

PAPER • OPEN ACCESS

Solving quantum statistical mechanics with variational autoregressive networks and quantum circuits

To cite this article: Jin-Guo Liu *et al* 2021 *Mach. Learn.: Sci. Technol.* **2** 025011

View the [article online](#) for updates and enhancements.



OPEN ACCESS

PAPER

Solving quantum statistical mechanics with variational autoregressive networks and quantum circuits

RECEIVED
16 March 2020

REVISED
6 June 2020

ACCEPTED FOR PUBLICATION
1 July 2020

PUBLISHED
3 February 2021

Original Content from
this work may be used
under the terms of the
Creative Commons
Attribution 4.0 licence.

Any further distribution
of this work must
maintain attribution to
the author(s) and the title
of the work, journal
citation and DOI.



Jin-Guo Liu¹ Liang Mao², Pan Zhang³ and Lei Wang^{4,5}

¹ Institute of Physics, Chinese Academy of Sciences, Beijing 100190, People's Republic of China

² Department of Physics, Tsinghua University, Beijing 100084, People's Republic of China

³ CAS Key Laboratory for Theoretical Physics, Institute of Theoretical Physics, Chinese Academy of Sciences, Beijing 100190, People's Republic of China

⁴ Institute of Physics, Chinese Academy of Sciences, Beijing 100190, People's Republic of China

⁵ Songshan Lake Materials Laboratory, Dongguan, Guangdong 523808, People's Republic of China

E-mail: wanglei@iphy.ac.cn

Keywords: quantum computing, betaVQE, thermal dynamics, NISQ

Abstract

We extend the ability of an unitary quantum circuit by interfacing it with a classical autoregressive neural network. The combined model parametrizes a variational density matrix as a classical mixture of quantum pure states, where the autoregressive network generates bitstring samples as input states to the quantum circuit. We devise an efficient variational algorithm to jointly optimize the classical neural network and the quantum circuit to solve quantum statistical mechanics problems. One can obtain thermal observables such as the variational free energy, entropy, and specific heat. As a byproduct, the algorithm also gives access to low energy excitation states. We demonstrate applications of the approach to thermal properties and excitation spectra of the quantum Ising model with resources that are feasible on near-term quantum computers.

1. Introduction

Quantum statistical mechanics concerns the physical properties of quantum systems at non zero temperatures. These problems pose two sets of challenges to classical computational approaches. First of all, classical algorithms generally encounter the difficulties of diagonalizing exponentially large Hamiltonians or the sign problem originating from the quantum nature of the problem. Moreover, even on the eigenbasis of the Hamiltonian one still faces the problem of the intractable partition function, which involves summation of an exponentially large number of terms.

A straightforward way to address these difficulties is to directly implement the physical Hamiltonian on analog quantum devices and study the system at thermal equilibrium; for example, see references [1, 2]. On the other hand, a potentially more general approach would be to study problems in quantum statistical mechanics with a universal gate model quantum computer. However, this strategy still calls for algorithmic innovations to prepare thermal quantum states given the unitary nature of quantum circuits. There have been quantum algorithms to prepare thermal Gibbs states on quantum computers [3–7]. Unfortunately, these approaches may not be feasible on near-term noisy quantum computers with limited circuit depth. While variational quantum algorithm for preparing thermofield double states [8, 9] requires additional quantum resources such as ancilla qubits, as well as measuring and extrapolating Renyi entropies. The quantum imaginary-time evolution [10] relies on exponentially difficult tomography on a growing number of qubits and synthesizes of general multi-qubit unitaries.

Recently, references [11, 12] proposed practical approaches to prepare the thermal density matrix as a classical probabilistic mixture of quantum pure states in the eigenbasis of the density matrix. In these proposals, the classical part is either assumed to be a factorized probability distribution or as an energy-based model [13]. However, the factorized distribution is generally a crude approximation for the Gibbs distribution in the eigenbasis. While the energy-based model still faces the problem of the intractable

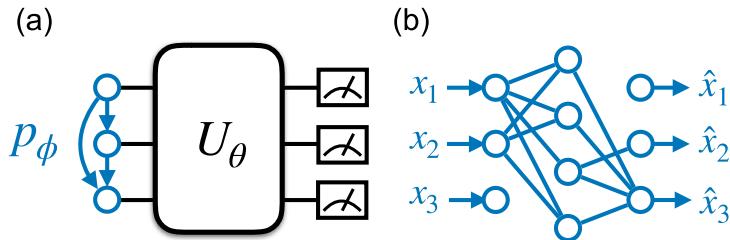


Figure 1. (a) The autoregressive network shown in blue is a classical probabilistic model that parametrizes a joint distribution in the form of equation (2). The model generates bit string as easy to prepare input product states to the quantum circuit. The neural network and the circuit produce a parametrized density matrix equation (1). (b) An implementation of the autoregressive model p_ϕ using the masked autoencoder [14]. The neural network maps bit strings to real-valued outputs which parametrizes the conditional probabilities in equation (2).

partition function, which inhibits efficient and unbiased sampling, learning, or even evaluating the model likelihood.

