commonly used in the literature as an objective function which measures the discrepancy between the target data distribution and the approximation given by the neural architecture; the smaller the discrepancy, the smaller its value. For energy based models like the RBM, the probability of the model depends on the accurate evaluation of the partition function in Eq. (1), which is intractable as it requires all possible data samples. The inability to have a reliable estimate of the relative entropy makes it impossible to get its gradients with respect to the RBM parameters. These gradients are needed to update the parameters toward optimal values. In practice, approximate gradients are obtained through contrastive divergence (CD) [73]. This procedure was employed in [50] to learn the Ising model from samples generated with the Swendsen-Wang algorithm. However, CD might not always be a good approximation [74,75] and the lack of a point estimator for an objective function makes it harder to assess whether training has converged. In addition, in many cases of interest, Monte Carlo sampling can be highly autocorrelated, so we might not be able to find algorithms that provide samples which reliably represent the physical distribution.

In this section we provide several suggestions for how to improve training. We can use our knowledge of the original Hamiltonian to design a network with only a small number of parameters, independent of system size. As a first step, borrowing from the results presented in previous sections, we construct a sparse RBM whose architecture is based on the interaction order in the physical Hamiltonian. For each interaction in the physical system there is a corresponding auxiliary variable. This architecture is compared to an RBM with the number of auxiliary variables equal to the number of visible variables with all-to-all connection between the two layers.

We then devise an objective function that does not depend on the partition function and is inspired by Monte Carlo acceptance and rejection method. As we demonstrate through experiments, this objective function makes training much easier as it makes use of our understanding of the physical system. Furthermore, since it can be easily evaluated during training, this objective function can serve as an indicator for convergence.

Lastly, we explore the scenario when we cannot rely on Monte Carlo samples collected for training and show that learning is still possible.

TABLE I. Numerical values of the physical couplings for the Ising model of Eq. (50).

Coupling	Numerical Value
Λ_X	0.017857
Λ_T	2.01273

As an illustration, we consider a classical anisotropic 2D Ising model in the eigenbasis of $\sigma^z |s\rangle = s|s\rangle$,

$$Z_{2D} = \mathcal{N} \exp \left(\sum_{i=1}^{L_X} \sum_{j=1}^{L_T} \Lambda_X s_{i,j} s_{i+1,j} + \Lambda_T s_{i,j} s_{i,j+1} \right). \quad (49)$$

This partition function approximates the 1D Ising model with a transverse field through Trotter expansion [76],

$$H_{1D} = -J \sum_i \sigma_i^z \sigma_{i+1}^z - B \sum_i \sigma_i^x,$$

$$\Lambda_X = \frac{\beta}{L_T} J, \quad \Lambda_T = \frac{1}{2} \ln \left[\coth \left(\frac{\beta}{L_T} B \right) \right]. \quad (50)$$

The Hamiltonian of this system is a simpler version of Eq. (43). L_X is the length of the Ising chain, and L_T is the number of Trotter steps and is the second dimension in the classical model. From Appendix A 1 we can derive the mapping between the physical couplings $\Lambda_{X,T}$ and the respective RBM parameters.

To verify the RBM learning procedure we compare the relative error in reproducing these couplings, $\delta\Lambda = |\frac{\Lambda - \Lambda_{RBM}}{\Lambda}|$. The results from [50] show that CD training can have a good performance for the Ising model, but this is not guaranteed in more complicated settings.

Following the procedure outlined in [50], we generate data of 10^5 samples with $L_X = L_T = 28$ and $J = B = 2\beta = 1$ through local Monte Carlo sampling and train in batches of 500 samples with a learning rate of 0.001 using CD with 10 Gibbs updates. The values of the physical couplings are given in Table I. To make sure there is little autocorrelation in the data, we collected a sample every 10^3 Monte Carlo updates.

In Fig. 8 we plot both the average error in energy per configuration and the obtained couplings as function of
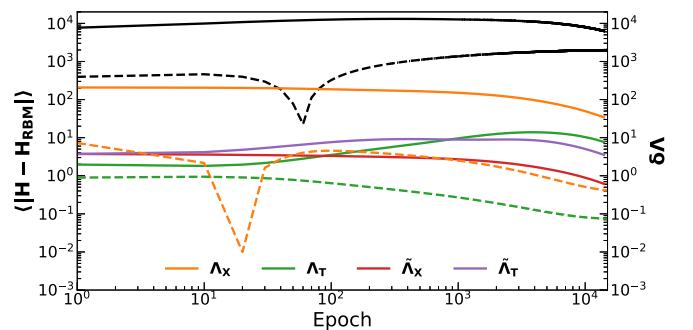

FIG. 8. Training by CD of fully connected (solid lines) and sparse (dashed lines) RBM; the average energy difference in black, and relative errors in couplings $\Lambda_{X,T}$. For the fully connected RBM we also show the average values for spurious couplings denoted by $\tilde{\Lambda}$.
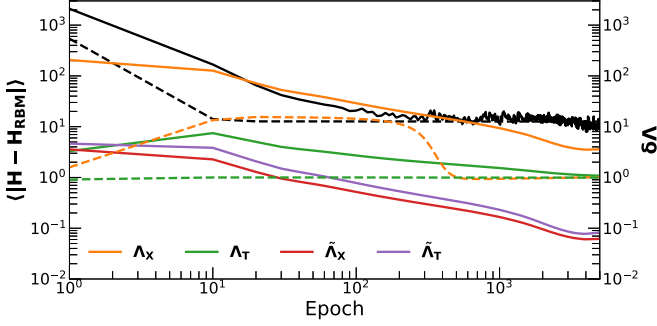
FIG. 9. Training by minimizing the difference in energy of fully connected (solid lines) and sparse (dashed lines) RBM; the average energy difference in black, and relative errors in couplings $\Lambda_{X,T}$. For the fully connected RBM we also show the average values for spurious couplings denoted by $\tilde{\Lambda}$.
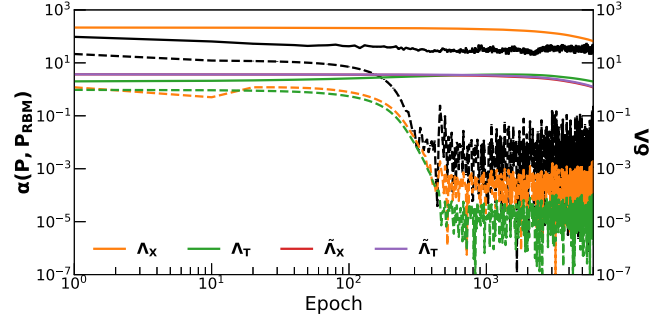

