Identifying quantum phases via disentanglement based on deep reinforcement learning

Zheng An,1, 2 Chenfeng Cao,3 and D. L. Zhou1, 2, 4, 5, *

1 Institute of Physics, Beijing National Laboratory for Condensed Matter Physics, Chinese Academy of Sciences, Beijing 100190, China
2 School of Physical Sciences, University of Chinese Academy of Sciences, Beijing 100049, China
3 Department of Physics, The Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, China
4 Collaborative Innovation Center of Quantum Matter, Beijing 100190, China
5 Songshan Lake Materials Laboratory, Dongguan, Guangdong 523808, China

(Dated: August 5, 2021)

Identifying phases of matter is a complicated process, especially in quantum theory, where the complexity of the ground state appears to rise exponentially with system size. Traditionally, physicists have been responsible for identifying suitable order parameters for the identification of the various phases. The entanglement of quantum many-body systems exhibits rich structures and can be determined over the phase diagram. The intriguing question of the relationship between entanglement and quantum phase transition has recently been addressed. As a method that works directly on the entanglement structure, disentanglement can provide factual information on the entanglement structure. In this work, we follow a radically different approach to identifying quantum phases: we utilize reinforcement learning (RL) approaches to develop an efficient variational quantum circuit that disentangles the ground-state of Ising spin chain systems. We show that the specified quantum circuit structure of the tested models correlates to a phase transition in the behaviour of the entanglement. We show a similar universal quantum circuit structure for ground states in the same phase to reduce entanglement entropy. This study sheds light on characterizing quantum phases with the types of entanglement structures of the ground states.

**Introduction.**—Entanglement is an important quantum mechanical phenomenon where the many-body quantum states cannot be described separately [1]. Motivated by the development of quantum resource theory and quantum computation, entanglement has become a fundamental resource in quantum technologies such as quantum teleportation, quantum cryptography, and quantum metrology [2–6]. The manipulation of entanglement, especially the destruction of entanglement via general operations, is still under study [7, 8].

A phase diagram depicts qualitative changes in many-body systems as a function of physical system parameters. The objective of a classifier is to determine the appropriate order parameters. For example, in the Landau theory of phase transitions, a discontinuity of the local order parameter or of one of its derivatives indicates a phase transition. In more exotic systems, the order parameters are global as it is the case for topological phases or topological insulators. Recently, it was shown that quantum phase transition and topological orders can be characterized by entanglement [9–12]. So it is believed that the entanglement of the ground state plays a crucial role in the understanding of quantum phase transition.

Disentanglement is a process in which an initial entangled state of a composite system is transformed into a separable state without affecting the reduced density matrices of the subsystems. However, a universal disentangling machine does not exist [13–15]. Nevertheless, dropping the constraint that the reduced density matrices of subsystems are perfectly unaffected, approximate disentanglement can be realized by local operations [16–18]. The central idea of disentanglement is to decouple the correlations among the subsystems. Thus, it is natural for us to investigate whether a unitary transformation realized by a quantum circuit, can destroy the entanglement within a quantum state. Further, we want to investigate the relationship between the entanglement structure and the circuit structure, especially for the states in the same phase.

Variational quantum circuit is a powerful tool to approximate a known or unknown unitary evolution, relevant hybrid quantum-classical algorithms have been demonstrated experimentally on noisy intermediate-scale quantum (NISQ) devices [19–21]. Generally, we can encode any optimization problem, e.g., MaxCut, to a specific Hamiltonian with the ground state being the solution [22], then optimize a parameterized quantum circuit to find the ground state and ground state energy. Variational quantum algorithms have been designed for a wide range of problems, such as quantum data compression [23–25], quantum fidelity estimation [26, 27], state diagonalization [28, 29], Hamiltonian diagonalization [30], etc.

