Variational Gibbs state preparation on noisy intermediate-scale quantum devices

Mirko Consiglio [1,*] Jacopo Settino [2,3,4] Andrea Giordano [4] Carlo Mastroianni [4] Francesco Plastina [2,3]
Salvatore Lorenzo [5] Sabrina Maniscalco,[6,7] John Goold,[8] and Tony J. G. Apollaro [1]

[1]*Department of Physics, University of Malta, Msida MSD 2080, Malta*
[2]*Dipartimento di Fisica, Università della Calabria, 87036 Arcavacata di Rende (CS), Italy*
[3]*Gruppo Collegato di Cosenza, INFN, 87036 Arcavacata di Rende (CS), Italy*
[4]*ICAR-CNR, 87036 Rende (CS), Italy*
[5]*Dipartimento di Fisica e Chimica - Emilio Segrè, Università degli Studi di Palermo, via Archirafi 36, 90123 Palermo, Italy*
[6]*QTF Centre of Excellence, Department of Physics, University of Helsinki, P.O. Box 43, 00014 Helsinki, Finland*
[7]*Algorithmiq Ltd., Kanavakatu 3 C, 00160 Helsinki, Finland*
[8]*School of Physics, Trinity College Dublin, College Green, Dublin 2, Ireland*

The preparation of an equilibrium thermal state of a quantum many-body system on noisy intermediate-scale quantum (NISQ) devices is an important task in order to extend the range of applications of quantum computation. Faithful Gibbs state preparation would pave the way to investigate protocols such as thermalization and out-of-equilibrium thermodynamics and provide useful resources for quantum algorithms, where sampling from Gibbs states constitutes a key subroutine. We propose a variational quantum algorithm (VQA) to prepare Gibbs states of a quantum many-body system. The novelty of our VQA consists of implementing a parameterized quantum circuit acting on two distinct, yet connected (via CNOT gates), quantum registers. The VQA evaluates the Helmholtz free energy, where the von Neumann entropy is obtained via postprocessing of computational-basis measurements on one register, while the Gibbs state is prepared on the other register via a unitary rotation in the energy basis. Finally, we benchmark our VQA by preparing Gibbs states of the transverse-field Ising and Heisenberg *XXZ* models and achieve remarkably high fidelities across a broad range of temperatures in state-vector simulations. We also assess the performance of the VQA on IBM quantum computers, showcasing its feasibility on current NISQ devices.

I. INTRODUCTION

An integral task in quantum state preparation is the generation of finite-temperature thermal states of a given Hamiltonian on a quantum computer. Indeed, Gibbs states (also known as thermal states) can be used for quantum simulation [1], quantum machine learning [2,3], quantum optimization [4], and the study of open quantum systems [5]. In particular, combinatorial optimization problems [4], semidefinite programming [6], and training of quantum Boltzmann machines [2] can be tackled by sampling from well-prepared Gibbs states.

The preparation of an arbitrary initial state is a challenging task in general, with finding the ground-state of a Hamiltonian being a quantum Merlin–Arthur (QMA)-hard problem [7]. Preparing Gibbs states, specifically at low temperatures, could be as hard as finding the ground state of that Hamiltonian [8]. The first algorithms for preparing Gibbs states were based on the idea of coupling the system to a register of ancillary qubits and letting the system and environment evolve under a joint Hamiltonian, simulating the physical process of thermalization, such as in Refs. [5,9,10], while others relied on dimension reduction [11].

The algorithm proposed in this paper for preparing Gibbs states can be placed in the category of VQAs, such as in Refs. [12–16], and can similarly be used for preparing thermofield double (TFD) states [17–19]. Variational ansätze based on multiscale entanglement renormalization [20] and the product-spectrum ansatz [21] have also been proposed to prepare Gibbs states.

Alternative algorithms prepare thermal states through quantum imaginary-time evolution, such as in Refs. [22–26], starting from a maximally mixed state, while others start from a maximally entangled state [27]. Reference [28] proposed quantum-assisted simulation to prepare thermal states, which does not require a hybrid quantum-classical feedback loop. In addition, methods exist to sample Gibbs state expectation values, rather than prepare the Gibbs state directly, such as in quantum metropolis methods [29,30], imaginary-time evolution applied to pure states [31], and random quantum circuits using intermediate measurements [32].

Recent methods also proposed using rounding promises [33], fluctuation theorems [34], pure thermal shadow tomography [35], and minimally entangled typical thermal states for finite-temperature simulations [36].

The goal of this work is to propose a VQA that efficiently prepares Gibbs states on NISQ computers, employing the free energy as a (physically motivated) objective function. This requires the evaluation of the von Neumann entropy

---

*Contact author: mirko.consiglio@um.edu.mt

[37], which is generally hard to obtain from a quantum register. In contrast to some of the currently employed VQAs [10,12–15,17,19,32], which employ truncated equations to approximate it, we directly estimate the von Neumann entropy without any truncation and with an error solely dependent on the number of shots, using sufficiently expressible ansätze capable of preparing the Boltzmann distribution. Our VQA, in fact, is composed of two ansätze: a heuristic, shallow one that prepares the Boltzmann distribution for a given temperature and another one, possibly designed with a problem-inspired approach, which depends on the Hamiltonian while being independent of the temperature.

This paper is organized as follows: in Sec. II, we present the VQA for preparing Gibbs states; in Sec. III, we apply the algorithm to the Ising model using both state-vector and noisy simulations, as well as running the algorithm on IBM quantum hardware. In Sec. IV, we apply the algorithm to the Heisenberg model, using both state-vector and shot-based simulations. Finally, in Sec. V, we draw our conclusions and discuss the future prospects of this work.

II. VARIATIONAL GIBBS STATE PREPARATION

Consider a Hamiltonian $\mathcal{H}$, describing n interacting qubits; then the Gibbs state at inverse temperature $\beta \equiv 1/(k_B T)$, where k_B is the Boltzmann constant and T is the temperature, is defined as

$$\rho(\beta, \mathcal{H}) = \frac{e^{-\beta \mathcal{H}}}{\mathcal{Z}(\beta, \mathcal{H})}, \qquad (1)$$

where the partition function $\mathcal{Z}(\beta, \mathcal{H})$ is

$$\mathcal{Z}(\beta, \mathcal{H}) = \text{Tr}\{e^{-\beta \mathcal{H}}\} = \sum_{i=0}^{d-1} e^{-\beta E_i}. \qquad (2)$$

Here $d = 2^n$, while $\{E_i\}$ are the eigenenergies of $\mathcal{H}$, with $\{|E_i\rangle\}$ denoting the corresponding eigenstates, i.e., $\mathcal{H}|E_i\rangle = E_i|E_i\rangle$.

Fixing a Hamiltonian $\mathcal{H}$ and inverse temperature β, for a general state ρ, one can define a generalized Helmholtz free energy as

$$\mathcal{F}(\rho) = \text{Tr}\{\mathcal{H}\rho\} - \beta^{-1}\mathcal{S}(\rho), \qquad (3)$$

where the von Neumann entropy $\mathcal{S}(\rho)$ can be expressed in terms of the eigenvalues p_i of ρ,

$$\mathcal{S}(\rho) = -\sum_{i=0}^{d-1} p_i \ln p_i. \qquad (4)$$

Since the Gibbs state is the unique state that minimizes the free energy of $\mathcal{H}$ [38], a variational procedure can be put forward that takes Eq. (3) as an objective function, such that

$$\rho(\beta, \mathcal{H}) = \underset{\rho}{\arg\min} \, \mathcal{F}(\rho). \qquad (5)$$

In this case, $p_i = \exp(-\beta E_i)/\mathcal{Z}(\beta, \mathcal{H})$ is the probability of getting the eigenstate $|E_i\rangle$ from the ensemble $\rho(\beta, \mathcal{H})$.
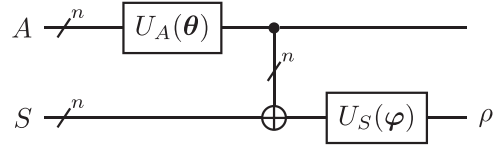

FIG. 1. PQC for Gibbs state preparation, with systems A and S each carrying n qubits. CNOT gates act between each qubit A_i and corresponding S_i.

A. Framework of the algorithm

The difficulty in measuring the von Neumann entropy [defined by Eq. (4)] of a quantum state on a NISQ device is typically the challenging part of variational Gibbs-state-preparation algorithms, as $\mathcal{S}(\rho)$ is not an observable. With this in mind, we present a VQA that avoids the direct measurement of the von Neumann entropy on a quantum computer by using a carefully constructed parameterized quantum circuit (PQC).

When preparing an n-qubit state starting from the input state $|0\rangle^{\otimes n}$, given that a quantum computer operates using only unitary gates, the final quantum state of the entire register will be pure. As a result, in order to prepare an n-qubit Gibbs state on the system register, we require an $m \leqslant n$ qubit ancillary register. For example, in the case of the infinite-temperature Gibbs state, which is the maximally mixed state, we require $m = n$ qubits in the ancillary register to achieve maximal von Neumann entropy. In order to evaluate the von Neumann entropy, without any truncation, we need to be able to prepare the entire Boltzmann distribution on the ancillary register; hence, we set $m = n$, irrespective of the temperature.

We denote the ancillary register as A, while the preparation of the Gibbs state is carried out on the system register S. The purpose of the VQA is to effectively create the Boltzmann distribution on A, which is then imposed on S via intermediary CNOT gates, to generate a diagonal mixed state. In the ancillary register we can choose a unitary ansatz capable of preparing such a probability distribution. Thus, the ancillary qubits are responsible for mixing in the probabilities of the thermal state while also being able to access these probabilities via measurements in the computational basis. On the other hand, the system register will host the preparation of the Gibbs state as well as the measurement of the expectation value of our desired Hamiltonian.

The specific design of the PQC instead allows classical postprocessing of simple measurement results carried out on ancillary qubits in the computational basis to determine the von Neumann entropy. A diagrammatic representation of the structure of the PQC is shown in Fig. 1. Note that while the PQC of the algorithm has to have a particular structure—a unitary acting on the ancillae and a unitary acting on the system, connected by intermediary CNOT gates—it is not dependent on the choice of Hamiltonian $\mathcal{H}$; inverse temperature β; or the variational ansätze, U_A and U_S, employed within.

