Supplemental Materials for: Purifying Deep Boltzmann Machines for Thermal Quantum States

Yusuke Nomura,[1,*] Nobuyuki Yoshioka,[2,3,†] and Franco Nori[4,3,5]

[1] *RIKEN Center for Emergent Matter Science, 2-1 Hirosawa, Wako, Saitama 351-0198, Japan*
[2] *Department of Applied Physics, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan*
[3] *Theoretical Quantum Physics Laboratory, RIKEN Cluster for Pioneering Research (CPR), Wako-shi, Saitama 351-0198, Japan*
[4] *RIKEN Center for Quantum Computing (RQC), Wako-shi, Saitama 351-0198, Japan*
[5] *Physics Department, University of Michigan, Ann Arbor, Michigan 48109-1040, USA*

(Dated: June 9, 2021)

S1. PRACTICAL DETAILS FOR METHOD (I)

A. Analytical imaginary time evolution for transverse-field Ising model

In Method (I), we construct a DBM such that the imaginary-time evolution is realized analytically within the Trotter error. As a concrete example, let us consider the transverse-field Ising (TFI) model on the $N_{\rm site}$ spin chain with periodic boundary condition,

$$\mathcal{H} = \mathcal{H}_1 + \mathcal{H}_2, \tag{S1}$$

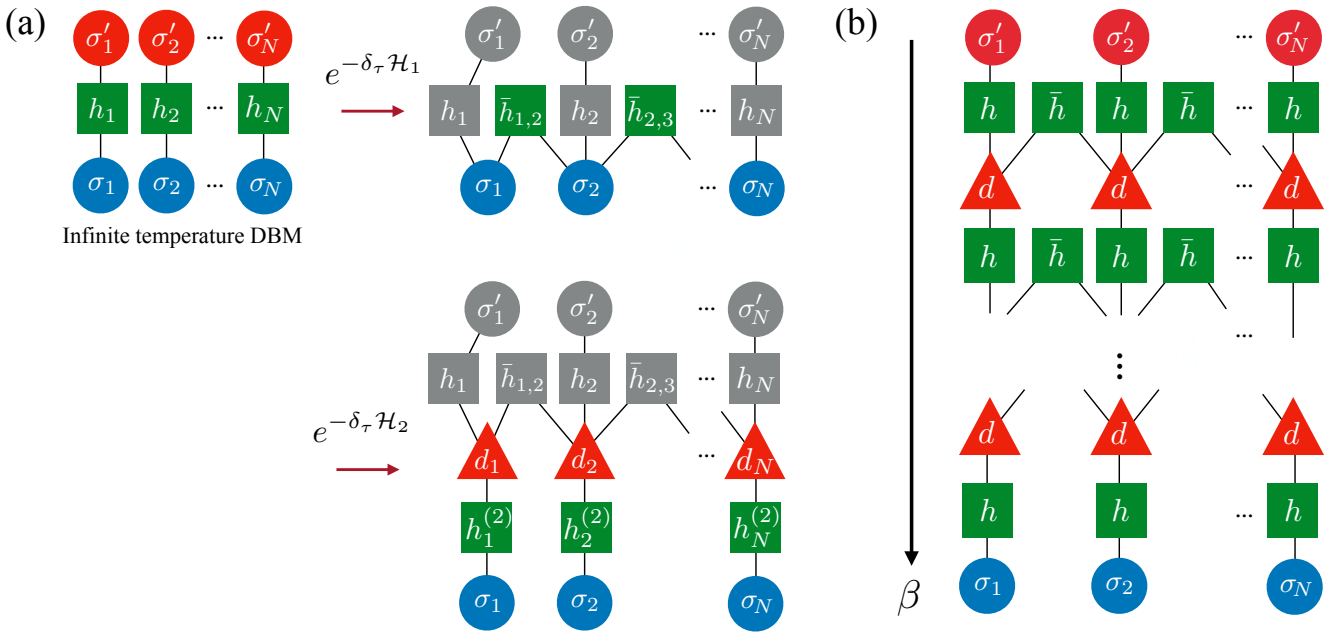

FIG. S1. Analytical construction of a DBM which realizes the imaginary-time evolution of the transverse-field Ising model. (a) Starting from the DBM representing the infinite-temperature state, we first encode the interaction propagator ($\nu = 1$) by introducing a single hidden spin for every interacting visible spins. Then, for the magnetic-field propagator ($\nu = 2$), we add deep- and hidden-spin layers in between the visible layer and neighboring hidden layer. (b) As we repetitively encode the propagators, the network structure grows vertically, with the depth increasing in proportion to the number of Trotter steps.

* These two authors contribute equally to the paper. E-mail at: yusuke.nomura@riken.jp
† These two authors contribute equally to the paper. E-mail at: nyoshioka@ap.t.u-tokyo.ac.jp

where $\mathcal{H}_1 = -J\sum_i \sigma_i^z\sigma_{i+1}^z$ is the Ising-type interaction with amplitude J and $\mathcal{H}_2 = -\Gamma\sum_i \sigma_i^x$ is the transverse magnetic field with amplitude Γ, where σ_i^a $(a = x, y, z)$ denotes the Pauli matrix operating on the i-th site. In the following, we provide the solutions to realize a small-time propagation δ_τ by updating the DBM parameters from $\mathcal{W}$ to $\bar{\mathcal{W}}$, i.e., $e^{-\delta_\tau\mathcal{H}_\nu}|\Psi_{\mathcal{W}}\rangle = C|\Psi_{\bar{\mathcal{W}}}\rangle$ (C is a constant) [1].