FIG. 10. Training by minimizing the relative importance of fully connected (solid lines) and sparse (dashed lines) RBM; the average energy difference in black, and relative errors in couplings $\Lambda_{X,T}$. For the fully connected RBM we also show the average values for spurious couplings denoted by $\tilde{\Lambda}$.

training epochs. One epoch consists of batch training the entire data once. As Fig. 8 shows, while the error in the couplings decreases with training, it is about two orders of magnitude lower for the sparse RBM. In addition, while in the Ising model there are only interactions among neighboring spins, the fully connected RBM includes couplings between spins that are distant and which should decrease to 0 when the training converges. These are denoted by $\tilde{\Lambda}$ in Fig. 8 and are set to 0 by default for the sparse RBM which has auxiliary variables coupled only to neighboring spins. While the architecture choice greatly helped the training process, we still need to find an objective function we can easily compute and optimize for.

Having access to the probability (up to an overall norm) of a given configuration S for the physical Hamiltonian allows us to cast the optimization process as supervised learning by considering this probability as a label to be learned by the neural architecture. In [11] the authors train the RBM by minimizing the difference in free energies, which is the difference between the Hamiltonian of the physical system and the RBM Hamiltonian for the visible layer, $\sum_S |H(S) - H_{\text{RBM}}(S)|$. Due to the similarity in functional form between the two, for the system studied in [11] the authors were able to engineer the one-body coupling for the visible layer by directly matching to the physical system and the other parameters were optimized by minimizing the objective function. As an overall constant shift in the energy has no impact on the physics, the authors added a constant to the energies of the physical system to make them non-negative. In general, one is not able to perform a direct matching like in [11], so we opted to implement the same objective function and training procedure, but we do not engineer any RBM parameter. We trained the RBM using various constant energy shifts and the results displayed here are the best ones obtained, for which no energy shift was used.

The training in Fig. 9 is faster than CD training in Fig. 8 as the average energy difference decreases rapidly until it reaches a steady value. In addition, spurious interactions $\tilde{\Lambda}_{X,T}$ decrease rapidly as well. However, the errors in the couplings $\Lambda_{X,T}$ do not decrease to 0. To improve convergence using this objective function, one would need to also optimize for the constant energy difference, either through grid search through all possible values or by other means. We recall that the con-

stant energy difference is related to an overall normalization factor in the RBM which can be computed exactly only by summing the probabilities of all configurations. In our experiment there are many of them, $2^{L_X \times L_T} \gtrsim 10^{236}$, and systems of physical interest are typically much larger.

To remove any dependence on the normalization factors, we introduce an objective function which we call relative importance,

$$
\alpha(P, P_{\text{RBM}}) = \sum_{S \epsilon \pi_S} \sum_{S' \epsilon \pi_{S'}} \left| \ln \left(\frac{P(S') P_{\text{RBM}}(S)}{P_{\text{RBM}}(S') P(S)} \right) \right|
$$
$$
= \sum_{S \epsilon \pi_S} \sum_{S' \epsilon \pi_{S'}} |\Delta H(S, S') - \Delta H_{\text{RBM}}(S, S')|,
$$
(51)

where $\Delta H(S, S') = H(S) - H(S')$ is the difference in energy between two configurations. We sample S, S' from two separate uniform distributions $\pi_S \neq \pi_{S'}$ defined over disjoint subsets of the training data set Ω,

$$
\Omega_S \cup \Omega_{S'} = \Omega, \quad \Omega_S \cap \Omega_{S'} = \emptyset. \tag{52}
$$

If the two distributions were equal, the objective function would be identically 0, so they must be chosen to be different. In our experimental setup, for simplicity, we pick a random batch in the beginning of the training to be the support of $\pi_{S'}$ and the rest of the data is reserved for π_S. The ratio inside the logarithmic function is the Metropolis-Hastings acceptance rate to apply the Monte Carlo update from S to S' using the RBM distribution to sample the new configuration [11]. In Fig. 10 we plot the error in relative importance and errors in couplings during training. Similarly to the previous training experiments, the sparse RBM produces much smaller errors than the fully connected RBM. In addition, the training for the sparse RBM is quite rapid and unlike the free energy difference, the relative importance error decreases rapidly to 0 (numerically almost 0) at the same time as the errors in the coupling decrease to 0. By combining a sparse RBM with our new objective function we have obtained complete learning and done so rather quickly in the number of training epochs.

To compare how the 3 optimization methods perform, in Fig. 11 we plot the relative error in the physical couplings for the sparse RBM for each of them. In all cases we have
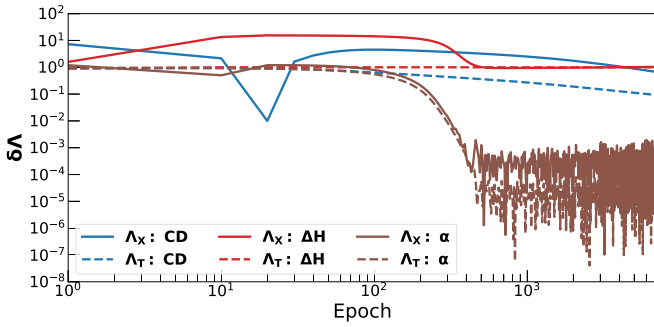
FIG. 11. Relative errors in couplings $\Lambda_{X,T}$ for the sparse RBM for each of the 3 optimization methods discussed in this section.

employed the same batch size and learning rate. Relative importance training converges quickly to the exact values while the two other methods either fail to converge or require training for many more epochs.