B. Modular structure of the PQC

The PQC, as shown in Fig. 1 for the VQA, is composed of a unitary gate U_A acting on the ancillary qubits and a unitary gate U_S acting on the system qubits, with CNOT gates in between. Note that the circuit notation we are using here

means that there are n qubits for both the system and the ancillae, as well as n CNOT gates that act in parallel and are denoted as

$$\text{CNOT}_{AS} \equiv \bigotimes_{i=0}^{n-1} \text{CNOT}_{A_i S_i}. \qquad (6)$$

The parameterized unitary U_A acting on the ancillae, followed by CNOT gates between the ancillary and system qubits, is responsible for preparing a probability distribution on the system. The parameterized unitary U_S is then applied on the system qubits to transform the computational-basis states into the eigenstates of the Hamiltonian.

A general unitary gate of dimension $d = 2^n$ is given by

$$U_A = \begin{pmatrix} u_{0,0} & u_{0,1} & \cdots & u_{0,d-1} \\ u_{1,0} & u_{1,1} & \cdots & u_{1,d-1} \\ \vdots & \vdots & \ddots & \vdots \\ u_{d-1,0} & u_{d-1,1} & \cdots & u_{d-1,d-1} \end{pmatrix}. \qquad (7)$$

Starting with the initial state of the $2n$-qubit register $|0\rangle_{AS}^{\otimes 2n}$, we apply the unitary gate U_A on the ancillae to get a quantum state $|\psi\rangle_A$, such that

$$(U_A \otimes \mathbb{1}_S)|0\rangle_{AS}^{\otimes 2n} = |\psi\rangle_A \otimes |0\rangle_S^{\otimes n}, \qquad (8)$$

where

$$|\psi\rangle_A = \sum_{i=0}^{d-1} u_{i,0}|i\rangle_A \qquad (9)$$

and $\mathbb{1}_S$ is the identity operator acting on the system. The next step is to prepare a probability mixture on the system qubits, which can be done by applying CNOT gates between each ancilla and system qubit. This results in a state

$$\text{CNOT}_{AS}\left(|\psi\rangle_A \otimes |0\rangle_S^{\otimes n}\right) = \sum_{i=0}^{d-1} u_{i,0}|i\rangle_A \otimes |i\rangle_S. \qquad (10)$$

By then tracing out the ancillary qubits, we arrive at

$$\text{Tr}_A \left\{ \left(\sum_{i=0}^{d-1} u_{i,0}|i\rangle_A \otimes |i\rangle_S \right)\left(\sum_{j=0}^{d-1} u_{j,0}^*\langle j|_A \otimes \langle j|_S \right) \right\}$$

$$= \sum_{i,j=0}^{d-1} u_{i,0}u_{j,0}^*\langle i|j\rangle|i\rangle\langle j|_S = \sum_{i=0}^{d-1} |u_{i,0}|^2|i\rangle\langle i|_S, \qquad (11)$$

ending up with a diagonal mixed state on the system, with probabilities given directly by the absolute square of the entries of the first column of U_A, that is, $p_i = |u_{i,0}|^2$. If the system qubits were traced out instead, we would end up with the same diagonal mixed state, but on the ancillary qubit register:

$$\text{Tr}_S \left\{ \left(\sum_{i=0}^{d-1} u_{i,0}|i\rangle_A \otimes |i\rangle_S \right)\left(\sum_{j=0}^{d-1} u_{j,0}^*\langle j|_A \otimes \langle j|_S \right) \right\}$$

$$= \sum_{i,j=0}^{d-1} u_{i,0}u_{j,0}^*\langle i|j\rangle|i\rangle\langle j|_A = \sum_{i=0}^{d-1} |u_{i,0}|^2|i\rangle\langle i|_A. \qquad (12)$$

This implies that by measuring in the computational basis of the ancillary qubits, we can determine the probabilities p_i, which can then be postprocessed to determine the von Neumann entropy $\mathcal{S}$ of the state ρ via Eq. (4) (since the entropy of A is the same as that of S). As a result, since U_A serves only to create a probability distribution from the entries of the first column, we can do away with a parameterized orthogonal (real unitary) operator, thus requiring fewer gates and parameters for the ancillary ansatz.

The unitary gate U_S then serves to transform the computational-basis states of the system qubits to the eigenstates of the Gibbs state, such that

$$\rho = U_S \left(\sum_{i=0}^{d-1} |u_{i,0}|^2|i\rangle\langle i|_S\right) U_S^\dagger$$

$$= \sum_{i=0}^{d-1} p_i|\psi_i\rangle\langle\psi_i|, \qquad (13)$$

where the expectation value $\text{Tr}\{\mathcal{H}\rho\}$ of the Hamiltonian can be measured. Ideally, at the end of the optimization procedure, $p_i = \exp(-\beta E_i)/\mathcal{Z}(\beta, \mathcal{H})$ and $|\psi_i\rangle = |E_i\rangle$, so that we get

$$\rho(\beta, \mathcal{H}) = \sum_{i=0}^{d-1} \frac{e^{-\beta E_i}}{\mathcal{Z}(\beta, \mathcal{H})}|E_i\rangle\langle E_i|. \qquad (14)$$

The VQA therefore avoids the entire difficulty of measuring the von Neumann entropy of a mixed state on a quantum computer and instead transfers the task of postprocessing computational-basis-measurement results to the classical computer, which is much more tractable.

C. Objective function

Finally, we can define the objective function of our VQA to minimize the free energy (3) via our constructed PQC to obtain the Gibbs state

$$\rho(\beta, \mathcal{H}) = \underset{\theta, \varphi}{\arg\min} \, \mathcal{F}(\rho(\theta, \varphi))$$

$$= \underset{\theta, \varphi}{\arg\min}[\text{Tr}\{\mathcal{H}\rho_S(\theta, \varphi)\} - \beta^{-1}\mathcal{S}(\rho_A(\theta))]. \qquad (15)$$

It is worth mentioning that while the energy expectation depends on both sets of angles θ (as U_A is responsible for parametrizing the Boltzmann distribution) and φ (as U_S is responsible for parametrizing the eigenstates of the Gibbs state), the calculation of the von Neumann entropy depends on only θ.

Furthermore, once we obtain the optimal parameters θ^* and φ^*, preparing the Gibbs state $\rho(\beta, \mathcal{H})$ on the system qubits S, we can place the same unitary U_S with optimal parameters φ^* on the ancillary qubits to prepare the TFD state on the entire qubit register, as shown in Fig. 2. A TFD state [12,17,19] is defined as

$$|\text{TFD}(\beta)\rangle = \sum_{i=0}^{d-1} \sqrt{\frac{e^{-\beta E_i}}{\mathcal{Z}(\beta, \mathcal{H})}}|i\rangle_A \otimes |i\rangle_S, \qquad (16)$$

and tracing out either the ancilla or system register yields the same Gibbs state on the other register. Equation (16) is equivalent to Eq. (10) after applying U_S on both registers.
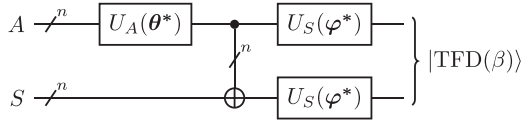
FIG. 2. Optimal PQC for TFD state preparation, with systems A and S each carrying n qubits. CNOT gates act between each qubit A_i and corresponding S_i.

D. Alternative implementations of the algorithm

There are several adjustments that could be applied to the PQC to modify the procedure. One specific example is replacing the intermediary CNOT gates with midcircuit measurements and implementing classically controlled NOT gates since no subsequent unitary gates act on the control qubits, as shown in Fig. 3. This method has a few benefits:

(1) Since the ancillary system needs to be measured to compute the von Neumann entropy, utilizing midcircuit measurements followed by classically controlling the system qubits is a natural approach to the algorithm.

(2) The two registers A and S can be made fully distinct in terms of the device topology, as well as reducing the depth of the entire circuit, leading to less overall decoherence affecting the protocol.

(3) Once optimization is carried out, the classically controlled NOT gates can still be kept in the circuit, yet if the experimentalist ignores the measurement information (equivalent to tracing out), then there is no operational difference between preparing an ensemble of pure states and preparing a mixed state using quantum CNOT gates.

The only downside occurs when the ancillary qubits are intended to be used again, such as when preparing the TFD state. In this case, the optimization for finding optimal parameters to prepare the Gibbs state can still be carried out using classically controlled NOT gates. However, at the end of the optimization procedure, the classically controlled NOT gates can be replaced with CNOT gates, followed by the optimized system unitary with the same structure as in Fig. 2, to obtain the TFD state.

The VQA can be further adapted so that U_A is replaced by a classical procedure that generates the probability distribution and prepares pure states of the Gibbs state ensemble $\{p_i, |E_i\rangle\}$, where $p_i = \exp(-\beta E_i)/\mathcal{Z}$ on the system qubits S. This procedure can be carried out by parametrizing a classical probability distribution $p(\boldsymbol{\theta})$ by $O(\text{poly}(n))$ parameters $\boldsymbol{\theta}$, using methods such as Markov chains composed of a sequence of local stochastic matrices, among others. The probability distribution will output bit strings $|i\rangle$ that can be fed as a computational input state to the unitary U_S that prepares the
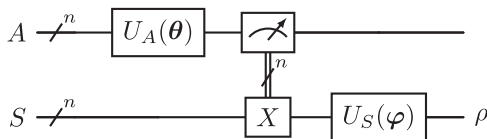

FIG. 3. PQC for Gibbs state preparation using midcircuit measurements, with systems A and S each carrying n qubits. Classically controlled NOT gates act between each qubit A_i and corresponding S_i.

eigenstates $|E_i\rangle$ of the Hamiltonian. By reducing the number of qubits and eliminating the requirement for intermediary CNOT gates, this process may result in a less expressible probability distribution function because entanglement is not used as a resource.

Furthermore, if the parametrization of the probability distribution corresponds to the output distribution of a known unitary circuit of a sufficiently shallow depth and expressibility, then the optimization can be carried using only the classical subroutine of sampling bit strings from the probability distribution $p(\boldsymbol{\theta})$ and feeding them to U_S to compute the free energy. Once the VQA is trained, U_A can be introduced with the optimized parameters $\boldsymbol{\theta^*}$ to prepare the mixed Gibbs state on the quantum computer. Nevertheless, finding such a parametrization that corresponds to a shallow, but expressible enough, unitary is a nontrivial task.

III. PERFORMANCE OF THE VQA IN THE ISING MODEL

In this section we assess the performance of the VQA for Gibbs state preparation of an Ising model. The Ising model is defined as

$$\mathcal{H} = -\sum_{i=1}^{n} \sigma_i^x \sigma_{i+1}^x - h \sum_{i=1}^{n} \sigma_i^z. \qquad (17)$$

The Ising Hamiltonian is a widely investigated model [39], and here we report only one relevant property for implementing a problem-inspired ansatz for U_S. The Hamiltonian in Eq. (17) commutes with the parity operator $\mathcal{P} = \bigotimes_{i=0}^{n-1} \sigma_i^z$. As a consequence, the eigenstates of $\mathcal{H}$ have definite parity, and so will the eigenstates of ρ_β.

