# scientific reports

Check for updates

**OPEN**

# Noise-assisted variational quantum thermalization

Jonathan Foldager [1]✉, Arthur Pesah[2] & Lars Kai Hansen[1]

Preparing thermal states on a quantum computer can have a variety of applications, from simulating many-body quantum systems to training machine learning models. Variational circuits have been proposed for this task on near-term quantum computers, but several challenges remain, such as finding a scalable cost-function, avoiding the need of purification, and mitigating noise effects. We propose a new algorithm for thermal state preparation that tackles those three challenges by exploiting the noise of quantum circuits. We consider a variational architecture containing a depolarizing channel after each unitary layer, with the ability to directly control the level of noise. We derive a closed-form approximation for the free-energy of such circuit and use it as a cost function for our variational algorithm. By evaluating our method on a variety of Hamiltonians and system sizes, we find several systems for which the thermal state can be approximated with a high fidelity. However, we also show that the ability for our algorithm to learn the thermal state strongly depends on the temperature: while a high fidelity can be obtained for high and low temperatures, we identify a specific range for which the problem becomes more challenging. We hope that this first study on noise-assisted thermal state preparation will inspire future research on exploiting noise in variational algorithms.

Noise is often considered to be one of the strongest adversaries of practical quantum computation. Decoherence effects due to a noisy environment can create errors in the final output of a circuit, destroying the advantage of many quantum algorithms. In contrast, noise is also what underlies stochastic processes, and is therefore a crucial element in classical computing, solving tasks such as sampling and optimization. In quantum systems, noise has also been shown to be a useful resource in several applications: carefully engineered dissipative processes can lead to universal quantum computation[1], shot-noise in the measurement process can drive variational algorithms out of local minima[2,3], and amplitude-damping channels can significantly improve quantum autoencoders for mixed states[4]. We investigate in the present paper a novel way to exploit noise in near-term quantum devices, with the objective of studying a central task in quantum computing: thermal state preparation.

Placing a quantum system driven by a Hamiltonian $H$ and weakly-coupled to a reservoir with an effective temperature $T = \frac{1}{\beta}$, the system will asymptotically reach a thermal equilibrium state, given by the quantum Gibbs distribution

$$\rho_\beta = \frac{1}{Z} e^{-\beta H} \tag{1}$$

where $Z = \mathrm{Tr}[e^{-\beta H}]$ is the partition function[5]. Efficiently preparing a thermal state on a quantum computer is a problem of broad practical importance, with applications ranging from quantum chemistry and many-body physics simulations in an open environment[6–8] to semi-definite programming[9,10] and quantum machine learning[11,12]. However, sampling from a general Gibbs distribution is a computationally hard task for classical computers, due to the complexity of calculating the partition function[13]. Most techniques rely on Monte-Carlo Markov Chain (MCMC) algorithms, which are often provably efficient only above a certain threshold temperature[14].

Many algorithms have been proposed to prepare the thermal state on a quantum computer. A growing body of work has suggested using variational algorithms to solve the task of thermal state preparation on Noisy Intermediate Scale Quantum (NISQ) devices. Since a unitary circuit acting on the zero-state cannot directly output a mixed state, most variational thermalization methods consist either in preparing a purification of the thermal state and tracing out the ancillary qubits at the end of the circuit[15–18], or in choosing an appropriate mixed state as input[19–21].

[1]Department of Applied Mathematics and Computer Science, Technical University of Denmark, 2800 Kongens Lyngby, Denmark. [2]Department of Physics and Astronomy, University College London, London WC1E 6BT, UK. ✉email: jonf@dtu.dk
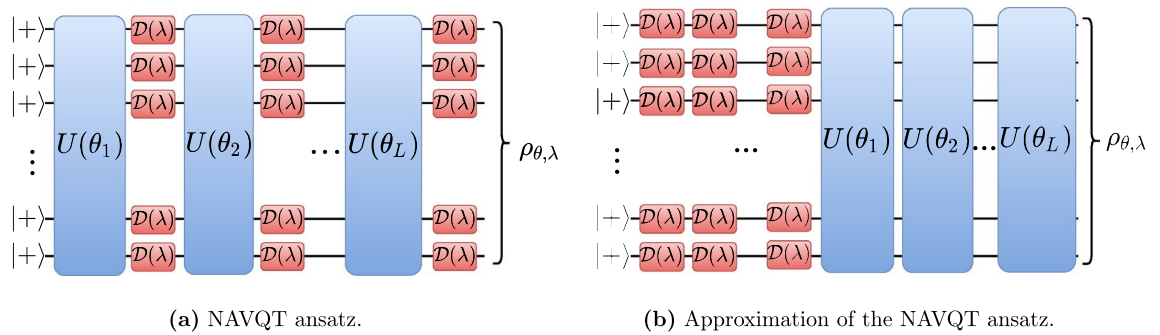
**Figure 1.** Illustration of circuit components used in NAVQT. (**a**) General NAVQT ansatz: a sequence of unitary layers $U(\theta_i)$ followed by depolarizing gates $\mathcal{D}(\lambda)$ on each qubit. (**b**) Approximation used in the free-energy calculations.

One of the main challenges associated to those methods is to design an appropriate cost function to be minimized during the variational training loop. While the ground-state of a Hamiltonian can be prepared by minimizing the average energy of the state, the thermal state can be prepared by minimizing the *free energy* $F = H - TS$ of the state, where $S = -\text{Tr}[\rho \log(\rho)]$ is its Von Neumann entropy. However, the Von Neumann entropy is not an observable and can often only be computed approximately[18,22]. A second problem is the need for additional qubits, which can be costly in near-term devices. Finally, none of those methods take into account the noise of the circuit, which can change the spectrum of the final state and affect the performance of the preparation algorithm[23].

In this paper, we propose a new method that we call *Noise-Assisted Variational Quantum Thermalizer* (NAVQT). Our algorithm assumes the ability to control the noise in the system down to some minimal noise level determined by the hardware. This type of control has been demonstrated in the context of error mitigation, where noise is increased in order to perform zero-noise extrapolation[24,25]. More precisely, we construct a variational circuit with a parametrized depolarizing channel after each layer of unitary gates, as illustrated in Fig. 1(a). To simplify the optimization process, we have only considered the case where all the depolarizing parameters take the same value. By varying both gate and noise parameters, we seek to minimize the free energy of the final state.

In order to compute the free energy (and its gradient), we derive an analytical expression for the entropy of a slightly different circuit: one where all the depolarizing gates have been displaced at the beginning of the circuit, as shown in Fig. 1(b). Using this approximation, we can compute the gradient of the free energy with respect to both the noise and the unitary parameters. While this might be a rough estimate of the actual gradient, we show that this approximate optimization problem exhibits similar performance as when minimizing the true free energy.