We conclude this section with a consideration on the data used for training, which is usually assumed to represent the physical distribution one is trying to learn. However, Monte Carlo sampling used to generate the data might suffer from autocorrelation which invalidates this assumption. In [50] this was circumvented by employing the the Swendsen-Wang cluster algorithm. In our case we had to run very long local Monte Carlo chains and discard more than 99% of the samples. In cases of practical interest, one might not have any clustering algorithm available and running long Monte Carlo chains to be discarded afterwards is computationally expensive. Since in our calculations we can evaluate the target distribution for any configuration S (up to some normalization factor), the role of the training set is redundant when using the relative importance as objective function. In fact, we can even take the sampling distributions π_S and $\pi_{S'}$ that define the relative importance to be uniform over the whole configuration space instead of just the training set Ω. As a proof of concept, we perform the optimization by uniformly sampling the full configuration space with $\pi_{S'}$ having a smaller support than π_S. In Fig. 12 we plot the error in relative importance for a sparse RBM by uniform sampling the configuration space. The inset displays the objective function and Gaussian fit ($\mathcal{N}e^{-ax^2}$) for training from the data collected with local Monte Carlo up-

dates and from uniform sampling.[2] As the figure shows, this procedure works quite well and the results are very similar (almost identical) to those obtained by the the sparse RBM with preprocessed data displayed in Fig. 10. The Hamiltonian studied in this section is rather simple and mainly serves as an illustration since we can derive an exact RBM mapping of the physical couplings. In complicated settings where such mapping might not be feasible, the RBM can still be a useful approximation and the respective parameters can be obtained by the optimization procedure outlined here.

V. CONCLUSION

Computational methods based on neural networks have seen many successes in computer vision and natural language processing in recent years. In this work we used a particularly simple architecture, the restricted Boltzmann machine, as a tool to represent the partition function of fermionic systems with many-body interactions in terms of noninteracting fermions coupled to an external auxiliary field. This generalized Hubbard-Stratonovich transformation is obtained by introducing a hybrid quantum-classical RBM model where the visible layer is composed by quantum operators while the hidden layer represents the (classical) auxiliary fields.

Due to their importance in effective field theories in nuclear physics, we focus most of our discussion in Sec. II to contact interactions among fermions like Eq. (5) but also discuss in detail how this can be generalized to any interaction of the form Eq. (4) (see, e.g., the discussion at the end of Sec. II A).

In order to provide support to the implementation of the RBM-based scheme we propose here, we provide details of the construction for the most common situation of two- and three-body interactions in Secs. II A and II B, respectively (see also Appendix A for the general case).

Thanks to the generality of the scheme, we expect the identities we have obtained here to prove useful in applications to quantum Monte Carlo simulations based on auxiliary fields like lattice-EFT and AFDMC.

An interesting application of these ideas is in finding efficient mappings of optimization problems into the native Hamiltonian of a quantum annealer as we discuss in Sec. III. Since the mapping is at the level of partition functions, robust techniques to extract the effective temperature will need to be used in this case.

In situations where exact representations cannot be used, we hope the optimization procedure outlined in Sec. IV could be helpful in finding useful approximations.
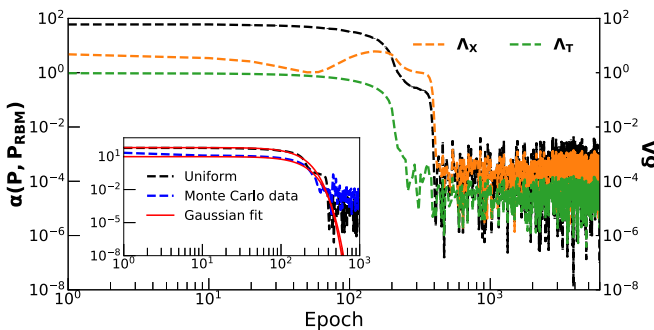

FIG. 12. Training by minimizing the relative importance of the sparse RBM; the average energy difference in black, and relative errors in couplings $\Lambda_{X,T}$.

---

[2]We found $a = 6.5 \times 10^{-5}$ for uniform sampling and $a = 5.3 \times 10^{-5}$ for data collected from local MC updates. These values depend on both the underlying Hamiltonian and optimization hyperparameters such as batch size and learning rate.

APPENDIX A: MAPPING BETWEEN THE RBM AND MANY-BODY INTERACTIONS

In this Appendix we provide further details on the approach we followed in producing the RBM representations for specific many-body interactions. Due to the fact that our visible layer is composed of operators, and therefore the RBM mapping is at the level of the quantum partition function, it is useful to fix a basis to work with. After the mapping is established for every element of the complete basis, we are guaranteed that it will work on any superposition state living in the same Hilbert space. A convenient basis to carry out the calculations is, not surprisingly, the eigenbasis of the operators that define the visible layer. For fermionic density operators, which are the main focus of this work, this is the Fock, or occupation number, basis for which

$$\hat{\rho}_k|\boldsymbol{\rho}\rangle = \rho_k|\boldsymbol{\rho}\rangle, \quad \rho_k = 0, 1. \quad \text{(A1)}$$

Here the number of possible orbitals, elements in the real vector $\boldsymbol{\rho}$, is M as in the main text. As we have discussed in Sec. II of the main text it is easy to see that, upon tracing out the hidden layer of the RBM, the induced Hamiltonian in the visible layer,

$$H_{\text{rbm}}(\hat{\boldsymbol{\rho}}) = -\ln\{\text{Tr}_{\mathbf{h}}\exp[-F_{\text{RBM}}(\hat{\boldsymbol{\rho}}, \mathbf{h})]\}, \quad \text{(A2)}$$

contains all possible terms up to the maximum M-body interaction. This can be shown for instance using the cumulant generating function. Here we will rewrite $H_{\text{rbm}}(\hat{\boldsymbol{\rho}})$ in a form which will simplify further manipulations. If we denote by

$C(M, k)$ the set of k combinations from the set of orbital indices $I = \{1, \dots, M\}$, we can express the Hamiltonian as

$$H_{\text{rbm}}^A(\hat{\boldsymbol{\rho}}) = \sum_{\mu=1}^{M} \sum_{P \in C(M,\mu)} A_P^{(\mu)} \prod_{\nu \in P} \hat{\rho}_\nu, \quad \text{(A3)}$$

which makes evident the presence of $2^M - 1$ coupling terms. Note that we removed an irrelevant constant energy shift from Eq. (A3) since it only affects the overall normalization of the partition function.