Modern probabilistic generative models offer solutions to these problems about expressibility and tractability [15] since the goals of generative modeling are exactly to represent, learn and sample from complex high-dimensional probability distributions efficiently. Popular generative models include autoregressive models [14, 16, 17], variational autoencoders [18], generative adversarial networks [19], and flow-based models [20]. For the purpose of this study, the autoregressive models stand out since they support unbiased gradient estimator for discrete variables, direct sampling, and tractable likelihood *at the same time*. The autoregressive models have reached state-of-the-art performance in modeling realistic data and found real-world applications in synthesizing natural speech and images [16, 17]. Variational optimization of the autoregressive network has been used for classical statistical physics problems [21, 22]. Quantum generalization of the network was also employed for the ground state of quantum many-body systems [23].

In this paper, we combine quantum circuits with autoregressive probabilistic models to study thermal properties of quantum systems at finite temperature. The resulting model allows one to perform variational free energy calculation over density matrices efficiently. We demonstrate applications of the approach to quantum Ising model on a square lattice, where we obtain its variational free energy, thermal entropy, specific heat and excitation energies.

By leveraging the recent advances in deep probabilistic generative models, the proposed approach extends the variational quantum eigensolver (VQE) [24] to thermal quantum states with essentially no overhead. Thus, the present algorithm is also feasible for near-term quantum computers [25–31]. The only practical difference to the ground state VQE is that one needs to sample input states to the quantum circuit from a classical distribution, and one has an additional term in the objective function to account for the entropy of the input distribution.

For the classical simulation of the proposed algorithm, we use Yao.jl, an extensible and efficient framework for quantum algorithm design [32]. Yao.jl's batched quantum register and automatic differentiation via reversible computing [33] makes it an ideal tool for differentiable programming models which combine classical neural networks and quantum circuits. Our code implementation can be found at [34].

2. Model architecture, objective function, and optimization scheme

Figure 1(a) shows the architecture of the variational ansatz. A classical probabilistic model generates binary random variables x according to a classical distribution $p_\phi(x)$, where ϕ are the network parameters. It is straightforward to prepare qubits to the classical product state $|x\rangle$. Then, a parametrized quantum circuit performs unitary transformation to the input states $U_\theta|x\rangle$, where the circuit parameters θ do not depend on the inputs. Overall, the model produces a classical mixture of quantum states. The density matrix of the ensemble reads [11, 12]

$$\rho = \sum_x p_\phi(x) U_\theta |x\rangle \langle x| U_\theta^\dagger. \quad (1)$$

The density matrix is hermitian and positive definite. Moreover, given a normalized classical probability, one has $\text{Tr}(\rho) = \sum_x p_\phi(x) = 1$. The density matrix depends both on parameters ϕ and θ . We omit the explicit dependence in the notation to avoid cluttering in the notations.

The parametrized quantum circuit performs a unitary transformation to the diagonal density matrix $\sum_x p_\phi(x)|x\rangle\langle x|$, whose diagonal elements are parametrized by a neural network. Using a quantum circuit for the unitary transformation [35] is more general than the classical flow model [36]. Moreover, it automatically ensures physical constraints such as the orthogonality of the eigenstates. The classical distribution $p_\phi(x)$ is in general nontrivial since it is not necessarily factorized for each dimension of x [11, 12]. Thus, exact representation of the classical distribution on the eigenbasis $p_\phi(x)$ may also incur exponential resources. Parametrizing the probability distribution using a classical Boltzmann distribution has the problem of intractable partition functions. Hence, we employ an autoregressive network to produce the input states of the quantum circuit.

The autoregressive network models the joint probability distribution as a product of conditional probabilities

$$p_\phi(x) = \prod_i p_\phi(x_i|x_{<i}) = p_\phi(x_1)p_\phi(x_2|x_1)p_\phi(x_3|x_1, x_2)\dots, \quad (2)$$

where one has assumed an order of each dimension of the variables. $x_{<i}$ denotes the set of variables that are before x_i . The autoregressive network is a special form of Bayesian network, which models conditional dependence of random variables as a directed acyclic graph shown in figure 1(a). The model can capture high-dimensional multimode distribution with complex correlations. One can also directly draw uncorrelated samples from the joint distribution via ancestral sampling, which follows the order of the conditional probabilities.

The practical implementation of the autoregressive networks largely benefits from rapid development of deep learning architectures such as the recurrent or convolutional neural networks [16, 17] and autoencoders [14]. In this paper, we employ the masked autoencoder shown in figure 1(b). The autoencoder network transforms bit string x to real-valued vector \hat{x} of the same dimension, where each element satisfies $0 < \hat{x}_i < 1$, e.g. outputs of sigmoid activation functions [13]. We mask out some connections in the autoencoder network so the connectivity ensures that \hat{x}_i only depends on the binary variable $x_{<i}$. Thus, each element of the output defines a conditional Bernoulli distribution $p_\phi(x_i|x_{<i}) = \hat{x}_i^{x_i}(1-\hat{x}_i)^{1-x_i}$ for the binary variable x_i . In this way, the joint probability for all binary variables satisfies the autoregressive property equation (2). Since each conditional probability is normalized $\sum_{x_i} p_\phi(x_i|x_{<i}) = 1$, the joint distribution is normalized by construction. The probability distribution is parameterized by the network parameters ϕ . In a simple limit where the network is disconnected, $\hat{x} = \text{sigmoid}(\phi)$ and one restores the product state ansatz considered in references [11, 12].