We use the Uhlmann-Josza fidelity [40], defined as $F(\rho, \sigma) = (\text{Tr}\{\sqrt{\sqrt{\rho}\sigma\sqrt{\rho}}\})^2$, as a figure of merit for the performance of our VQA since it describes how "close" the prepared state is to the Gibbs state and it is also the most commonly employed measure of distinguishability. However, other measures can be used which have different interpretations. One example is the trace distance [41], which has the property that, if its value between the two states is bounded by ϵ, expectation values computed on the effectively prepared state differ from those taken on the Gibbs state by an amount that is, at most, proportional to ϵ [34]. Another choice is the relative entropy [41], which describes the distinguishability between the two states as the surprise that occurs when an event that is not possible with the true Gibbs state happens [42].

We use a simple, linearly entangled PQC for the unitary U_A, with parameterized $R_y(\theta_i)$ gates and CNOT gates as the entangling gates. This ansatz is hardware efficient and is sufficient to produce real amplitudes for preparing the probability distribution. Note that we require the use of entangling gates because, otherwise, we will not be able to prepare any arbitrary probability distribution, including the Boltzmann distribution of the Ising model. A proof of this is given in Appendix B.

For the unitary U_S, we choose a parity-preserving PQC. We employ a brick-wall structure, with the gates being $R_{xy}(\varphi_i) \equiv \exp[-\iota\varphi_i(\sigma^x \otimes \sigma^y)/2]$ followed by $R_{yx}(\varphi_j) \equiv \exp[-\iota\varphi_j(\sigma^y \otimes \sigma^x)/2]$ gates.
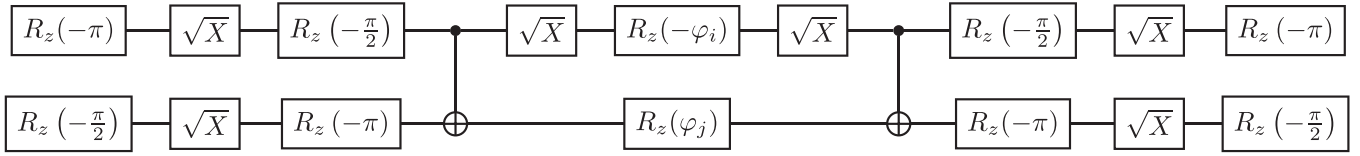
FIG. 4. Decomposed R_p gate in Eq. (18).

If we combine the two gates, which we denote as $R_p(\varphi_i, \varphi_j)$, we get

$$R_p(\varphi_i, \varphi_j) = R_{yx}(\varphi_j)R_{xy}(\varphi_i) = \begin{pmatrix} \cos\left(\frac{\varphi_i+\varphi_j}{2}\right) & 0 & 0 & \sin\left(\frac{\varphi_i+\varphi_j}{2}\right) \\ 0 & \cos\left(\frac{\varphi_i-\varphi_j}{2}\right) & -\sin\left(\frac{\varphi_i-\varphi_j}{2}\right) & 0 \\ 0 & \sin\left(\frac{\varphi_i-\varphi_j}{2}\right) & \cos\left(\frac{\varphi_i-\varphi_j}{2}\right) & 0 \\ -\sin\left(\frac{\varphi_i+\varphi_j}{2}\right) & 0 & 0 & \cos\left(\frac{\varphi_i+\varphi_j}{2}\right) \end{pmatrix}, \tag{18}$$

which can be decomposed into two CNOT gates, six $\sqrt{X}$ gates, and 10 R_z gates. One layer of the unitary acting on the system qubits consists of a brick-wall structure composed of an even-odd sublayer of R_p gates, followed by an odd-even sublayer of R_p gates. The decomposed unitary is shown in Fig. 4. An example of a PQC split into a four-qubit ancillary register and a four-qubit system register is shown in Fig. 5. Table I shows the scaling of the VQA assuming a closed ladder connectivity for $n > 2$.

A. State-vector results

Figure 6 shows the fidelity of the generated mixed state when compared with the exact Gibbs state of the Ising model

with $h = 0.5, 1.0, 1.5$ across a range of temperatures for system size between two and six qubits. The VQA was carried out using state-vector simulations with the Broyden-Fletcher-Goldfarb-Shanno (BFGS) optimizer [43]. We used one layer for the ancilla ansatz and $n - 1$ layers for the system ansatz with the scaling highlighted in Table II. The number of layers was heuristically chosen to satisfy, at most, a polynomial scaling in quantum resources while achieving a fidelity higher than 98% in state-vector simulations. Furthermore, in order to alleviate the issue of getting stuck in local minima, the optimizer is embedded in a Monte Carlo framework, that is, taking multiple random initial positions and carrying out a local optimization from each position, which we call a "run," and, finally, taking the global minimum to be the minimum over all runs.
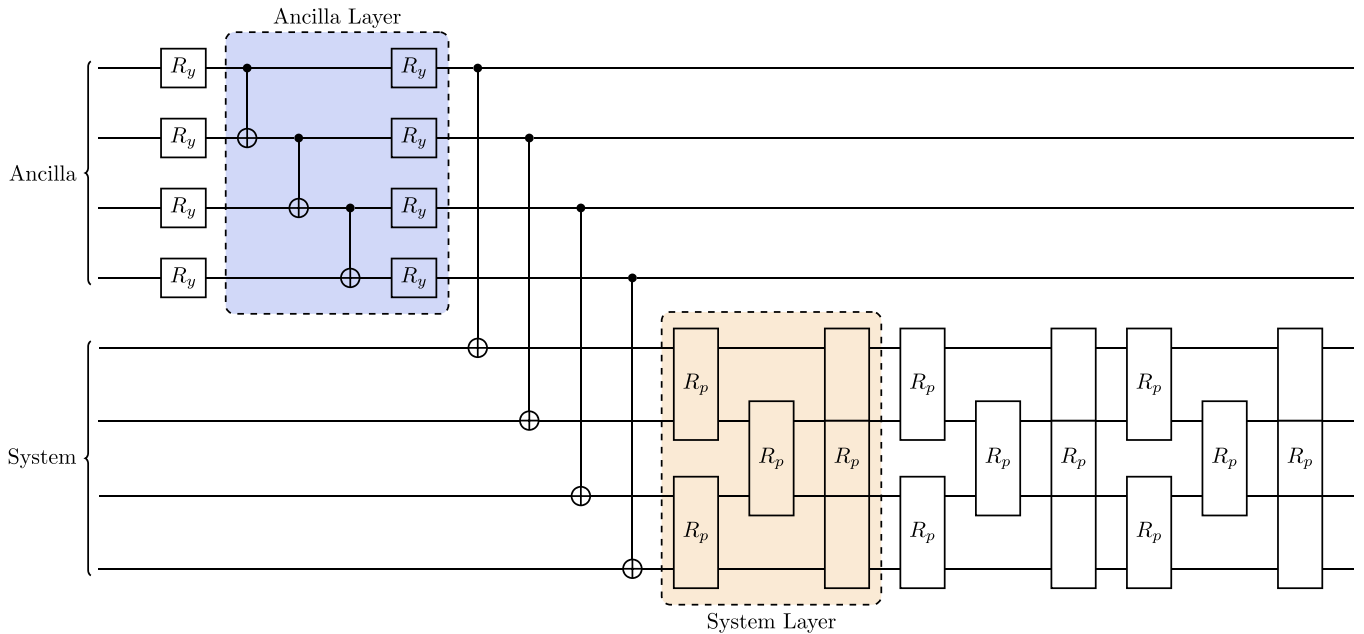

FIG. 5. Example of an eight-qubit PQC, consisting of one ancilla layer acting on a four-qubit register and three $(n-1)$ system layers acting on another four-qubit register. Each R_y gate is parameterized with one parameter, θ_i, while each R_p gate has two parameters, φ_i and φ_j. The R_p gate is defined in Eq. (18). Note that the intermediary CNOT gates, as well as the R_p gates acting on qubits 2 and 3 and on qubits 1 and 4 of the system, can be carried out in parallel.
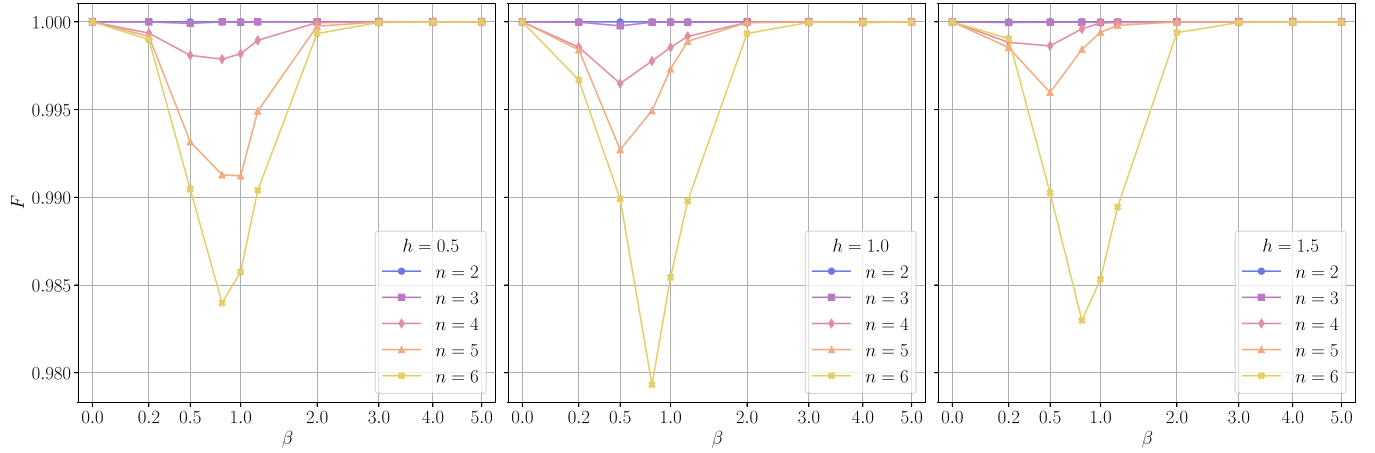
FIG. 6. Fidelity F of the obtained state via state-vector simulations (using BFGS) with the exact Gibbs state vs inverse temperature β for two to six qubits of the Ising model with $h = 0.5, 1.0, 1.5$. A total of 100 runs is made for each point, with the optimal state taken to be the one that maximizes the fidelity.