We then empirically investigate our algorithm on three different types of Hamiltonians: the Ising chain, with and without a transverse field, and the Heisenberg model. For each model, we consider both uniform coefficients and coefficients drawn from a standard normal distribution, and train our variational algorithm for several choices of hyperparameters (number of layers, learning rates, initialization, etc.). To study the performance of our approach, we extract the fidelity of the prepared state compared to the actual thermal state for a range of different temperatures.

Our results reveal different patterns. On the one hand, fidelities above 0.9 are reached for uniform Ising chains, with and without a transverse field, for all temperatures and system sizes up to 7 qubits. On the other hand, the performance tend to decrease with the system size and for specific ranges of temperatures, with fidelities that can get below 0.7 for some of the most complex systems tested in this work.

Our paper is organized as follows. We start by reviewing previous work on variational thermalization in "Related work" section. We then introduce NAVQT in "Noise-assisted variational quantum thermalization" section. We follow this up by a description of our experiments in "Methods" section, and present our results in "Results" section. Finally, we discuss our work and provide ideas for future studies in "Discussion" section.

## Related work
Variational circuits have recently been proposed for thermal state preparation, due to the existence of a natural cost function for this task: the free energy. Using variational circuits to prepare a thermal state presents two challenges specific to this task: (1) finding an ansatz that can prepare mixed states, (2) finding a scalable optimization strategy.

**Choice of the ansatz.**    A common approach to VQT consists in preparing a purification of the thermal state using a variational circuit that acts on $2N$ qubits—$N$ system qubits and $N$ ancilla/environment qubits—, and tracing the ancilla qubits out at the end of the circuit[15–18]. An example of purification often considered in the literature is the thermofield double (TFD) state[15,16]. For a Hamiltonian $H$ and an inverse temperature $\beta$, it is given by

$$|\text{TFD}\rangle = \frac{1}{\sqrt{Z}} \sum_n e^{-\beta E_n/2} |n\rangle_S \otimes |n\rangle_E \qquad (2)$$

where the $\{E_n, |n\rangle\}_n$ are pairs of eigenvalue/eigenvector of $H$, and subscript $S$ and $E$ refers to the system and environment, respectively. For instance, Refs.[15,16] use a Quantum Approximate Optimization Ansatz (QAOA) ansatz acting on $2N$ qubits to prepare the TFD state of the transverse-field Ising model, the XY chain, and free fermions. One advantage of this approach is the ability to simulate the TFD, which can be interesting in in its own right, for instance for studying black holes[26]. The obvious disadvantage is that it requires twice as many qubits that the thermal state we want to simulate. A converse approach consists in starting with a mixed state $\rho_0$ and applying a unitary circuit ansatz on the $N$ qubits of the system. The initial $\rho_0$ can either be fixed[19] or modified during the optimization process[20,21]. In Ref.[19], $\rho_0$ is the fixed thermal state of $H_I = \sum_{i=1}^{N} Z_i$, where $Z_i$ is the Pauli $Z$ operator applied to qubit $i$ of the system. It can easily be prepared using the purification

$$\bigotimes_j \sqrt{2\cosh(\beta)} \sum_{b \in \{0,1\}^N} e^{(-1)^{1+b}\beta/2} |b\rangle_S |b\rangle_E. \qquad (3)$$

However, since the spectrum does not change when we apply the unitary ansatz, having a static $\rho_0$ freezes the spectrum of the final state. Therefore, if the spectrum of the thermal state we want to approximate is far from the spectrum of $\rho_0$, this approach will fail. In Ref.[20], they use the thermal state $\rho_0(\boldsymbol{\varepsilon})$ of $H = \sum_{i=1}^{n} \varepsilon_i P_i$, where $P_i = \frac{1-Z_i}{2}$ as an initial state and $\boldsymbol{\varepsilon} = \{\varepsilon_1, \dots, \varepsilon_n\}$ are parameters optimized during the training process. Finally, Ref.[21] proposes to use a unitary with stochastic parameters to prepare $\rho_0$. More precisely,

$$\rho_0(\boldsymbol{\theta}) = V(X_{\boldsymbol{\theta}}) |0\rangle\langle 0| V(X_{\boldsymbol{\theta}})^\dagger \qquad (4)$$

where $V(\boldsymbol{x})$ is a unitary ansatz and $X_{\boldsymbol{\theta}} \sim p_{\boldsymbol{\theta}}$ is a random vector with parametrized density $p_{\boldsymbol{\theta}}$. The density $p_{\boldsymbol{\theta}}$ can be given by a classical model, such as an energy-based model (e.g. restricted Boltzmann machine) or a normalizing flow, which will be trained to get a $\rho_0$ with a spectrum close to the thermal state of interest.

**Optimization strategies.** Once the ansatz has been fixed, the parameters within needs to be optimized. Two main approaches have been proposed in the literature: (1) explicitly minimizing the free energy, (2) using imaginary-time evolution. In the following, we describe both these methods.

*Free energy methods.* The thermal state is the density matrix that minimizes the free energy. Therefore, in the same way as VQE uses the energy as a cost function, any thermal state preparation method can use the free energy as its cost function[15,16,19,21]. However, one main difference with VQE is that the free energy cannot be easily estimated. Indeed, the Von Neumann entropy term, as a non-linear function of $\rho$, cannot be turned into an observable, and doing a full quantum state tomography would be very costly. Several methods have been proposed to solve this challenge:

- Computing several Renyi entropies $S_\alpha = \frac{1}{1-\alpha} \text{Tr}[\rho^\alpha]$ (using multiple copies of $\rho$) and approximating the Von Neumann entropy with them[15,27].
- Computing the Von Neumann entropies locally on a small subsystem[15]
- Approximate the Von Neumann entropy by truncating its Taylor[18] or Fourier[22] decomposition.

In our work, the entropy term does not come from a purification procedure, but from the presence of depolarizing gates in the circuits. This led us to propose a different type of approximation that we will study in "Noise-assisted variational quantum thermalization".