For $\nu = 1$ (interaction propagator), a solution is to add one hidden spin per bond and put the couplings with strength $\frac{1}{2}\text{arcosh}(e^{2J\delta_\tau})$ to the visible spins on each bond [See Fig S1(a)].

This is followed by the propagator for $\nu = 2$ (transverse-field propagator), for which we perform the following for each site: (1) cut the existing couplings between the hidden and visible spins and (2) add a new deep spin d_{new} that is coupled with the hidden spins.

After applying (1) and (2) for each site, we furthermore introduce a new hidden spin which is coupled to the visible and the newly added deep spins $\{d_{\text{new}}\}$ with the interaction $\frac{1}{2}\text{arcosh}\left(\frac{1}{\tanh(\Gamma\delta_\tau)}\right)$.

By repeating these operations until the desired inverse temperature is reached, we obtain a DBM that exactly encodes the Gibbs state via the quantum-to-classical correspondence [See Fig S1(b)]. Here, the number of the hidden and deep spins scales as $O(N_\tau N_{\text{site}})$, where N_τ is the number of Trotter steps and N_{site} is the number of physical spins.

While we have described the case for the lowest-order Suzuki-Trotter decomposition, we may alternatively consider higher-order ones so that the errors are suppressed. In the actual numerical implementation for the 1D TFI model, we employ the 2nd-order expansion

$$\exp\left[-\delta_\tau(\mathcal{H}_1+\mathcal{H}_2)\right] = \exp(-\delta_\tau\mathcal{H}_1/2)\exp(-\delta_\tau\mathcal{H}_2)\exp(-\delta_\tau\mathcal{H}_1/2) + \mathcal{O}(\delta_\tau^3). \tag{S2}$$

B. Monte Carlo sampling scheme for Method (I)

Here, we describe how we calculate the expectation value of a physical observable $\langle\mathcal{O}\rangle$ from the constructed DBM states. Considering that the norm of the DBM states is given by

$$\langle\Psi|\Psi\rangle = \sum_{\sigma,\sigma'}\left|\Psi(\sigma,\sigma')\right|^2 = \sum_{\sigma,\sigma'}\sum_{h_1,d_1,h_2,d_2}\phi^*(\sigma,\sigma';h_1,d_1)\phi(\sigma,\sigma';h_2,d_2) = \sum_{\sigma,\sigma'}\sum_{h,d}w(\sigma,\sigma';h,d) \tag{S3}$$

