

ARTICLE

DOI: 10.1038/s41467-018-07520-3

OPEN

Constructing exact representations of quantum many-body systems with deep neural networks

Giuseppe Carleo^{1,2}, Yusuke Nomura³ & Masatoshi Imada³

Obtaining accurate properties of many-body interacting quantum matter is a long-standing challenge in theoretical physics and chemistry, rooting into the complexity of the many-body wave-function. Classical representations of many-body states constitute a key tool for both analytical and numerical approaches to interacting quantum problems. Here, we introduce a technique to construct classical representations of many-body quantum systems based on artificial neural networks. Our constructions are based on the deep Boltzmann machine architecture, in which two layers of hidden neurons mediate quantum correlations. The approach reproduces the exact imaginary-time evolution for many-body lattice Hamiltonians, is completely deterministic, and yields networks with a polynomially-scaling number of neurons. We provide examples where physical properties of spin Hamiltonians can be efficiently obtained. Also, we show how systematic improvements upon existing restricted Boltzmann machines *ansatze* can be obtained. Our method is an alternative to the standard path integral and opens new routes in representing quantum many-body states.

¹Center for Computational Quantum Physics, Flatiron Institute, 162 5th Avenue, New York, NY 10010, USA. ²Institute for Theoretical Physics, ETH Zurich, Wolfgang-Pauli-Str. 27, 8093 Zurich, Switzerland. ³Department of Applied Physics, The University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan. Correspondence and requests for materials should be addressed to G.C. (email: gcarleo@flatironinstitute.org)

A tremendous amount of successful developments in quantum physics builds upon the mapping between many-body quantum systems and effective classical theories. The probably most well-known mapping is due to Feynman, who introduced an exact representation of many-body quantum systems in terms of statistical summations over classical particles trajectories¹. Effective classical representations of quantum many-body systems are however not unique, and other approaches rely on different inspiring principles, such as perturbative expansions², or decomposition of interactions with auxiliary degrees of freedom^{3,4}. The classical representations of quantum states allow both for novel conceptual developments and efficient numerical simulations. On one hand, perturbative approaches based on the graphical resummation of classes of diagrams are at the heart of many-body analytical approaches in various fields of research, ranging from particle to condensed-matter physics⁵. On the other hand, several non-perturbative numerical methods for many-body quantum systems are also based on these mappings. Quantum Monte Carlo (QMC) methods are among the most successful numerical techniques, relying on continuous-space polymer representations^{6–9}, world-line lattice path integrals^{10,11}, continuous time algorithm¹², and summation of perturbative diagrams^{13,14}. Effective classical representations are also the building block of variational methods based on correlated many-body wave-functions¹⁵. Several successful variational techniques make extensive use of parametric representations of quantum states, where the effective parameters are determined by means of the variational principle^{16–19}. In matrix-product and tensor-network-states the ground-state is expressed as a classical network^{20,21}. In general, finding alternative, efficient classical representations of quantum states can help establishing novel numerical and analytical techniques to study challenging open issues.

Recently, an efficient variational representation of many-body systems in terms of artificial neural networks, which consists of classical degrees of freedom, has been introduced²². Numerical results have shown that artificial neural networks can represent many-body states with high accuracy^{22–31}. The majority of the variational approaches adopted so-far are based on shallow neural networks, called restricted Boltzmann machines (RBM), in which the physical degrees of freedom interact with an ensemble of hidden degrees of freedom (neurons). While shallow RBM states have promising features in terms of entanglement capacity^{25,32–34}, only deep networks are guaranteed to provide a complete and efficient description of the most general quantum states^{35,36}.

In this work, we introduce a constructive approach to explicitly generate deep network structures corresponding to exact quantum many-body ground states. We demonstrate this construction for interacting lattice spin models, including the transverse-field Ising and Heisenberg models. Our constructions are fully deterministic, in stark contrast to the shallow RBM case, in which the numerical optimization of the network parameters is inevitable. The number of neurons required in the construction scales only polynomially with the system size, thus the present approach constitutes a new family of efficient quantum-to-classical mappings exhibiting a prominent representational flexibility. Given as a simple set of iterative rules, these constructions can be used both as a self-standing tool, or to systematically improve results obtained with variational shallow networks. The latter improves the efficiency of the method because the numerically optimized shallow RBM states are already good approximations for ground states. Finally, we discuss sampling strategies from the generated deep networks and show numerical results for one-dimensional spin models.

Results

General scheme of constructing deep neural states. The ground state of a generic Hamiltonian, \mathcal{H} , can be found through imaginary-time evolution, $|\Psi(\tau)\rangle = e^{-\tau\mathcal{H}}|\Psi_0\rangle$, for a sufficiently large $\tau \gg \Delta E^{-1}$. Here ΔE is the energy gap between the ground and the first excited state, $|\Psi_0\rangle$ is an arbitrary initial state non-orthogonal to the exact ground state, and we work in units where $\hbar = 1$. For a finite system, the energy gap is typically finite, and the total propagation time needed to reach the ground state within an arbitrary given accuracy is expected to grow at most polynomially with the system size (for systems becoming gapless in the thermodynamic limit).

Here, we introduce a representation of the wave-function coefficients in terms of a deep Boltzmann machine (DBM)³⁷. For the sake of concreteness, let us consider the case of N spins, described by the quantum numbers $|\sigma^z\rangle = |\sigma_1^z \dots \sigma_N^z\rangle$. Then, we represent generic many-body amplitudes $\langle \sigma_1^z \dots \sigma_N^z | \Psi \rangle \equiv \Psi(\sigma^z)$ in the two-layer DBM form:

$$\Psi_{\mathcal{W}}(\sigma^z) = \sum_{\{h,d\}} \exp \left[\sum_i a_i \sigma_i^z + \sum_{ij} \sigma_i^z W_{ij} h_j + \sum_j b_j h_j + \sum_{jk} h_j d_k W'_{jk} + \sum_k b'_k d_k \right] \quad (1)$$

where we have introduced M hidden units h , M' deep units d , and a set of couplings and bias terms $\mathcal{W} \equiv (a, b, b', W, W')$. A sketch of the DBM architecture is shown in Fig. 1.

In the following, we specialize to the case of spin 1/2, thus all the units are taken to be $\sigma^z, h, d = \pm 1$. This representation is the natural deep-network generalization of the shallow RBM, introduced as variational ansatz in ref. 22. As for the RBM form, also in this case direct connections between variables in the same layer are not allowed. A crucial difference is however that the layer of deep variables makes, in general, the evaluation of the wave-function amplitudes not possible analytically. At variance with RBM, the DBM form is known to be universal, as proven by Gao and Duan recently³⁵.

Our key finding is that, thanks to the much more flexible representability, the DBM wave function can reproduce the Hamiltonian imaginary-time evolution exactly by changing its form dynamically, and that the parameters for ground state DBM network can be derived analytically. In order to find explicit expressions for the parameters \mathcal{W} that represent $|\Psi(\tau)\rangle$ for arbitrary imaginary time, we start considering a second-order

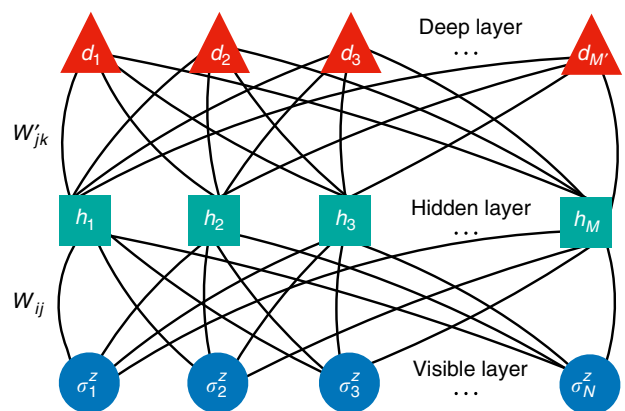


Fig. 1 Structure of deep Boltzmann machine. Dots, squares, and triangles represent physical degrees of freedom (σ_i^z), hidden units (h_i), deep units (d_k), respectively. Solid curves represent interlayer couplings (W_{ij} and W'_{jk})

Trotter–Suzuki decomposition^{10,38}:

$$|\Psi(\tau)\rangle = \mathcal{G}_1(\delta_\tau/2)\mathcal{G}_2(\delta_\tau)\dots\mathcal{G}_1(\delta_\tau)\mathcal{G}_2(\delta_\tau)\mathcal{G}_1(\delta_\tau/2)|\Psi_0\rangle, \quad (2)$$

where we have decomposed the Hamiltonian into two non-commuting parts, $\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2$, and introduced the short-time propagators $\mathcal{G}_\nu(\delta_\tau) = e^{-\mathcal{H}_\nu\delta_\tau}$. The problem of finding an exact representation for $|\Psi(\tau)\rangle$ then reduces to finding a rule to construct the building blocks of the time-evolution, namely representing the state after two-types of propagators by DBM with new parameters \mathcal{W} :

$$e^{-\mathcal{H}_\nu\delta_\tau}|\Psi_{\mathcal{W}}\rangle = C|\Psi_{\mathcal{W}}\rangle. \quad (3)$$

In practice, this is achieved either by changing parameters \mathcal{W} , or by introducing additional parameters in \mathcal{W} , adding new neurons and creating new connections in the network.

In the following, we show concrete examples for paradigmatic spin hamiltonians, namely the transverse-field Ising and Heisenberg models. The rest of this section provides a general overview of how the DBM constructions are derived (how Eq. (3) is satisfied) for these models. The next section (Sampling strategies) discusses how they can be used in numerical schemes. A complete, in-depth derivation of the representations and algorithms can be found both in Methods and in the Supplementary Notes, as referred to at each step in this section. Furthermore, we provide computer codes to create the DBM network for each model as Supplementary Software 1–4.

Transverse-field Ising model (TFIM). We start considering the TFIM on an arbitrary interaction graph. In this case, we decompose the Hamiltonian into two parts: $\mathcal{H}_1 = -\sum_l \Gamma_l \sigma_l^x$, and $\mathcal{H}_2 = \sum_{l<m} V_{lm} \sigma_l^z \sigma_m^z$, where σ denote Pauli matrices, $\Gamma_l (>0)$ are site-dependent transverse fields, and V_{lm} are arbitrary coupling constants.