Given a Hamiltonian H at inverse temperature β , the density matrix $\sigma = e^{-\beta H}/Z$ plays a central role in the quantum statistical mechanics problem, where $Z = \text{Tr}(e^{-\beta H})$ is an intractable partition function. One can perform the variational calculation over the parametrized density matrix equation (1) by minimizing the objective function

$$\mathcal{L} = \text{Tr}(\rho \ln \rho) + \beta \text{Tr}(\rho H), \quad (3)$$

which follows the Gibbs–Delbrück–Molière variational principle of quantum statistical mechanics [37]. The two terms of equation (3) correspond to the entropy and the expected energy of the variational density matrix respectively. The objective function is related to the quantum relative entropy $S(\rho\|\sigma) = \mathcal{L} + \ln Z$ between the variational and the target density matrices. Since the relative entropy is non-negative [38], one has $\mathcal{L} \geq -\ln Z$, i.e. the loss function is lower bounded by the physical free energy. The equality is reached only when the variational density matrix reaches the physical one $\rho = \sigma$.

To estimate the objective function equation (3), one can sample a batch of input states $|x\rangle$ from the autoregressive network, then apply the parametrized circuit and measure the following estimator:

$$\mathcal{L} = \mathbb{E}_{x \sim p_\phi(x)} \left[\ln p_\phi(x) + \beta \left\langle x \left| U_\theta^\dagger H U_\theta \right| x \right\rangle \right]. \quad (4)$$

The first term depends solely on the classical probabilistic model, which can be directly computed via equation (2) on the samples. Note that the entropy of the autoregressive model is known exactly rather than being intractable in the energy-based models [12]. Moreover, having direct access to the entropy avoids the difficulties of extrapolating the Renyi entropies measured on the quantum processor [8, 9]. The second term of equation (4) involves the expected energy of Hamiltonian operators $\langle H \rangle$, where we denote $\langle O \rangle = \mathbb{E}_{x \sim p_\phi(x)} \left[\left\langle x \left| U_\theta^\dagger O U_\theta \right| x \right\rangle \right]$. The classical neural network and the quantum circuit perform classical and quantum average respectively. The equation (4) shows zeros variance property, i.e. when the variational density matrix exactly reaches to the physical one, the variance of the estimator equation (4) reduces to zero. This can be used as a self-verification of the variational ansatz and minimization procedure [30].

We would like to utilize the gradient information to train the hybrid model which consists of neural networks and quantum circuits efficiently. Moreover, random sampling of the autoregressive net and the quantum circuit suggest that one should employ stochastic optimization with noisy gradient estimators [13]. First, the gradient with respect to the circuit parameters reads

$$\nabla_{\theta} \mathcal{L} = \beta \mathbb{E}_{x \sim p_{\phi}} \left[\nabla_{\theta} \langle x | U_{\theta}^{\dagger} H U_{\theta} | x \rangle \right]. \quad (5)$$

The term inside the square bracket is a gradient of a quantum expected value. To evaluate the expectation on an actual quantum device, one can employ the parameter shift rule of [39–42]. These approaches estimate the gradient of each circuit parameter using the difference of two sets of measurement on the quantum circuit with the same architecture. While in the classical simulation of the quantum algorithm one can employ the automatic differentiation [43] to evaluate the gradient efficiently.

Next, the gradients of the neural network parameters can be evaluated using the REINFORCE algorithm [45]

$$\nabla_{\phi} \mathcal{L} = \mathbb{E}_{x \sim p_{\phi}} [(f(x) - b) \nabla_{\phi} \ln p_{\phi}(x)], \quad (6)$$

where the term $\nabla_{\phi} \ln p_{\phi}(x)$, known as the score-function gradient in the machine learning literature [46], can be efficiently evaluated via backpropagation through the probabilistic model equation (2) [43]. In this regard, $f(x) = \ln p_{\phi}(x) + \beta \langle x | U_{\theta}^{\dagger} H U_{\theta} | x \rangle$ can be viewed as the ‘reward signal’ given the policy $p_{\phi}(x)$ for generating bit string samples. We have introduced the baseline $b = \mathbb{E}_{x \sim p_{\phi}} [f(x)]$ which does not affect the expectation of equation (6) since $\mathbb{E}_{x \sim p_{\phi}} [\nabla_{\phi} \ln p_{\phi}(x)] = 0$. However, the baseline helps to reduce the variance of the gradient estimator [47].