*Imaginary-time evolution.* Thermal state preparation can be seen as the application of imaginary-time evolution during a time $\Delta t = i\beta/2$ on the maximally-mixed state $\rho_m = \frac{1}{d}I$, using the decomposition

$$\rho_\beta = \left(\frac{1}{C} e^{-\beta H/2}\right) \left(\frac{1}{d}I\right) \left(\frac{1}{C} e^{-\beta H/2}\right)$$

This imaginary-time evolution can be simulated using a variational circuit and a specific update rule[28,29]. In Ref.[17], the authors use a variational circuit $U(\boldsymbol{\theta})$ on $2N$ qubits, initialized such that

$$U(\boldsymbol{\theta}_0)|0\rangle^{\otimes 2N} \approx |\Phi^+\rangle$$

where $\Phi^+$ is a maximally-entangled state. An imaginary-time update rule with a small learning rate $\tau$ will lead to a unitary $U(\boldsymbol{\theta}_0)$ such that:

$$U(\boldsymbol{\theta}_1)|0\rangle^{\otimes 2N} \approx \frac{1}{C} e^{-\tau H} |\Phi^+\rangle$$

Repeating it during $k = \frac{\beta}{2}$ steps will give the state

$$U(\boldsymbol{\theta}_k)|0\rangle^{\otimes 2N} \approx \frac{1}{C} e^{-\beta H/2} |\Phi^+\rangle$$

which will be the thermal state after tracing out the environment. In Ref.[30], the authors also use imaginary-time evolution to prepare the thermal state, but manage to reduce the number of qubits to $N$ when the Hamiltonian is diagonal in the $Z$-basis. Finally, an ansatz-independent imaginary-time evolution method has been proposed for thermal state preparation[31,32].

In this work, we optimize the ansatz parameters using the free energy approach. Adapting imaginary-time evolution to a noisy ansatz could however be an interesting alternative, that we let for future work.

## Noise-assisted variational quantum thermalization

We introduce here the *Noise-Assisted Variational Quantum Thermalizer* (NAVQT), a variational algorithm where depolarizing noise is used as the source of entropy for preparing the thermal state. We consider a noise model where each layer of unitary gates is followed by a one-qubit depolarizing channel

$$\mathcal{D}(\lambda)(\rho) = (1 - \lambda)\rho + \lambda\frac{\mathrm{I}}{2}, \tag{5}$$

where I is the identity matrix. The channel is represented in Fig. 1. For the purpose of this work, we consider that we have the same noise value $\lambda \in [\lambda_{\min}, 1]$ everywhere in the circuit, where $\lambda_{\min}$ is the minimum noise reachable by the hardware. We note $\rho_{\boldsymbol{\theta}, \lambda}$ the output of the noisy circuit with unitary parameters $\boldsymbol{\theta}$ and noise parameter $\lambda$, and want to find the optimal parameters $\{\boldsymbol{\theta}^*, \lambda^*\}$ such that $\rho_{\boldsymbol{\theta}^*, \lambda^*} \approx \rho_\beta$ where the latter is given by Eq. (1).

The thermal state $\rho_\beta$ can be approximated by minimizing the free energy of the system, given by:

$$F(\boldsymbol{\theta}, \lambda) = E(\boldsymbol{\theta}, \lambda) - \frac{1}{\beta}S(\boldsymbol{\theta}, \lambda) \tag{6}$$

where

$$E(\boldsymbol{\theta}, \lambda) = \mathrm{Tr}[H\rho_{\boldsymbol{\theta}, \lambda}] \tag{7}$$

is the energy and

$$S(\boldsymbol{\theta}, \lambda) = -\mathrm{Tr}[\rho_{\boldsymbol{\theta}, \lambda} \log(\rho_{\boldsymbol{\theta}, \lambda})] \tag{8}$$

is the Von Neumann entropy of the state.

The energy term and its gradient are easy to evaluate: we can use the parameter shift-rule[33] to compute $\nabla_{\boldsymbol{\theta}} E(\boldsymbol{\theta}, \lambda)$, and the finite-difference method to calculate $\partial_\lambda E(\boldsymbol{\theta}, \lambda)$. The entropy term is much harder to evaluate as it is a non-linear function of the state. To approximate it, we consider the circuit where all the noise has been put at the beginning, as shown in Fig. 1(b). While the resulting free energy will not be equal to the free energy of our original circuit in general, they tend to follow similar trajectories when varying the noise level (see Supplementary Fig. S1). The new entropy does not depend on $\boldsymbol{\theta}$ and can be calculated analytically as if there were no unitary gates. For a circuit with $N$ qubits and $m$ layers, this approximate entropy $\widetilde{S}(\lambda)$ is given by

$$\begin{aligned} \widetilde{S}(\lambda) = &-N\left((1-\lambda)^m + \frac{(1-(1-\lambda)^m)}{d}\right) \cdot \\ &\ln\left((1-\lambda)^m + \frac{(1-(1-\lambda)^m)}{d}\right) \\ &+ \frac{(d-1)(1-(1-\lambda)^m)}{d}\ln\left(\frac{(1-(1-\lambda)^m)}{d}\right) \end{aligned} \tag{9}$$

where $d = 2^N$. The full derivation is given in the Supplementary material. Using this approximation, we get the following gradient-based update rule at each optimization step:

$$\boldsymbol{\theta}^{(n+1)} = \boldsymbol{\theta}^{(n)} - \eta_\theta \nabla_{\boldsymbol{\theta}} E(\boldsymbol{\theta}, \lambda) \tag{10}$$

$$\lambda^{(n+1)} = \lambda^{(n)} - \eta_\lambda\left(\nabla_\lambda E(\boldsymbol{\theta}, \lambda) - \frac{1}{\beta}\nabla_\lambda \widetilde{S}(\lambda)\right) \tag{11}$$

where $\eta_\theta$ and $\eta_\lambda$ are the learning rates for $\boldsymbol{\theta}$ and $\lambda$, respectively.

## Methods

In this section, we will briefly describe the basis of conducted experiments. All quantum circuit simulations are done in Cirq[34] and TensorFlow-Quantum[35].

**Ansatz.** For the unitary layers of our circuit, we chose an ansatz inspired by the Quantum Approximate Optimization Ansatz (QAOA) applied to the Ising chain Hamiltonian[36]. More precisely, if we define a problem Hamiltonian

$$H_P = -\sum_i Z_i Z_{i+1} - \sum_i Z_i \tag{12}$$

and a mixing Hamiltonian
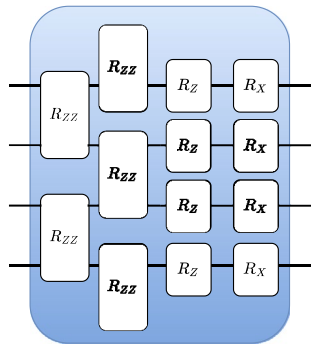
**Figure 2.** A layer of the unitary ansatz used in our experiments, inspired by QAOA for the 1D Ising model. $R_Z$ and $R_X$ are parametrized rotations around the corresponding axis, and $R_{ZZ} = e^{-i\theta Z_i Z_j}$.

$$H_M = -\sum_i X_i, \tag{13}$$

the QAOA ansatz with $p$ layers is given by

$$U(\boldsymbol{\gamma}, \boldsymbol{\beta}) = e^{i\beta_p H_M} e^{i\gamma_p H_P} \cdots e^{i\beta_1 H_M} e^{i\gamma_1 H_P} \tag{14}$$

This ansatz, whose explicit construction is represented in Fig. 2, has been well-studied in the context of ground-state preparation[37] and has been shown to be universal[38] in the limit $p \to \infty$. We test two different versions of this ansatz. In the first one, denoted *restricted QAOA*, gates of the same type from a given layer share the same parameters $\beta_i$ and $\gamma_i$. In the second version, which we call *flexible QAOA*, every gate has its own parameter.