In order to implement the mapping to a DBM, we first consider the action of the diagonal propagator $e^{-\delta_\tau V_{lm} \sigma_l^z \sigma_m^z}$, acting on a bond V_{lm} . In this case, the goal of finding an exact DBM representation can be rephrased as finding solutions to

$$\langle \sigma^z | e^{-\delta_\tau V_{lm} \sigma_l^z \sigma_m^z} | \Psi_{\mathcal{W}} \rangle = C \Psi_{\mathcal{W}}(\sigma^z), \quad (4)$$

i.e. finding a set of new parameters $\bar{\mathcal{W}}$ that exactly reproduces the imaginary time evolution on the left-hand side. Here C is an arbitrary finite normalization constant. The diagonal propagator introduces an interaction between two visible, physical spins, which is not directly available in the DBM architecture. This interaction can be mediated by a new hidden unit in the first layer, $h_{[lm]}$ which is only connected to the visible spins on that bond, i.e. $\bar{W}_{l[lm]}$ and $\bar{W}_{m[lm]}$ are finite, but $\bar{W}_{i[lm]} = 0, \forall i \neq l, m$ and $\bar{W}'_{j[lm]} = 0, \forall j$ (see Fig. 2a).

More concretely, the new wave function has then the form:

$$\begin{aligned} \Psi_{\bar{\mathcal{W}}}(\sigma^z) &= \sum_{h_{[lm]}} e^{\sigma_l^z \bar{W}_{l[lm]} h_{[lm]} + \sigma_m^z \bar{W}_{m[lm]} h_{[lm]}} \Psi_{\mathcal{W}}(\sigma^z) \\ &= 2 \cosh\left(\sigma_l^z \bar{W}_{l[lm]} + \sigma_m^z \bar{W}_{m[lm]}\right) \Psi_{\mathcal{W}}(\sigma^z). \end{aligned} \quad (5)$$

Equation (4) is then satisfied if

$$e^{-\delta_\tau V_{lm} \sigma_l^z \sigma_m^z} = 2C \cosh\left(\sigma_l^z \bar{W}_{l[lm]} + \sigma_m^z \bar{W}_{m[lm]}\right) \quad (6)$$

for all the possible values of σ_l^z and σ_m^z . By means of a useful identity [Eq. (21) in Methods], the new parameters $\bar{W}_{l[lm]}$ and

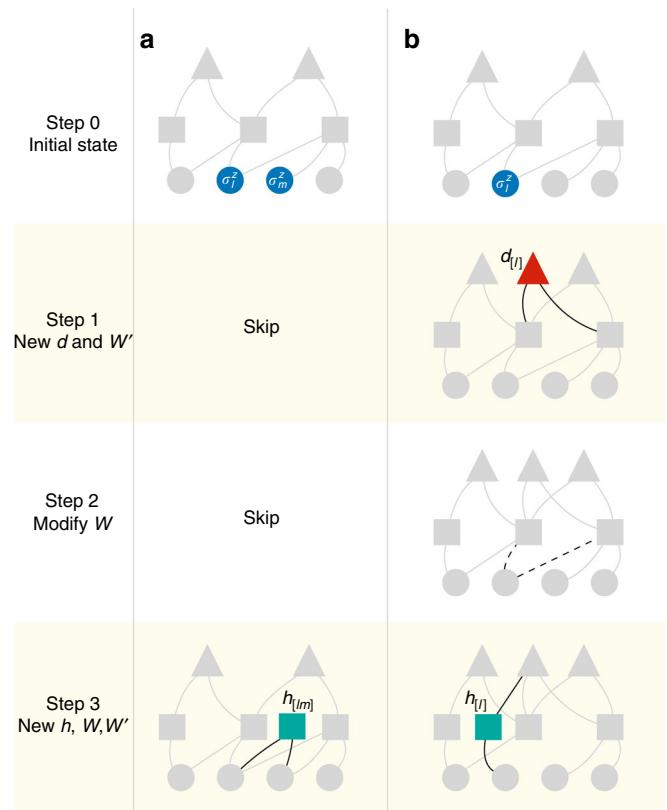


Fig. 2 Construction of exact DBM representations of the transverse-field Ising model. In this example, a step of imaginary-time evolution is shown, for the case of the one-dimensional transverse-field Ising model. Dots represent physical degrees of freedom (σ_i^z), squares represent hidden units (h_j), triangles represent deep units (d_k). In each panel, upper networks are the initial state with arbitrary network form, and the bottom networks are the final states, after application of the propagator. Intermediate steps illustrate how the network is modified, where the relevant modified couplings at each step are highlighted in black. The highlighted solid and dashed curves indicate new and vanishing couplings, respectively. **a** Shows the diagonal (interaction) propagator being applied to the highlighted blue spins. This introduces a hidden unit (green) connected only to the two physical spins. In **(b)** the off-diagonal (transverse-field) propagator is applied, acting on the blue physical spin. Here, we then add one deep unit (red triangle), and a hidden unit (green) mediating visible–deep interactions

$W_{m[lm]}$ are given by

$$W_{l[lm]} = \frac{1}{2} \operatorname{arcosh}(e^{2|V_{lm}|\delta_\tau}) \quad (7)$$

$$W_{m[lm]} = -\operatorname{sgn}(V_{lm}) \times W_{l[lm]}. \quad (8)$$

In this way the classical two-body interaction can, in general, be represented exactly by the shallow RBM.

Next, to exactly represent the off-diagonal propagator $e^{\delta_\tau \Gamma_l \sigma_l^x} |\Psi_{\mathcal{W}}\rangle$, we must solve:

$$\begin{aligned} \cosh(\Gamma_l \delta_\tau) \Psi_{\mathcal{W}}(\sigma^z) + \sinh(\Gamma_l \delta_\tau) \Psi_{\mathcal{W}}(\sigma_l^z \rightarrow -\sigma_l^z) \\ = C \Psi_{\bar{\mathcal{W}}}(\sigma^z) \end{aligned} \quad (9)$$

for the new weights $\bar{\mathcal{W}}$, and for an appropriate finite normalization constant C . In this case, one possible solution is obtained by adding one deep $d_{[l]}$ and one hidden $h_{[l]}$ neurons. For $d_{[l]}$, we create new couplings $\bar{W}'_{j[l]}$ to the existing hidden neurons h_j which are connected to σ_l^z . We simultaneously allow for changes in the

existing parameters. By the procedure given in Methods, after applying the off-diagonal propagator for the site l , a solution of Eq. (9) is found by the matching condition of the hidden unit interactions on the left and the right hand sides of Eq. (9). Overall, the solution results in a three-step process (Fig. 2b): First, the hidden units attached to σ_i^z are connected to the newly introduced deep unit $d_{[l]}$ as

$$W'_{j[l]} = -W_{lj} \quad (10)$$

(see Eq. (35)). Second, all the hidden units previously connected to the spin σ_i^z lose their connection, i.e., $\bar{W}_{lj} = 0, \forall j$. Third, the spin σ_i^z and the deep unit $d_{[l]}$ are connected to the new hidden unit $h_{[l]}$, through the interaction $W_{l[l]}$ and $W'_{[l][l]}$, respectively, as

$$W_{l[l]} = \frac{1}{2} \operatorname{arccosh} \left(\frac{1}{\tanh(\Gamma_l \delta_\tau)} \right), \quad (11)$$

$$W'_{[l][l]} = -W_{l[l]}. \quad (12)$$

Using the given expressions for the parameters $\bar{\mathcal{W}}$ we can then exactly implement a single step of imaginary-time evolution. The full imaginary-time evolution is achieved by applying the above procedure for \mathcal{H}_1 and \mathcal{H}_2 alternately and repeatedly. Example applications of these rules, for both the diagonal and the off-diagonal propagators are shown in Fig. 2.

Approximate RBM from DBM for transverse ising model.

From the previous discussion, we have seen that the action of the off-diagonal propagator is responsible for the introduction of deep units in the network, thus breaking the shallow RBM structure. An interesting question is whether, in some limit, it is possible to stay within the RBM structure even for the off-diagonal propagator. The action of the off-diagonal propagator onto an RBM state can be then systematically expanded in powers of the weights:

$$\langle \sigma^z | e^{\delta_\tau \Gamma_l \sigma_i^z} | \Psi_{\bar{\mathcal{W}}}^{\text{RBM}} \rangle \propto \sum_{\{h\}} e^{\sum_j W_{ij} \sigma_i^z h_j} \left\{ 1 + \tanh(\Gamma_l \delta_\tau) \left(1 - 2\sigma_i^z \sum_j h_j W_{lj} \right) \right\} + \mathcal{O}(W_{lj}^2). \quad (13)$$

In the case of small weights, we can then exactly reproduce the off-diagonal propagator upon imposing a small change in the parameters $W_{lj} \rightarrow W_{lj} + \Delta W_{lj}$ and keeping an RBM structure. If we expand the new RBM with modified weights, we get

$$\langle \sigma^z | e^{\delta_\tau \Gamma_l \sigma_i^z} | \Psi_{\bar{\mathcal{W}}}^{\text{RBM}} \rangle \propto \sum_{\{h\}} e^{\sum_j W_{ij} \sigma_i^z h_j} \left\{ 1 + \sigma_i^z \sum_j \Delta W_{lj} h_j \right\} + \mathcal{O}(\Delta W_{lj}^2). \quad (14)$$

Comparing Eqs. (13) and (14), it follows that (apart from an irrelevant global normalization) the state after the off-diagonal propagator is still an RBM, with weights equal to:

$$W_{lj} \rightarrow W_{lj} - 2 \tanh(\Gamma_l \delta_\tau) W_{lj}, \quad (15)$$

and an error proportional to the square of the weights at that time step. In general, we expect that this kind of approximate updates is accurate in perturbative regimes (for example in the limit of small Γ_l) or in the limit of small imaginary time evolution. A similar approximation scheme has been derived in ref. 39. Numerical results for this approximation are discussed in a dedicated section before the Discussion.