We can also express the right hand side of Eq. (A2) explicitly in terms of the RBM parameters; for a single hidden variable we find

$$H_{\text{rbm}}^B(\hat{\boldsymbol{\rho}}) = -\ln\left(\sum_{h=0}^{\mathcal{K}-1} e^{h\left(C + \sum_{\mu=1}^{M} W_\mu \hat{\rho}_\mu\right)}\right), \quad \text{(A4)}$$

where we set the visible biases $\mathbf{B}$ to 0 as it is easy to reintroduce them if need be. As discussed at the beginning of this section, we will now match these two energy functionals H_{rbm}^A and H_{rbm}^B on the states that span the fermionic Fock space. This allows us to replace the vector of operators $\hat{\boldsymbol{\rho}}$ with a vector of eigenvalues $\boldsymbol{\rho}$. Furthermore, if we want the RBM to represent the physical system exactly, the two energy functionals can only differ by a constant. In order to take this possible difference into account, we introduce the shift $A^{(0)}$ in Eq. (A3) and fix its value by performing the matching in the vacuum state $|0\rangle$. This immediately yields

$$A^{(0)} = -\ln\left[\sum_{h=0}^{\mathcal{K}-1} \exp(hC)\right]. \quad \text{(A5)}$$

The matching condition becomes now

$$\sum_{\mu=1}^{M} \sum_{P \in C(M,\mu)} A_P^{(\mu)} \prod_{\nu \in P} \rho_\nu = -\ln\left\{\frac{\sum_{h=0}^{\mathcal{K}-1} \exp\left[h\left(C + \sum_{\mu=1}^{M} W_\mu \rho_\mu\right)\right]}{\sum_{h=0}^{\mathcal{K}} \exp(hC)}\right\}, \quad \text{(A6)}$$

where for simplicity we have subtracted the constant term $A^{(0)}$ on both sides. For practical purposes we can use this compact relation to construct a simple linear system for the unknown induced couplings $A_P^{(\mu)}$ using the remaining $2^M - 1$ basis states (note that for the applications discussed in the main text M is usually not large) and considering the right hand side as fixed for a given choice of RBM parameters. If we want a specific value for the some induced coupling we can then use the solution to this simple linear system (which is guaranteed to be nonsingular) together with a root-finding algorithm to determine a possible solution for the RBM parameters that satisfy the constrain. This is the procedure we used throughout our work.

In the next subsection we show in more detail this procedure for the physically relevant cases of the two and three body forces we mention in the main text. This procedure can be repeated for higher many-body forces, and at each step, we can represent the lower-order couplings in term of RBM parameters from the identities already found for them.

Similar relations can be obtained when the visible layer is composed by (for instance) spin operators by ensuring that the matching condition is enacted for every state in the corresponding eigenbasis. For Pauli matrices this will lead to the change $\rho_i \rightarrow \sigma_i = \pm 1$.

1. Derivation of RBM mapping for two- and three-body interactions

Having set up the system of linear equations that allows one to solve for the many-body couplings in the previous section, we proceed to derive the identities presented in the main text for the two- and three-body interactions in a way that makes easy the generalization to arbitrary categorical hidden variable $h = \{0, \dots, \mathcal{K} - 1\}$.

We start by denoting by Z_2 the general partition function with one- and two-body interactions,

$$Z_2 = \exp\left(-A^{(2)}\rho_1\rho_2 - A_1^{(1)}\rho_1 - A_2^{(1)}\rho_2\right). \quad \text{(A7)}$$

We can express the equation that maps the physical system to an RBM with free energy given by Eq. (19) as

$$Z_{\text{rbm}}^{(2)} = \mathcal{N}_2 \sum_{h=0}^{\mathcal{K}-1} e^{-h(W_1\rho_1 + W_2\rho_2 + C)}, \tag{A8}$$

where $\mathcal{N}_2$ is an overall normalization factor. As discussed above, the matching condition Eq. (A6) can be converted to a system of linear equations. The result is

$$\begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} A^{(2)} \\ A_1^{(1)} \\ A_2^{(1)} \end{pmatrix} = \begin{pmatrix} L^{(2)}(0,1) \\ L^{(2)}(1,0) \\ L^{(2)}(1,1) \end{pmatrix}, \tag{A9}$$

where the entries of the right hand side vector are

$$L^{(2)}(\rho_1, \rho_2) = -\ln\left[\sum_{h=0}^{\mathcal{K}-1} \exp\left(-hC - h\sum_{\mu}^{2} W_\mu \rho_\mu\right)\right]$$
$$+ \ln\left[\sum_{h=0}^{\mathcal{K}-1} \exp\left(-hC\right)\right].$$

By solving this linear system we can find that the general expression for the one-body term

$$A_\mu^{(1)} = -\ln\left\{\frac{\sum_{h=0}^{\mathcal{K}-1} \exp\left[-h(C + W_\mu)\right]}{\sum_{h=0}^{\mathcal{K}-1} \exp\left(-hC\right)}\right\}, \tag{A10}$$

and the two-body coupling is obtained as

$$A^{(2)} = L^{(2)}(1,1) - L^{(2)}(1,0) - L^{(2)}(0,1)$$
$$= -\ln\left\{\frac{\sum_{h=0}^{\mathcal{K}-1} \exp\left[-h(C + W_1 + W_2)\right]}{\sum_{h=0}^{\mathcal{K}-1} \exp(-hC)}\right\}$$
$$- A_1^{(1)} - A_2^{(1)}. \tag{A11}$$

Note that the three parameters (W_1, W_2, C) define a hypersurface of equivalent RBMs corresponding to the same two-body coupling and different one-body counterterms.