with $w(\sigma,\sigma';h,d) \equiv \phi^*(\sigma,\sigma';h_1,d_1)\phi(\sigma,\sigma';h_2,d_2)$, the expectation value $\langle\mathcal{O}\rangle$ can be evaluated as

$$\langle\mathcal{O}\rangle = \frac{\langle\Psi(T)|\mathcal{O}\otimes\mathbb{1}'|\Psi(T)\rangle}{\langle\Psi(T)|\Psi(T)\rangle} = \frac{\sum_{\sigma,\sigma'}\sum_{h,d}w(\sigma,\sigma';h,d)O_{\text{loc}}(\sigma,\sigma';h,d)}{\sum_{\sigma,\sigma'}\sum_{h,d}w(\sigma,\sigma';h,d)}, \tag{S4}$$

where the sum over σ, σ', h, and d is numerically approximated by MC sampling with weight $w(\sigma,\sigma';h,d)$. Here, the "local" observable $O_{\text{loc}}(\sigma,\sigma';h,d)$ reads

$$O_{\text{loc}}(\sigma,\sigma';h,d) = \frac{1}{2}\sum_{\varsigma}\left(\langle\varsigma|\,\mathcal{O}\,|\sigma\rangle\,\frac{\phi^*(\varsigma,\sigma';h_1,d_1)}{\phi^*(\sigma,\sigma';h_1,d_1)} + \langle\sigma|\,\mathcal{O}\,|\varsigma\rangle\,\frac{\phi(\varsigma,\sigma';h_2,d_2)}{\phi(\sigma,\sigma';h_2,d_2)}\right), \tag{S5}$$

where the sum over $\varsigma = (\varsigma_1, ..., \varsigma_{N_{\text{site}}})$ is taken over all configurations. Note that non-diagonal elements $\langle\varsigma|\mathcal{O}|\sigma\rangle$ $(\varsigma \neq \sigma)$ are mostly zero when the size of the support of $\mathcal{O}$ is finite. For instance, if $\mathcal{O}$ is a product of Pauli operators with the total number of bit-flipping operators (i.e., σ_x or σ_y) given as k, the number of non-zero elements contributing in Eq. (S5) is 2^k.

Alternatively, we can use a marginal probability as the weight for the Monte Carlo (MC) method by tracing out one of h and d degrees of freedom. Here, as an example, we show the case where the h spins are traced out (we can also trace out the d spins and sample the h spins). In this case, the marginal probability is given by

$$\tilde{w}(\sigma,\sigma';d) \equiv \sum_{h_1,h_2}\phi^*(\sigma,\sigma';h_1,d_1)\phi(\sigma,\sigma';h_2,d_2) = \tilde{\phi}^*(\sigma,\sigma';d_1)\tilde{\phi}(\sigma,\sigma';d_2), \tag{S6}$$

with

$$\tilde{\phi}(\sigma,\sigma';d) = \prod_j 2\cosh\left[b_j + \sum_i(W_{ji}\sigma_i + W'_{ji}\sigma'_i) + \sum_k W'_{jk}d_k\right]. \tag{S7}$$

The formula for the expectation value $\langle \mathcal{O} \rangle$ is recast as

$$\langle \mathcal{O} \rangle = \frac{\sum_{\sigma,\sigma'} \sum_d \tilde{w}(\sigma,\sigma';d) \tilde{O}_{\rm loc}(\sigma,\sigma';d)}{\sum_{\sigma,\sigma'} \sum_d \tilde{w}(\sigma,\sigma';d)}, \tag{S8}$$

where the local observable $\tilde{O}_{\rm loc}(\sigma,\sigma';d)$ is given by

$$\tilde{O}_{\rm loc}(\sigma,\sigma';d) = \frac{1}{2} \sum_{\varsigma} \left(\langle \varsigma | \mathcal{O} | \sigma \rangle \frac{\tilde{\phi}^*(\varsigma,\sigma';d_1)}{\tilde{\phi}^*(\sigma,\sigma';d_1)} + \langle \sigma | \mathcal{O} | \varsigma \rangle \frac{\tilde{\phi}(\varsigma,\sigma';d_2)}{\tilde{\phi}(\sigma,\sigma';d_2)} \right). \tag{S9}$$

In the present study, we evaluated the physical observables based on Eq. (S8), since the degrees of freedom participating in the sampling are mitigated.