Heisenberg model. We now consider the anti-ferromagnetic Heisenberg model (AFHM), on bipartite lattices. In one dimension, we decompose the Hamiltonian into odd and even bonds: $\mathcal{H}_1 = \sum_{(l,m)}^{\text{odd}} \mathcal{H}_{lm}^{\text{bond}}$ and $\mathcal{H}_2 = \sum_{(l,m)}^{\text{even}} \mathcal{H}_{lm}^{\text{bond}}$, with $\mathcal{H}_{lm}^{\text{bond}} = J(\sigma_l^x \sigma_m^x + \sigma_l^y \sigma_m^y + \sigma_l^z \sigma_m^z)$, where σ denote Pauli matrices. Because the bond Hamiltonian $\mathcal{H}_{lm}^{\text{bond}}$ is a building block also in higher dimensional models, construction of an exact DBM representation of the ground states can be achieved by finding solutions for the bond-propagator $\langle \sigma^z | e^{-\delta_\tau \mathcal{H}_{lm}^{\text{bond}}} | \Psi_{\bar{\mathcal{W}}} \rangle = C \langle \sigma^z | \Psi_{\bar{\mathcal{W}}} \rangle$, where the parameters $\bar{\mathcal{W}}$ are such that the previous equation is satisfied for all the possible $\langle \sigma^z |$, and for an arbitrary finite normalization constant C . More explicitly, we need to satisfy

$$\delta_{\sigma_i^z, \sigma_m^z} e^{-J\delta_\tau} \Psi_{\bar{\mathcal{W}}}(\sigma^z) + (1 - \delta_{\sigma_i^z, \sigma_m^z}) e^{J\delta_\tau} \cosh(2J\delta_\tau) \times (\Psi_{\bar{\mathcal{W}}}(\sigma^z) - \tanh(2J\delta_\tau) \Psi_{\bar{\mathcal{W}}}(\sigma_i^z \leftrightarrow \sigma_m^z)) = C \Psi_{\bar{\mathcal{W}}}(\sigma^z). \quad (16)$$

The basic strategy of finding a solution for Eq. (16) is similar to that for Eq. (9) in the transverse Ising model. Several possibilities arise when looking for solutions of the bond-propagator equation, Eq. (16). The existence of non-equivalent solutions prominently shows the non-uniqueness of DBM structure to represent the very same state and, at the same time, provides us flexibility in designing DBM architectures. Here, we show three concrete constructions. See Methods and Supplementary Note 2 for a detailed derivation of the DBM construction for the Heisenberg model, including anisotropic and bond-disordered coupling cases.

1 deep + 3 hidden variables construction for Heisenberg model.

The first construction is dubbed “1 deep, 3 hidden” (1d-3h). It amounts to adding an extra deep neuron, $d_{[lm]}$, and three more hidden neurons to satisfy Eq. (16). A crucial difference with respect to the TFIM is that the introduced deep spin $d_{[lm]}$ has a constraint depending on the state of the spins on the bond: σ_i^z and σ_m^z . Specifically, when $\sigma_i^z = \sigma_m^z$ the deep spin is constrained to be $d_{[lm]} = \sigma_i^z = \sigma_m^z$, whereas when $\sigma_i^z \neq \sigma_m^z$, its value is unconstrained. From a pictorial point of view, the action of the bond propagator is a four-step process (see Fig. 3a). Starting from a given initial network (uppermost structures in Fig. 3), $d_{[lm]}$ is added and connected, through $W'_{j[lm]}$ given in Eq. (43), to the existing hidden units h_j connected to σ_i^z and σ_m^z . Second, spin σ_i^z is disconnected to all hidden units and reconnected to those hidden units the spin σ_m^z is attached to [see Eq. (42)]. Third, two new hidden units are introduced. One of the hidden units, $h_{[lm1]}$, mediates the interaction between σ_i^z and $d_{[lm]}$ [Eq. (47)], and the other hidden unit $h_{[lm2]}$ mediates a direct spin-spin interaction between σ_i^z and σ_m^z [Eq. (49)]. Fourth, a further hidden unit connected to σ_i^z , σ_m^z , and $d_{[lm]}$ is inserted, in such a way that the constraint previously described is satisfied. For all but the last step, the DBM weights are real-valued. In the last step instead the constraint is enforced by introducing imaginary-valued interactions (dotted lines in Fig. 3), referred to the “ $i\pi/6$ ” trick, resulting in a sign-problem-free global term $\cos(\pi/6(\sigma_i^z + \sigma_m^z - d_{[lm]}))$ after the summation over ± 1 for the lastly added hidden unit $h_{[lm3]}$: $\sum_{h_{[lm3]} = \pm 1} \exp[i\pi/6(\sigma_i^z + \sigma_m^z - d_{[lm]})h_{[lm3]}]$. The constraint mentioned above is assured by this cosine term.

2 deep + 6 hidden variables construction for Heisenberg model.

The second construction is dubbed “2 deep, 6 hidden” (2d-6h), and is more similar to the lattice path-integral formulation. In this representation, we introduce two auxiliary deep spins per bond, $d_{[l]}$ and $d_{[m]}$ with constraint $d_{[l]} + d_{[m]} = \sigma_i^z + \sigma_m^z$,

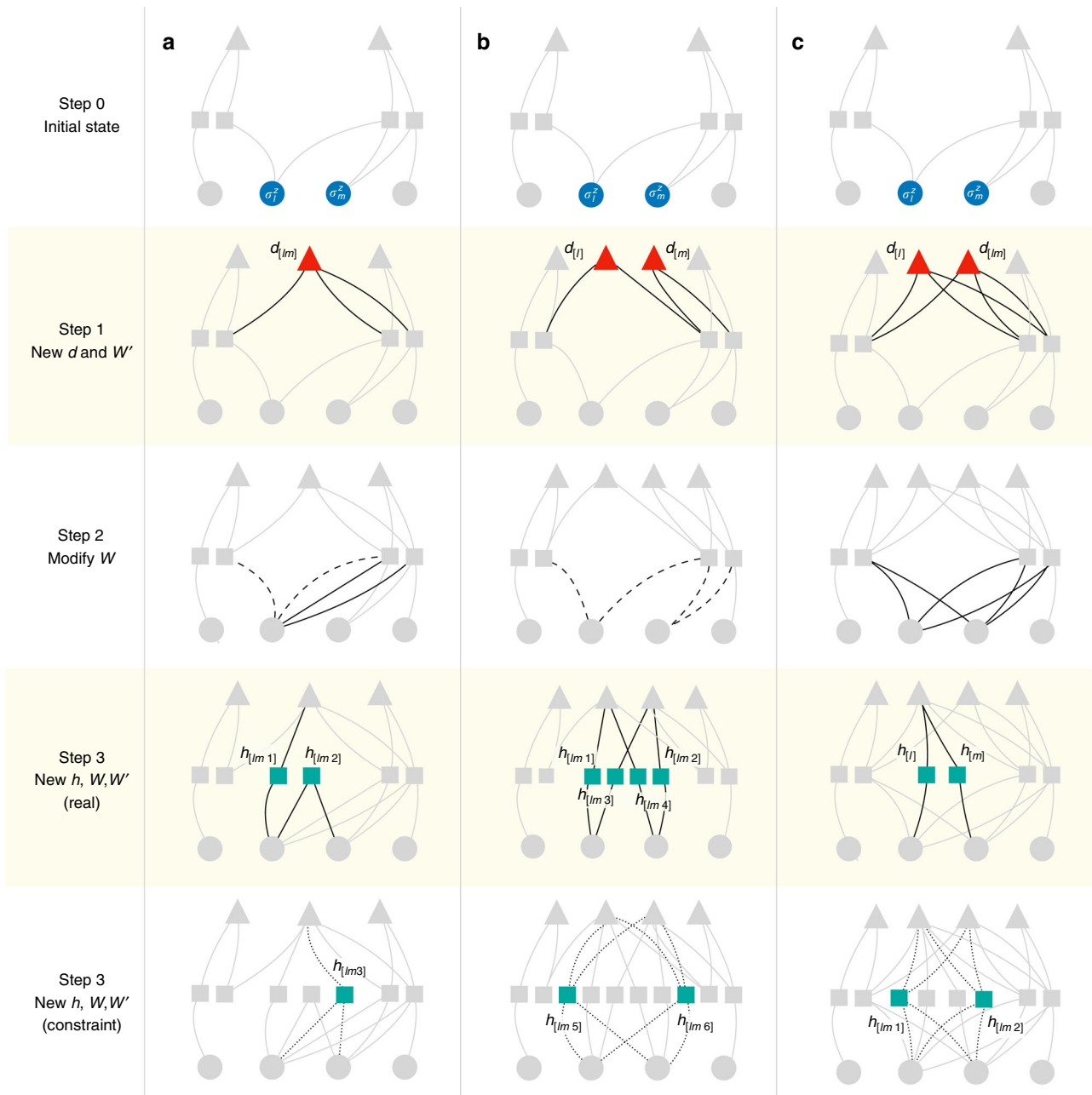


Fig. 3 Construction of exact DBM representations of Heisenberg models. In this example, a time step of imaginary-time evolution is shown, for the case of the one-dimensional antiferromagnetic Heisenberg model. Dots represent physical degrees of freedom (σ_i^z), squares represent hidden units (h_j), triangles represent deep units (d_k). The three panels (a–c) represent different possible explicit constructions. In each panel, upper networks are the initial state with arbitrary network form, and the bottom networks are the final states, after application of the propagator. Intermediate steps illustrate how the network is modified, where the relevant modified weights at each step are highlighted in black. In those diagrams, dashed lines indicate that the corresponding weights are set to zero, and dotted lines indicate complex-valued weights. The three panels correspond to the (a) “1 deep, 3 hidden” (1d-3h), (b) “2 deep, 6 hidden” (2d-6h), and (c) “2 deep, 4 hidden” (2d-4h) constructions (see text for a more detailed explanation of the individual steps characteristic of each construction)

and six hidden neurons. The action of the bond propagator is schematically illustrated in Fig. 3b: first, two deep units $d_{[l]}$ and $d_{[m]}$ are introduced, connecting, respectively, to the hidden units spins σ_i^z and σ_m^z are attached to [see Eqs. (51) and (52)]. Second, all the connections between spins σ_i^z , σ_m^z , and hidden units h_j are cut off [Eqs. (53) and (54)]. Third, four hidden units $h_{[lm1]}$, ..., $h_{[lm4]}$ are introduced, to mediate interactions between the two deep units and the physical spins l, m [Eqs. (61) and (62)]. Finally, two hidden units $h_{[lm5]}$ and $h_{[lm6]}$ are introduced, connecting both to $d_{[l]}$, $d_{[m]}$ and σ_i^z , σ_m^z with imaginary-valued

weights. The last step realizes the constraint $d_{[l]} + d_{[m]} = \sigma_i^z + \sigma_m^z$, through the “ $i\pi/4$, $i\pi/8$ ” trick discussed in Methods and the discussion of the 2d–6h representation in Supplementary Note 2.