A total of 100 runs of BFGS per β was carried out to verify the reachability of the PQC, with Fig. 6 showcasing the maximal fidelity achieved for each β out of all runs. The results show that, indeed, our VQA is able to reach a very high fidelity $F \gtrsim 0.98$ for up to six-qubit Gibbs states of the Ising model. In the case of the extremal points, that is, $\beta \to 0$ and $\beta \to \infty$, the fidelity reaches unity for all investigated system sizes. However, for intermediary temperatures $\beta \sim 1$, the fidelity decreases with the number of qubits, which can be attributed to one layer of U_A not being expressible enough to prepare the Boltzmann distribution around intermediary temperatures [since the von Neumann entropy depends solely on $U_A(\boldsymbol{\theta})$ as in Eq. (15)]. Moreover, at intermediary temperatures, most of the Boltzmann probabilities p_i are still relatively small, resulting in a larger error in obtaining the correct eigenstate. On the other hand, at high temperatures, all probabilities are equally likely, and preparing the maximally mixed state is a straightforward task. However, at low temperatures, the VQA effectively reduces to a variational quantum eigensolver (VQE), i.e., finding the ground state of the Hamiltonian.

B. Noisy simulation results

The next step was to carry out noisy simulations of the VQA. We took the noise model of IBM Quantum (IBMQ) Guadalupe [44] for the Ising model with $h = 0.5$, which similarly has one layer for the ancilla ansatz and $n - 1$ layers for the system ansatz. However, it must be noted that in this case,

the scaling of the algorithm does not follow Table II because IBMQ Guadalupe does not have a closed ladder connectivity. As a result, transpilation was carried out by the QISKIT transpiler using the SWAP-based BiDiREctional heuristic search (SABRE) algorithm [45]. Apart from this, because the BFGS optimizer is incapable of optimizing a noisy objective function, an optimizer that accommodates noisy measurements was chosen: simultaneous perturbation stochastic approximation (SPSA) [46].

Using SPSA, 10 runs were carried out for each β, while the number of iterations was taken to be $100n$ for each run, with only two measurements at each iteration to estimate the gradient in a random direction, i.e., $200n$. As a consequence, a total of $2000n$ function evaluations was used to obtain the fidelity for each β shown in Fig. 7 (with an extra 50 function evaluations at each run to calibrate the hyperparameters of SPSA). Similar to the number of layers, the choice of the number of iterations was heuristically chosen so that, at most, the scaling is linear while still retaining a fidelity greater than 95% for the two- and three-qubit noisy simulation cases.

To measure the energy expectation value of the Ising model $\text{Tr}\{\mathcal{H}\rho\}$, we need to split the Ising Hamiltonian into its constituent Pauli strings, whose number scales linearly with the number of qubits as $2n$. However, we can group the $\sigma^x \sigma^x$ terms, as well as the σ^z terms, and measure them simultaneously, reducing the number of measurement circuits to two. Each circuit was also measured with 1024 shots. The M3 package [47] was also utilized to perform basic error

TABLE I. Scaling of the VQA assuming a closed ladder connectivity for $n > 2$ of the Ising model, where l_A and l_S are the number of ancilla ansatz and system ansatz layers, respectively, and P is 12 when n is even and 18 when n is odd.

	Total	Order
No. of parameters	$n(l_A + 1) + 2nl_S$	$O(n(l_A + l_S))$
No. of CNOT gates	$(n - 1)l_A + 2nl_S + n$	$O(n(l_A + l_S))$
No. of $\sqrt{X}$ gates	$2n(l_A + 1) + 6nl_S$	$O(n(l_A + l_S))$
Circuit depth	$(n + 1)l_A + Pl_S + 3$	$O(nl_A + l_S)$

TABLE II. Scaling of the VQA assuming a closed ladder connectivity for $n > 2$, with $l_A = 1$ and $l_S = n - 1$; P is 12 when n is even and 18 when n is odd. The depth counts both CNOT and $\sqrt{X}$ gates.

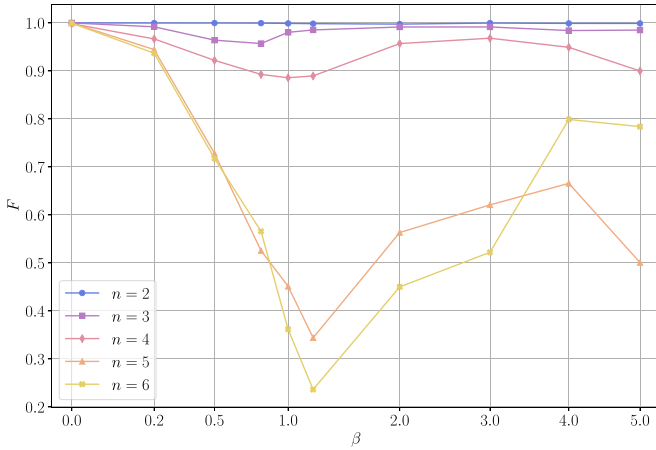|	Total	Order
No. of parameters	$2n^2$	$O(n^2)$
No. of CNOT gates	$2n^2 - 1$	$O(n^2)$
No. of $\sqrt{X}$ gates	$2n(3n - 2)$	$O(n^2)$
Circuit depth	$(P + 1)n - P + 4$	$O(n)$

FIG. 7. Fidelity F of the obtained state via noisy simulations (using SPSA) of IBMQ Guadalupe with the exact Gibbs state vs inverse temperature β for two to six qubits of the Ising model with $h = 0.5$. A total of 10 runs is made for each point, with the optimal state taken to be the one that maximizes the fidelity.
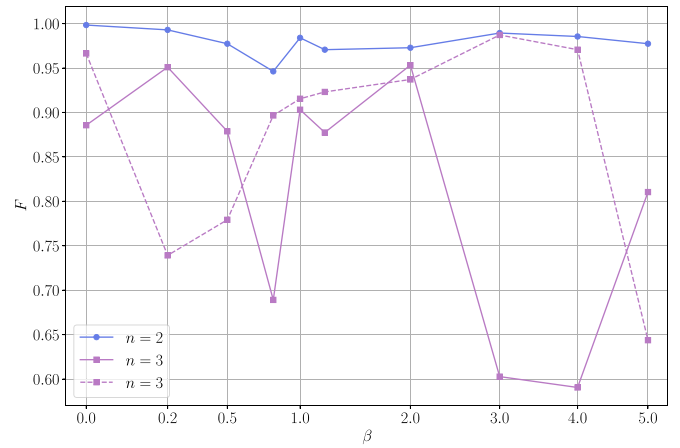

FIG. 8. Fidelity F of the obtained state (using SPSA) running directly on IMBQ Nairobi with the exact Gibbs state vs inverse temperature β for two and three qubits of the Ising model with $h = 0.5$. The dashed line represents the run with no R_p gate between nonadjacent qubits in the system layers. One run is carried out for $n = 2$ and for $n = 3$ denoted by the dashed line, and two runs are carried out for $n = 3$ denoted by the solid line.

mitigation. It operates by using a matrix-free preconditioned iterative-solution method to mitigate measurement error that does not form the full assignment matrix or its inverse. A summary of the optimization scaling is shown in Table III.

From Fig. 7, one can see that the fidelity is significantly high in the case of $n = 2, 3, 4$. However, in the case of $n = 5, 6$, the Gibbs state is faithfully prepared only for low β. This can be attributed to the level of noise present in the optimization procedure, with the transpiled circuits going well beyond the quantum volume of IBMQ Guadalupe. There are also some points which obtain a worse fidelity for five qubits than for six qubits, which could be due to the larger depth acquired by an odd number of qubits in the system ansatz, as highlighted in Table II.

The performance of VQAs is heavily impacted by the presence of noise-induced barren plateaus [48]. While the analysis of barren plateaus for Gibbs state preparation is beyond the scope of this paper, which aims at providing an alternative approach for variationally preparing Gibbs states and avoiding any estimations of the entropy using Taylor expansions or other truncations, we still carry out brief analyses as starting points for future works. In particular, we discuss the implications of barren plateaus in Appendix C, and we also carry out analysis of the error from estimating the entropy in Appendix A.

C. IBM quantum device results

Finally, the VQA was carried out on an actual quantum device. Figure 8 displays the fidelity of Gibbs states obtained using IBM quantum hardware [44], specifically IMBQ Nairobi. Like for the noisy simulations, SPSA was used, but this time with only one run for each β in the case $n = 2$ and two runs in the case $n = 3$, with $100n$ iterations and 1024 shots. The Gibbs states were obtained by taking the optimal parameters from the optimization carried out on IMBQ Nairobi and determining the state vector on a classical computer.

The solid lines in Fig. 8 represent the two- and three-qubit results. At all points, the two-qubit Gibbs state shows excellent fidelity. On the other hand, the three-qubit Gibbs state is remarkably reproduced at certain temperatures, while it is lacking at other points. Since IMBQ Nairobi does not have a closed ladder connectivity, several SWAP gates are necessary for carrying out transpilation. In an attempt to reduce the number of SWAP gates, we carried out another run at each β, where we removed the R_p gate acting on nonadjacent qubits in the system layers, with the result shown by the dashed line in Fig. 8 (note that this also resulted in fewer parameters and less depth of the PQC). A considerable improvement in fidelity is achieved at the points where fidelity was lacking in the previous case. Since the available run time on the quantum device was limited, the number of runs is still too low to determine the reason why the Gibbs state was not achieved with higher fidelity. Nevertheless, comparing the results in Fig. 8 with the state-vector results in Fig. 6 and with the noise-simulated results in Fig. 7, we conclude that limited connectivity, combined with device noise, is severely hampering the effectiveness of the VQA.