We ran some preliminary tests to verify that this unitary ansatz is at least able to express the ground-state of all the systems tested in our work, and found it to be the case when the number of layers is fixed at $\lceil \frac{N}{2} \rceil$. Hence the noisy ansatz should in principle be able to represent the correct thermal state for large $\beta$, by setting $\lambda = 0$ and fitting the unitary parameters corresponding to the ground-state. Moreover, NAVQT is also able to represent the maximally-mixed state, corresponding to a low $\beta$, by setting $\lambda = 1$. In Supplementary Figure S3, we provide some results for a varying number of layers $L \in \{2, 3, 4, 5, 6, 7, 8, 9, 10\}$ at $\beta = 1$ for the three-qubit Heisenberg Hamiltonian with random coefficients, showing that the fidelity does not improve significantly compared to our heuristic number of layers. Hence we find evidence to rule out the number of ansatz layers as a limiting factor to achieve better performance. The ability of the ansatz to learn intermediate temperatures is an open question, that we tackle in our numerical analysis.

**Hyperparameters.** Since the choice of hyperparameters can have a substantial impact on the performance of variational circuits[37], we perform a grid-search to reduce the potential negative effects resulting from a single design choice. Hence we try all combinations in the search space defined by

- Restricted QAOA and flexible QAOA
- Initial noise level: $\lambda = \{10^{-8}, 0.001, 0.1\}$
- Unitary learning rate: $\eta_\theta = \{0.01, 0.4\}$
- Noise learning rate: $\eta_\lambda = \{0.0001, 0.1\}$
- Seeds for unitary parameters: $[0; 4]$.

We run our algorithm for $N \in [3; 7]$ qubits and for maximum 1000 iterations. To test the performance across temperatures, we take 10 different betas in the interval $\beta \in [10^{-3}; 10^2]$, namely $\{0.001, 0.1, 0.25, 0.5, 0.75, 1.0, 2.0, 5.0, 10.0, 100.0\}$. We initialize the unitary parameters by sampling from a uniform distribution in the interval $[0.0001, 0.05]$ as done in[37]. Finally, we extract the solution that gives the lowest (approximated) free energy among all the tested hyperparameters and initializations. We also include the same grid-search using finite-difference on the true free-energy in Supplementary Fig. S2.

**Noisy circuit simulation.** To simulate the noise in our circuit, we use the fact that depolarizing gates can also be written as[39]

$$\mathcal{D}(\lambda)(\rho) = \left(1 - \frac{3\lambda}{4}\right)\rho + \frac{\lambda}{4}(X\rho X + Y\rho Y + Z\rho Z) \tag{15}$$

which can be interpreted as applying a random Pauli error with probability $p = \frac{3\lambda}{4}$ and nothing with probability $p = 1 - \frac{3\lambda}{4}$. We can therefore simulate depolarizing gates as stochastic mixtures over unitary circuits containing errors. More precisely, if we sample $K$ unitaries $U^{(k)}$, each being a combination of the unitary part of the ansatz and some random errors, we can extract the corresponding density matrix as:

$$\rho_{\text{out}} \approx \frac{1}{K} \sum_{k=1}^{K} U^{(k)} \rho_{in} \left( U^{(k)} \right)^{\dagger} \tag{16}$$

We found that taking a sample size of $K = 500N$ was sufficient to get stable gradients and reach the maximum entropy $S \leq \log 2^N$. However, we also found that $K$ could be smaller, especially when $\beta$ was large and hence the target entropy was low.

**Performance metric.** For each experiment, we report the fidelity

$$F(\rho_1, \rho_2) = \text{Tr}\left[ \sqrt{\sqrt{\rho_1} \rho_2 \sqrt{\rho_1}} \right] \tag{17}$$

between the thermal state and the output state of the trained circuit. Tracking the fidelity requires us to compute the true thermal state $\rho_\beta$ for each Hamiltonian $H$ and temperature $\beta$. In practice, taking the exponential of a matrix containing potentially large numerical values (e.g. when $\beta$ is large) can result in numerical issues. To alleviate those issues, we calculate the thermal state density matrix $\rho_\beta$ by taking the log on both sides of Eq. (1) and using the log-sum-exp trick[40]:

$$\begin{aligned} \log \rho_\beta &= \log e^{-\beta H} - \log \text{Tr}[e^{-\beta H}] \\ &= -\beta H - \log \sum_i e^{-\beta \lambda_i} \\ &= -\beta H - \left( -\beta c + \log \sum_i e^{-\beta(\lambda_i - c)} \right) \end{aligned} \tag{18}$$

where $c$ is the largest eigenvalue of $H$.

**Models.** We evaluated our algorithm on three different models: the Ising chain, with and without a transverse field, and the Heisenberg model. For each model, we considered two cases: when the coefficients $J_i = h_i = 1$ for all $i$, denoted the *uniform* version, and when $J_i, h_i \sim \mathcal{N}(0, 1)$ for all $i$, denoted the *random* version. Between five seeds for the random version, we pick the Hamiltonian with the lowest spectral gap as this could be considered the hardest Hamiltonian. In the case for Hamiltonians with random coefficients, we normalized the set of all coefficients such that the vector containing all coefficients had unit length. See Supplementary Fig. S4 for a plot of the model energy scales.

*Ising chain.* The 1D Ising model, or Ising chain (IC), considers a set of spins on a chain such that all spins have exactly two coupled neighbors when considering $N > 2$. The Hamiltonian associated with such system is given by

$$H_{\text{IC}} = -\sum_i J_i Z_i Z_{i+1} - \sum_i h_i Z_i \tag{19}$$

where $Z_i$ is the Pauli $Z$ operator acting on qubit $i$.

*Transverse field Ising chain.* The transverse-field Ising chain (TFI) adds quantum effects to the previous model by including some non-diagonal terms in its Hamiltonian. It is defined as

$$H_{\text{TFI}} = -\sum_i J_i^Z Z_i Z_{i+1} - \sum_i h_i^Z Z_i - \sum_i h_i^X X_i \tag{20}$$

where $X_i$ is the Pauli $X$ operator acting on qubit $i$.