In this representation, if the hidden neurons are traced out, the imaginary-time evolution becomes equivalent to that of the path-integral Monte Carlo method. More specifically, the number of deep neurons introduced at each time slice is exactly the same as the number of visible spins, and the deep neurons at each time slice can be regarded as additional classical spin degrees of freedom in the path-integral. Moreover, the constraint

$d_{[l]} + d_{[m]} = \sigma_l^z + \sigma_m^z$ ensures that the total magnetization is conserved at each time slice. Finally, the W and W' interactions reproduce the matrix element of $\exp(-\delta_\tau \mathcal{H}_{lm}^{\text{bond}})$ between neighboring time slices. See Supplementary Note 2 for more detail on this point.

2 deep + 4 hidden variables construction for Heisenberg model. A further possible solution to Eq. (16) is dubbed “2 deep, 4 hidden” (2d–4h) construction. In this case, we introduce two auxiliary deep variables $d_{[l]}$ and $d_{[m]}$. We also introduce four hidden units $h_{[l]}$, $h_{[m]}$, $h_{[lm1]}$, and $h_{[lm2]}$. Before the imaginary time evolution, $e^{-\delta_\tau \mathcal{H}_{lm}^{\text{bond}}}$, the physical variables σ_n^z ($n = l$ or m) are already coupled to each hidden variable h_j with a coupling W_{nj} . After the time evolution $e^{-\delta_\tau \mathcal{H}_{lm}^{\text{bond}}}$, as shown schematically in Fig. 3c, the coupling parameters are updated in the following way based on the old W_{nj} : First, the first deep unit $d_{[l]}$ becomes coupled to the already existing hidden variables h_j through the coupling $W'_{j[l]}$ given in Eq. (67). The second deep unit $d_{[m]}$ becomes similarly coupled to h_j through a term Z_{lmj} given in Eq. (67). Second, W_{nj} is updated to $\bar{W}_{nj} = W_{nj} + \Delta W_{nj}$ [see Eq. (66)]. Third, newly introduced $h_{[n]}$ ($n = l$ or m) gets coupled to $d_{[l]}$ through $W'_{[n][l]}$, and also to σ_n^z through $W_{n[n]}$ [Eqs. (71) and (73)].

Finally, as clarified in Methods, we also need to satisfy the constraint $d_{[l]}d_{[m]} = \sigma_l^z \sigma_m^z$. Such a constraint is represented in DBM form as

$$\sum_{h_{[m1]}, h_{[m2]}} \exp \left[\frac{i\pi}{4} \left(h_{[lm1]} + h_{[lm2]} \right) \left(\sigma_l^z + \sigma_m^z + d_{[l]} + d_{[m]} \right) \right], \quad (17)$$

which ensures $d_{[l]}d_{[m]} = \sigma_l^z \sigma_m^z$ after explicit summation of $h_{[lm1]}$ and $h_{[lm2]}$. Finally, we remark that the three constructions presented here have different intrinsic network topologies. In particular, 2d–6h gives rise to a local topology (because of the equivalence with the path-integral construction), 1d–3h has a local structure in the first layer and non-local in the second one, and 2d–4h is purely non-local in both layers.

Sampling strategies. With network structures explicitly determined, we now focus on the problem of extracting meaningful physical quantities from them. To this end, it is convenient to decompose the DBM weight into two parts, such that

$$\Psi_{\mathcal{W}}(\sigma^z) = \sum_{\{h,d\}} P_1(\sigma^z, h) P_2(h, d), \quad (18)$$

where $P_1(\sigma^z, h) = e^{\sigma^z \cdot a + \sigma^z \cdot W \cdot h + h \cdot b}$, and $P_2(h, d) = e^{h \cdot W' \cdot d + d \cdot b'}$. The expectation value of an arbitrary (few-body) operator \mathcal{O} can then be computed through the expression

$$\langle \mathcal{O} \rangle = \frac{\sum_{\{\sigma^z, h, h', d, d'\}} \Pi(\sigma^z, h, h', d, d') \mathcal{O}_{\text{loc}}(\sigma^z, h, h')}{\sum_{\{\sigma^z, h, h', d, d'\}} \Pi(\sigma^z, h, h', d, d')}, \quad (19)$$

where we have introduced the pseudo-probability density $\Pi(\sigma^z, h, h', d, d') \equiv P_1(\sigma^z, h) P_2(h, d) P_1^*(\sigma^z, h') P_2^*(h', d')$, and the “local” estimator

$$\mathcal{O}_{\text{loc}}(\sigma^z, h, h') = \frac{1}{2} \sum_{\sigma^z} \langle \sigma^z | \mathcal{O} | \sigma^z \rangle \left(\frac{P_1(\sigma^z, h)}{P_1(\sigma^z, h)} + \frac{P_1(\sigma^z, h')^*}{P_1(\sigma^z, h')^*} \right). \quad (20)$$

For the sampling over the Π distribution, a block Gibbs sampling analogous to what performed in standard DBM

architectures can be performed^{37,40}. Alternatively, it is possible to devise a set of Metropolis local updates sampling the exactly known marginals $\bar{\Pi}(\sigma^z, h, h') = \sum_{\{d, d'\}} \Pi(\sigma^z, h, h', d, d')$ or $\bar{\Pi}'(\sigma^z, d, d') = \sum_{\{h, h'\}} \Pi(\sigma^z, h, h', d, d')$.

In general, we have found that efficiently sampling the DBMs arising from the Heisenberg model constructions is typically more challenging than for the TFIM. This circumstance is a consequence of the imaginary couplings which set constraints on the value of hidden/deep units. These constraints typically make local Metropolis updates inefficient. With the notable exception of the 2d–6h representation, for which loop updates can be readily implemented, we leave the problem of designing efficient Monte Carlo sampling for the other Heisenberg constructions open. The sampling strategies adopted in our numerica are discussed more in detail in Supplementary Note 3.

Numerical results. We have implemented numerical algorithms to sample and obtain physical properties from the DBM previously derived. In Fig. 4a we show results for the one-dimensional TFIM. Specifically, we show the expectation value of the energy following the imaginary-time evolution starting from a fully polarized (in the x direction) initial state. The initial state corresponds to an empty network, where all the DBM parameters are set to zero. The DBM results closely match the exact imaginary-time evolution, thus verifying the correctness of our construction.

In Fig. 4a we also show the corresponding imaginary-time evolution as obtained from the approximate RBM construction, Eq. (15). As expected, this approximation is very accurate for short times, and breaks at later times.

Numerical results for the one-dimensional Heisenberg model are shown in Figs. 4b and 5a. Specifically, 4b shows the numerical check for the DBM (construction 2d–6h) time evolution for one-dimensional Heisenberg model for $N = 16$. As expected, the DBM results also in this case follow the exact time evolution. Figure 5a shows the dependence of the energy from the initial state, for $N = 80$ case. Specifically, by taking a pre-optimized variational RBM as an initial state, we can significantly decrease the time τ needed to reach the ground state.

Results for two-dimensional models are shown in Fig. 5b, both for the two-dimensional Heisenberg model, and for the frustrated $J_1 - J_2$ model, on 4×4 lattice with periodic boundary conditions.

In the case of the TFIM, sampling from the DBM is realized through the Gibbs scheme previously sketched, in conjunction with a parallel tempering scheme, to improve ergodicity in the sampling.

For the AFHM and for the $J_1 - J_2$ model with 2d–6h representation, we adopt loop updates⁴¹ used in the path-integral QMC method, because the imaginary-time evolution in the 2d–6h representation has a direct correspondence to the path-integral formulation, allowing for an efficient handling of the constraint $d_{[l]} + d_{[m]} = \sigma_l^z + \sigma_m^z$.

All the simulations carried here are sign-problem free, with the notable exception of the simulations carried on the two-dimensional $J_1 - J_2$ model. In this case, we start the imaginary-time evolution from a pre-optimized variational wave function, thus setting the fully evolved state as product of a DBM and the initial state. Because of the quality of the initial guess, a moderate sign problem can be numerically afforded for short time evolutions, and in this case it is enough to converge to the exact ground state (see Fig. 5b).

Discussion

We have shown how exact ground states of interacting spin Hamiltonians can be explicitly constructed using artificial

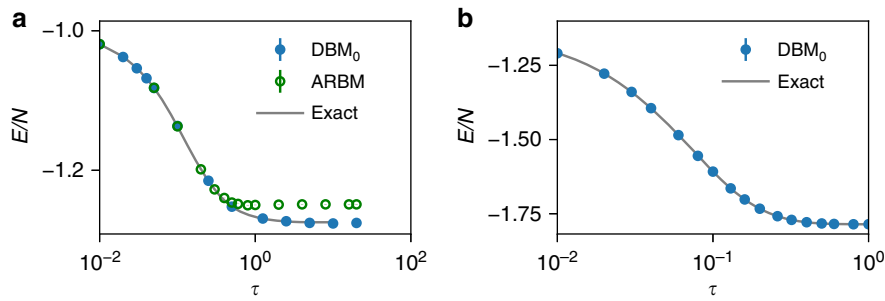


Fig. 4 Imaginary-time evolution with a DBM for 1D spin models. **a** Expectation value of energy of the transverse-field Ising Hamiltonian in the exact imaginary-time evolution (continuous line) is compared to the stochastic result obtained with a DBM (filled circles) ($\delta_\tau = 0.01$). Empty circles correspond to the approximate RBM evolution scheme, Eq. (15). We consider the critical point ($\Gamma_I = V_{Im}$), periodic boundary conditions, and $N = 20$ sites. **b** Expectation value of the isotropic antiferromagnetic Heisenberg Hamiltonian (AFHM) in the exact imaginary-time evolution (continuous line) is compared to the stochastic result obtained with a DBM ($\delta_\tau = 0.01$) following the 2d-6h construction. We consider periodic boundary conditions, $N = 16$ sites. The subscript α in DBM_α in panels (**a**, **b**) specifies a different initial state $|\Psi_0\rangle$: $\alpha = 1$ means that the initial state is an RBM state with hidden-unit density $M/N = 1$, whereas when $\alpha = 0$ the initial state is the empty-network state ($M = 0$). All energies are in units of the transverse field ($\Gamma_I = 1$) for the TFIM, and of the exchange ($J = 1$) for the AFHM