We now turn to the more involved three-body case, as we will see the structure of the mapping between physical and RBM coupling follows a similar pattern to the two-body case. Let us first introduce, in analogy to Eq. (A7), the three-body partition function

$$Z_3 = e^{-A^{(3)}\rho_1\rho_2\rho_3 - \sum_{\mu<\nu} A_{\mu\nu}^{(2)}\rho_\mu\rho_\nu - \sum_\mu A_\mu^{(1)}\rho_\mu}. \tag{A12}$$

The relation between physical and RBM parameters for the three-body case is now given by

$$Z_3 = Z_{\text{rbm}}^{(3)} = \mathcal{N}_3 \sum_{h=0}^{\mathcal{K}-1} e^{-h(W_1\rho_1 + W_2\rho_2 + W_3\rho_3 + C)}, \tag{A13}$$

with $\mathcal{N}_3$ an overall normalization factor. As before we can convert this to a system of linear equations,

$$\begin{pmatrix} 0 & 0 & 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 & 1 & 0 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} A^{(3)} \\ A_{12}^{(2)} \\ A_{13}^{(2)} \\ A_{23}^{(2)} \\ A_1^{(1)} \\ A_2^{(1)} \\ A_3^{(1)} \end{pmatrix} = \begin{pmatrix} L^{(3)}(0,0,1) \\ L^{(3)}(0,1,0) \\ L^{(3)}(0,1,1) \\ L^{(3)}(1,0,0) \\ L^{(3)}(1,0,1) \\ L^{(3)}(1,1,0) \\ L^{(3)}(1,1,1) \end{pmatrix},$$
$$\tag{A14}$$

where, similarly to before, we have defined

$$L^{(3)}(\rho_1, \rho_2, \rho_3) = -\ln\left[\sum_{h=0}^{\mathcal{K}-1} \exp\left(-hC - h\sum_{\mu=1}^{3} W_\mu\rho_\mu\right)\right]$$
$$- \ln\left[\sum_{h=0}^{\mathcal{K}-1} \exp\left(-hC\right)\right].$$

The solution for the one-body couplings is the same as Eq. (A10), while the two-body term is generalized to

$$A_{\mu\nu}^{(2)} = -\ln\left\{\frac{\sum_{h=0}^{\mathcal{K}-1} \exp\left[-h(C + W_\mu + W_\nu)\right]}{\sum_{h=0}^{\mathcal{K}-1} \exp\left(-hC\right)}\right\}$$
$$- A_\mu^{(1)} - A_\nu^{(1)}. \tag{A15}$$

We can now move to the solution for three-body term which reads

$$A^{(3)} = -\ln\left\{\frac{\sum_{h=0}^{\mathcal{K}-1} \exp\left[-h\left(C + \sum_{\mu=1}^{3} W_\mu\right)\right]}{\sum_{h=0}^{\mathcal{K}-1} \exp\left(-hC\right)}\right\}$$
$$- \sum_{\mu<\nu} A_{\mu\nu}^{(2)} - \sum_\mu A_\mu^{(1)}, \tag{A16}$$

and, again, the four parameters (W_1, W_2, W_3, C) define a hypersurface of equivalent RBMs. Note that the sums that appear in the definition of the physical couplings can be summed directly using standard relations for geometric progressions as we did for the two-body case in Eqs. (34) and (38) of the main text. We will make use of this in the next section.

APPENDIX B: DERIVATION OF THE BOUNDS FOR GENERAL CATEGORICAL HIDDEN UNITS

In this section we provide the proof for the two lower bounds Eq. (37) and Eq. (39) presented in the main text to explain the rate of growth of the induced couplings with the maximal value $\mathcal{K}$ that defines the range of the categorical auxiliary variable $h \in \{0, \ldots, \mathcal{K} - 1\}$.

Given the fact that for a given target value of the physical couplings like $A_{\mathcal{K}}^{(2)}$ and $A_{\mathcal{K}}^{(3)}$ there is a continuous space of solutions for the RBM parameters, in this section we will restrict the discussion to just a one-parameter subspace for convenience. For the two-body case we choose the RBM parameters $(W_1, W_2.C)$ according to the following:

(i) for attractive interactions $A_{\mathcal{K}}^{(2)} < 0$ we take

$$W_1 = W_2 = -C = \alpha_2 > 0; \tag{B1}$$

(ii) for repulsive interactions $A_{\mathcal{K}}^{(2)} > 0$ we take

$$W_1 = -W_2 = \alpha_2 > 0, \quad C = 0. \tag{B2}$$

For the three-body case,

(iii) for attractive interactions $A_{\mathcal{K}}^{(3)} < 0$ we take

$$W_1 = W_2 = W_3 = \alpha_3 > 0, \quad C = -2\alpha_3; \tag{B3}$$

(iv) for repulsive interactions $A_{\mathcal{K}}^{(3)} > 0$ we take

$$W_1 = W_2 = W_3 = \alpha_3 > 0, \quad C = -\alpha_3. \tag{B4}$$

Note that, as we mention in the main text, these parametrization choices do not have any particularly useful property other than being able to generate the desired couplings with the adjustment of a single real parameter α. Depending on the particular application one is interested in, different choices might be closer to optimal but this will need to be assessed on a case-by-case basis.