In addition, quantum state tomography for the two-qubit case was carried out on IMBQ Nairobi with 1024 shots for the cases of $\beta = 0, 1, 5$, where the fidelities obtained were 0.992, 0.979, and 0.907, respectively. A three-dimensional
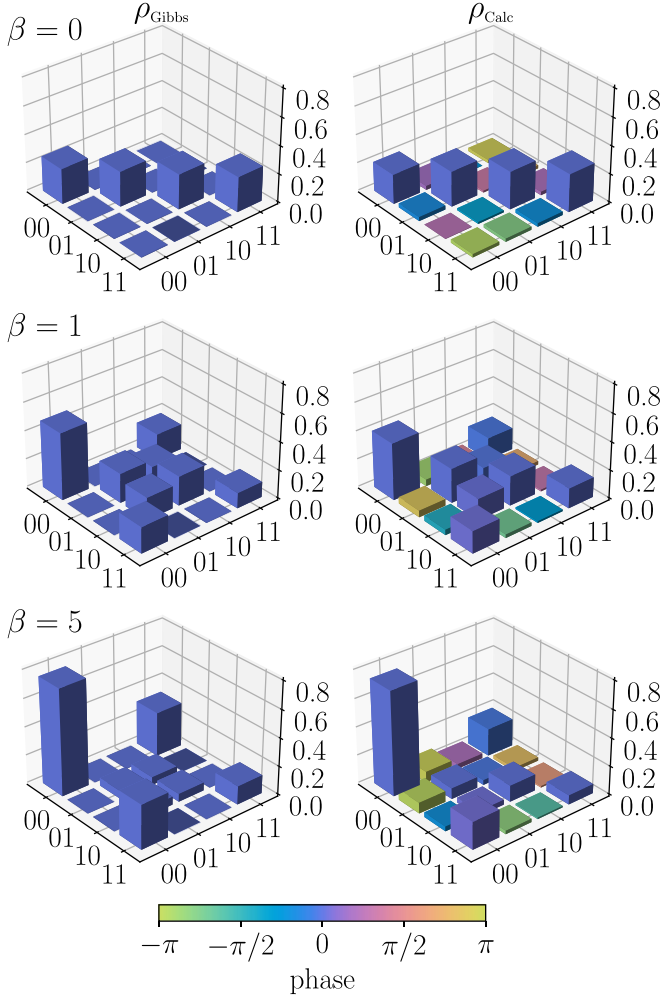
TABLE III. Scaling of SPSA for noisy simulations on quantum hardware.

	Total	Order
No. of iterations for each run	$100n$	$O(n)$
No. of function evaluations for each run	$200n$	$O(n)$
No. of circuits per function evaluation	2	$O(1)$
No. of circuit evaluations for each run	$400n$	$O(n)$
No. of shots for each circuit evaluation	1024	$O(1)$

FIG. 9. Three-dimensional bar plot of the two-qubit results from IMBQ Nairobi for $\beta = 0, 1, 5$ for the Ising model with $h = 0.5$. The analytical Gibbs states are shown in the left column, while the tomographically obtained Gibbs states are shown in the right column.

(3D) bar plot of the tomographic results is given in the right column Fig. 9 and can be compared with the analytical form of the Gibbs state in the left column. The largest discrepancies can be seen in the off-diagonal terms, which increase as β increases, showcasing symptoms of amplitude damping. This can be attributed to the thermal relaxation and dephasing noise present in the quantum devices, leading to an overall decoherence in the Gibbs state.

While noisy simulations were run using the calibration data of IBMQ Guadalupe—since it has access to up to 16 qubits—the actual hardware used for the two- and three-qubit Gibbs state preparation was IMBQ Nairobi due to its accessibility.

IV. PERFORMANCE OF THE VQA IN THE HEISENBERG *XXZ* MODEL

In this section, to further explore the feasibility of our VQA for few-body thermal state preparation on NISQ devices, we assess its performance on a more complex, interacting system: the Heisenberg model. The Heisenberg *XXZ* model is

TABLE IV. Scaling of the VQA assuming a closed ladder connectivity for $n > 2$ of the Heisenberg model, where l_A and l_S are the number of ancilla ansatz and system ansatz layers, respectively, and P is 2 when n is even and 3 when n is odd.

	Total	Order
No. of parameters	$n(l_A + 1) + 2nl_S$	$O(n(l_A + l_S))$
No. of CNOT gates	$nl_A + 2nl_S + n$	$O(n(l_A + l_S))$
No. of $\sqrt{X}$ gates	$2n(l_A + 1) + 6nl_S$	$O(n(l_A + l_S))$
Circuit depth	$Pl_A + 2Pl_S + 1$	$O(l_A + l_S)$

defined as

$$\mathcal{H} = -\frac{1}{4} \sum_{i=1}^{n} \left(\sigma_i^x \sigma_{i+1}^x + \sigma_i^y \sigma_{i+1}^y + \Delta \sigma_i^z \sigma_{i+1}^z \right) - h \sum_{i=1}^{n} \sigma_i^z. \tag{19}$$

At variance with the Ising model investigated in Sec. III, the so-called *XXZ* model in a transverse magnetic field is an interacting Hamiltonian once mapped into spinless fermions. The phase diagram is much more complex and exhibits a paramagnetic-to-ferromagnetic phase transition at $h = \frac{1}{2}(1 - \Delta)$ [39].

The Heisenberg model also commutes with the parity operator. As such, U_S is the same as in Sec. III. On the other hand, we use an alternating layered ansatz for U_A, with parameterized $R_y(\theta_i)$ gates and CNOT gates as the entangling gates. Once again, this ansatz is hardware efficient and is sufficient to produce real amplitudes for preparing the probability distribution. We utilize the fidelity as our figure of merit for quantifying the performance of the algorithm. In this case, the scaling of U_A and U_S is given in Table IV. Moreover, an example of a PQC, split into a four-qubit ancilla register and a four-qubit system register, is shown in Fig. 10.

A. State-vector results

Figure 11 shows the fidelity of the generated mixed state compared with the exact Gibbs state of the Heisenberg model with $h = 0.5$ and $\Delta = -0.5, 0.0, 0.5$ across a range of temperatures for system size between two and six qubits. The VQA was carried out using state-vector simulations with the BFGS optimizer [43]. We used $n - 1$ layers for the ancilla ansatz and $n - 1$ layers for the system ansatz, with the scaling highlighted in Table V. The number of layers was heuristically chosen to satisfy, at most, a polynomial scaling in quantum

TABLE V. Scaling of the VQA assuming a closed ladder connectivity, for $n > 2$ for the Heisenberg model, with $l_A = n - 1$, and $l_S = n - 1$, and P is 2 when n is even and 3 when n is odd. The depth counts only CNOT gates.

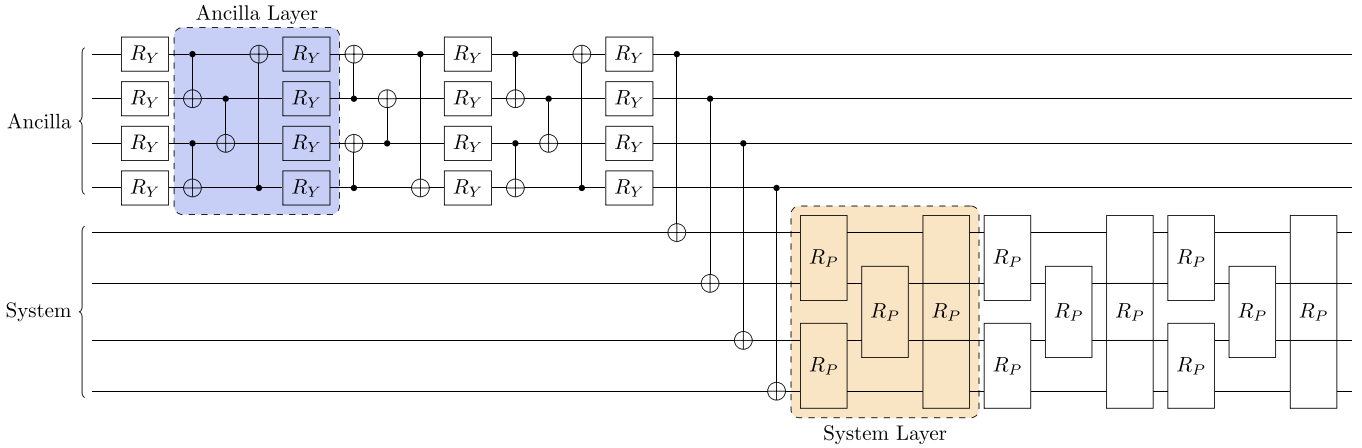|	Total	Order
No. of parameters	$3n^2 - 2n$	$O(n^2)$
No. of CNOT gates	$3n^2 - 2n$	$O(n^2)$
No. of $\sqrt{X}$ gates	$8n^2 - 6n$	$O(n^2)$
Circuit depth	$3Pn - 3P + 1$	$O(n)$

FIG. 10. Example of an eight-qubit PQC, consisting of three $(n-1)$ ancilla layers acting on a four-qubit register and three $(n-1)$ system layers acting on another four-qubit register. Each R_y gate is parameterized with one parameter, θ_i, while each R_p gate has two parameters, φ_i and φ_j. The R_p gate is defined in Eq. (18). Note that the intermediary CNOT gates, as well as the R_p gates acting on qubits 2 and 3 and on qubits 1 and 4 of the system, can be carried out in parallel.

resources while achieving a fidelity higher than 98% in state-vector simulations. Furthermore, in order to alleviate the issue of getting stuck in local minima, the optimizer was embedded in a Monte Carlo framework.

Like for the state-vector results for the Ising model in Sec. III A, we obtained very high fidelities $F > 0.98$ for the number of qubits ranging from two to six across a broad range of temperatures of the Heisenberg *XXZ* model. It must be noted that the paramagnetic-to-ferromagnetic transition point lies at $\Delta = 0$ for $h = 0.5$, resulting in the noninteracting *XX* model and achieving a much better performance. The same dip in fidelity at intermediary temperatures reappears at around $\beta \sim 1$ for all the plots in Fig. 11.

B. Shot-based results

The next step was to carry out shot-based simulations for the Heisenberg model with $h = 0.5$ and $\Delta = -0.5, 0.0, 0.5$, as shown in Fig. 12. Using SPSA, 10 runs were carried out

for each β, while the number of iterations was taken to be $100n$ for each run, with $2n$ function evaluations at each iteration to estimate the gradient in n random directions, i.e., $200n^2$. Similar to the number of layers, the choice of the number of function evaluations was heuristically chosen so that the scaling was polynomial. The number of commuting sets of Pauli strings of the Heisenberg model is three. Furthermore, each circuit was also measured with 1024 shots, and the M3 package [47] was also utilized to perform error mitigation. Table VI shows a summary of the optimization scaling.