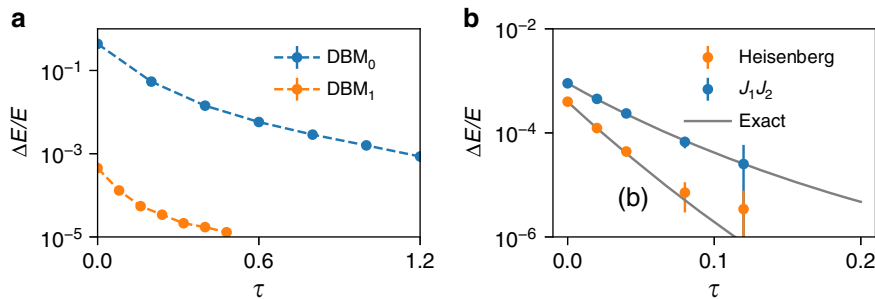


Fig. 5 Approaching the exact ground-state energy. **a** Relative error on the ground-state energy for the 1D AFHM as a function of the imaginary time. Here we consider periodic boundary conditions, $N = 80$ sites, and $\delta_\tau = 0.01$, in units of the exchange $J = 1$. The subscript α in DBM_α specifies a different initial state $|\Psi_0\rangle$: $\alpha = 1$ means that the initial state is an RBM state with hidden-unit density $M/N = 1$, whereas when $\alpha = 0$ the initial state is the empty-network state ($M = 0$). **b** Relative error on the ground-state energy for the two-dimensional $J_1 - J_2$ AFHM as a function of the imaginary time. As an energy unit, we consider $J_1 = 1$, and take $J_2 = 0.0$ and 0.4 , periodic boundary conditions, $N = 4 \times 4 = 16$ sites, and $\delta_\tau = 0.001$. Initial states are pre-optimized pair-product (geminal) state $|\psi_{pp}\rangle$ supplemented by Gutzwiller factor $P_G^\infty = \prod_l (1 - n_{l\uparrow} n_{l\downarrow})$ prohibiting double occupancy and quantum number projection onto the singlet state $\mathcal{L}^{S=0}$, i.e., $|\Psi_0\rangle = \mathcal{L}^{S=0} P_G^\infty |\psi_{pp}\rangle$. The PP states are given by $|\psi_{pp}\rangle = \left(\sum_{l,m=1}^N f_{lm}^\dagger c_{l\uparrow}^\dagger c_{m\downarrow}^\dagger \right)^{N/2} |0\rangle$, where f_{lm}^\dagger are variational parameters and $c_{l\sigma}^\dagger$ are the operators creating the electron with spin σ at l th site

neural networks comprising only two layers of hidden variables. In contrast to approaches based on one-layer RBMs, the constructions we have derived here do not require further variational optimization of the network parameters, and the exact representation of many-body ground states can be achieved with only polynomially many neurons. In the case of the Heisenberg model, all of the explicit algorithms presented here give rise to sign-problem-free representations, if the lattice is bipartite.

The DBM representation has an intrinsic conceptual value, as an alternative quantum-to-classical mapping to the path-integral representation. In the path-integral formalism, the addition of an extra dimension (the imaginary time direction) is needed to exactly represent the quantum many-body state. In our case, the DBM deep hidden layer plays a similar role as the additional dimension in the path integral. As argued in Methods [see Eq. (28)], a single-layer RBM is indeed sufficient to exactly, and efficiently describe the state of arbitrary classical spin systems. On the other hand, a second, deep layer is necessary for the efficient, and exact construction of compact networks describing quantum mechanical states.

DBM-based schemes can be further used to systematically improve upon existing RBM variational results. More generally, the initial state for the present DBM scheme can be generic variational states or even combinations of RBMs and more conventional wave functions^{24,33}. We have shown that, by starting the DBM construction from a pre-optimized variational state, a fast convergence to the exact ground state is observed. As shown in Fig. 5b, this kind of scheme opens the possibility of characterizing the ground state even in the case of non-bipartite lattices with frustration effects, exploiting the transient regime in which the sign problem can be still efficiently handled numerically, as for example discussed in ref. 42.

Methods

Useful identities. It is useful to introduce several identities, which can be used when more complicated interactions between the visible spins σ^z , hidden variables h and deep variables d beyond the standard form Eq. (1) are needed. The first identity reads

$$e^{s_1 s_2 V} = C \sum_{s_3 = \pm 1} e^{s_1 s_3 \tilde{V}_1 + s_2 s_3 \tilde{V}_2} = 2C \cosh(s_1 \tilde{V}_1 + s_2 \tilde{V}_2). \tag{21}$$

with

$$C = \frac{1}{2} e^{-|V|} \quad (22)$$

$$\tilde{V}_1 = \frac{1}{2} \operatorname{arccosh}\left(e^{2|V|}\right) \quad (23)$$

$$\tilde{V}_2 = \operatorname{sgn}(V) \times \tilde{V}_1 \quad (24)$$

for Ising variables s_1 and s_2 , and a real interaction V . This is a gadget for decomposing two-body interactions, and can be proven by examining all the cases of s_1 and s_2 .

By taking s_1 and s_2 as visible (physical) variables σ^z and s_3 as a hidden variable h , the direct classical two-body interaction between physical variables [the leftmost part in Eq. (21)] is cut and instead mediated by the hidden neuron h . Furthermore, a direct interaction between σ^z and d can also be decomposed: In the following derivations for the DBM wave constructions, for convenience, we sometimes introduce the direct interaction between σ^z and d , which is not allowed in the DBM structure. However, by taking s_1 as a visible spin σ^z , s_2 as a deep variable d , and s_3 as a hidden variable h in Eq. (21), one can eliminate the direct interaction between σ^z and d and decompose it into the interaction mediated only by h with trade-off of the summation over the hidden variable h . With this trick, one can recover the standard DBM form in Eq. (1).

Another identity (decomposition of four-body interaction) is

$$\begin{aligned} e^{s_1 s_2 s_3 s_4 V} &= \frac{1}{4} \sum_{s_5, s_6, s_7} \exp\left[i\frac{\pi}{4}(s_5 + s_6)(s_1 + s_2 + s_3 + s_7)\right] \\ &\quad \times \exp(s_4 s_7 V) \\ &= \sum_{s_7} \cos^2\left[\frac{\pi}{4}(s_1 + s_2 + s_3 + s_7)\right] \exp(s_4 s_7 V) \end{aligned} \quad (25)$$

for Ising variables s_i with $i = 1, \dots, 4$. Although we have introduced complex couplings in the first line, each term in the summation in the second line of Eq. (25) is positive definite if V is real. The second line remains nonzero only if $s_1 s_2 = s_3 s_7$, which proves the identity. This identity with s_1 and s_2 as physical variables, s_4, s_5 , and s_6 as hidden variables, and s_3 and s_7 as deep variables, reads

$$\begin{aligned} e^{\sigma_1 \sigma_2 d_1 h_1 V} &= \frac{1}{4} \sum_{h_2, h_3, d_2} \exp\left[i\frac{\pi}{4}(h_2 + h_3)(\sigma_1 + \sigma_2 + d_1 + d_2)\right] \\ &\quad \times \exp(h_1 d_2 V), \end{aligned} \quad (26)$$

Note that the right-hand side fits the DBM structure.

General three-body and two-body interactions can also be represented by the two-body form just by putting some of s_1, \dots, s_4 as constants in Eq. (25). These could be used instead of Eq. (21), although we employ Eq. (21) in the formalism below for the decoupling of the two-body interaction.

Finally, we discuss the gadgets for decomposing general N -body classical interactions using complex bias term b_j in addition to the couplings W and W' , whereas the gadgets Eqs. (21) and (26) are represented only by W and W' interactions. The gadget reads

$$e^{\sigma_1 \sigma_2 \dots \sigma_N V} = C \cos^2\left(b + \frac{\pi}{4} \sum_{i=1}^N \sigma_i\right) \quad (27)$$

$$= \frac{C}{4} \sum_{h_1, h_2} e^{ib(h_1 + h_2)} e^{i\frac{\pi}{4}(h_1 + h_2)(\sigma_1 + \sigma_2 + \dots + \sigma_N)} \quad (28)$$

with

$$b = \arctan(e^{-V}) - \frac{\pi}{4} \operatorname{mod}(N, 4), \quad (29)$$

$$C = \frac{1}{\cos(\arctan(e^{-V})) \times \sin(\arctan(e^{-V}))}. \quad (30)$$

This fact suggests that any classical partition function defined for Ising spins can be written exactly in terms of an RBM. Although the RBM is shown to be powerful in representing also the quantum states, there is no analytical way to map quantum states to the RBM and one must rely on numerical optimizations to get the RBM parameters. In the present study, we show analytical mappings from quantum states to the DBM, which has additional hidden layer. In the statistical mechanics, it is known that quantum systems with D dimension can be mapped on $(D + 1)$ -dimensional classical systems. Therefore, having additional hidden layer in neural network language is equivalent to acquiring additional dimension in statistical mechanics.

Transverse-field ising model. The solution of Eq. (9) is found in the following way. The left-hand side of Eq. (9) can be rewritten by using the notation Eq. (18) as

$$\sum_{\{h, d\}} P_1(\sigma^z, h) P_2(h, d) \left[1 + \tanh(\Gamma_l \delta_\tau) e^{-2\sigma_l^z \sum_j h_j W_{jl}} \right] = C \Psi_{\mathcal{W}}(\sigma^z). \quad (31)$$

We look for a solution by adding one deep neuron $d_{[l]}$ and creating new couplings $W'_{j[l]}$ to the existing hidden neurons h_j which are connected to σ_l^z . We also allow for changes in the existing interaction parameters. In particular we set the new couplings to be $\tilde{W}_{lj} = W_{lj} + \Delta W_{lj}$ (with ΔW_{lj} to be determined). Moreover, we introduce one hidden neuron $h_{[l]}$ coupled to σ_l^z and $d_{[l]}$ through the interactions $W_{l[l]}$ and $W'_{l[l]}$, respectively. If we trace out $h_{[l]}$, the hidden neuron $h_{[l]}$ mediates the interaction between σ_l^z and $d_{[l]}$ (denoted as $W''_{l[l]}$).

