Learning quantum phases via single-qubit disentanglement

Zheng An[1,2,3], Chenfeng Cao[3], Cheng-Qian Xu[1,2], and D. L. Zhou[1,2]

[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

July 19, 2024

arXiv:2107.03542v4 [quant-ph] 18 Jul 2024

Identifying phases of matter presents considerable challenges, particularly within the domain of quantum theory, where the complexity of ground states appears to increase exponentially with system size. Quantum many-body systems exhibit an array of complex entanglement structures spanning distinct phases. Although extensive research has explored the relationship between quantum phase transitions and quantum entanglement, establishing a direct, pragmatic connection between them remains a critical challenge. In this work, we present a novel and efficient quantum phase transition classifier, utilizing disentanglement with reinforcement learning-optimized variational quantum circuits. We demonstrate the effectiveness of this method on quantum phase transitions in the transverse field Ising model (TFIM) and the XXZ model. Moreover, we observe the algorithm's ability to learn the Kramers-Wannier duality pertaining to entanglement structures in the TFIM. Our approach not only identifies phase transitions based on the performance of the disentangling circuits but also exhibits impressive scalability, facilitating its application in larger and more complex quantum systems. This study sheds light on the characterization of quantum phases through the entanglement structures inherent in quantum many-body systems.

1 Introduction

A phase diagram illustrates qualitative variations in many-body systems as a function of physical system parameters [1]. The purpose of a phase classifier is to ascertain the phase diagram by examining measurement results of observables. In the Landau theory of phase transitions, for instance, only the average measurement value of a single local variable, known as the local order parameter, is necessary to determine the phase diagram. Indeed, a discontinuity in a local order parameter or one of its derivatives signifies a phase transition.

In addition to order parameters, entanglement, an essential quantum mechanical nonlocal property, has been proposed to characterize quantum phase transitions and to identify topological orders [2–7]. Entanglement entropy is powerful for detecting critical points and studying universal properties in quantum phase transitions, especially in one-dimensional systems. At critical points, entanglement entropy exhibits logarithmic scaling with subsystem size, revealing crucial information about the underlying critical system [4, 5]. Topological entanglement entropy characterizes topologically ordered states, capturing non-local entanglement and serving as an effective diagnostic tool for detecting and characterizing these exotic phases [2, 6]. Quantum entanglement is believed to be crucial for understanding quantum phase transitions [5, 8, 9], with distinct entanglement structures characterizing different quantum phases.

In our examination of quantum entanglement, we discovered a strong correlation between disentanglement and entanglement structure. Disentanglement is a process that transforms an initially entangled state of a composite system into a separable state. However, a universal disentangling machine is non-existent [10–12]. Despite this, by relaxing the constraint that the reduced density matrices of subsystems remain

D. L. Zhou: zhoudl72@iphy.ac.cn

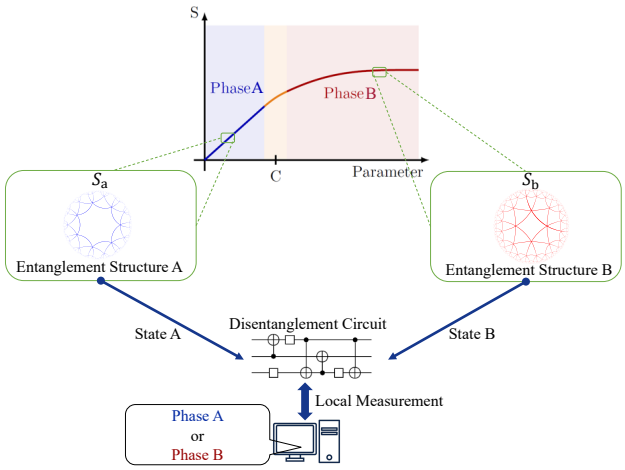


Figure 1: Classifying quantum phases through single-qubit disentanglement. The local entanglement of a quantum many-body system's ground state may encompass intricate structures, with some of the primary structures potentially being closely related to the quantum phase. The RL agent's objective is to reduce the local entanglement of the ground state within a given quantum phase by employing limited operations and observations. Consequently, we can deduce the predominant entanglement structure present in the local entanglement, facilitating the classification of the quantum phase.

perfectly unaffected, local operations can achieve approximate disentanglement [13–15]. Disentanglement's central concept is to decouple quantum correlations among subsystems. Consequently, we aim to investigate whether a quantum circuit-implemented unitary transformation can eradicate local entanglement within a quantum state. Moreover, we seek to explore the relationship between entanglement structure and circuit structure, particularly for states within the same phase.