Given the gradient information we train the autoregressive network and the quantum circuit jointly with the stochastic gradient descend method. The training procedure finds out the circuit U_{θ} which approximately diagonalizes the density matrix and brings the negative log-likelihood $-\ln p_{\phi}(x)$ closer to the energy spectrum of the system. In principle, the same circuit can diagonalize the density matrices at all temperatures if U_{θ} fully diagonalize the Hamiltonian. However, in the practical variational calculation, this does not need to be the case to achieve good variational free energy since the temperature selects the relevant low-energy spectra which contributes mostly to the objective function.

After training, one can sample a batch of input states $|x\rangle$ and treat them as approximations of the eigenstates of the system. Since the unitary circuit preserves orthogonality of the input states, the sampled quantum states span a low energy subspace of the Hamiltonian. For example, measuring the expected energy $\langle x | U_{\theta}^{\dagger} H U_{\theta} | x \rangle$ reveals the excitation energies of the system. In this respect, the objective function equation (3) is related to the weighted subspace-search VQE algorithm for the excited states [48]. Different from the weighted subspace-search VQE, a single physical parameter inverse temperature β controls the relative weights on the input states. Adaptive sampling of the autoregressive model provides the correct weights that spans the relevant low energy space. Due to its close connection to the original VQE algorithm [24], we denote the present approach as the β -VQE algorithm. While suppose the Hamiltonian is diagonal in the computational basis, i.e. a classical Hamiltonian, one can leave out the quantum circuit and the approach falls back to the variational autoregressive network approach of reference [21]. In the classical limit it is also obvious that the autoregressive ansatz is advantageous than a simple product ansatz.

3. Numerical simulations

We demonstrate application of β -VQE to thermal properties of quantum lattice problems. Despite that most of the efforts on VQE have been devoted to quantum chemistry problems [25–29], quantum lattice problems are more native applications on near-term quantum computers for two reasons. First, typical problems with local interaction one does not suffer from unfavorable scaling of a large number of Hamiltonian terms. Second, quantum lattice models that only involve spins and bosons do not invoke the overhead of mapping from fermion to qubits. Therefore, it is anticipated that near-term devices should already produce valuable results for quantum lattice problems before they are impactful for quantum chemistry problems [49].

We consider the prototypical transverse field Ising model on a square lattice with open boundary conditions:

$$H = - \sum_{\langle i,j \rangle} Z_i Z_j - \Gamma \sum_i X_i, \quad (7)$$

where Z_i and X_i are Pauli operators acting on the lattice sites. The model exhibits a quantum critical point at zero temperature at $\Gamma_c = 3.044\ 38(2)$. While for $\Gamma < \Gamma_c$ the model exhibits a thermal phase transition from a

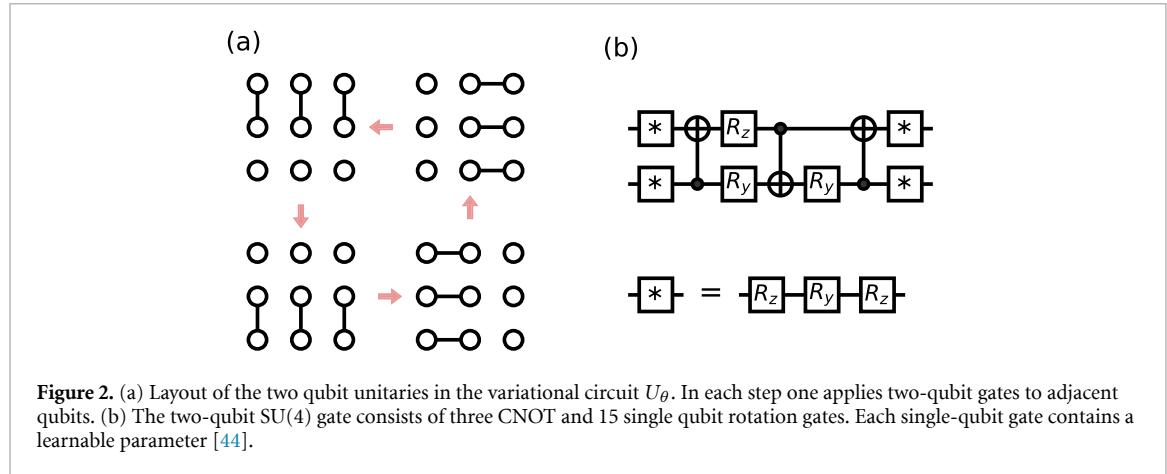


Figure 2. (a) Layout of the two qubit unitaries in the variational circuit U_θ . In each step one applies two-qubit gates to adjacent qubits. (b) The two-qubit SU(4) gate consists of three CNOT and 15 single qubit rotation gates. Each single-qubit gate contains a learnable parameter [44].