The main challenge of designing a desire quantum circuit is to discover suitable sequences of unitary gates; the size of the search space and the lack of simple strategies in many situations make this a challenging problem. Reinforcement learning (RL) is an area of machine learning concerned with finding optimal strategies for complex decision-making problems [31]. RL has recently been employed in broad applications, such as video games [32, 33], board games [34–36], and neural architecture search [37]. Primarily, it has been widely applied to a wide array of physics problems, such as quantum control [38–40],...
quantum error correction [41], and quantum circuit optimization [42–44]. Also, the quantum machine learning, especially quantum reinforcement learning, may produce quantum advantages with machine learning [45–47]. Those successes naturally raise the question of how much quantum circuit design might benefit from reinforcement learning.

In our paper, we discuss a novel method of classifying phase transitions, i.e., design the disentanglement ansatz of the ground states to identify different phases of Ising spin chain systems. To this end, we apply deep reinforcement learning to variational quantum circuit design. To explore the optimal strategy of the problem in this scenario, we use the Deep Q-network (DQN) [32] to study this problem.

**Methodology.**—The reinforcement learning process is a finite Markov decision process [31]. As shown in Fig. 1, this Markov process is described as a state-action-reward sequence, a state $S_t$ at time $t$ is transmitted into a new state $S_{t+1}$ together with giving a scalar reward $R_{t+1}$ at time $t+1$ by the action $A_t$ with the transmission probability $p(S_{t+1}; R_{t+1}|S_t; A_t)$. For a finite Markov decision process, the sets of the states, the actions and the rewards are finite. The total discounted return at time $t$ is

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}, \quad (1)$$

where $\gamma$ is the discount rate and $0 \leq \gamma \leq 1$.

![Figure 1: A schematic diagram of Markov decision process.](image)

The goal of an RL algorithm is to maximize the total discounted return for each state and action selected by the policy $\pi$, which is specified by a conditional probability of selecting an action $A$ for each state $S$, denoted as $\pi(A|S)$.

In this work, the state space of the RL environment is the space of quantum circuit structure. In the modified input scheme of the learning stage, the agent observes the entire state space (Fig. 2), and we, therefore, deal with a fully observed Markov process. As we described above, the action set is defined by the available hardware instructions (unitary gates).

In particular, we consider the gate set consisting of single-qubit rotation (along x, y, or z-direction), Hadamard, and Controlled-Not (CNOT) gates. Together, they form a universal gate set. Whereas continuous variables parameterize rotation gates, respectively, the CNOT and Hadamard are fixed gates. All of the parameters of single-qubit rotation gates are initialized to $\pi$.

The reward of the agent received in each step is

$$R_t = \begin{cases} 0, & t \in \{0,1,\ldots,N-1\} \\ \mathcal{R}, & t = N \end{cases}, \quad (2)$$

where $\mathcal{R}$ is the modified reward function for our learning task. The agent will get a reward not immediately, but at time $N$. In particular, we define $\mathcal{R}$ as

$$\mathcal{R} = \begin{cases} 0, & S_{RL} \leq S_0 \\ S_0 - S_{RL}, & \text{otherwise} \end{cases}, \quad (3)$$

where $S_{RL}$ is the final result entanglement entropy given by our RL-assistant quantum algorithm, and $S_0$ is the ground-state entanglement entropy given by the physical systems.

Now, the whole frame of our learning task is introduced in Fig. 3. For the classical optimizer, we employ the BFGS algorithm implemented in SciPy [48].

In order to measure how much entanglement in the ground state of quantum systems, we use the entanglement entropy $S$. When a bipartite quantum system $AB$ is in a pure state, the entanglement entropy is calculated from the reduced density matrix $\rho_A$ or $\rho_B$ according to the formula

$$S \equiv -\text{tr}(\rho_A \log_2 \rho_A) = -\text{tr}(\rho_B \log_2 \rho_B). \quad (4)$$

When either subsystem $A$ or $B$ is a spin-\(\frac{1}{2}\) system, $S$ varies from $S = 0$ (product state) to $S = 1$ (maximally entangled state). For the ground state of the two studied models, we regard a half sites as subsystem $A$ and the rest of the sites as subsystem $B$.