Reinforcement learning (RL) is a branch of machine learning that focuses on training agents to make optimal decisions in complex environments by interacting with these environments and receiving feedback in the form of rewards or penalties [16]. Over the years, RL has been successfully applied to a wide range of applications, from video games [17, 18] and board games [19–21] to neural architecture search [22]. In recent years, RL has emerged as a powerful tool in quantum information science, where it has been applied to various quantum-related problems including quantum control [23–25], quantum error correction [26], and quantum circuit optimization [27–29].

Variational quantum circuits have become instrumental in approximating both known and unknown unitary evolutions, as well as for studying ground state properties, as demonstrated by hybrid quantum-classical algorithms on noisy intermediate-scale quantum devices [30–35]. These circuits have also been used to characterize quantum phase transitions, applying established order parameters [36–39] or identifying unknown ones through machine learning [40]. In this work, we explore from the disentanglement perspective and introduce an reinforcement learning approach to classify quantum phase transitions. Our method employs RL algorithm [17] to design and optimize variational quantum circuits, which efficiently disentangle the ground state of quantum systems (Fig. 1). By assessing the performance of these disentangling circuits over a range of system parameters, we can identify the critical point that demarcates distinct quantum phases. For practical purposes, the circuit strives to disentangle a single qubit from the rest of the subsystems within a constrained operation space.

We demonstrate the effectiveness of our approach on two paradigmatic models of quantum magnetism: the transverse field Ising model (TFIM) and the XXZ model. For both models, our method accurately identifies the critical point and phase transition. Moreover, we find similarities in the disentangling circuit structures for states within the same phase, shedding light on the relation between entanglement structure and circuit design. Our approach not only classifies quantum phases but also provides insight into their inherent entanglement properties. We observe the algorithm's ability to learn the Kramers-Wannier duality related to entanglement structures in the TFIM. Compared to existing methods like the variational quantum classifier [41], our approach requires significantly fewer resources, using only local measurements and a few layers of variational quantum circuits. We also show that disentangling circuits trained on small system sizes can be applied to much larger systems, exhibiting the scalability of our method.

This paper is organized as follows: Section 2 provides a succinct introduction to the identification of quantum phase transitions through disentanglement and delves into the reinforcement learning framework underpinning our approach. Section 3 showcases the numerical results of our

method, highlighting its ability to classify phases of the TFIM and XXZ models while illustrating the properties of the derived disentangling circuits. In Section 4, we draw comparisons between our method and the variational quantum classifier. Section 5 examines the scalability of our approach. Section 6 discusses the scope of our work, its limitations, and potential extensions to its application. Finally, Section 7 offers concluding remarks and explores prospective avenues for future research in this domain.

2 Methdology

2.1 Phase transitions via disentanglement

In the study of quantum phase transitions, disentanglement plays a crucial role in understanding the intricate nature of entangled states and their transformations across critical points. These transitions, characterized by non-analytic changes in ground state properties, are often accompanied by alterations in entanglement structures. Designing disentangling circuits poses a significant challenge, as it involves creating a specific unitary transformation as a finite sequence of unitary operators A_t chosen from a universal set of gates $\mathcal{G}$. In this study, we task the agent with developing a disentanglement unitary operation U that can effectively eliminate a target local entanglement entropy. Consequently, the agent's objective is to identify a unitary gate U, derived from the composition of gates in the sequence, which eradicates the target entanglement entropy.

We employ entanglement entropy S to quantify the entanglement present in quantum systems' ground states. When a bipartite quantum system AB exists in a pure state, entanglement entropy can be calculated using the reduced density matrices ρ_A or ρ_B, as shown in the following formula:

$$S \equiv -\operatorname{tr}(\rho_A \log_2 \rho_A) = -\operatorname{tr}(\rho_B \log_2 \rho_B). \quad (1)$$

For subsystems A or B represented as spin-$\frac{1}{2}$ systems, S ranges from $S = 0$ (product state) to $S = 1$ (maximally entangled state). For the ground state of the two investigated models, we consider one site as subsystem A and the remainder as subsystem B.

In this work, we introduce a training scheme for identifying distinct phases and their critical point. For a family of ground states ρ_α, the primary concept involves training a reinforcement learning (RL) agent on two states $\rho_{\alpha=a}$ and $\rho_{\alpha=b}$, while presuming the existence of an undetermined critical point $a < c < b$ that categorizes the states into two groups. The aim is to pinpoint the accurate critical point c by training the RL agent to achieve disentangling circuits until convergence with circuit layer p.