Naturally, the choice of optimizer, along with the finite number of measurements used to reconstruct both the von Neumann entropy and the expectation value, are shown to affect the performance of the VQA. Nevertheless, while the results in Fig. 12 exhibit a lower, albeit relatively high, fidelity $F \gtrsim 0.93$, the VQA shows notable promise in being able to produce Gibbs states of complex interacting Hamiltonians, such as the Heisenberg *XXZ* model.
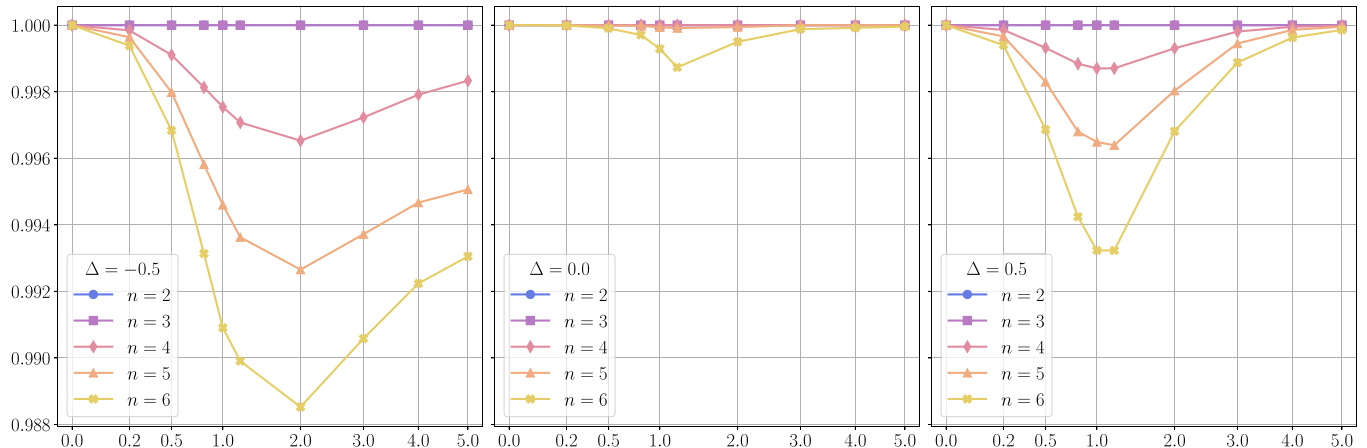

FIG. 11. Fidelity F of the obtained state via state-vector simulations (using BFGS) with the exact Gibbs state vs inverse temperature β for two to six qubits of the Heisenberg model with $h = 0.5$ and $\Delta = -0.5, 0.0, 0.5$. A total of 100 runs is made for each point, with the optimal state taken to be the one that maximizes the fidelity.
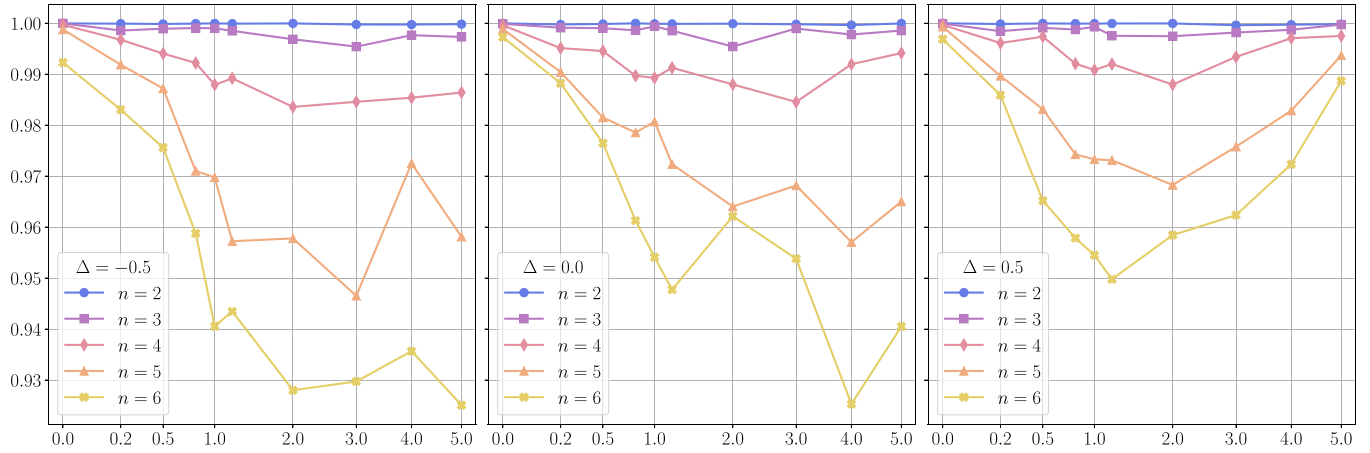
FIG. 12. Fidelity F of the obtained state via shot-based simulations (using SPSA) with the exact Gibbs state vs inverse temperature β for two to six qubits of the Heisenberg model with $h = 0.5$ and $\Delta = -0.5, 0.0, 0.5$. A total of 100 runs is made for each point, with the optimal state taken to be the one that maximizes the fidelity.

V. CONCLUSION

We addressed the preparation of a thermal equilibrium state of a quantum many-body system on a NISQ device. We exploited the uniqueness of the Gibbs state as the state that minimizes the Helmholtz free energy, thus providing a faithful objective function for a VQA.

The novelty of the proposed VQA consisted of splitting the PQC into two parameterized unitaries, one acting on an ancillary register and one acting on a system register. The former is tasked with determining the Boltzmann weights of the Gibbs distribution, corresponding to a given temperature, while the latter performs the rotation from the computational basis to the energy basis of a given Hamiltonian.

We benchmarked our VQA preparing the Gibbs state of the transverse-field Ising model and obtained fidelities $F \simeq 1$ for system sizes of up to six qubits in state-vector simulations across a broad range of temperatures, with a slight dip at intermediate ones. Moreover, we tested our VQA on the Heisenberg model with a transverse field, similarly obtaining fidelities $F \simeq 1$ in state-vector simulations. However, performance on current NISQ devices, investigated by both noisy simulations and real-hardware execution on IBMQ devices, showed a degradation in the results of the VQA with increasing system size. This may have been caused by the limited connectivity and the noise present in the device. Nevertheless, executing our VQA on NISQ devices still provides an improvement upon the recent developments in variational Gibbs state preparation (see e.g., Ref. [19]).

TABLE VI. Scaling of SPSA for shot-based simulations.

	Total	Order
No. of iterations for each run	$100n$	$O(n)$
No. of function evaluations for each run	$200n^2$	$O(n^2)$
No. of circuits per function evaluation	3	$O(1)$
No. of circuit evaluations for each run	$600n^2$	$O(n^2)$
No. of shots for each circuit evaluation	1024	$O(1)$

It is important to note that the structure of our VQA does not depend on the specific Hamiltonian to be tackled or on any prior knowledge of its spectrum. For example, the structure of the unitary $U_S(\boldsymbol{\varphi})$ could be adjusted in order to match some specific features of the eigenstates (if they are known), or the parameterized unitary $U_A(\boldsymbol{\theta})$ could be replaced by a deterministic procedure (e.g., the one reported in [49]) if the probabilities of the Boltzmann distribution are known.

However, even without requiring any such knowledge, our "Hamiltonian-agnostic" variational approach gives an effective way to prepare Gibbs states of arbitrary quantum many-body systems on a quantum computer, providing an advancement over previous methods, especially thanks to the modular structure of our PQC. This could significantly contribute to both performing quantum thermodynamical experiments on a quantum computer and faithfully preparing Gibbs states to be used in a great variety of computational tasks. Furthermore, preparing moderately sized many-body systems with our VQA may also be sufficient for exploring finite-size effects of certain physical models [50].

Some final remarks are in order: many modular elements of the VQA have the capacity to be significantly improved. While the scope of this paper was to provide a proof-of-concept VQA for preparing Gibbs states by estimating the entropy directly without any truncation, we will mention potential avenues for future research. In particular, more robust error mitigation techniques could be implemented, such as those present in the MITIQ library [51]. In addition to this, we delve into the consequences of working with a limited number of samples when attempting to estimate the entropy in Appendix A. Moreover, we explore an alternative entropy estimation technique that exhibits promise in its potential to scale beyond the NISQ era of computing. Barren plateaus in deep PQCs and noisy devices are also a considerable challenge to address. In Appendix C, we qualitatively discuss the requirements needed to investigate barren plateaus for our algorithm, as well as the PQCs that stem from its particular structure. Last, it is worth noting that the choice of optimizer plays a pivotal role in the performance of the VQA, particularly in the presence of noise. An in-depth analysis of various

optimizers could significantly enhance the reliability of the VQA.

The PYTHON code for running the state-vector simulations, using QULACS [52], and the noisy simulations, as well as the run-time program, using QISKIT [53], can be found on GitHub [54].

APPENDIX A: ERROR ANALYSIS OF ENTROPY ESTIMATION

In general, reconstructing the probability distribution faithfully requires an exponential number of shots, and particularly, free-fermion distributions can be hard to learn [55]. However, let us look at estimating the entropy using the maximum-likelihood (ML) estimator, rather than focusing on the reconstruction of the probability distribution. The ML estimator was shown to have a bias and a variance that, in general, decreases as $O(N^{-1})$ for $N \gg M$ [56]. The outcome of one shot of a quantum circuit can be described by a multinomial distribution $\vec{p}$, where p_i is the probability of observing a bit string i. Given N shots filling M bins, the ML estimator [56,57] of the von Neumann entropy is given by

$$\mathcal{S}_{\mathrm{ML}} = -\sum_{i=1}^{M} q_i \ln q_i, \tag{A1}$$

were $q_i = n_i/N$, such that n_i represents the number of times the bit string i appears after N shots. The variance of the entropy can be easily computed as

$$\mathbb{V}(\mathcal{S}_{\mathrm{ML}}) = \sum_{i}^{M} (1 + \ln q_i)^2 \mathbb{V}(q_i). \tag{A2}$$

where $\mathbb{V}(q_i) \approx q_i(1 - q_i)/N$. In fact, it can be shown that for all N and all possible distributions, the variance of the ML estimator for entropy is bounded above as

$$\mathbb{V}(\mathcal{S}_{\mathrm{ML}}) \leqslant \frac{[\ln(N)]^2}{N}, \tag{A3}$$

as proven in Ref. [58], and that

$$P(|\mathcal{S}_{\mathrm{ML}} - \mathbb{E}(\mathcal{S}_{\mathrm{ML}})| > \epsilon) \leqslant 2e^{-\frac{N\epsilon^2}{2[\ln(N)]^2}}. \tag{A4}$$

It is important to note that this bound is not particularly tight, and it is independent of M and the probability distribution. Moreover, the ML estimator was proven to be negatively biased everywhere, and

$$\mathbb{E}_{\vec{p}}(\mathcal{S}_{\mathrm{ML}}) \leqslant \mathcal{S}(\vec{p}), \tag{A5}$$

where $\mathbb{E}_{\vec{p}}$ denotes the conditional expectation given $\vec{p}$; that equality is achieved only when $\mathcal{S}(\vec{p}) = 0$, meaning the distribution is supported on a single point (or in the case of Boltzmann distributions for $\beta \to \infty$). In the case of $N \gg M$, the Miller-Madow bias correction [56,57] gives

$$\mathcal{S}(\vec{p}) = \mathcal{S}_{\mathrm{ML}}(\vec{p}) + \frac{M - 1}{2N} + O(N^{-1}). \tag{A6}$$

In the case of NISQ and future quantum algorithms, given that the Hilbert space of qubits grows as $M = 2^n$, we can reasonably assume that $N \ll M$ as soon as $n > 20$. As a result, we need to look towards entropy estimation techniques when we are in a heavily undersampled regime. Bayesian inference is typically employed in these situations. While learning a probability distribution might generally require an exponential number of samples [55], computing functionals of such distributions might not. As a result, the Nemenman-Shafee-Bialek (NSB) estimator employs Bayesian inference to obtain both the entropy and its *a posteriori* standard deviation. We utilize the PYTHON package NDD [59] to compute $\mathcal{S}_{\mathrm{NSB}}$; interested readers are referred to Refs. [60,61] for the details.