**Transverse-field Ising model.**—We first apply our method to one-dimensional transverse-field Ising
The Hamiltonian for the Ising model on a 1D lattice with \( N \) sites in a transverse field is given by

\[
H = - \sum_{j=0}^{N-1} (\lambda \sigma_j^x \sigma_{j+1}^x + \sigma_j^z),
\]

where \( \sigma_j^a \) is the \( a \) component of Pauli matrix (\( a = x, y \) or \( z \)) at site \( j \), and \( \lambda \) is the inverse strength of the external field. We assume the periodic boundary conditions, i.e., the \( N \)th site is identified with the 0th site. As is well known, there is a quantum phase transition at the critical point \( \lambda = 1 \) for the transverse Ising model, which means the entanglement properties of the ground state changes dramatically when crossing this critical point.

When \( \lambda \) approaches zero, the TFIM ground state becomes a product of spins pointing in the positive \( z \) direction,

\[
|\psi_{\text{TFIM}}\rangle_{\lambda \to 0} = |0\rangle^\otimes N,
\]

where \(|0\rangle\) and \(|1\rangle\) denote the single spin state with spin up and with spin down along \( z \)-direction respectively.

In the \( \lambda \to \infty \) limit the ground state becomes the GHZ state [49]:

\[
|\psi_{\text{TFIM}}\rangle_{\lambda \to \infty} = \frac{1}{\sqrt{2}}(|+\rangle^\otimes N + |−\rangle^\otimes N),
\]

where \(|+\rangle\) and \(|−\rangle\) denote the single spin state with spin up and with spin down along \( x \)-direction respectively.

Under the above two limits, the entanglement entropy of the ground state behaves as

\[
S_{\lambda \to 0} = 0,
\]

\[
S_{\lambda \to \infty} = 1.
\]

At the critical point, \( \lambda = 1 \), there is a phase transition in the ground state structure in the thermodynamic limit. The correlation function \( \langle \sigma_i^\alpha \sigma_j^\beta \rangle - \langle \sigma_i^\alpha \rangle \langle \sigma_j^\beta \rangle \) decays polynomially as a function of separation at this point, while for all other values of \( \lambda \), this decay is exponential.
two subsystems, so there has a universal circuit structure with different $\lambda$. Because we only measure the half-sites of the state, disentangling the interaction between the two half-sites is an useful way to minimize entanglement entropy with $p = 1$. This interpretation is identical with Fig. 5 in that the circuit entangles the first half-sites. Notice that the single rotation gate $R$ and Hadamard gate in the circuit have a negligible effect with entanglement entropy.

Figure 5: The circuit structure designed by our RL agent to disentangle TFIM model with layer number $p = 1$. The $R$ gate means a single rotation gate which has different direction with different numbers of $\lambda$.

However, with the layer number getting more extensive than one, the results show different structures of circuit with different regions of $\lambda$. Fig. 6 and Fig. 7 show the designed circuit structures with 2 and 3 layers, respectively. We find that the most significant entanglement is present in the parameter region close to the critical point. This region corresponds with the quantum critical region [50], and this region becomes smaller as the layer number getting large. This is indicated that the designed circuit structure has deep relation with the phase transition region.

**XXZ model.**—The second model we apply our method to is the XXZ model. The Hamiltonian of the antiferromagnetic XXZ model on a 1D lattice with $N$ sites is

$$H = \sum_{j=0}^{N-1} \left( \sigma_j^x \sigma_{j+1}^x + \sigma_j^y \sigma_{j+1}^y + \Delta \sigma_j^z \sigma_{j+1}^z \right), \quad (10)$$

where $\Delta$ is a dimensionless parameter characterizing the anisotropy of the model. The XXZ spin chain has been widely studied [51–55]. For the XXZ model, there exist two different phases at the ground state where $\Delta > 0$, i.e., metallic phase, $0 < \Delta \leq 1$, and antiferromagnetic phase, $\Delta > 1$, which is resulted from that the former is gap-less while the latter is gapful. The critical point of quantum phase transition locates at the isotropic point $\Delta = 1$ which is the Kosterlitz-Thouless quantum phase transitions (KT-QPTs).