Using the general expressions found in the previous section, Eqs. (A10) and (A11), together with the choice of parametrization for attractive interactions Eq. (B1), we find that the induced one-body couplings are all equal to

$$A_{\mathcal{K}}^{(1)} = -\ln(\mathcal{K}) + \ln\left[\frac{1 - \exp(\mathcal{K}\alpha_2)}{1 - \exp(\alpha_2)}\right], \tag{B5}$$

while we find for the two-body term

$$A_{\mathcal{K}}^{(2)} = (\mathcal{K} - 1)\alpha_2 + 2\ln(\mathcal{K}) - 2\ln\left(\frac{1 - e^{\mathcal{K}\alpha_2}}{1 - e^{\alpha_2}}\right)$$
$$\lesssim -(\mathcal{K} - 1)\alpha_2 + 2\ln(\mathcal{K}), \tag{B6}$$

where we used the fact that for $\alpha_2 > 0$ we have

$$\ln\left(\frac{1 - e^{\mathcal{K}\alpha_2}}{1 - e^{\alpha_2}}\right) \geqslant (\mathcal{K} - 1)\alpha_2. \tag{B7}$$

In the case of a repulsive interaction, the parametrization Eq. (B2) leads to the following one-body terms

$$A_1^{(1)} = -\ln\left[\frac{\exp(-\mathcal{K}\alpha_2) - 1}{\exp(-\alpha_2) - 1}\right] + \ln(\mathcal{K}),$$
$$A_2^{(1)} = -\ln\left[\frac{\exp(\mathcal{K}\alpha_2) - 1}{\exp(\alpha_2) - 1}\right] + \ln(\mathcal{K})$$
$$= A_1^{(1)} - (\mathcal{K} - 1)\alpha_2. \tag{B8}$$

For the induced two-body coupling, we first note that the first term in Eq. (A11) vanishes when taking the parametrization

Eq. (B2). The final result reads

$$A_{\mathcal{K}}^{(2)} = -A_1^{(1)} - A_2^{(1)} = -2A_2^{(1)} - (\mathcal{K} - 1)\alpha_2$$
$$= (1 - \mathcal{K})\alpha_2 - 2\ln(\mathcal{K}) + 2\ln\left(\frac{1 - e^{\mathcal{K}\alpha_2}}{1 - e^{\alpha_2}}\right)$$
$$\gtrsim (\mathcal{K} - 1)\alpha_2 - 2\ln(\mathcal{K}). \tag{B9}$$

These results can then be summarized in a single lower bound for the magnitude of the two-body coupling,

$$\left|A_{\mathcal{K}}^{(2)}\right| \gtrsim (\mathcal{K} - 1)\alpha_2 - 2\ln(\mathcal{K}). \tag{B10}$$

As we mention in the main text, Eq. (B10) shows that we have an almost linear dependence of the magnitude of the two-body coupling on the maximal value of the auxiliary variable.

Next, we turn our attention to the three-body coupling. Using the parametrization for the attractive case Eq. (B3) we find the one-body terms are all equal to

$$A_{\mathcal{K}}^{(1)} = -\ln\left[\frac{1 - e^{\mathcal{K}\alpha_3}}{1 - e^{\alpha_3}}\right] + \ln\left[\frac{1 - e^{2\mathcal{K}\alpha_3}}{1 - e^{2\alpha_3}}\right]$$
$$= \ln\left[\frac{1 + \exp(\mathcal{K}\alpha_3)}{1 + \exp(\mathcal{K}\alpha_3)}\right], \tag{B11}$$

and similarly for the two-body terms we find

$$A_{\mathcal{K}}^{(2)} = -\ln(\mathcal{K}) + \ln\left(\frac{1 - e^{2\mathcal{K}\alpha_3}}{1 - e^{2\alpha_3}}\right) - 2A_{\mathcal{K}}^{(1)}, \tag{B12}$$

for all pair interactions. The final result for the induced three-body term in the attractive case is then

$$A_{\mathcal{K}}^{(3)} = 3\ln(\mathcal{K}) + (\mathcal{K} - 1)\alpha_3$$
$$- 3\ln\left(\frac{1 - e^{\mathcal{K}\alpha_3}}{1 - e^{\alpha_3}}\right) + \ln\left(\frac{1 + e^{\mathcal{K}\alpha_3}}{1 + e^{\alpha_3}}\right). \tag{B13}$$

Moving now to the repulsive case, the parametrization Eq. (B4) implies that the one- and two-body couplings are the same as what we found for the attractive two-body case in Eq. (B5) and Eq. (B6) but with α_2 replaced with α_3. The three-body interaction is instead given by

$$A_{\mathcal{K}}^{(3)} = -3\ln(\mathcal{K}) - (\mathcal{K} - 1)\alpha_3$$
$$+ 3\ln\left(\frac{1 - e^{\mathcal{K}\alpha_3}}{1 - e^{\alpha_3}}\right) - \ln\left(\frac{1 + e^{\mathcal{K}\alpha_3}}{1 + e^{\alpha_3}}\right), \tag{B14}$$

and together with the result for the attractive case we have

$$\left|A_{\mathcal{K}}^{(3)}\right| = -3\ln(\mathcal{K}) - (\mathcal{K} - 1)\alpha_3$$
$$+ 3\ln\left(\frac{1 - e^{\mathcal{K}\alpha_3}}{1 - e^{\alpha_3}}\right) - \ln\left(\frac{1 + e^{\mathcal{K}\alpha_3}}{1 + e^{\alpha_3}}\right)$$
$$\gtrsim (\mathcal{K} - 1)\alpha_3 - 3\ln(\mathcal{K}). \tag{B15}$$

To obtain the lower bound, we used the fact that

$$\ln\left(\frac{1 - e^{\mathcal{K}\alpha_3}}{1 - e^{\alpha_3}}\right) - \ln\left(\frac{1 + e^{\mathcal{K}\alpha_3}}{1 + e^{\alpha_3}}\right) \geqslant 0, \tag{B16}$$

together with the inequality Eq. (B7). This concludes the proof for the bounds referenced in the main text.