*Heisenberg model.* Finally, we consider the 1D Heisenberg model, whose Hamiltonian is given by

$$\begin{aligned} H_{\text{Heisenberg}} = &-\sum_i J_i^Z Z_i Z_{i+1} - \sum_i J_i^X X_i X_{i+1} \\ &-\sum_i J_i^Y Y_i Y_{i+1} - \sum_i h_i^X X_i \end{aligned} \tag{21}$$

The Heisenberg model is of fundamental importance in the study of quantum materials[41–44] and is therefore a standard benchmark for thermal state preparation methods[31,32,45].
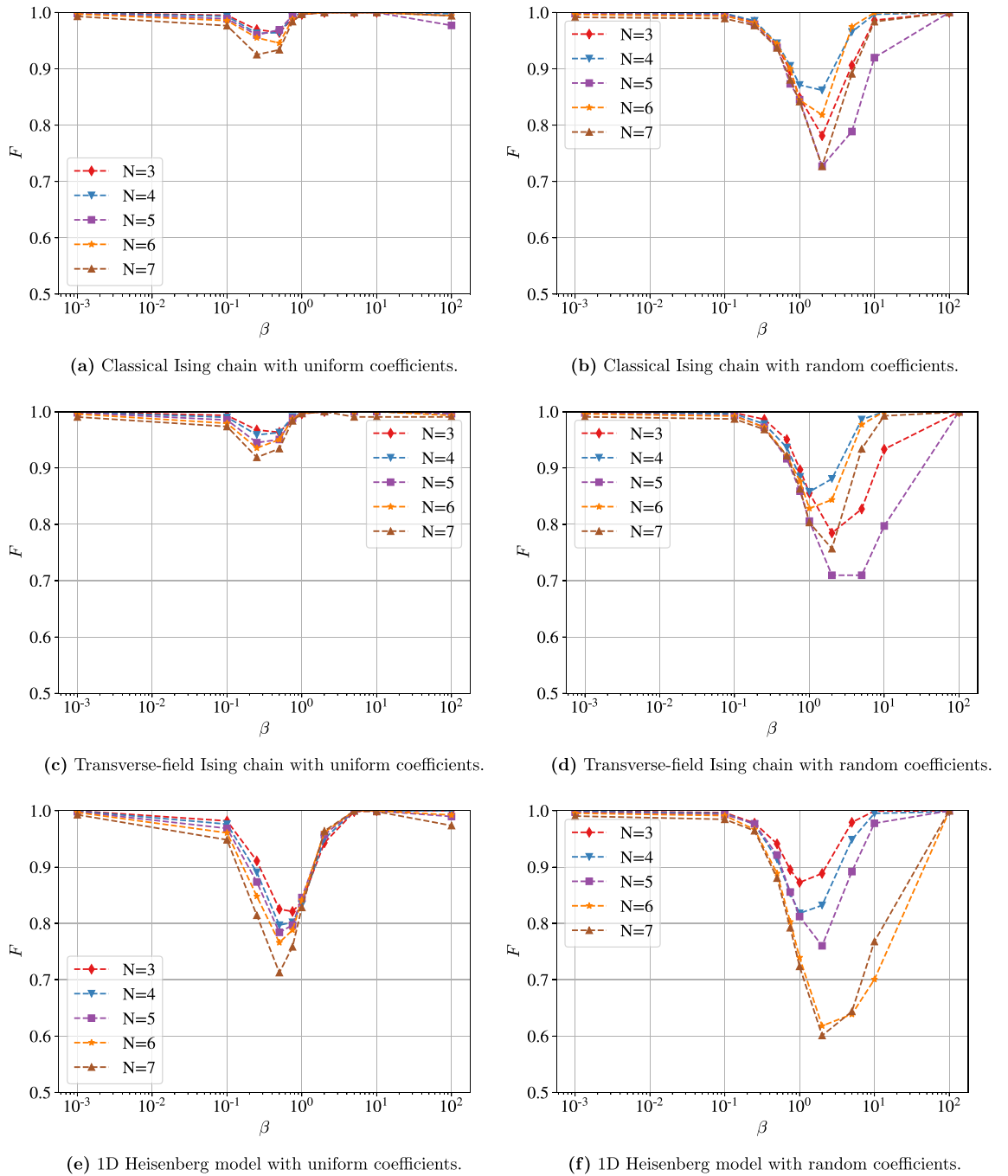
## Results

We first present the optimization curves for N = 4, at three different temperatures $\beta \in \{0.1, 0.5, 10\}$ in Fig. 3. We report the fidelity between the learned state and the thermal state as a function of the inverse temperature $\beta$ for all the different models in Fig. 4. Finally, we also report the final noise level $\lambda$ as a function of $\beta$ for all models in Fig. 5. We can notice a few phenomena from those figures:

**(a)** $\beta = 0.1$



**(b)** $\beta = 0.5$



**(c)** $\beta = 10$
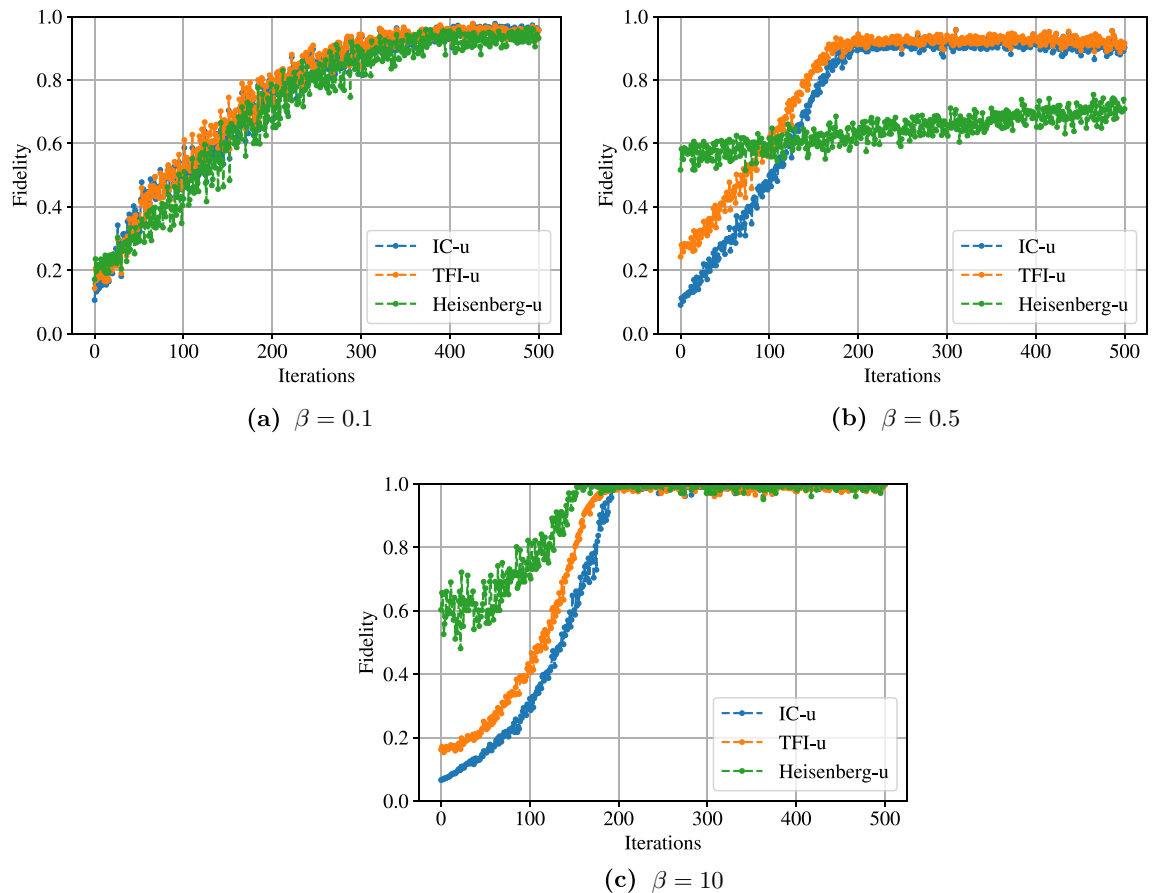
**Figure 3.** Optimization curves for the three models with uniform coefficients and $N = 4$. We observe in all the cases a constant increase of the fidelity, showing that minimizing the approximate free energy cost function tends to result in a maximization of the fidelity. It also shows that the final result found by the algorithm is always significantly better than the random initialization.