S2. PRACTICAL DETAILS OF METHOD (II)

A. Stochastic Reconfiguration method

The Stochastic Reconfiguration (SR) method [2, 3] employs a variational principle such that the exact imaginary-time evolution is approximated within the expressive power of the wave function ansatz. In other words, the SR method provides an update rule of variational parameters so that the distance, whose definition relies on the employed variational principle, between states following exact and approximate imaginary-time evolution is minimized. In the following, we provide a concise derivation of the parameter update rule used in the SR method.

Let $|\Psi_\theta\rangle$ be a variational wave function with a set of complex variational parameters θ. In the SR method, we update the variational parameters as $\theta \leftarrow \theta + \delta\widetilde{\theta}$ according to a variational principle based on the Fubini-Study metric $\mathcal{F}$:

$$\delta\widetilde{\theta} = \arg\min_{\delta\theta} \mathcal{F}[e^{-\delta_\tau \mathcal{H}} |\Psi_\theta\rangle, |\Psi_{\theta+\delta\theta}\rangle], \tag{S10}$$

$$\mathcal{F}[|\psi\rangle, |\phi\rangle] := \arccos \sqrt{\frac{\langle\psi|\phi\rangle \langle\phi|\psi\rangle}{\langle\psi|\psi\rangle \langle\phi|\phi\rangle}}. \tag{S11}$$

Here, $\mathcal{H}$ is the Hamiltonian of the system and δ_τ is a small step of imaginary-time evolution. Note that the metric $\mathcal{F}$ measures the distance between two quantum states $|\psi\rangle$ and $|\phi\rangle$ after imposing an appropriate normalization, and therefore it is closely related with the fidelity between pure states F as $\mathcal{F} = \arccos\sqrt{F^2}$, where the fidelity is defined as $F^2[|\psi\rangle, |\phi\rangle] = \frac{\langle\psi|\phi\rangle\langle\phi||\psi\rangle}{\langle\psi|\psi\rangle\langle\phi||\phi\rangle}$. Consequently, the parameter update $\delta\widetilde{\theta}$ also satisfies

$$\delta\widetilde{\theta} = \arg\max_{\delta\theta} F^2[e^{-\delta_\tau \mathcal{H}} |\Psi_\theta\rangle, |\Psi_{\theta+\delta\theta}\rangle]. \tag{S12}$$

After some algebraic calculations, we obtain the explicit expression, up to the second order, of the squared fidelity F^2as

$$F^2 = 1 - \left(\sum_{k,l} \delta\theta_k^* S_{kl} \delta\theta_l + \delta_\tau \sum_k (f_k \delta\theta_k^* + \text{c.c.}) + \delta_\tau^2 (\langle \mathcal{H}^2 \rangle - \langle \mathcal{H} \rangle^2) \right), \tag{S13}$$

where the variance of the energy, contributing to the intrinsic algorithm error, is estimated using the MC sampling as $\langle\cdot\rangle := \frac{\langle\Psi_\theta|\cdot|\Psi_\theta\rangle}{\langle\Psi_\theta||\Psi_\theta\rangle}$. [See S2 C for further information.] We have introduced an element of the covariance matrix S_{kl} as

$$S_{kl} = \frac{\langle \partial_k \Psi_\theta | \partial_l \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle} - \frac{\langle \partial_k \Psi_\theta | \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle} \frac{\langle \Psi_\theta | \partial_l \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle} \tag{S14}$$

$$= \langle O_k^\dagger O_l \rangle - \langle O_k^\dagger \rangle \langle O_l \rangle, \tag{S15}$$

and the generalized force f_k as

$$f_k = \frac{\langle \partial_k \Psi_\theta | \mathcal{H} | \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle} - \frac{\langle \partial_k \Psi_\theta | \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle} \frac{\langle \Psi_\theta | \mathcal{H} | \Psi_\theta \rangle}{\langle \Psi_\theta | \Psi_\theta \rangle} \tag{S16}$$

$$= \langle O_k^\dagger \mathcal{H} \rangle - \langle O_k^\dagger \rangle \langle \mathcal{H} \rangle, \tag{S17}$$

TABLE S1. Character table of the C_{4v} point group.