Figure 13 shows the results of using the ML and NSB estimators using a finite number of shots $N = 1024$ for the Ising model with $h = 0.5, 1.0, 1.5$. In particular we show how the additive error (bias) of the entropy scales as the number of qubits increases. Each point in Fig. 13 is obtained by averaging the entropy over 100 samples, with the error bars representing the standard deviation.

As one can expect, for a number of qubits n such that $2^n \ll N$, both ML and NSB estimators obtain a bias and standard deviation close to zero for all values of β. There is a transition region where $2^n \sim N$, where the bias, particularly for low values of β, starts to increase. In the region where $2^n \gg N$, the ML estimator is valid in only the large-β regime since there is usually a finite number of nonzero probabilities which is much smaller than N. It is important to note that although the error increases linearly with the number of qubits, the number of shots N needed to reduce the bias to
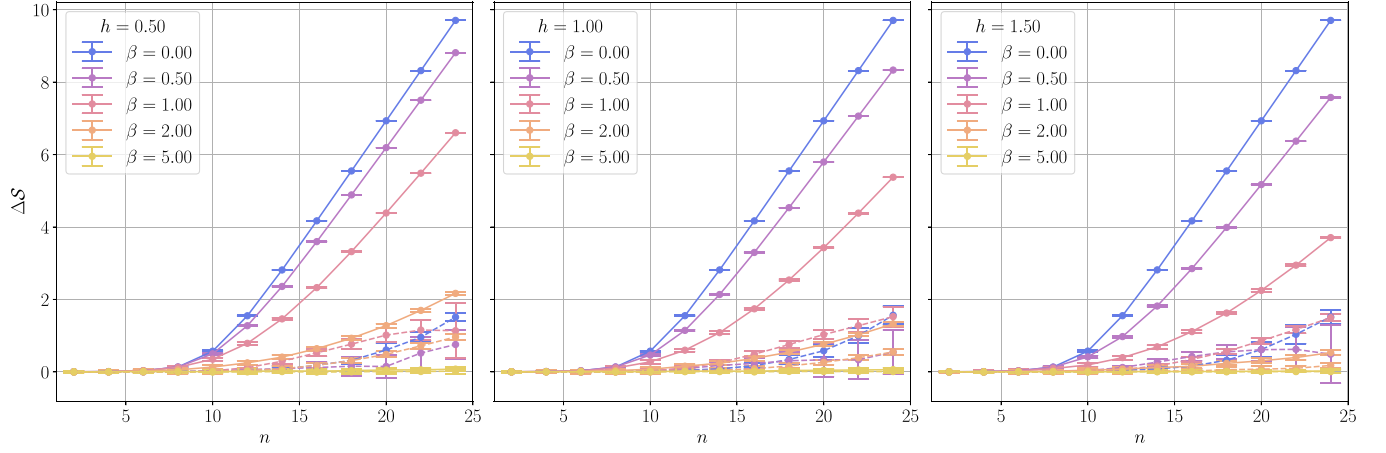
FIG. 13. Average additive error $\Delta\mathcal{S}$ (bias) in the entropy estimation as a function of the number of qubits n, using the ML (solid lines) and NSB (dashed lines) estimators, where the error bars represent the standard deviation. The number of shots is $N = 1024$, and the number of samples for each point is 100, with $h = 0.5, 1.0, 1.5$ for the Ising model.

a constant error increases exponentially as a function of the number of qubits n. Specifically, in the region $2^n \gg N$, the ML estimator reaches the upper bound of $\ln(N)$, and so for $\beta = 0$, $\Delta\mathcal{S}_{\mathrm{ML}} = \ln(2^n) - \ln(N) = \ln(2^n/N)$.

On the other hand, the NSB estimator obtains a much lower bias at the cost of a slightly higher standard deviation. While the behavior of the NSB estimator is hard to surmise given that exact diagonalization results allowed us to see until $n = 24$, there are a few instances where for intermediate and high β the estimator seems to flatten or even decrease as n increases. Proving that, for a particular β, the NSB estimator reliably acquires a bias that scales as $O(\mathrm{poly}[\ln(n)])$ would mean that using a number of shots N that scales polynomially would achieve a constant additive error, implying scalability in entropy estimation.

APPENDIX B: NECESSITY OF ENTANGLING GATES IN U_A

In the main text, we specified that we required entanglement in the ancillary register to be able to prepare the Boltzmann distribution of the Ising model. While we used only one layer of a hardware-efficient ansatz, we concluded that least one entangling layer is necessary for preparing the Boltzmann distribution of the Ising model, and we will show this by considering the converse. Suppose the ancilla ansatz is composed of only local R_y gates; then we get

$$
\begin{aligned}
\bigotimes_{i=0}^{n-1} R_y(\theta_i)|0\rangle_i &= \bigotimes_{i=0}^{n-1} \left[\cos\left(\frac{\theta_i}{2}\right)|0\rangle_i + \sin\left(\frac{\theta_i}{2}\right)|1\rangle_i \right] \\
&= \sum_{i=0}^{d-1} \prod_{j\in S_{i=0}} \cos\left(\frac{\theta_j}{2}\right) \prod_{k\in S_{i=1}} \sin\left(\frac{\theta_k}{2}\right)|i\rangle \\
&= \sum_{i=0}^{d-1} p_i|i\rangle,
\end{aligned}
\tag{B1}
$$

where $S_{i=0} \equiv \{j \,|\, j = 0 \,\forall\, \text{bits } j \in i\}$, that is, the set of bits in i which are equal to zero, with similar notation for $S_{i=1}$, and $|i\rangle$

is the computational-basis state. This implies that

$$
\prod_{j\in S_{i=0}} \cos\left(\frac{\theta_j}{2}\right) \prod_{k\in S_{i=1}} \sin\left(\frac{\theta_k}{2}\right) = p_i.
\tag{B2}
$$

Now, without loss of generality, consider these specific cases:

$$
\prod_{j=0}^{n-1} \cos\left(\frac{\theta_j}{2}\right) = p_0,
\tag{B3a}
$$

$$
\prod_{j=0}^{n-2} \cos\left(\frac{\theta_j}{2}\right) \sin\left(\frac{\theta_{n-1}}{2}\right) = p_1,
\tag{B3b}
$$

$$
\prod_{j=0}^{n-3} \cos\left(\frac{\theta_j}{2}\right) \cos\left(\frac{\theta_{n-1}}{2}\right) \sin\left(\frac{\theta_{n-2}}{2}\right) = p_2,
\tag{B3c}
$$

$$
\prod_{j=0}^{n-3} \cos\left(\frac{\theta_j}{2}\right) \sin\left(\frac{\theta_{n-2}}{2}\right) \sin\left(\frac{\theta_{n-1}}{2}\right) = p_3.
\tag{B3d}
$$

Combining the above equations results in

$$
\frac{p_0}{p_1} = \frac{\cos\left(\frac{\theta_{n-1}}{2}\right)}{\sin\left(\frac{\theta_{n-1}}{2}\right)},
\tag{B4a}
$$

$$
\frac{p_0}{p_2} = \frac{\cos\left(\frac{\theta_{n-2}}{2}\right)}{\sin\left(\frac{\theta_{n-2}}{2}\right)},
\tag{B4b}
$$

$$
\frac{p_0}{p_3} = \frac{\cos\left(\frac{\theta_{n-2}}{2}\right)\cos\left(\frac{\theta_{n-1}}{2}\right)}{\sin\left(\frac{\theta_{n-2}}{2}\right)\sin\left(\frac{\theta_{n-1}}{2}\right)},
\tag{B4c}
$$

and finally, combining the above equations implies that

$$
\frac{p_0}{p_1}\frac{p_0}{p_2} = \frac{p_0}{p_3} \Rightarrow \frac{p_0}{p_1 p_2} = \frac{1}{p_3}
$$

$$
\Rightarrow p_0 p_3 = p_1 p_2
$$

$$
\Rightarrow e^{-\beta E_0} e^{-\beta E_3} = e^{-\beta E_1} e^{-\beta E_2}.
\tag{B5}
$$

Applying logs to both sides and simplifying, we get

$$
E_0 + E_3 = E_1 + E_2,
\tag{B6}
$$

which is not, in general, true for the Ising model. The above reasoning can be adjusted to obtain further constraints in the manner of Eq. (B6).

APPENDIX C: BARREN-PLATEAU ANALYSIS

By following the analysis carried out in Ref. [62], we qualitatively discuss the trainability of our VQA. We can decompose our cost function as

$$\mathcal{F}(\rho_S) = \text{Tr}\{\mathcal{H}\rho_S\} - \beta^{-1}\mathcal{S}(\rho_A)$$

$$= \text{Tr}\{\mathcal{H}U_S\rho'_S U_S^\dagger\} + \beta^{-1}\sum_{i=0}^{d-1}\text{Tr}\{O_i U_A|0\rangle\langle 0|_A^{\otimes n}U_A^\dagger\}$$

$$\times \ln\text{Tr}\{O_i U_A|0\rangle\langle 0|_A^{\otimes n}U_A^\dagger\}, \qquad (C1)$$

where $\rho'_S = \text{Tr}_A\{V|0\rangle\langle 0|_{AS}^{\otimes 2n}V^\dagger\}$, $V = \text{CNOT}_{AS}(U_A \otimes \mathbb{1}_S)$m and $O_i = |i\rangle\langle i|$. Reference [62] considered cost functions only of the form

$$C = \text{Tr}\{OU\rho U^\dagger\}, \qquad (C2)$$

where ρ is an arbitrary quantum state of n qubits, O is any operator, and U is an alternating layered ansatz. Given that our cost function in Eq. (C1) is not in the form of Eq. (C2) because of the logarithm in the von Neumann entropy, our comparison should be taken solely as a qualitative discussion of the possibility of barren plateaus.