1. The optimization curves presented in Fig. 3 show that the optimization procedure improves the solution compared to a random initialization, both when a very high fidelity is reached at the end and when the fidelity is lower. It eliminates the possibility that random states being closed to the desired thermal states would explain our results. Moreover, the fidelity tends to increase with the number of iterations, showing that our approximate cost-function might be well-suited to our optimization goal.
2. Thermal states at low and high temperatures are easily approximated by our method, for all models and system sizes. Looking at the $\lambda$ curves, we see that the optimizer is indeed able to find $\lambda = 0$ for very large $\beta$ and $\lambda = 1$ for very low $\beta$. Hence, when the thermal state gets close to a maximally-mixed state or to a pure state, the algorithm learns to respectively maximize or minimize the noise, independently of the initial noise level.
3. The performance tends to degrade at intermediate temperatures, reaching for instance a fidelity of 0.6 for the Heisenberg model with random coefficients. However, there are several temperatures for which a non-trivial noise level is learned and the fidelity remains high, such as the same model at $\beta = 10^{-1}$, for which a fidelity above 96% is reached for all system sizes with a noise level between 0.5 and 0.8. Hence the algorithm can actually find the correct thermal state in non-trivial temperature regimes.

From those results, an important question to consider is whether the low fidelity obtained for some systems is due to a failure of the optimization procedure or to the potentially low expressibility of our noisy ansatz. To tackle this question, we tested different methods to optimize the parameters of the ansatz, including a grid-search in the parameter space for systems that are small enough to allow it to run in a reasonable time. We found no significant improvement in the fidelity compared to the original optimization method. We also tried to initialize the unitary ansatz to the ground-state solution before turning on the noise, but it did not result in a significant increase of fidelity neither. Finally, to evaluate the effect of our free energy approximation, we performed all the experiments previously mentioned using finite-difference on the true free energy. The corresponding results can be found in Supplementary Figure S2, where we observe very similar fidelities as with the approximate free energy method. It means that for the hardest systems tested in this work, the noisy ansatz was probably not expressible enough to output an accurate approximation of the thermal state, independently of the optimization algorithm.
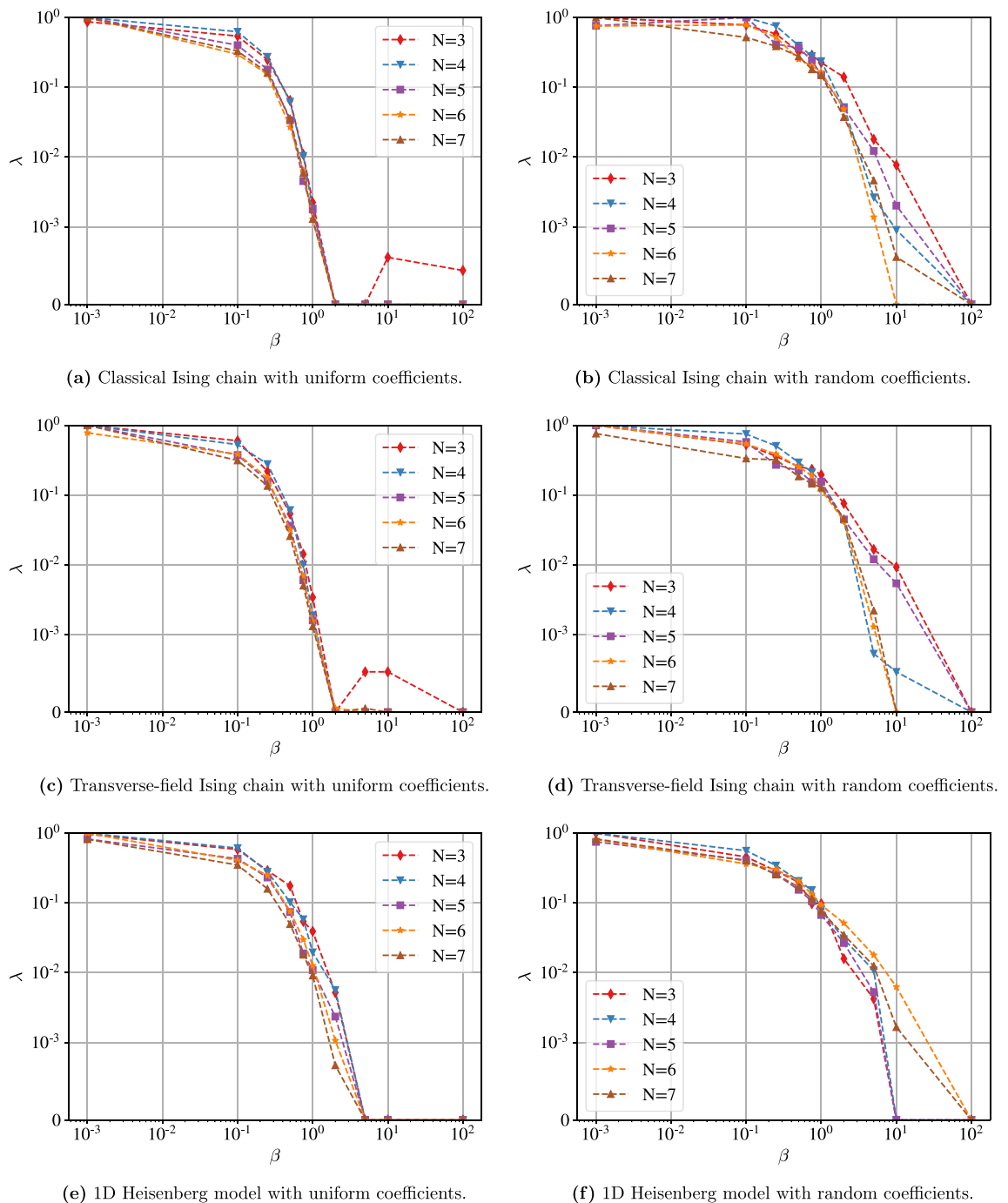
**(a)** Classical Ising chain with uniform coefficients.