	E	$2C_4$	C_2	$2\sigma_v$	$2\sigma_d$
A_1	1	1	1	1	1
A_2	1	1	1	−1	−1
B_1	1	−1	1	1	−1
B_2	1	−1	1	−1	1
E	2	0	−2	0	0

where O_k denotes a diagonal matrix with each element given as $O_k(\sigma) = \frac{\partial_k \Psi_\theta(\sigma)}{\Psi_\theta(\sigma)}$. Equations. (S15) and (S17) indicates that all the elements S_{kl} and f_k can be estimated by MC sampling. Finally, by using the stationary condition for the optimal solution of Eq. (S13), the expression for the parameter update can be obtained as

$$\delta\widetilde{\theta}_k = -\delta_\tau \sum_l S^{-1}_{kl} f_l. \tag{S18}$$

In the actual calculation, we add a stabilization factor to the diagonal elements of S as $S_{kk} + \epsilon_k$, where ϵ_k is typically taken uniformly as $\sim 10^{-4}\bar{S}_{\rm diag}$, where $\bar{S}_{\rm diag}$ is the mean value of diagonal elements. We observe that the calculation becomes more reliable when the time step δ_τ is initially taken to be small; $O(10^{-4})$. After tens to hundreds of iterations, δ_τ is increased gradually up to $O(10^{-2})$.

B. Symmetrization

In Method (II), we numerically optimize a purified DBM wave function with the form:

$$\Psi(\sigma,\sigma') = \prod_j 2\cosh\Big[b_j + \sum_i (W_{ji}\sigma_i + W'_{ji}\sigma'_i)\Big]. \tag{S19}$$

Here, b, W, and W' are variational parameters. We assign the site index i for the pair of physical σ and ancilla σ' spins.

To improve the quality of the calculation, we consider the symmetry of the system. In the case of the 2D J_1-J_2 Heisenberg model, we encode the translational and point-group symmetry in the purified DBM wave function. We observe that the initial state, i.e., the purified infinite-temperature state, is invariant under certain symmetry operations: (1) translation $T_{\mathbf{R}}$ that shifts all the spins by the amount $\mathbf{R}$ as $(\sigma,\sigma') \mapsto (T_{\mathbf{R}}\sigma, T_{\mathbf{R}}\sigma')$, and (2) symmetry operation R of C_{4v} point group that maps a spin configuration as $(\sigma,\sigma') \mapsto (R\sigma, R\sigma')$. Alternatively, we can understand that the purified state is in the zero wave-number sector and belongs to the A_1 representation of the C_{4v} point group (see also Table S1) of a bilayer system. Along the imaginary time evolution, the purified DBM wave function stays within the identical symmetry sector at arbitrary temperature. Hence, we impose such a condition by symmetrizing the purified DBM wave function as

$$\Psi_{\rm sym.}(\sigma,\sigma') = \sum_{R,\mathbf{R}} \Psi(T_{\mathbf{R}}R\sigma, T_{\mathbf{R}}R\sigma'). \tag{S20}$$

We employ this formula to impose the symmetry on the wave function on the left-hand side $\Psi_{\rm sym.}(\sigma,\sigma')$. Observe that $\Psi_{\rm sym.}(\sigma,\sigma')$ satisfies the symmetry even when the bare DBM wave function $\Psi(\sigma,\sigma')$ on the right-hand side does not preserve the symmetry. We emphasize that the corresponding Gibbs state of the target system consists of contributions from all symmetry sectors of the original Hamiltonian.