With this choice, we have (in the representation where $h_{[l]}$ is traced out):

$$\Psi_{\mathcal{W}}(\sigma^z) = \sum_{\{h, d\}} \sum_{d_{[l]}} P_1(\sigma^z, h) P_2(h, d) e^{\sigma_l^z \sum_j \Delta W_{lj} h_j + d_{[l]} \sum_j h_j W'_{j[l]} + \sigma_l^z d_{[l]} W''_{l[l]}}. \quad (32)$$

The equations to be verified are obtained considering the two possible values of $\sigma_l^z = \pm 1$:

$$e^{\sum_j h_j (\Delta W_{lj} + W'_{j[l]}) + W''_{l[l]}} + e^{\sum_j h_j (\Delta W_{lj} - W'_{j[l]}) - W''_{l[l]}} = C \times \left(1 + \tanh(\Gamma_l \delta_\tau) e^{-2 \sum_j h_j W_{lj}} \right) \quad (33)$$

$$e^{\sum_j h_j (-\Delta W_{lj} + W'_{j[l]}) - W''_{l[l]}} + e^{\sum_j h_j (-\Delta W_{lj} - W'_{j[l]}) + W''_{l[l]}} = C \times \left(1 + \tanh(\Gamma_l \delta_\tau) e^{2 \sum_j h_j W_{lj}} \right). \quad (34)$$

This equation has a solution from the requirement that the hidden unit interactions on the left and right hand sides match, thus we require

$$\Delta W_{lj} + W'_{j[l]} = -2W_{lj} \quad (35)$$

$$\Delta W_{lj} - W'_{j[l]} = 0, \quad (36)$$

and

$$W''_{l[l]} = \frac{\log \tanh(\Gamma_l \delta_\tau)}{2}. \quad (37)$$

Notice that when $\Gamma_l > 0$, $W''_{l[l]}$ is also real. By using Eq. (21) with the following replacement $s_1 \rightarrow \sigma_l^z$, $s_2 \rightarrow d_{[l]}$, $s_3 \rightarrow h_{[l]}$, $V \rightarrow W'_{l[l]}$, $\tilde{V}_1 \rightarrow W_{l[l]}$, and $\tilde{V}_2 \rightarrow W''_{l[l]}$, the last condition determines the real couplings $W_{l[l]}$ and $W'_{l[l]}$ as Eqs. (11) and (12).

Heisenberg model. Here, we show the derivation for the general form of bond Hamiltonian allowing anisotropy and bond-disorder: $\mathcal{H}_{lm}^{\text{bond}} = J_{lm}^{xy} (\sigma_l^x \sigma_m^x + \sigma_l^y \sigma_m^y) + J_{lm}^z \sigma_l^z \sigma_m^z$. In the case of the bipartite lattice and the antiferromagnetic exchange $J_{lm}^z, J_{lm}^{xy} > 0$, we further apply a local gauge transformation by a π rotation around the z -axis in the spin space as $\sigma^x \rightarrow -\sigma^x$ and $\sigma^y \rightarrow -\sigma^y$ on one of the sublattices, which gives a $-$ sign for $\sigma_l^x \sigma_m^x$ and $\sigma_l^y \sigma_m^y$ interactions. This transformation is equivalent to taking

$$J_{lm}^{xy} \rightarrow -J_{lm}^{xy}. \quad (38)$$

The gauge transformation enables to design a DBM neural network with real couplings $\{W, W'\}$ except for those to put ‘constraint’ on the values of deep neuron spins (see more detail about the constraint in the following sections). It ensures that the DBM algorithm has no negative sign problems.

In the case of the antiferromagnetic Heisenberg model after the gauge transformation on the bipartite lattice, we must solve, for each bond,

$$\begin{aligned} \delta_{\sigma_l^z, \sigma_m^z} e^{-\delta_l J_{lm}^z} \Psi_{\mathcal{W}}(\sigma^z) + \left(1 - \delta_{\sigma_l^z, \sigma_m^z}\right) e^{\delta_l J_{lm}^z} \\ (\Psi_{\mathcal{W}}(\sigma^z) \cosh(2J_{lm}^{xy} \delta_\tau) + \Psi_{\mathcal{W}}(\sigma_l^z \leftrightarrow \sigma_m^z) \sinh(2J_{lm}^{xy} \delta_\tau)) \\ = C \langle \sigma^z | \Psi_{\mathcal{W}} \rangle. \end{aligned} \quad (39)$$

It is also useful to explicitly write the expression for the exchange term in the

second line above:

$$\begin{aligned} & \Psi_{\mathcal{W}}(\sigma^z) \cosh(2J_{lm}^{xy} \delta_\tau) + \Psi_{\mathcal{W}}(\sigma_l^z \leftrightarrow \sigma_m^z) \sinh(2J_{lm}^{xy} \delta_\tau) \\ &= \sum_{\{h,d\}} P_1(\sigma^z, h) P_2(h, d) \left[\cosh(2J_{lm}^{xy} \delta_\tau) \right. \\ & \quad \left. + \sinh(2J_{lm}^{xy} \delta_\tau) e^{\sum_j (\sigma_m^z - \sigma_l^z) h_j (W_{ij} - W_{mj})} \right]. \end{aligned} \quad (40)$$

In the following derivations, for the antiferromagnetic Hamiltonian ($J_{lm}^z, J_{lm}^{xy} > 0$) after the gauge transformation, we look for a solution with zero bias terms

($a_i, b_j, b'_k = 0, \forall i, j, k$). We can also derive a sign-problem-free solution for the imaginary time evolution in the absence of the explicit gauge transformation by introducing a complex bias term a_i . Indeed, in the “2 deep, 4 hidden” representation, we will explicitly show that taking a specific set of complex bias term a_i on physical spins is equivalent to the gauge transformation, making a solution free from the sign problem.

In a way similar to the TFIM, solutions of Eq. (39) can be found by specifying the structure of the DBM and the three examples are the following.

1d-3h construction for Heisenberg model. We assume the structure of the updated wave function (corresponding to Eq. (32) for the TFIM) to be

$$\begin{aligned} \Psi_{\mathcal{W}}(\sigma^z) &= \sum_{\{h,d\}} \sum_{d_{[lm]} = \pm 1} P_1(\sigma^z, h) P_2(h, d) \\ & \quad d_{[lm]} = \sigma_l^z \text{ if } \sigma_l^z = \sigma_m^z \\ & \quad e^{\sum_j \Delta W_{ij} h_j + d_{[lm]} \sum_j h_j W'_{j[lm]} + d_{[lm]} \sigma_l^z W''_{l[lm]} + V_{[lm]} \sigma_l^z \sigma_m^z}. \end{aligned} \quad (41)$$

Similarly to the case of the TFIM, a solution of Eq. (39) is given by

$$\Delta W_{lj} = -W_{lj} + W_{mj} \quad (42)$$

$$W'_{j[lm]} = W_{lj} - W_{mj}. \quad (43)$$

and

$$W''_{l[lm]} = -(\log \tanh(2J_{lm}^{xy} \delta_\tau))/2 \quad (44)$$

$$V_{[lm]} = -(\log \cosh(2J_{lm}^{xy} \delta_\tau))/2 - J_{lm}^z \delta_\tau \quad (45)$$

Notice that the first condition is equivalent to cutting all connections from spin l to the hidden units and attaching the spin l to all the hidden units connected to spin m , with an interaction W_{mj} .

Although the terms proportional to $W'_{j[lm]}$ and V_{lm} do not satisfy the standard DBM form, they can be transformed to the DBM form by introducing new hidden neurons $h_{[lm1]}$ and $h_{[lm2]}$ [see the gadget Eq. (21)]:

$$e^{\sigma_l^z d_{[lm]} W'_{l[lm]}} = C_{[lm1]} \sum_{h_{[lm1]}} e^{\sigma_l^z h_{[lm1]} W_{l[lm1]} + h_{[lm1]} d_{[lm]} W'_{[lm1][lm]}}, \quad (46)$$

with

$$W_{l[lm1]} = W'_{[lm1][lm]} = \frac{1}{2} \operatorname{arccosh} \left(\frac{1}{\tanh(2J_{lm}^{xy} \delta_\tau)} \right). \quad (47)$$

Similarly, the coupling $V_{[lm]}$ is decomposed as

$$e^{\sigma_l^z \sigma_m^z V_{[lm]}} = C_{[lm2]} \sum_{h_{[lm2]}} e^{\sigma_l^z h_{[lm2]} W_{l[lm2]} + \sigma_m^z h_{[lm2]} W_{m[lm2]}}, \quad (48)$$

with

$$W_{l[lm2]} = -W_{m[lm2]} = \frac{1}{2} \operatorname{arccosh}(\cosh(2J_{lm}^{xy} \delta_\tau) e^{2J_{lm}^z \delta_\tau}). \quad (49)$$

Finally, as discussed in the main text, the constraint $d_{[lm]} = \sigma_l^z$ when $\sigma_l^z = \sigma_m^z$ can be satisfied by adding the third neuron $h_{[lm3]}$, introducing pure complex couplings

2d-6h construction for Heisenberg model. In this case, the form of the new wave function reads

$$\begin{aligned} \Psi_{\mathcal{W}}(\sigma^z) &= \sum_{\{h,d\}} \sum_{d_{[l]}, d_{[m]}} P_1(\sigma^z, h) P_2(h, d) \\ & \quad d_{[l]} + d_{[m]} = \sigma_l^z + \sigma_m^z \\ & \quad e^{\sum_j \sum_{n=l,m} h_j (\Delta W_{nj} \sigma_n^z + W'_{j[n]} d_{[n]}) + \sum_{n=l,m} \sigma_n^z (W''_{n[l]} d_{[l]} + W''_{n[m]} d_{[m]})}. \end{aligned} \quad (50)$$