In Fig. 8 we show the results of the optimal circuits to disentangle the ground state with different values of $\Delta$. Similar with the TFIM model mentioned above, the performance of the designed circuits with the average entanglement entropy ($\bar{S}^{p=1} \approx 0.450, \bar{S}^{p=2} \approx 0.191$ and $\bar{S}^{p=3} \approx 0.120$) is proportional to the circuit depth $p$.

In Fig. 9 and Fig. 10, we show the results of the optimal circuits to disentangle the ground state with different values of $\Delta$. Similar with the TFIM model mentioned above, the performance of the designed circuits with the average entanglement entropy ($\bar{S}^{p=1} \approx 0.450, \bar{S}^{p=2} \approx 0.191$ and $\bar{S}^{p=3} \approx 0.120$) is proportional to the circuit depth $p$. 

![Diagram](attachment:image.png)
circuits to disentangle the ground state with one and two layers. Both cases have a universal circuit structure for different values of $\Delta$, which differ from the TFIM result where only one layer has the same results. The above situation indicated that the ground-state entanglement structure of the XXZ model might be more complex than the TFIM case.

Nevertheless, unlike the TFIM case, the disentangle circuit entropy shows two regions with different values of $\Delta$. The performance is different between the 3 layer case and the others. Fig. 11 shows the optimal circuits structure with 3 layers. From Fig. 11 we observe interesting differences for disentangle quantum circuit structure at the critical points. The numerical results indicate that the entanglement structure of the ground state of the XXZ model in different values of $\Delta$ differs not only quantitatively but also qualitatively. Also, the result is consistent with what we discussed in former section.

**Conclusions & Outlook.**– In this work, we show a new approach to identifying quantum phase transition. The universal disentanglement circuits found by the method is in very good agreement with the phase diagram. Therefore it allows the characterization of classical, quantum, and topological phase transitions. This phenomenon shows a new perspective to study the re-
We have proposed an RL-based framework to design a particular variational quantum circuit. RL agent is capable of finding a particular target operation by building a circuit gate by gate. For tested models, the RL-designed circuit can successfully disentangle the ground state entanglement. Our results also show that the designed circuit structure correlates with the phase region of the ground state. Moreover, RL can find the disentanglement circuit, which would be very hard for traditional optimization methods.

Although the circuit design problems dealt with a huge optimization space, the RL agent can find high-quality solutions. This suggests that learning the circuit landscape can be performed with other circuit design problems. Thus our results suggest that the RL-based methods can be a powerful method capable of circuit design problems. The current approach is flexible enough to be applied to different problems, such as quantum circuit optimization, ground state preparation, and novel quantum algorithm design.

In summary, we associate the entanglement structure of a quantum phase with an optimal disentanglement circuit, and show that the disentanglement circuit can be used to characterize different phases in two typical models. We expect that our approach may be extended to characterize more exotic quantum phases, e.g., a phase with topological order and a spin liquid, and that it provides a universal way to quantify entanglement structures of quantum phases.

Acknowledgments.— This work is supported by NSF of China (Grants No. 11775300 and No. 12075310), the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDB28000000), and the National Key Research and Development Program of China (Grant No. 2016YFA0300603).