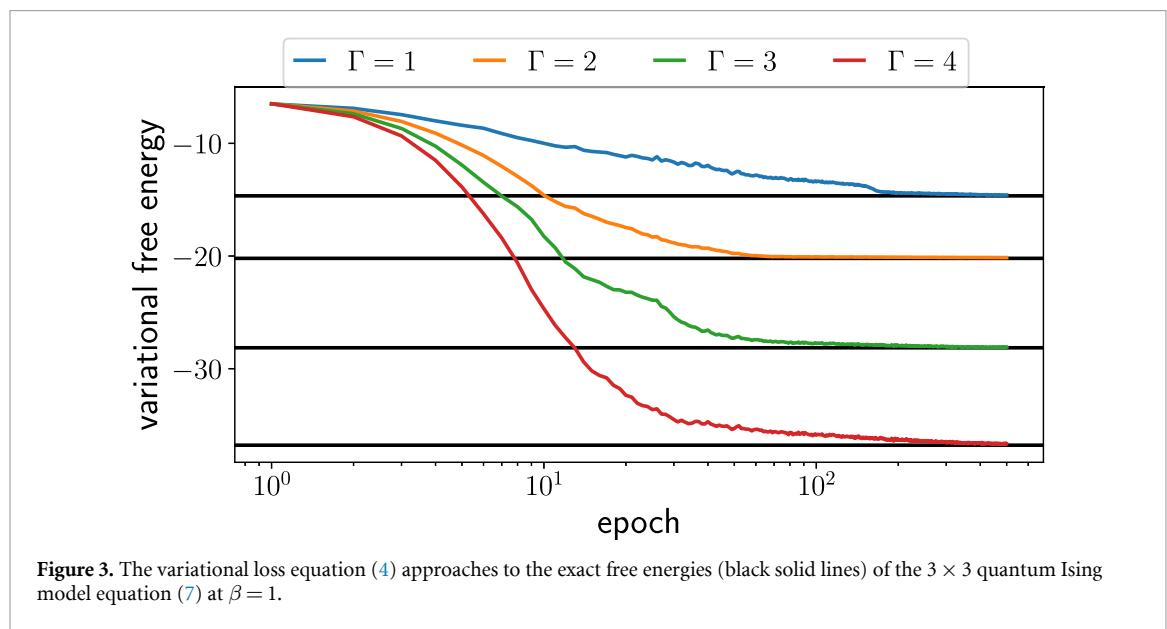


Figure 3. The variational loss equation (4) approaches to the exact free energies (black solid lines) of the 3×3 quantum Ising model equation (7) at $\beta = 1$.

ferromagnetic phase to a disordered phase. All of these rich physics can be studied unbiasedly with sign-problem free quantum Monte Carlo approach, e.g. see [50]. Having abundant established knowledge makes the problem equation (7) an ideal benchmark problem for the β -VQE algorithm on near-term quantum computers.

For the autoregressive network equation (2) we employ the masked autoencoder architecture [14] shown in figure 1(b). We arrange the qubits on the two dimensional grid following the typewriter order. The autoencoder has a single hidden layer of 500 hidden neurons with rectified linear unit activation. For the variational quantum circuit, we employ the setup shown in figure 2, which arranges the qubits on a two dimensional grid [51] and apply general two qubit gates [44] on the neighboring sites in each layer. The general gate consists of 15 single-qubit gates and three CNOT gates. Each two qubit unitary is parametrized by 15 parameters in the rotational gates, which parametrizes the SU(4) group. The circuit architecture enjoys a balanced expressibility and hardware efficiency. We repeat the pattern for d times, which we denote as the depth d of the variational quantum circuit. Therefore, for the 3×3 system considered in figure 2(a) with $d = 5$, there are $15 \times 12 \times 5 = 900$ circuit parameters. Initially we set all the circuit parameters to be zero. We estimate the gradients equations (5) and (6) on batch of 1000 samples, and we use the Adam algorithms [13] to optimize the parameters ϕ and θ jointly.

Figure 3 shows that the objective function decreases towards the exact values as a function of training epochs. We measure physical observables on the trained model and compare them with exact results. For example, figures 4(a) and (b) show the expected energy $\langle H \rangle$ and the specific heat $\beta^2 (\langle H^2 \rangle - \langle H \rangle^2)$ computed by measuring Hamiltonian expectation and its variance. Moreover, one sees in figure 4(c) that the entropy $\mathbb{E}_{x \sim p_\phi(x)} [-\ln p_\phi(x)]$ changes from $\ln 2$ per site in the high temperature limit towards zero at zero temperature. While the purity of the system $\text{Tr}(\rho^2) = \mathbb{E}_{x \sim p_\phi(x)} [p_\phi(x)]$ shown in figure 4(d) increases from zero towards one as the temperature decreases. All these observables can be directly measured on an actual