**(b)** Classical Ising chain with random coefficients.

**(c)** Transverse-field Ising chain with uniform coefficients.

**(d)** Transverse-field Ising chain with random coefficients.

**(e)** 1D Heisenberg model with uniform coefficients.

**(f)** 1D Heisenberg model with random coefficients.

**Figure 4.** Fidelities obtained using NAVQT as a function of the inverse temperature $\beta$, for various models and system sizes. For all the models, we observe that the algorithms reaches a high fidelity for low and high temperature, while it tends to decrease at intermediate temperatures. Overall, good performance is obtained at all temperatures for the two types of uniform Ising chains, while lower fidelities are reached with the other models.

Changing the depolarizing gates to more general noise channels could help improve the expressibility of the ansatz and is let for future work.

## Discussion

In this paper, we introduced a novel type of variational algorithms, in which the noise is parameterized and optimized together with the unitary gates. We used this architecture to prepare thermal states, overcoming some of the most common challenges for this task, such as the need of ancilla qubits and the adverse effect of noise. To optimize our ansatz, we used a closed-form approximation of the free-energy and performed gradient-descent with it. We investigated various Hamiltonians and deduced that the ability of our method to learn the correct thermal states strongly depends on the model, the temperature and the system size. While we systematically

**(a)** Classical Ising chain with uniform coefficients.

**(b)** Classical Ising chain with random coefficients.

**(c)** Transverse-field Ising chain with uniform coefficients.

**(d)** Transverse-field Ising chain with random coefficients.

**(e)** 1D Heisenberg model with uniform coefficients.

**(f)** 1D Heisenberg model with random coefficients.

**Figure 5.** Final noise level $\lambda$ as a function of the inverse temperature $\beta$ for various models and system sizes. We used a *symlog* scale for the y-axis, hence the scale becomes linear below $10^{-3}$. We observe a clear decrease of the noise level with $\beta$, with $\lambda \approx 1$ for $\beta = 10^{-3}$ (corresponding to the maximally-mixed state) and $\lambda \approx 0$ for $\beta \approx 10^2$ (corresponding to the ground-state). It shows that the general relationship between the noise and the temperature has overall been correctly learned by our model.

obtained fidelities above 0.9 for both the transverse-field and the classical Ising chain, we had fidelities below 0.7 at some temperatures for the 1D Heisenberg model with random coefficients. We also identified a specific range of temperatures for each model, for which the task is harder for NAVQT to solve. Our experiments with different optimization algorithms reveal that the failure of the ansatz to learn the correct thermal state in those cases is probably an expressibility rather than an optimization issue.

This paper serves as a starting point in the study of noise-assisted thermalization, and many avenues are still open for future work. For instance, we only considered a single shared parameter $\lambda$ for all the depolarizing gates, as it allowed us to derive an approximation of the free energy, which simplified the optimization process. Varying the noise across each layer and each qubit independently could significantly increase the expressibility of

the ansatz. More generally, replacing the depolarizing gates by channels that are more tailored for thermal state preparation would be an interesting avenue to improve our method. For instance, Davies maps are non-unital channels that can model the evolution of quantum systems weakly-coupled to a thermal reservoir, making them particularly adapted to thermal state preparation[46]. Moreover, their unitary and dissipative parts commute, making the calculation of the entropy potentially easier than for our ansatz.

A second important aspect for future work would be to better understand the theory behind noise-assisted variational circuits. For instance, what are the conditions on the Hamiltonian and the temperature under which NAVQT can approximate the thermal state with an arbitrary high fidelity? How does our method scale with the system size? What type of noise is necessary to approximate a given thermal state?

Finally, it could be interesting to study the optimization landscape of NAVQT and potentially come up with optimization algorithms that are more tailored to this problem. For instance, it has been shown that a barren plateau phenomenon occurs in noisy circuits that are similar to our ansatz[47]. It can potentially hinder the scalability of our method, as it relies explicitly on increasing the noise. Finding the relationship between the temperature $\beta$, the system size $N$ and the magnitude of the gradient could be an interesting direction for future research.