APPENDIX C: GENERAL QUANTUM OPERATORS IN THE VISIBLE LAYER

The RBM network structure in our work is a hybrid quantum-classical architecture with quantum operators in the visible layer and classical categorical variables in the hidden layer. Up to now our main focus has been the fermionic density operator $\hat{\rho}$, and we have also provided an example for Pauli matrices in Eq. (25). In the first case, the RBM mapping was made possible due to the idempotency of the operator, $\hat{\rho}^2 = \hat{\rho}$, and in the second case the Pauli matrix is involutory, $\sigma^2 = I$. These properties are special cases of operators with generalized idempotency

$$\hat{O}^R = \hat{O}, \tag{C1}$$

for some integer $R > 1$ (for Pauli matrices $R = 3$). Due to Eq. (C1), the highest exponent present in the Taylor series expansion of the cumulant generating function in Eq. (17) is bounded above, $k_\mu \leqslant R - 1$, effectively truncating the order of operator products that can be generated by the RBM. Note that for more general operators not satisfying the idempotency condition of Eq. (C1), the induced operators may not have any particularly useful structure to be exploited to match to some target partition function. However, this is not a problem since, as already mentioned in the main text, any operator on a finite Hilbert space can be represented by direct products of Pauli matrices. The general idempotency described here, while not necessary, could be useful if present in a given Hamiltonian.

[1] M. Jobes and H. R. Shaylor, Rep. Prog. Phys. **35**, 1077 (1972).

[2] P. Olmos, J. C. Diaz, J. M. Perez, P. Gomez, V. Rodellar, P. Aguayo, A. Bru, G. Garcia-Belmonte, and J. L. de Pablos, IEEE Trans. Nucl. Sci. **38**, 971 (1991).

[3] B. Denby, Comput. Phys. Commun. **119**, 219 (1999).

[4] T. Ciodaro, D. Deva, J. M. de Seixas, and D. Damazio, J. Phys.: Conf. Ser. **368**, 012030 (2012).

[5] A. L. Edelen, S. G. Biedron, B. E. Chase, D. Edstrom, S. V. Milton, and P. Stabile, IEEE Trans. Nucl. Sci. **63**, 878 (2016).

[6] A. Radovic, M. Williams, D. Rousseau, M. Kagan, D. Bonacorsi, A. Himmel, A. Aurisano, K. Terao, and T. Wongjirad, Nature (London) **560**, 41 (2018).

[7] A. Scheinker, A. Edelen, D. Bohler, C. Emma, and A. Lutman, Phys. Rev. Lett. **121**, 044801 (2018).

[8] L. Wang, Phys. Rev. B **94**, 195105 (2016).

[9] J. Carrasquilla and R. G. Melko, Nat. Phys. **13**, 431 (2017).

[10] G. Carleo and M. Troyer, Science **355**, 602 (2017).

[11] L. Huang and L. Wang, Phys. Rev. B **95**, 035105 (2017).

[12] J. Liu, Y. Qi, Z. Y. Meng, and L. Fu, Phys. Rev. B **95**, 041101(R) (2017).

[13] H. Shen, J. Liu, and L. Fu, Phys. Rev. B **97**, 205140 (2018).

[14] J. Nelson, R. Tiwari, and S. Sanvito, Phys. Rev. B **99**, 075132 (2019).

[15] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Nat. Phys. **14**, 447 (2018).

[16] J. Carrasquilla, G. Torlai, R. G. Melko, and L. Aolita, Nat. Mach. Intell. **1**, 155 (2019).

[17] Y. Bengio, A. Courville, and P. Vincent, IEEE Trans. Pattern Anal. Mach. Intell. **35**, 1798 (2013).

[18] P. Mehta, M. Bukov, C.-H. Wang, A. G. Day, C. Richardson, C. K. Fisher, and D. J. Schwab, Phys. Rep. **810**, 1 (2019).

[19] P. Smolensky, in *Parallel Distributed Processing, Foundations*, edited by D. E. Rumelhart (MIT Press, Cambridge, MA, 1987), Vol. 1, pp. 194–281.

[20] Y. Freund and D. Haussler, in *Advances in Neural Information Processing Systems 4*, edited by J. E. Moody, S. J. Hanson, and R. P. Lippmann (MIT Press, Cambridge, MA, 1992), pp. 912–919.

[21] N. Le Roux and Y. Bengio, Neural Comput. **20**, 1631 (2008).

[22] G. F. Montúfar, J. Rauh, and N. Ay, in *Advances in Neural Information Processing Systems 24 (NIPS 2011)* (Curran Associates, Red Hook, NY, 2011), pp. 415–423.

[23] G. E. Hinton, S. Osindero, and Y.-W. Teh, Neural Comput. **18**, 1527â1554 (2006).

[24] R. Salakhutdinov and G. Hinton, in *Proceedings of the Twelfth International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research, edited by D. van Dyk and M. Welling (Microtome Publishing, Brookline, MA, 2009), Vol. 5, pp. 448–455.

[25] G. Montúfar, in *Information Geometry and Its Applications*, edited by N. Ay, P. Gibilisco, and F. Matúš (Springer International Publishing, Cham, 2018), pp. 75–115.

[26] G. Carleo, Y. Nomura, and M. Imada, Nat. Commun. **9**, 5322 (2018).

[27] W. M. C. Foulkes, L. Mitas, R. J. Needs, and G. Rajagopal, Rev. Mod. Phys. **73**, 33 (2001).

[28] J. Carlson, S. Gandolfi, F. Pederiva, S. C. Pieper, R. Schiavilla, K. E. Schmidt, and R. B. Wiringa, Rev. Mod. Phys. **87**, 1067 (2015).

[29] R. L. Stratonovich, Sov. Phys. Dokl. **2**, 416 (1957).

[30] J. Hubbard, Phys. Rev. Lett. **3**, 77 (1959).

[31] J. E. Hirsch, Phys. Rev. B **28**, 4059 (1983).

[32] C. Körber, E. Berkowitz, and T. Luu, Europhys. Lett. **119**, 60006 (2017).

[33] R. B. Wiringa, V. G. J. Stoks, and R. Schiavilla, Phys. Rev. C **51**, 38 (1995).

[34] A. Gezerlis, I. Tews, E. Epelbaum, M. Freunek, S. Gandolfi, K. Hebeler, A. Nogga, and A. Schwenk, Phys. Rev. C **90**, 054323 (2014).

[35] D. Lee, Prog. Part. Nucl. Phys. **63**, 117 (2009).