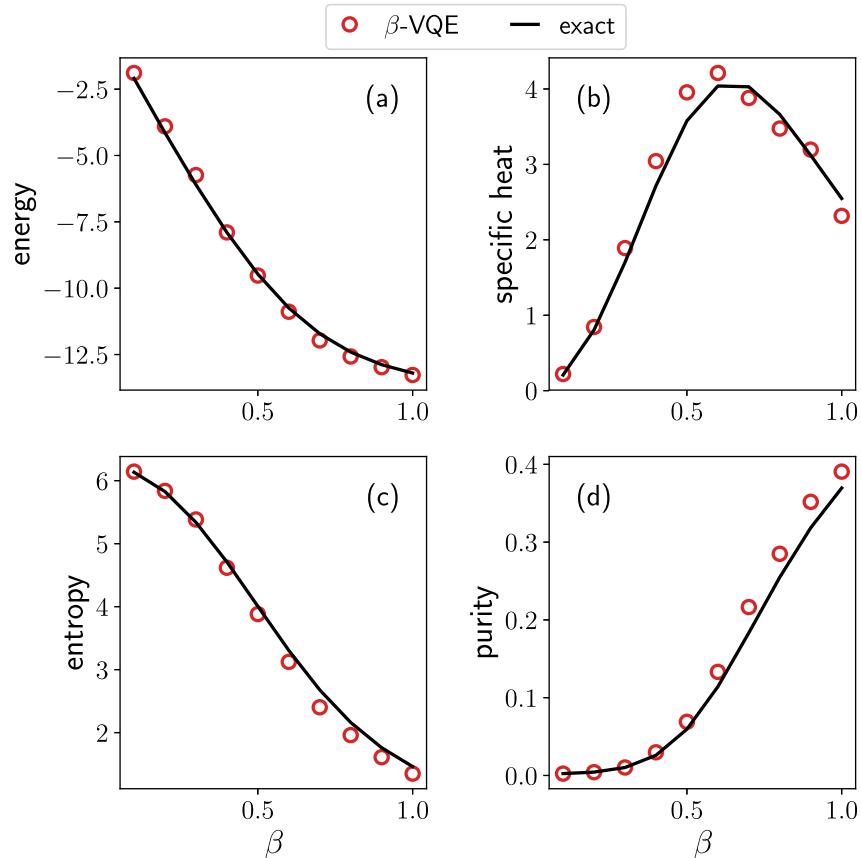


Figure 4. Physical quantities obtained via the β -VQE algorithm for the 3×3 quantum Ising model equation (7) at $\Gamma = 3$. The black solid lines are exact solutions.

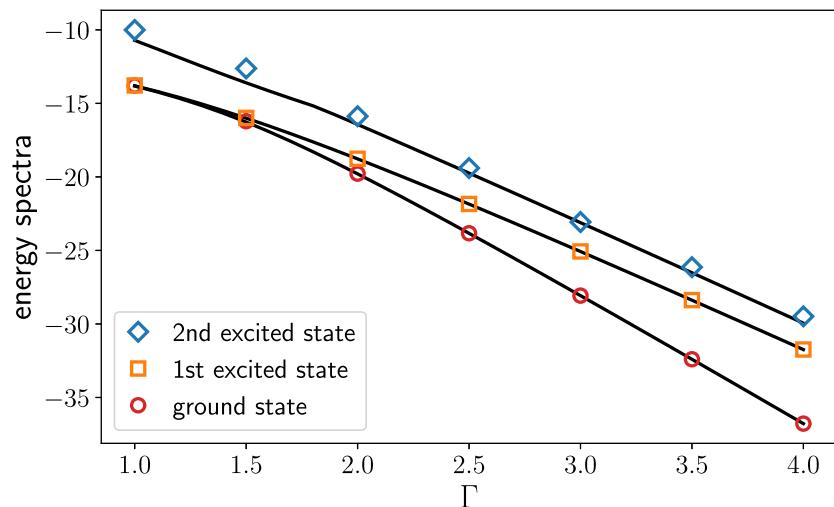


Figure 5. Low energy excitation spectra of 3×3 quantum Ising model. The symbols denote energy expectations $\langle x | U_\theta^\dagger H U_\theta | x \rangle$ measured on the samples of the autoregressive model trained at inverse temperature $\beta = 0.5$. The solid lines are exact solutions.

quantum device. Overall, one sees the autoregressive model equation (2) combined with variational quantum circuit yields accurate results over all temperatures. Note that in both the simulation and the actual experiment there will be statistical error bars due to random sampling of the autoregressive model. However, given the Hilbert space size ($2^9 = 512$) and the batch size used in the calculation (1000), the remaining deviation from the exact results is due to limits of representability and optimization, but not the statistical noises.

Figure 5 shows the low energy spectrum of the quantum Ising model obtained from β -VQE at $\beta = 0.5$. One sees that the approach provides low energy spectrum of the problem at various strength of the

transverse field. The approach works nicely even when the first excited state becomes nearly degenerated with the ground state.

4. Outlooks

The present approach would be most useful for studying the thermal properties of frustrated quantum systems which are prevented by the sign problem [52]. Moreover, one can further employ the qubit efficient VQE scheme [53, 54], where one can study the thermal properties of quantum many-body systems on a quantum computer with the number of qubits smaller than problem size. In that scenario, the ansatz for the density matrix is a classical mixture of matrix product states. The variational ansatz for density matrix can also be used in a quantum algorithm for non-equilibrium dynamics [55] and steady states [56].

The quantum circuit also acts as a canonical transformation that brings the density matrix to a diagonal representation. Combined with the fact that one can obtain the marginal likelihood of the leading bits in the autoregressive models, the setup may be useful for deriving effective models with less degrees of freedom similar to the classical case [57]. Therefore, one can envision using the present setup to derive effective models by using a quantum circuit for renormalization group transformation. Moreover, since the circuit approximately diagonalizes the density matrix, one also can make use of it for later purposes, such as accelerated time evolution [58].