A solution of Eq. (39) is given by

$$W'_{j[l]} = W_{lj}, \quad (51)$$

$$W'_{j[m]} = W_{mj}, \quad (52)$$

$$\Delta W_{lj} = -W_{lj}, \quad (53)$$

$$\Delta W_{mj} = -W_{mj}, \quad (54)$$

and

$$W''_{l[l]} = W''_{m[m]} = -\frac{J_{lm}^z \delta_\tau}{2} - \frac{1}{4} \log \sinh(2J_{lm}^{xy} \delta_\tau), \quad (55)$$

$$W''_{l[m]} = W''_{m[l]} = -\frac{J_{lm}^z \delta_\tau}{2} - \frac{1}{4} \log \cosh(2J_{lm}^{xy} \delta_\tau). \quad (56)$$

The direct interactions between $(\sigma_l^z, d_{[l]})$, $(\sigma_m^z, d_{[m]})$, $(\sigma_l^z, d_{[m]})$, and $(\sigma_m^z, d_{[l]})$, are mediated by $h_{[lm1]}$, $h_{[lm2]}$, $h_{[lm3]}$, and $h_{[lm4]}$, respectively, as follows:

$$e^{\sigma_l^z d_{[l]} W'_{l[l]}} = C_{[lm1]} \sum_{h_{[lm1]}} e^{\sigma_l^z h_{[lm1]} W_{l[lm1]} + h_{[lm1]} d_{[l]} W'_{[lm1][l]}}, \quad (57)$$

$$e^{\sigma_m^z d_{[m]} W'_{m[m]}} = C_{[lm2]} \sum_{h_{[lm2]}} e^{\sigma_m^z h_{[lm2]} W_{m[lm2]} + h_{[lm2]} d_{[m]} W'_{[lm2][m]}}, \quad (58)$$

$$e^{\sigma_l^z d_{[m]} W'_{l[m]}} = C_{[lm3]} \sum_{h_{[lm3]}} e^{\sigma_l^z h_{[lm3]} W_{l[lm3]} + h_{[lm3]} d_{[m]} W'_{[lm3][m]}}, \quad (59)$$

$$e^{\sigma_m^z d_{[l]} W'_{m[l]}} = C_{[lm4]} \sum_{h_{[lm4]}} e^{\sigma_m^z h_{[lm4]} W_{m[lm4]} + h_{[lm4]} d_{[l]} W'_{[lm4][l]}}. \quad (60)$$

By applying the gadget Eq. (21), the new W and W' interactions are given by, for small δ_τ (such that $\frac{e^{-J_{lm}^z \delta_\tau}}{\sqrt{\sinh(2J_{lm}^{xy} \delta_\tau)}} > 1$),

$$\begin{aligned} W_{l[lm1]} &= W'_{[lm1][l]} = W_{m[lm2]} = W'_{[lm2][m]} \\ &= \frac{1}{2} \operatorname{arccosh} \left(\frac{e^{-J_{lm}^z \delta_\tau}}{\sqrt{\sinh(2J_{lm}^{xy} \delta_\tau)}} \right) \end{aligned} \quad (61)$$

and

$$\begin{aligned} W_{l[lm3]} &= -W'_{[lm3][m]} = W_{m[lm4]} = -W'_{[lm4][l]} \\ &= \frac{1}{2} \operatorname{arccosh} \left(\sqrt{\cosh(2J_{lm}^{xy} \delta_\tau)} \times e^{J_{lm}^z \delta_\tau} \right). \end{aligned} \quad (62)$$

Finally, the constraint $d_{[l]} + d_{[m]} = \sigma_l^z + \sigma_m^z$ can be put by introducing additionally two hidden neurons $h_{[lm5]}$ and $h_{[lm6]}$, and by introducing complex couplings

$$\begin{aligned} & \sum_{h_{[lm5]}, h_{[lm6]}} e^{i\pi} (\sigma_l^z + \sigma_m^z) h_{[lm5]} - h_{[lm5]} (d_{[l]} + d_{[m]}) \\ & \quad \times e^{i\pi} ((\sigma_l^z + \sigma_m^z) h_{[lm6]} - h_{[lm6]} (d_{[l]} + d_{[m]})) \end{aligned} \quad (63)$$

This term gives interactions among $d_{[l]}$, $d_{[m]}$, σ_l^z and σ_m^z : $4 \cos\left(\frac{\pi}{4} (\sigma_l^z + \sigma_m^z - d_{[l]} - d_{[m]})\right) \cos\left(\frac{\pi}{8} (\sigma_l^z + \sigma_m^z - d_{[l]} - d_{[m]})\right)$, which realize the constraint.

2d-4h construction for Heisenberg model. For this construction, we assume the following structure for the wave-function after the propagator:

$$\Psi_{\mathcal{W}}(\sigma^z) = \sum_{\{h,d\}} \sum_{d_{[l]}} P_1(\sigma^z, h) P_2(h, d) e^{i \sum_{n=l,m} \sigma_n^z h_j \Delta W_{nj}} \times e^{\sum_j h_j d_{[l]} W'_{j[l]} + \sum_{n=l,m} \sigma_n^z d_{[l]} W'_{n[l]} + \sum_j \sigma_j^z \sigma_m^z h_j d_{[l]} Z_{lmj}} \quad (64)$$

In this case, we also look for a solution for the bond operator without the gauge transformation. This shows that the introduction of a complex bias term a_i can play the same role as the gauge transformation. Then, we need to solve

$$\delta_{\sigma_l^z, \sigma_m^z} e^{-\delta_l \tau J_{lm}} \Psi_{\mathcal{W}}(\sigma^z) + \left(1 - \delta_{\sigma_l^z, \sigma_m^z}\right) e^{\delta_l \tau J_{lm}} \left(\Psi_{\mathcal{W}}(\sigma^z) \cosh(2J_{lm}^{\text{xy}} \delta_\tau) - \Psi_{\mathcal{W}}(\sigma_l^z \leftrightarrow \sigma_m^z) \sinh(2J_{lm}^{\text{xy}} \delta_\tau)\right) = C(\sigma^z | \Psi_{\mathcal{W}}). \quad (65)$$

Note that the sign for $\Psi_{\mathcal{W}}(\sigma_l^z \leftrightarrow \sigma_m^z) \sinh(2J_{lm}^{\text{xy}} \delta_\tau)$ term is different from that in Eq. (39).

A solution of Eq. (65) is obtained as

$$\Delta W_{lj} = -\Delta W_{mj} = -\frac{1}{2} (W_{lj} - W_{mj}), \quad (66)$$

where W_{nj} ($n = l, m$) is updated to \bar{W}_{nj} with the increment ΔW_{nj} as $\bar{W}_{nj} = W_{nj} + \Delta W_{nj}$. The new couplings $W'_{j[l]}$, Z_{lmj} and $W'_{n[l]}$ are also given by

$$W'_{j[l]} = -Z_{lmj} = -\frac{1}{2} (W_{lj} - W_{mj}) \quad (67)$$

and

$$W'_{l[l]} = \frac{1}{4} \left[\log \left[-e^{-2a_{l-m}} \tanh(2J_{lm}^{\text{xy}} \delta_\tau) \right] + 2 \operatorname{arccosh} \left[\frac{e^{-2J_{lm}^{\text{xy}} \delta_\tau}}{\sqrt{-2e^{-2a_{l-m}} \sinh(4J_{lm}^{\text{xy}} \delta_\tau)}} \right] \right] \quad (68)$$

$$W'_{m[l]} = \frac{1}{4} \left[-\log \left[-e^{-2a_{l-m}} \tanh(2J_{lm}^{\text{xy}} \delta_\tau) \right] + 2 \operatorname{arccosh} \left[\frac{e^{-2J_{lm}^{\text{xy}} \delta_\tau}}{\sqrt{-2e^{-2a_{l-m}} \sinh(4J_{lm}^{\text{xy}} \delta_\tau)}} \right] \right] \quad (69)$$

with $a_{l-m} = a_l - a_m$. On a bipartite lattice, to avoid the negative sign (or complex phase) problem we need to keep $W'_{l[l]}$ and $W'_{m[l]}$ real. This can be achieved by choosing $a_l = 0$ for any l if $J_{lm} < 0$ (ferromagnetic case). For $J_{lm} > 0$ (antiferromagnetic case), $a_l = n\pi i$ with an arbitrary integer n if the site l belongs to the sub-lattice A and $a_l = (n + 1/2)\pi i$ if l belongs to the sub-lattice B. This local gauge for $J_{lm} > 0$ is equivalent to the transformation $J_{lm}^{\text{xy}} \rightarrow -J_{lm}^{\text{xy}}$ and $a_l = 0$ for any site l . We further notice that $W'_{m[l]}$ can be taken positive if we take a sufficiently small δ_τ in Eq. (69), with the leading order term $-\log(2J_{lm}^{\text{xy}} \delta_\tau)/2$. On the other hand, in Eq. (68), the leading order term is negative ($= -J_{lm} \delta_\tau$).

To recover the original form of the DBM, we first use Eq. (21) with the replacement $s_1 \rightarrow \sigma_n^z$, $s_2 \rightarrow d_{[l]}$, $s_3 \rightarrow h_{[n]}$, $C \rightarrow D_n$, $V \rightarrow W'_{n[l]}$, $\bar{V}_1 \rightarrow W_{n[n]}$, and $\bar{V}_2 \rightarrow W'_{n[l]}$ for $n = l, m$. Then a solution for D_n , $W_{n[n]}$, and $W'_{n[l]}$ are represented by using $W'_{n[l]}$ as

$$D_n = \frac{1}{2} \exp \left[-W'_{n[l]} \right] \quad (70)$$

$$W_{n[n]} = W'_{n[l]} = \frac{1}{2} \operatorname{arccosh} \left(\exp \left[2W'_{n[l]} \right] \right), \quad (71)$$

for positive $W'_{n[l]}$ and

$$D_n = \frac{1}{2} \exp \left[W'_{n[l]} \right] \quad (72)$$

$$W_{n[n]} = -W'_{n[l]} = \frac{1}{2} \operatorname{arccosh} \left(\exp \left[-2W'_{n[l]} \right] \right), \quad (73)$$

for negative $W'_{n[l]}$ to give real $W_{n[n]}$ and $W'_{n[l]}$.