## Data availability
All code is available at https://github.com/jfold/navqt.

## References

1. Verstraete, F., Wolf, M. M. & Cirac, J. I. Quantum computation and quantum-state engineering driven by dissipation. *Nat. Phys.* **5**, 633–636 (2009).
2. Kübler, J. M., Arrasmith, A., Cincio, L. & Coles, P. J. An adaptive optimizer for measurement-frugal variational algorithms. *Quantum* **4**, 263 (2020).
3. Sweke, R. *et al.* Stochastic gradient descent for hybrid quantum-classical optimization. *Quantum* **4**, 314 (2020).
4. Cao, C. & Wang, X. Noise-assisted quantum autoencoder. arXiv preprint arXiv arXiv:2012.08331 (2020).
5. Spohn, H. & Lebowitz, J. L. Irreversible thermodynamics for quantum systems weakly coupled to thermal reservoirs. *Adv. Chem. Phys.* **38**, 109–142 (1978).
6. Cao, Y. *et al.* Quantum chemistry in the age of quantum computing. *Chem. Rev.* **119**, 10856–10915 (2019).
7. Whitfield, J. D. *et al.* Introduction to quantum algorithms for physics and chemistry. *Advances in Chemical Physics* (2013).
8. Lee, C.-K., Patil, P., Zhang, S. & Hsieh, C.-Y. A neural-network variational quantum algorithm for many-body dynamics. arXiv preprint arXiv:2008.13329 ( 2020).
9. Brandao, F. G. & Svore, K. M. Quantum speed-ups for solving semidefinite programs. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, 415–426 (IEEE, 2017).
10. Van Apeldoorn, J., Gilyén, A., Gribling, S. & de Wolf, R. Quantum sdp-solvers: Better upper and lower bounds. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, 403–414 (IEEE, 2017).
11. Kieferová, M. & Wiebe, N. Tomography and generative training with quantum Boltzmann machines. *Phys. Rev. A* **96**, 062327 (2017).
12. Amin, M. H., Andriyash, E., Rolfe, J., Kulchytskyy, B. & Melko, R. Quantum Boltzmann machine. *Phys. Rev. X* **8**, 021050 (2018).
13. Jerrum, M. & Sinclair, A. Polynomial-time approximation algorithms for the Ising model. *SIAM J. Comput.* **22**, 1087–1116 (1993).
14. Eldan, R., Koehler, F. & Zeitouni, O. A spectral condition for spectral gap: Fast mixing in high-temperature Ising models. arXiv preprint arXiv:2007.08200 (2020).
15. Wu, J. & Hsieh, T. H. Variational thermal quantum simulation via thermofield double states. *Phys. Rev. Lett.* **123**, 220502 (2019).
16. Zhu, D. *et al.* Generation of thermofield double states and critical ground states with a quantum computer. arXiv preprint arXiv:1906.02699 (2019).
17. Zoufal, C., Lucchi, A. & Woerner, S. Variational quantum Boltzmann machines. arXiv preprint arXiv:2006.06004 (2020).
18. Wang, Y., Li, G. & Wang, X. Variational quantum Gibbs state preparation with a truncated taylor series. arXiv preprint arXiv:2005.08797 (2020).
19. Verdon, G., Broughton, M. & Biamonte, J. A quantum algorithm to train neural networks using low-depth circuits. arXiv preprint arXiv:1712.05304 (2017).
20. Martyn, J. & Swingle, B. Product spectrum ansatz and the simplicity of thermal states. *Phys. Rev. A* **100**, 032107 (2019).
21. Verdon, G., Marks, J., Nanda, S., Leichenauer, S. & Hidary, J. Quantum Hamiltonian-based models and the variational quantum thermalizer algorithm. arXiv preprint arXiv:1910.02071 (2019).
22. Chowdhury, A. N., Low, G. H. & Wiebe, N. A variational quantum algorithm for preparing quantum Gibbs states. arXiv preprint arXiv:2002.00055 ( 2020).
23. Franca, D. S. & Garcia-Patron, R. Limitations of optimization algorithms on noisy quantum devices. arXiv preprint arXiv:2009.05532 (2020).
24. Temme, K., Bravyi, S. & Gambetta, J. M. Error mitigation for short-depth quantum circuits. *Phys. Rev. Lett.* **119**, 180509 (2017).
25. Endo, S., Cai, Z., Benjamin, S. C. & Yuan, X. Hybrid quantum-classical algorithms and quantum error mitigation. arXiv preprint arXiv:2011.01382 (2020).
26. Cottrell, W., Freivogel, B., Hofman, D. M. & Lokhande, S. F. How to build the thermofield double state. *J. High Energy Phys.* **2019**, 58 (2019).
27. D'Hoker, E., Dong, X. & Wu, C.-H. An alternative method for extracting the von Neumann entropy from Rényi entropies. *J. High Energy Phys.* **2021**, 1–23 (2021).
28. McArdle, S. *et al.* Variational ansatz-based quantum simulation of imaginary time evolution. *NPJ Quantum Inf.* **5**, 1–6 (2019).
29. Benedetti, M., Fiorentini, M. & Lubasch, M. Hardware-efficient variational quantum algorithms for time evolution. *Phys. Rev. Res.* **3**, 033083 (2021).
30. Shingu, Y. *et al.* Boltzmann machine learning with a variational quantum algorithm. arXiv preprint arXiv:2007.00876 (2020).
31. Sun, S.-N. *et al.* Quantum computation of finite-temperature static and dynamical properties of spin systems using quantum imaginary time evolution. *PRX Quantum* **2**, 010317 (2021).
32. Motta, M. *et al.* Determining eigenstates and thermal states on a quantum computer using quantum imaginary time evolution. *Nat. Phys.* **16**, 205–210 (2020).
33. Schuld, M., Bergholm, V., Gogolin, C., Izaac, J. & Killoran, N. Evaluating analytic gradients on quantum hardware. *Phys. Rev. A* **99**, 032331 (2019).

34. Team, Q. A. & collaborators. Cirq. https://doi.org/10.5281/zenodo.4062499 (2020).
35. Broughton, M. *et al.* Tensorflow quantum: A software framework for quantum machine learning. arXiv preprint arXiv:2003.02989 (2020).
36. Farhi, E., Goldstone, J. & Gutmann, S. A quantum approximate optimization algorithm. arXiv preprint arXiv:1411.4028 (2014).
37. Wierichs, D., Gogolin, C. & Kastoryano, M. Avoiding local minima in variational quantum eigensolvers with the natural gradient optimizer. *Phys. Rev. Res.* **2**, 043246 (2020).
38. Morales, M. E., Biamonte, J. D. & Zimborás, Z. On the universality of the quantum approximate optimization algorithm. *Quantum Inf. Process.* **19**, 1–26 (2020).
39. Preskill, J. Lecture notes for physics 229: Quantum information and computation. *Calif. Inst. Technol.* **16**, 1–8 (1998).
40. Blanchard, P., Higham, D. J. & Higham, N. J. Accurately computing the log-sum-exp and softmax functions. *IMA J. Numer. Anal.* **41**, 2311–2330 (2021).
41. Billoni, O. V., Cannas, S. A. & Tamarit, F. A. Spin-glass behavior in the random-anisotropy Heisenberg model. *Phys. Rev. B* **72**, 104407 (2005).
42. Gong, S.-S., Zhu, W. & Sheng, D. Emergent chiral spin liquid: Fractional quantum hall effect in a Kagome Heisenberg model. *Sci. Rep.* **4**, 1–6 (2014).
43. Jepsen, P. N. *et al.* Spin transport in a tunable Heisenberg model realized with ultracold atoms. *Nature* **588**, 403–407 (2020).
44. Rodriguez-Nieva, J. F. Turbulent relaxation after a quench in the Heisenberg model. *Phys. Rev. B* **104**, L060302 (2021).
45. Powers, C., Bassman, L. & de Jong, W. A. Exploring finite temperature properties of materials with quantum computers. arXiv preprint arXiv:2109.01619 (2021).
46. Roga, W., Fannes, M. & Życzkowski, K. Davies maps for qubits and qutrits. *Rep. Math. Phys.* **66**, 311–329 (2010).
47. Wang, S. *et al.* Noise-induced barren plateaus in variational quantum algorithms. arXiv preprint arXiv:2007.14384 (2020).

## Acknowledgements

## Author contributions

J.F. and A.P. formulated, conceived and conducted the theory and experiments. J.F., A.P. and L.K.H. analyzed the results. All authors reviewed the manuscript.

## Competing interests

The authors declare no competing interests.

## Additional information

**Supplementary Information** The online version contains supplementary material available at https://doi.org/10.1038/s41598-022-07296-z.

**Correspondence** and requests for materials should be addressed to J.F.

**Reprints and permissions information** is available at www.nature.com/reprints.

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