Regarding further improvements of the algorithm, one may consider using tensor network probabilistic models [59, 60] instead of the autoregressive network to represent the classical distribution in the eigenbasis. Both models have the shortcoming that the sampling approach produces the bits sequentially. To address this issue, one may consider employing the recently proposed flow models for discrete variables [61, 62]. While to further improve the optimization efficiency, one may consider using the improved gradient estimator with even lower variances [63, 64]. To this end, differentiable programming of neural networks and quantum circuits shares a unified computational framework. Therefore, a seamless integration of models and techniques will enjoy advances of both worlds.

Acknowledgment

We thank useful discussion with Nobuyuki Yoshioka and Yuya O Nakagawa. We thank Xiu-Zhe Luo for contribution to Yao. *j1* [32] and Dian Wu for collaboration in reference [21]. The authors are supported by the National Natural Science Foundation of China under Grant No. 11774398. P Z is supported by the Key Research Program of Frontier Sciences, CAS, Grant No. QYZDB-SSW-SYS032, and Project Nos. 11 747 601 and 11 975 294 of National Natural Science Foundation of China.

All data that support the findings of this study are included within the article (and any supplementary information files).

ORCID iD

Jin-Guo Liu  <https://orcid.org/0000-0003-1635-2679>

References

- [1] Van Houcke K *et al* 2012 *Nat. Phys.* **8** 366–70
- [2] King A D *et al* 2018 *Nature* **560** 456–60
- [3] Terhal B M and DiVincenzo D P 2000 *Phys. Rev. A* **61** 022301
- [4] Poulin D and Wocjan P 2009 *Phys. Rev. Lett.* **103** 220502
- [5] Temme K, Osborne T J, Vollbrecht K G, Poulin D and Verstraete F 2011 *Nature* **471** 87
- [6] Riera A, Gogolin C and Eisert J 2012 *Phys. Rev. Lett.* **108** 080402
- [7] Brandao F G and Kastoryano M J 2019 *Commun. Math. Phys.* **365** 1
- [8] Wu J and Hsieh T H 2018 *Phys. Rev. Lett.* **123** 220502
- [9] Zhu D, Johri S, Linke N M, Landsman K A, Nguyen N H, Alderete C H, Matsuura A Y, Hsieh T H and Monroe C 2019 (arXiv:1906.02699)
- [10] Motta M, Sun C, Tan A T K, O'rourke M J, Ye E, Minnich A J, Brandao F G S L and Chan G K L 2019 (arXiv:1901.07653)
- [11] Martyn J and Swingle B 2019 *Phys. Rev. A* **100** 032107
- [12] Verdon G, Marks J, Nanda S, Leichenauer S and Hidary J 2019 (arXiv:1910.02071)
- [13] Goodfellow I, Bengio Y and Courville A 2016 *Deep Learning* (Cambridge, MA: MIT Press) (www.deeplearningbook.org)
- [14] Germain M, Gregor K, Murray I and Larochelle H 2015 (arXiv:1502.03509)
- [15] Wang L 2018 (<http://wangleiphy.github.io/lectures/PILtutorial.pdf>)
- [16] van den Oord A, Dieleman S, Zen H, Simonyan K, Vinyals O, Graves A, Kalchbrenner N, Senior A and Kavukcuoglu K 2016 (arXiv:1609.03499)
- [17] van den Oord A, Kalchbrenner N and Kavukcuoglu K 2016 (arXiv:1601.06759)
- [18] Kingma D P and Welling M 2013 (arXiv:1312.6114)

[19] Goodfellow I J, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A and Bengio Y 2014 (arXiv:[1406.2661](https://arxiv.org/abs/1406.2661))

[20] Kobyzev I, Prince S and Brubaker M A 2019 *IEEE Trans. Pattern Anal. Mach. Intelligence* **2020**

[21] Wu D, Wang L and Zhang P 2019 *Phys. Rev. Lett.* **122** 080602

[22] Pan F, Zhou H J and Zhang P 2019 (arXiv:[1906.10935](https://arxiv.org/abs/1906.10935))

[23] Sharir O, Levine Y, Wies N, Carleo G and Shashua A 2020 *Phys. Rev. Lett.* **124** 020503

[24] Peruzzo A, McClean J, Shadbolt P, Yung M H, Zhou X Q, Love P J, Aspuru-Guzik A and O'Brien J L 2014 *Nat. Commun.* **5** 4213

[25] Shen Y, Zhang X, Zhang S, Zhang J N, Yung M H and Kim K 2017 *Phys. Rev. A* **95** 020501

[26] O'Malley P J *et al* 2016 *Phys. Rev. X* **6** 031007

[27] Kandala A, Mezzacapo A, Temme K, Takita M, Brink M, Chow J M and Gambetta J M 2017 *Nature* **549** 242–6