[36] Y. Nishida and D. T. Son, Phys. Rev. A **82**, 043606 (2010).

[37] J. Carlson, S. Gandolfi, U. van Kolck, and S. A. Vitiello, Phys. Rev. Lett. **119**, 223002 (2017).

[38] R. Blankenbecler, D. J. Scalapino, and R. L. Sugar, Phys. Rev. D **24**, 2278 (1981).

[39] J. E. Hirsch, Phys. Rev. B **31**, 4403 (1985).

[40] Y. Alhassid, D. J. Dean, S. E. Koonin, G. Lang, and W. E. Ormand, Phys. Rev. Lett. **72**, 613 (1994).

[41] G. Sugiyama and S. Koonin, Ann. Phys. **168**, 1 (1986).

[42] C. W. Johnson, S. E. Koonin, G. H. Lang, and W. E. Ormand, Phys. Rev. Lett. **69**, 3157 (1992).

[43] S. Zhang, J. Carlson, and J. E. Gubernatis, Phys. Rev. Lett. **74**, 3652 (1995).

[44] S. Gandolfi, A. Y. Illarionov, K. E. Schmidt, F. Pederiva, and S. Fantoni, Phys. Rev. C **79**, 054005 (2009).

[45] J. Zhang, Spin orbit interactions in nuclear matter with auxiliary field diffusion Monte Carlo, Ph.D. thesis, Arizona State University, Tempe, 2014.

[46] F. F. Assaad, M. Imada, and D. J. Scalapino, Phys. Rev. B **56**, 15001 (1997).

[47] Y. Motome and M. Imada, J. Phys. Soc. Jpn. **66**, 1872 (1997).

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

[49] N. Yoshioka, Y. Akagi, and H. Katsura, Phys. Rev. E **99**, 032113 (2019).

[50] G. Cossu, L. Del Debbio, T. Giani, A. Khamseh, and M. Wilson, Phys. Rev. B **100**, 064304 (2019).

[51] L. H. Thomas, Phys. Rev. **47**, 903 (1935).

[52] H.-W. Hammer, A. Nogga, and A. Schwenk, Rev. Mod. Phys. **85**, 197 (2013).

[53] U. van Kolck, in *Effective Field Theory in Particle Physics and Cosmology*, Les Houches Summer School 2017 (Oxford University Press, Oxford, UK, 2020).

[54] J.-W. Chen, D. Lee, and T. Schaefer, Phys. Rev. Lett. **93**, 242302 (2004).

[55] A. Finnila, M. Gomez, C. Sebenik, C. Stenson, and J. Doll, Chem. Phys. Lett. **219**, 343 (1994).

[56] T. Kadowaki and H. Nishimori, Phys. Rev. E **58**, 5355 (1998).

[57] E. Farhi, J. Goldstone, S. Gutmann, and M. Sipser, arXiv:quant-ph/0001106.

[58] F. Gaitan and L. Clark, Phys. Rev. Lett. **108**, 010501 (2012).

[59] A. Perdomo-Ortiz, N. Dickson, M. Drew-Brook, G. Rose, and A. Aspuru-Guzik, Sci. Rep. **2**, 571 (2012).

[60] A. D. King, J. Carrasquilla, J. Raymond, I. Ozfidan, E. Andriyash, A. Berkley, M. Reis, T. Lanting, R. Harris, F. Altomare *et al.*, Nature (London) **560**, 456 (2018).

[61] E. Farhi, J. Goldstone, S. Gutmann, J. Lapan, A. Lundgren, and D. Preda, Science **292**, 472 (2001).

[62] A. Das and B. K. Chakrabarti, *Quantum Annealing and Related Optimization Methods* (Springer-Verlag Berlin Heidelberg, 2005).

[63] A. K. Chandra, A. Das, and B. K. Chakrabarti, *Quantum Quenching, Annealing and Computation*, Lecture Notes in Physics (Springer, Berlin, 2010).

[64] M. Johnson, M. Amin, S. Gildert, T. Lanting, F. Hamze, N. Dickson, R. Harris, A. Berkley, J. Johansson, P. Bunyk, E. Chapple, C. Enderud, J. Hilton, K. Karimi, E. Ladizinsky, N. Ladizinsky, T. Oh, I. Perminov, C. Rich, M. Thom, E. Tolkacheva, C. Truncik, S. Uchaikin, J. Wang, B. Wilson, and G. Rose, Nature (London) **473**, 194 (2011).

[65] A. Perdomo, C. Truncik, I. Tubert-Brohman, G. Rose, and A. Aspuru-Guzik, Phys. Rev. A **78**, 012320 (2008).

[66] Z. Bian, F. Chudak, R. Israel, B. Lackey, W. G. Macready, and A. Roy, Front. Phys. **2**, 56 (2014).

[67] N. Chancellor, S. Zohren, and P. A. Warburton, npj Quantum Inf. **3**, 21 (2017).

[68] N. Dattani and N. Chancellor, arXiv:1901.07676.

[69] T. Boothby, A. D. King, and A. Roy, Quantum Inf. Process. **15**, 495 (2016).

[70] M. H. Amin, Phys. Rev. A **92**, 052323 (2015).

[71] A. Perdomo-Ortiz, B. O'Gorman, J. Fluegemann, R. Biswas, and V. N. Smelyanskiy, Sci. Rep. **6**, 18628 (2016).

[72] M. Benedetti, J. Realpe-Gómez, R. Biswas, and A. Perdomo-Ortiz, Phys. Rev. A **94**, 022308 (2016).

[73] G. E. Hinton, Neural Comput. **14**, 1771 (2002).

[74] Y. Bengio and O. Delalleau, Neural Comput. **21**, 1601 (2009).

[75] I. Sutskever and T. Tieleman, in *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research, edited by Y. W. Teh and M. Titterington (Microtome Publishing, Brookline, MA, 2010), Vol. 9, pp. 789–795.

[76] M. Suzuki, Prog. Theor. Phys. **56**, 1454 (1976).