C. Computing expectation values of physical observables

As in Method (I), expectation values are numerically evaluated by the MC method. The formula for the expectation value $\langle\mathcal{O}\rangle$ in Method (II) reads

$$\langle\mathcal{O}\rangle = \frac{\sum_{\sigma,\sigma'} p(\sigma,\sigma') O_{\rm loc}(\sigma,\sigma')}{\sum_{\sigma,\sigma'} p(\sigma,\sigma')}, \tag{S21}$$

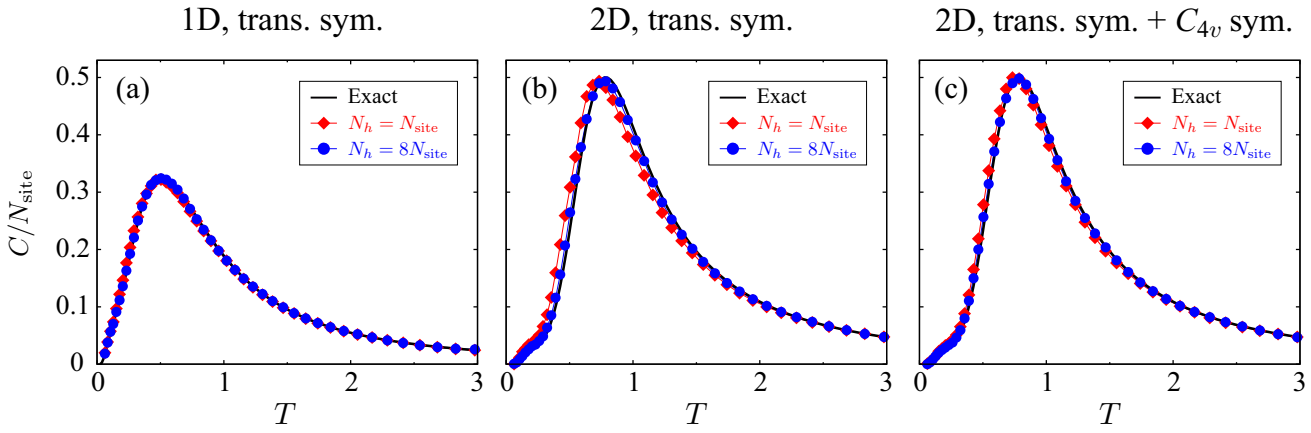

FIG. S2. Method (II) results (symbols) for the specific heat C for (a) the 1D Heisenberg model on the 16-site chain and (b,c) the 2D Heisenberg model on the 4x4 lattice. In all cases, we take the periodic boundary condition, and the number of sites is 16 ($N_{\rm site} = 16$). In (a) and (b), we impose translational symmetry on the purified DBM wave function for the extended system. In (c), we use both translational and C_{4v} point-group symmetry. The solid black curves are obtained by exact diagonalization.

where $p(\sigma, \sigma')$ is the weight $p(\sigma, \sigma') \equiv |\Psi_{\rm sym.}(\sigma, \sigma')|^2$, and the local observable $O_{\rm loc}(\sigma, \sigma')$ is given by

$$O_{\rm loc}(\sigma, \sigma') = \frac{1}{2} \sum_{\varsigma} \left(\langle \sigma | \mathcal{O} | \varsigma \rangle \frac{\Psi_{\rm sym.}(\varsigma, \sigma')}{\Psi_{\rm sym.}(\sigma, \sigma')} + c.c. \right). \tag{S22}$$

We perform the Metropolis sampling over the σ and σ' spins with the weight $p(\sigma, \sigma')$ to compute expectation values.

D. Calculation conditions

In the present calculations for the 2D J_1–J_2 Heisenberg model on the 6×6 lattice (Figure 3 in the paper), we introduce $8N_{\rm site}$ (=288) hidden spins. We set the b parameters (magnetic field) to zero, so that the "up" and "down" spins are equivalent, and optimize only the W and W' parameters. With this setting, the purified DBM wave function becomes even with respect to the global spin inversion.