* zhoudl72@iphy.ac.cn

[1] A. Einstein, B. Podolsky, and N. Rosen, Phys. Rev. 47, 777 (1935).
[2] D. Bouwmeester, J.-W. Pan, K. Mattle, M. Eibl, H. Weinfurter, and A. Zeilinger, Nature 390, 575 (1997).
[3] A. Furusawa, J. L. Sørensen, S. L. Braunstein, C. A. Fuchs, H. J. Kimble, and E. S. Polzik, Science 282, 706 (1998), https://science.sciencemag.org/content/282/5389/706.full.pdf.
[4] A. Karlsson and M. Bourennane, Phys. Rev. A 58, 4394 (1998).
[5] N. Gisin, G. Ribordy, W. Tittel, and H. Zbinden, Rev. Mod. Phys. 74, 145 (2002).
[6] T. Jennewein, C. Simon, G. Weihs, H. Weinfurter, and A. Zeilinger, Phys. Rev. Lett. 84, 4729 (2000).
[7] C. H. Bennett, D. P. DiVincenzo, J. A. Smolin, and W. K. Wootters, Phys. Rev. A 54, 3824 (1996).
[8] V. Vedral and M. B. Plenio, Phys. Rev. A 57, 1619 (1998).
[9] X. Chen, Z.-C. Gu, and X.-G. Wen, Phys. Rev. B 82, 155138 (2010).
[10] T. J. Osborne and M. A. Nielsen, Phys. Rev. A 66, 032110 (2002).
[11] G. Vidal, J. I. Latorre, E. Rico, and A. Kitaev, Phys. Rev. Lett. 90, 227902 (2003).
Reinforcement learning: Q-learning model

In the Q-learning algorithm [56], the objective is to maximize the value of state-action function, the conditional discount return

$$Q_\pi(s,a) = E_\pi(G_t|S_t = s, A_t = a).$$

(11)

If we have a policy $\pi$, then we calculate the value of state-action function $Q_\pi(S,A)$. For each state $s$, we take an action maximizing the value of state action $Q_\pi(s,A)$, which forms our new policy $\pi'$. Then we calculate the value of state-action function $Q_{\pi'}(S,A)$. Repeating the above procedure until the new policy equals the updated one, which is the optimal policy $\pi^*$ we are looking for.

The update of Q-learning algorithm is defined as

$$Q(S_t,A_t) \leftarrow Q(S_t,A_t) + \Delta Q$$

(12)

with

$$\Delta Q = \alpha[R_{t+1} + \gamma \max_a Q(S_{t+1},a) - Q(S_t,A_t)],$$

(13)

where $\alpha$ is the step size parameter.
As shown in Fig. 12, a deep Q-network is composed by a deep neural network, trained with a variant of Q-learning, whose input is state vector and whose output is a value function estimating future rewards. One significant improvement is that it designed a memory buffer to store the agents experiences and used them later. At each learning step, we fed the agent with a minibatch of experiences \( \{S_t, A_t, R_{t+1}, S_{t+1}\} \) with a modified prioritized experience replay (PER) method [57]. The state \( S_t \) is fed into the neural network to calculate the state-action value \( Q(S_t, A_t; \theta) \). At the same time, a target Q-network is to calculate \( \max_{a'} Q(S_{t+1}, a'; \theta^-) \) in Eq. (14). At the end of each step of training, the evaluation network is updated through the back-propagation by minimizing the loss. Based on Eq. (12), the loss is the mean square error (MSE) of the difference between the evaluation \( Q(S_t, A_t; \theta) \) and the target \( \max_{a'} Q(S_{t+1}, a'; \theta^-) \)

\[
\text{loss} = \text{MSE}((R_{t+1} + \gamma \max_{a'} Q(S_{t+1}, a'; \theta^-)) - Q(S_t, A_t; \theta)). \tag{14}
\]

During the learning episodes (see Fig. 3), the agent updates the parameters of the target network \( \theta^- \rightarrow \theta \) to make better decisions.

**Hyperparameters in RL algorithm**

Our RL agent makes use of a deep neural network to approximate the Q values for the possible actions of each state. The network consists of 4 layers of the neural network. All layers have ReLU activation functions except the output layer which has linear activation. The hyper-parameters of the network are summarized in Table I.
Table I: Training Hyper-Parameters

| Hyper-parameter               | Values                  |
|------------------------------|-------------------------|
| Neurons in neural network    | \{1024, 1024, 1024, 1024\} |
| Minibatch size               | 60                      |
| Replay memory size           | 100000                  |
| Learning rate                | 0.0001                  |
| Update period                | 100                     |
| Reward decay $\gamma$        | 0.99                    |