Now with reference to Eq. (C1), we have that $\mathcal{H}$ is 2-local, while O_i is 1-local, and both U_A and U_S are alternating layered ansätze. Theorem 2 of Ref. [62] gives a lower bound on the variance of the gradient of the cost function as a function of the number of layers and, as such, the trainability of the PQC. If the number of layers $l = O(\ln(n))$, then the variance vanishes no faster than polynomially, hence making the PQC trainable. If the number of layers $l = O(\text{poly}[\ln(n)])$, then the variance vanishes faster than polynomially, but no faster than exponentially, settling in a transition region between trainable and untrainable.

In the case of $\beta \to \infty$, our cost function equates directly to Eq. (C2) (since the VQA effectively reduces to a VQE), and thus, we require $l = O(\ln(n))$ for our circuit to be trainable. On the other hand, in the case of $\beta \to 0$, the cost function simplifies to maximizing the von Neumann entropy, that is, acquiring the maximally mixed state. While we cannot directly relate the von Neumann entropy as a cost function with Eq. (C2), we have numerically seen that preparing the mixed state is a relatively straightforward task. Nevertheless, analysis of the trainability of utilizing the von Neumann entropy as the (or part of the) cost function should be sought to be able to detect the presence of barren plateaus.

Now, for any finite $\beta > 0$, the problem of determining whether a barren plateau is possible for the generalized free energy is out of the scope of this work. Nevertheless, we can possibly surmise that, given U_A and U_S are alternating layered ansätze, with $\mathcal{H}$ being a 2-local Hamiltonian consisting of traceless operators, using a number of layers for both U_A and U_S that scales at most as $l = O(\ln(n))$ might result in a PQC that is trainable. This would hold if Theorem 2 of Ref. [62] also holds for cost functions in the form of Eq. (C1).

[1] A. M. Childs, D. Maslov, Y. Nam, N. J. Ross, and Y. Su, Proc. Natl. Acad. Sci. USA **115**, 9456 (2018).

[2] M. Kieferová and N. Wiebe, Phys. Rev. A **96**, 062327 (2017).

[3] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Nature (London) **549**, 195 (2017).

[4] R. D. Somma, S. Boixo, H. Barnum, and E. Knill, Phys. Rev. Lett. **101**, 130504 (2008).

[5] D. Poulin and P. Wocjan, Phys. Rev. Lett. **103**, 220502 (2009).

[6] F. G. S. L. Brandao and K. M. Svore, in *Proceedings of the 2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS), Los Alamitos, CA* (IEEE Computer Society, Piscataway, NJ, 2017), pp. 415–426.

[7] J. Watrous, Quantum computational complexity, in *Encyclopedia of Complexity and Systems Science*, edited by R. A. Meyers (Springer, New York, NY, 2009), pp. 7174–7201.

[8] D. Aharonov, I. Arad, and T. Vidick, SIGACT News **44**, 47 (2013).

[9] B. M. Terhal and D. P. DiVincenzo, Phys. Rev. A **61**, 022301 (2000).

[10] A. Riera, C. Gogolin, and J. Eisert, Phys. Rev. Lett. **108**, 080402 (2012).

[11] E. Bilgin and S. Boixo, Phys. Rev. Lett. **105**, 170405 (2010).

[12] J. Wu and T. H. Hsieh, Phys. Rev. Lett. **123**, 220502 (2019).

[13] A. N. Chowdhury, G. H. Low, and N. Wiebe, arXiv:2002.00055.

[14] Y. Wang, G. Li, and X. Wang, Phys. Rev. Appl. **16**, 054035 (2021).

[15] A. Warren, L. Zhu, N. J. Mayhall, E. Barnes, and S. E. Economou, arXiv:2203.12757.

[16] J. Foldager, A. Pesah, and L. K. Hansen, Sci. Rep. **12**, 3862 (2022).

[17] D. Zhu, S. Johri, N. M. Linke, K. A. Landsman, C. H. Alderete, N. H. Nguyen, A. Y. Matsuura, T. H. Hsieh, and C. Monroe, Proc. Natl. Acad. Sci. USA **117**, 25402 (2020).

[18] S. P. Premaratne and A. Y. Matsuura, in *2020 IEEE International Conference on Quantum Computing and Engineering (QCE)* (IEEE, Piscataway, NJ, 2020), pp. 278–285.

[19] R. Sagastizabal, S. P. Premaratne, B. A. Klaver, M. A. Rol, V. Negîrneac, M. S. Moreira, X. Zou, S. Johri, N. Muthusubramanian, M. Beekman, C. Zachariadis, V. P. Ostroukh, N. Haider, A. Bruno, A. Y. Matsuura, and L. DiCarlo, npj Quantum Inf. **7**, 130 (2021).

[20] T. J. Sewell, C. D. White, and B. Swingle, arXiv:2210.16419.

[21] J. Martyn and B. Swingle, Phys. Rev. A **100**, 032107 (2019).

[22] F. Verstraete, J. J. García-Ripoll, and J. I. Cirac, Phys. Rev. Lett. **93**, 207204 (2004).

[23] A. N. Chowdhury and R. D. Somma, Quantum Inf. Comp. **17**, 0041 (2017).

[24] C. Zoufal, A. Lucchi, and S. Woerner, Quantum Mach. Intell. **3**, 7 (2021).

[25] J. Gacon, C. Zoufal, G. Carleo, and S. Woerner, Quantum **5**, 567 (2021).

[26] X. Wang, X. Feng, T. Hartung, K. Jansen, and P. Stornati, Phys. Rev. A **108**, 022612 (2023).

[27] X. Yuan, S. Endo, Q. Zhao, Y. Li, and S. C. Benjamin, Quantum **3**, 191 (2019).

[28] T. Haug and K. Bharti, Quantum Sci. Technol. **7**, 045019 (2022).

[29] K. Temme, T. J. Osborne, K. G. Vollbrecht, D. Poulin, and F. Verstraete, Nature (London) **471**, 87 (2011).

[30] M.-H. Yung and A. Aspuru-Guzik, Proc. Natl. Acad. Sci. USA **109**, 754 (2012).

[31] M. Motta, C. Sun, A. T. K. Tan, M. J. O'Rourke, E. Ye, A. J. Minnich, F. G. S. L. Brandão, and G. K.-L. Chan, Nat. Phys. **16**, 205 (2020).

[32] O. Shtanko and R. Movassagh, arXiv:2112.14688.

[33] P. Rall, C. Wang, and P. Wocjan, Quantum **7**, 1132 (2023).

[34] Z. Holmes, G. Muraleedharan, R. D. Somma, Y. Subasi, and B. Şahinoğlu, Quantum **6**, 825 (2022).

[35] L. Coopmans, Y. Kikuchi, and M. Benedetti, PRX Quantum **4**, 010305 (2023).

[36] J. C. Getelina, N. Gomes, T. Iadecola, P. P. Orth, and Y.-X. Yao, SciPost Phys. **15**, 102 (2023).

[37] I. Bengtsson and K. Zyczkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, Cambridge, 2006).

[38] T. Matsui, Commun. Math. Phys. **162**, 321 (1994).

[39] F. Franchini, *An Introduction to Integrable Techniques for One-Dimensional Quantum Systems*, Lecture Notes in Physics Vol. 940 (Springer, Cham, 2017).

[40] A. Uhlmann, Found. Phys. **41**, 288 (2011).

[41] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information*, 10th anniversary ed. (Cambridge University Press, Cambridge, 2010).

[42] V. Vedral, Rev. Mod. Phys. **74**, 197 (2002).

[43] J. Nocedal and S. J. Wright, *Numerical Optimization*, 2nd ed. (Springer, New York, 2006).

[44] IBM Quantum, https://quantum-computing.ibm.com/services/resources.

[45] G. Li, Y. Ding, and Y. Xie, in *ASPLOS '19: Proceedings of the Twenty-Fourth International Conference on Architec-tural Support for Programming Languages and Operating Systems* (Association for Computing Machinery, New York, NY, 2019).

[46] J. Spall, IEEE Trans. Autom. Control **37**, 332 (1992).

[47] P. D. Nation, H. Kang, N. Sundaresan, and J. M. Gambetta, PRX Quantum **2**, 040326 (2021).

[48] S. Wang, E. Fontana, M. Cerezo, K. Sharma, A. Sone, L. Cincio, and P. J. Coles, Nat. Commun. **12**, 6961 (2021).

[49] A. Sannia, A. Giordano, N. Lo Gullo, C. Mastroianni, and F. Plastina, Sci. Rep. **13**, 3913 (2023).

[50] J. Um, B. J. Kim, and S.-I. Lee, J. Korean Phys. Soc. **50**, 285 (2007).

[51] R. LaRose, A. Mari, S. Kaiser, P. J. Karalekas, A. A. Alves, P. Czarnik, M. E. Mandouh, M. H. Gordon, Y. Hindy, A. Robertson, P. Thakre, M. Wahl, D. Samuel, R. Mistri, M. Tremblay, N. Gardner, N. T. Stemen, N. Shammah, and W. J. Zeng, Quantum **6**, 774 (2022).

[52] Y. Suzuki, Y. Kawase, Y. Masumura, Y. Hiraga, M. Nakadai, J. Chen, K. M. Nakanishi, K. Mitarai, R. Imai, S. Tamiya, T. Yamamoto, T. Yan, T. Kawakubo, Y. O. Nakagawa, Y. Ibe, Y. Zhang, H. Yamashita, H. Yoshimura, A. Hayashi, and K. Fujii, Quantum **5**, 559 (2021).

[53] IBM Quantum, QISKIT: An open-source framework for quantum computing, https://quantum-computing.ibm.com.

[54] M. Consiglio, Variational Gibbs state preparation, 2023, https://github.com/mirkoconsiglio/VariationalGibbsStatePreparation.

[55] A. Nietner, arXiv:2306.04731.

[56] L. Paninski, Neural Comput. **15**, 1191 (2003).

[57] M. S. Roulston, Phys. D (Amsterdam, Neth.) **125**, 285 (1999).

[58] A. Antos and I. Kontoyiannis, Random Struct. Algorithms **19**, 163 (2001).

[59] S. Marsili and C. Cattuto, Bayesian entropy estimation in Python - via the Nemenman-Schafee-Bialek algorithm, 2003, https://github.com/simomarsili/ndd.

[60] I. Nemenman, F. Shafee, and W. Bialek, arXiv:physics/0108025.

[61] I. Nemenman, Entropy **13**, 2013 (2011).

[62] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Nat. Commun. **12**, 1791 (2021).