The initial W and W' parameters are prepared to represent the infinite-temperature state. The infinite-temperature state can be reproduced exactly by setting $W_{ji} = W'_{ji} = \mathrm{i}\frac{\pi}{4}\delta_{ji}$, for $1 \le j \le N_{\rm site}$, and $W_{ji} = W'_{ji} = 0$, for $N_{\rm site} + 1 \le j \le 8N_{\rm site}$. In the actual calculations, to make the initial gradient of the parameter optimization finite, we put small perturbations to the above setting by adding small random numbers.

As we describe in the main text, starting from the initial W and W' values, we optimize the parameters with the SR method [3], which makes it possible to reproduce the imaginary time evolution as much as possible within the representability of the DBM. To reduce the number of variational parameters, we take half of the W and W' parameters to be complex and the rest real as in Ref. 4. We do not impose symmetry constraints on the W and W' parameters, but instead, the symmetry is restored with the projection [Eq. (S20) in the case of the 2D J_1–J_2 Heisenberg model]. We observe that the most time-consuming part of the calculation is the MC sampling to estimate the expectation values of the energy and the gradient of the parameter optimization; its time scales as $\mathcal{O}(N_h N_{\rm site}^2)$, where N_h is the number of hidden spins ($N_h = 8N_{\rm site}$ in the present case).

E. Benchmark calculations for 1D and 2D Heisenberg models

Here, we show the benchmark results of Method (II) for the 1D and 2D Heisenberg models. The Hamiltonian reads $\mathcal{H} = J_1 \sum_{\langle i,j \rangle} \mathbf{S}_i \cdot \mathbf{S}_j$, where $J_1 = 1$ and $\langle i, j \rangle$ denotes a pair of neighboring sites. In the 2D case, the Hamiltonian corresponds to the $J_2 = 0$ case for the 2D J_1–J_2 Heisenberg model. We take the 16-site chain for the 1D model, and the 4×4 square lattice for the 2D model. In both cases, periodic boundary conditions are assumed and the number of sites is 16 ($N_{\rm site} = 16$). We consider the zero magnetization sector ($\sum_i S_i^z = 0$) as in the main text.

As we have discussed above and in the main text, the quality of the Method (II) calculations can be improved by increasing the number of hidden units N_h and/or by imposing symmetry. For a numerical demonstration, we compare $N_h = N_{\rm site}$ and $N_h = 8N_{\rm site}$ cases. We follow the conditions described above (Sec. S2 D), except that all the parameters are taken to be complex in the $N_h = N_{\rm site}$ case (note that when $N_h = N_{\rm site}$, all the hidden units need to have complex couplings to prepare infinite-temperature states).

Figure S2 shows the Method (II) results for the specific heat C for (a) the 1D model when imposing translational symmetry, (b) the 2D model when imposing translational symmetry, and (c) the 2D model when imposing translational and point-group symmetries. We see that in the case of the 1D model, the Method (II) result agrees well with the exact result already at $N_h = N_{\rm site}$ [Figure S2(a)]. On the other hand, in the case of the 2D model, when only the translational symmetry is imposed as in the 1D model case, the result with $N_h = N_{\rm site}$ shows a visible (but small) deviation from the exact result [Figure S2(b)]. The result improves by increasing N_h [Figure S2(b)] or imposing an additional symmetry (point-group symmetry) [Figure S2(c)]. From this benchmark, we indeed observe that the number of hidden units N_h and the symmetrization are important factors to ensure the quality of the Method (II) calculations.

[1] G. Carleo, Y. Nomura, and M. Imada, Nature Communications **9**, 5322 (2018).
[2] S. Sorella, Phys. Rev. Lett. **80**, 4558 (1998).
[3] S. Sorella, Phys. Rev. B **64**, 024512 (2001).
[4] Y. Nomura, J. Phys.: Condens. Matter **33**, 174003 (2021).