[28] Colless J I, Ramasesh V V, Dahlen D, Blok M S, Kimchi-Schwartz M E, McClean J R, Carter J, De Jong W A and Siddiqi I 2018 *Phys. Rev. X* **8** 011021

[29] Hempel C *et al* 2018 *Phys. Rev. X* **8** 31022

[30] Kokail C *et al* 2019 *Nature* **569** 355

[31] Mazzola G, Ollitrault P J, Barkoutsos P K and Tavernelli I 2019 *Phys. Rev. Lett.* **123** 130501

[32] Luo R *et al* 2021 QuantumBFS (<https://github.com/QuantumBFS/Yao.jl>)

[33] Luo X Z, Liu J G, Zhang P and Wang L 2019 (arXiv:[1912.10877](https://arxiv.org/abs/1912.10877))

[34] Wang L 2019 Solving quantum statistical mechanics with variational autoregressive networks and quantum circuits (<https://github.com/wangleiphy/BetaVQE.jl>)

[35] Verstraete F, Cirac J I and Latorre J I 2009 *Phys. Rev. A* **79** 032316

[36] Cranmer K, Golkar S and Pappadopulo D 2019 (arXiv:[1904.05903](https://arxiv.org/abs/1904.05903))

[37] Huber A 1968 Variational principles in quantum statistical mechanics *Mathematical Methods in Solid State and Superfluid Theory* (Berlin: Springer) pp 364–92

[38] Nielsen M A and Chuang I L 2011 *Quantum Computation and Quantum Information: 10th Anniversary edition* 10th edn (New York, USA: Cambridge University Press)

[39] Li J, Yang X, Peng X and Sun C P 2017 *Phys. Rev. Lett.* **118** 150503

[40] Mitarai K, Negoro M, Kitagawa M and Fujii K 2018 *Phys. Rev. A* **98** 032309

[41] Schuld M, Bergholm V, Gogolin C, Izrae J and Killoran N 2019 *Phys. Rev. A* **99** 032331

[42] Nakanishi K M, Fujii K and Todo S 2019 (arXiv:[1903.12166](https://arxiv.org/abs/1903.12166))

[43] Baydin A G, Pearlmutter B A, Radul A A and Siskind J M 2015 *J. Mach. Learn.* **18** 1–43 (<http://jmlr.org/papers/v18/16-107.html>)

[44] Shende V V, Markov I L and Bullock S S 2004 *Phys. Rev. A* **69** 062321

[45] Williams R J 1992 *Mach. Learn.* **8** 229

[46] Mohamed S, Rosca M, Figurnov M and Mnih A 2019 (arXiv:[1906.10652](https://arxiv.org/abs/1906.10652))

[47] Mnih A and Gregor K 2014 (arXiv:[1402.0030](https://arxiv.org/abs/1402.0030))

[48] Nakanishi K M, Mitarai K and Fujii K 2018 (arXiv:[1810.09434](https://arxiv.org/abs/1810.09434))

[49] Wecker D, Hastings M B and Troyer M 2015 *Phys. Rev. A* **92** 042303

[50] Hesselmann S and Wessel S 2016 *Phys. Rev. B* **93** 155157

[51] Arute F *et al* 2019 *Nature* **574** 505

[52] Troyer M and Wiese U J 2005 *Phys. Rev. Lett.* **94** 170201

[53] Huggins W, Patel P, Whaley K B and Stoudenmire E M 2019 *Quantum Sci. Technol.* **4** 024001

[54] Liu J G, Zhang Y H, Wan Y and Wang L 2019 *Phys. Rev. Res.* **1** 23025

[55] Lamm H and Lawrence S 2018 *Phys. Rev. Lett.* **121** 170501

[56] Yoshioka N, Nakagawa Y O, Mitarai K and Fujii K 2019 (arXiv:[1908.09836](https://arxiv.org/abs/1908.09836))

[57] Li S H, Dong C X, Zhang L and Wang L 2019 (arXiv:[1910.00024](https://arxiv.org/abs/1910.00024))

[58] Cirstoiu C, Holmes Z, Iosue J, Cincio L, Coles P J and Sornborger A 2019 *npj Quantum Information* **6** 82

[59] Han Z Y, Wang J, Fan H, Wang L and Zhang P 2018 *Phys. Rev. X* **8** 031012

[60] Cheng S, Wang L, Xiang T and Zhang P 2019 *Phys. Rev. B* **99** 155131

[61] Tran D, Vafa K, Agrawal K K, Dinh L and Poole B 2019 (arXiv:[1905.10347](https://arxiv.org/abs/1905.10347))

[62] Hoogeboom E, Peters J W T, van den Berg R and Welling M 2019 (arXiv:[1905.07376](https://arxiv.org/abs/1905.07376))

[63] Tucker G, Mnih A, Maddison C J, Lawson D and Sohl-Dickstein J 2017 (arXiv:[1703.07370](https://arxiv.org/abs/1703.07370))

[64] Grathwohl W, Choi D, Wu Y, Roeder G and Duvenaud D 2017 (arXiv:[1711.00123](https://arxiv.org/abs/1711.00123))