To completely recover the original DBM form, we next use Eq. (26) by replacing σ_1 with σ_l^z , σ_2 with σ_m^z , d_1 with $d_{[l]}$, d_2 with $d_{[m]}$, h_1 with h_j , h_2 with $h_{[lm1]}$, h_3 with $h_{[lm2]}$, and V with Z_{lmj} .

With these solutions, by ignoring the trivial constant factors including D_l and D_m , the evolution is described by introducing two deep and four hidden additional

variables $d_{[l]}$, $d_{[lm]}$, $h_{[l]}$, $h_{[m]}$, $h_{[lm1]}$, and $h_{[lm2]}$ as

$$\Psi_{\mathcal{W}}(\sigma^z) = \sum_{\{h,d\}} P_1(\sigma^z, h) P_2(h, d) \exp \left[\sum_{j,n=l,m} \sigma_n^z h_j \Delta W_{nj} + \sum_j h_j d_{[l]} W'_{j[l]} + \sum_{n=l,m} h_{[n]} \left(\sigma_n^z W_{n[n]} + d_{[l]} W'_{n[l]} \right) + d_{[lm]} \sum_j h_j Z_{lmj} + \frac{i\pi}{4} \left(h_{[lm1]} + h_{[lm2]} \right) \left(\sigma_l^z + \sigma_m^z + d_{[l]} + d_{[m]} \right) \right], \quad (74)$$

where $\{\bar{h}, \bar{d}\}$ is a set consisting of the existing and new neurons.

Code availability. Computer codes to create the deep Boltzmann machine networks for each model are provided as Supplementary Software 1–4. Other code written for and used in this study is available from the corresponding author upon reasonable request.

Data availability

The datasets generated during and/or analyzed during the current study are available from the corresponding author on reasonable request.

Received: 26 February 2018 Accepted: 8 November 2018

Published online: 14 December 2018

References

- Feynman, R. P. Space-time approach to non-relativistic quantum mechanics. *Rev. Mod. Phys.* **20**, 367–387 (1948).
- Dyson, F. J. The S matrix in quantum electrodynamics. *Phys. Rev.* **75**, 1736–1755 (1949).
- Hubbard, J. Calculation of partition functions. *Phys. Rev. Lett.* **3**, 77–78 (1959).
- Stratonovich, R. L. On a method of calculating quantum distribution functions. *Sov. Phys. Dokl.* **2**, 416–419 (1957).
- Abrikosov, A. A. *Methods of Quantum Field Theory in Statistical Physics*. (Dover Publications, New York, 1975). revised edition.
- Binder, K. *Applications of the Monte Carlo Method in Statistical Physics*. (Springer Verlag, Berlin, 1984).
- Takahashi, M. & Imada, M. Monte carlo calculation of quantum systems. *J. Phys. Soc. Jpn.* **53**, 963 (1984).
- Takahashi, M. & Imada, M. Monte carlo calculation of quantum systems. II. Higher order correction. *J. Phys. Soc. Jpn.* **53**, 3765 (1984).
- Ceperley, D. Path-integrals in the theory of condensed helium. *Rev. Mod. Phys.* **67**, 279–355 (1995).
- Suzuki, M. Relationship between d-dimensional quantum spin systems and (d + 1)-dimensional ising systems: equivalence, critical exponents and systematic approximants of the partition function and spin correlations. *Prog. Theor. Phys.* **56**, 1454–1469 (1976).
- Hirsch, J. E., Sugar, R., Scalapino, D. & Blankenbecler, R. Monte carlo simulations of one-dimensional fermion systems. *Phys. Rev. B* **26**, 5033–5055 (1982).
- Beard, B. & Wiese, U.-J. Simulations of discrete quantum systems in continuous euclidean time. *J. Phys. Rev. Lett.* **77**, 5130–5133 (1996).
- Sandvik, A. W. Stochastic series expansion method with operator-loop update. *Phys. Rev. B* **59**, R14157–R14160 (1999).
- Prokofev, N. & Svistunov, B. Bold diagrammatic Monte Carlo technique: when the sign problem is welcome. *Phys. Rev. Lett.* **99**, 250201 (2007).
- Feynman, R. P. Atomic theory of the two-fluid model of liquid helium. *Phys. Rev.* **94**, 262–277 (1954).
- Gros, C. Physics of projected wavefunctions. *Ann. Phys.* **189**, 53–88 (1989).
- Kashima, T. & Imada, M. Path-integral renormalization group method for numerical study on ground states of strongly correlated electronic systems. *J. Phys. Soc. Jpn.* **70**, 2287–2299 (2001).
- Tahara, D. & Imada, M. Variational monte carlo method combined with quantum-number projection and multi-variable optimization. *J. Phys. Soc. Jpn.* **77**, 114701 (2008).
- Becca, F. & Sorella, S. *Quantum Monte Carlo Approaches for Correlated Systems*. (Cambridge University Press, Cambridge, UK; New York, NY, 2017).
- White, S. R. Density-matrix algorithms for quantum renormalization groups. *Phys. Rev. B* **48**, 10345–10356 (1993).
- Orús, R. A practical introduction to tensor networks: Matrix product states and projected entangled pair states. *Ann. Phys.* **349**, 117–158 (2014).
- Carleo, G. & Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **355**, 602–606 (2017).
- Torlai, G. et al. Many-body quantum state tomography with neural networks. *Nat. Phys.* **14**, 447–450 (2018).

24. Nomura, Y., Darmawan, A. S., Yamaji, Y. & Imada, M. Restricted boltzmann machine learning for solving strongly correlated quantum systems. *Phys. Rev. B* **96**, 205152 (2017).
25. Deng, D.-L., Li, X. & Das Sarma, S. Quantum entanglement in neural network states. *Phys. Rev. X* **7**, 021021 (2017).
26. Rocchetto, A., Grant, E., Strelchuk, S., Carleo, G. & Severini, S. Learning hard quantum distributions with variational autoencoders. *npj Quantum Inf.* **4**, 28 (2018).
27. Glasser, I., Pancotti, N., August, M., Rodriguez, I. D. & Cirac, J. I. Neural networks quantum states, string-bond states and chiral topological states. *Phys. Rev. X* **8**, 011006 (2018).
28. Kaubruegger, R., Pastori, L. & Budich, J. C. Chiral topological phases from artificial neural networks. *Phys. Rev. B* **97**, 195136 (2018).
29. Cai, Z. Approximating quantum many-body wave-functions using artificial neural networks. *Phys. Rev. B* **97**, 035116 (2018).
30. Saito, H. & Kato, M. Machine learning technique to find quantum many-body ground states of Bosons on a lattice. *J. Phys. Soc. Jpn.* **87**, 014001 (2017).
31. Saito, H. Solving the Bose–Hubbard model with machine learning. *J. Phys. Soc. Jpn.* **86**, 093001 (2017).
32. Chen, J., Cheng, S., Xie, H., Wang, L. & Xiang, T. Equivalence of restricted Boltzmann machines and tensor network states. *Phys. Rev. B* **97**, 085104 (2018).
33. Clark, S. R. Unifying neural-network quantum states and correlator product states via tensor networks. *J. Phys. A* **51**, 135301 (2018).
34. Deng, D.-L., Li, X. & Das Sarma, S. Machine learning topological states. *Phys. Rev. B* **96**, 195145 (2017).
35. Gao, X. & Duan, L.-M. Efficient representation of quantum many-body states with deep neural networks. *Nat. Commun.* **8**, 662 (2017).
36. Huang, Y. & Moore, J. E. Neural network representation of tensor network and chiral states. Preprint at <http://arxiv.org/abs/1701.06246> (2017).
37. Salakhutdinov, R. & Hinton, G. Deep Boltzmann machines. *Proc. Mach. Learn. Res.* **5**, 448–455 (2009).
38. Trotter, H. F. On the product of semi-groups of operators. *Proc. Am. Math. Soc.* **10**, 545–551 (1959).
39. Freitas, N., Morigi, G. & Dunjko, V. Neural network operations and Susuki–Trotter evolution of neural network states. Preprint at <http://arxiv.org/abs/1803.02118> (2018).
40. Salakhutdinov, R. & Hinton, G. An efficient learning procedure for deep Boltzmann machines. *Neural Comput.* **24**, 1967–2006 (2012).
41. Evertz, H. G., Lana, G. & Marcu, M. Cluster algorithm for vertex models. *Phys. Rev. Lett.* **70**, 875–879 (1993).
42. Ceperley, D. M. & Alder, J. Ground state of the electron gas by a stochastic method. *Phys. Rev. Lett.* **45**, 566–569 (1980).

Acknowledgements

G.C. acknowledges useful discussions with Xun Gao, and Markus Heyl. Y.N. and M.I. are grateful for the useful discussions with Youhei Yamaji and Andrew S. Darmawan. Y.N. was financially supported by Grant-in-Aids for Scientific Research (JSPS KAKENHI)

(No. 17K14336). M.I. and Y.N. were financially supported by a Grant-in-Aid for Scientific Research (No. 16H06345) from Ministry of Education, Culture, Sports, Science and Technology, Japan. Part of the calculations were done at Supercomputer Center, Institute for Solid State Physics, University of Tokyo. This work was also supported in part by MEXT as a social and scientific priority issue (Creation of new functional devices and high-performance materials to support next-generation industries CDMSI) to be tackled by using post-K computer. We also thank the support provided by the RIKEN Advanced Institute for Computational Science through the HPCI System Research project (hp170263) supported by Ministry of Education, Culture, Sports, Science, and Technology, Japan.

Author contributions

G.C. conceived the general idea and contributed the Ising model, and the approximate RBM construction. G.C., Y.N. and M.I. each contributed one of the three Heisenberg model representations. Numerical simulations were performed by Y.N. and G.C. All authors contributed equally to the manuscript preparation and presentation of the results.

Additional information

Supplementary Information accompanies this paper at <https://doi.org/10.1038/s41467-018-07520-3>.

Competing interests: The authors declare no competing interests.

Reprints and permission information is available online at <http://npg.nature.com/reprintsandpermissions/>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2018