We proceed to evaluate the performance of these disentangling circuits across the parameter range (a, b), considering the total performance (S_a and S_b) in relation to the proposed critical point c as $S_a(c)$ and $S_b(c)$. We will demonstrate that the performance of the two RL-ansatz disentangling circuits with deep layers is inversely related, with the cross point at the precise critical point c, i.e., $S_a(c) = S_b(c)$.

The circuit performance can be comprehended as follows: we assume that entanglement possesses two distinct structures in regimes below c (phase $\mathcal{A}$) and above c (phase $\mathcal{B}$), and that the circuits can differentiate and eliminate them. We refer to these diverse structures as features. When setting the states in the parameter range (a, c') where $c' < c$, the circuit identifies the feature of phase $\mathcal{A}$ and accurately removes those entanglements. Nonetheless, when in the range (c', c) of phase $\mathcal{A}$, the circuit encounters states with different features, but the majority feature persists from a to c'. The S_a will increase, signifying that some entanglements cannot be eradicated. When in the range (c, b), the majority feature shifts, and S_a swiftly elevates until saturation. The behavior of S_b is precisely the reverse of S_a.

If phases dictate entanglement structures, the feature count for both phases will be equal when situated at the critical point c, i.e., $S_a(c) = S_b(c)$. Conversely, if $S_a(c) \neq S_b(c)$, then the intersection cannot represent the critical point c, as the phase spanning from the cross point to the critical point would constitute another phase.

2.2 Reinforcement learning and variational quantum disentangler

Reinforcement learning (RL) is a machine learning paradigm in which an intelligent agent learns to make decisions by interacting with an environment (see Fig. 2). The agent performs ac-

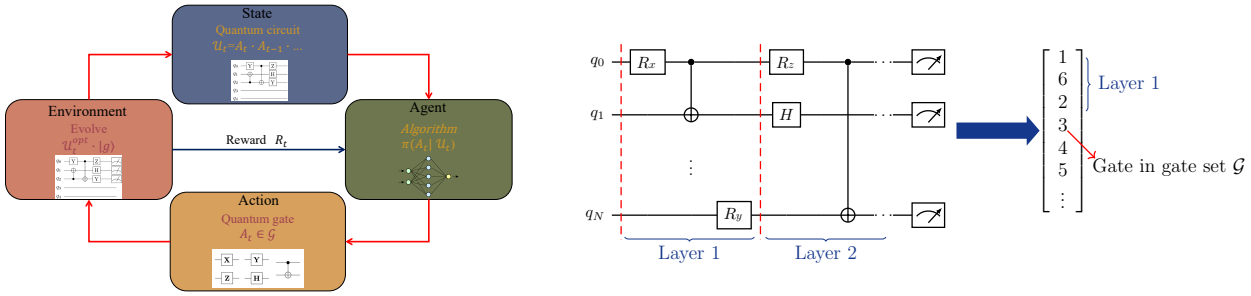


Figure 2: The deep reinforcement learning (DRL) architecture consists of two primary components: (a) The DRL environment, which is depicted as a quantum circuit modeled by the optimized variational quantum circuit U_t^{opt} with an initial ground state $|g\rangle$ at each episode. During each time step t, the agent acquires the current state of U_t and, based on this information, selects the subsequent gate A_t from the gate set $\mathcal{G}$ to apply to the quantum circuit. Consequently, the environment returns the real-valued reward R_t to the agent, which is a function of the entanglement entropy S. The agent's policy π is encoded within a deep neural network. (b) State representation for the RL agent: At each time step, the agent receives an input vector representing the architecture of the quantum circuit U_t. This circuit is depicted as an array, with each element symbolizing a quantum gate in the given gate set. A layer of a quantum circuit is defined as a collection of exactly one quantum gate acting on each qubit within the selected range. Two-qubit gates (CNOTs) are only counted for the target qubit and so each qubit in the selected range is either acted upon by a single-qubit gate or is the target qubit of a CNOT.

tions within the environment and receives feedback in the form of rewards and state (Representation of quantum circuit), which are used to adjust the agent's behavior. In this process, the agent seeks to maximize the reward per step. The agent's knowledge about the environment and its decision-making strategy, referred to as its policy, are typically represented by a mathematical function or a deep neural network (deep reinforcement learning). The policy dictates the agent's choice of actions based on the current state of the environment. By optimizing its policy to maximize the accumulated rewards, the agent learns to navigate the complex dynamics of the environment and make strategic decisions. In the context of our paper, reinforcement learning enables agents to efficiently select and apply quantum gates to optimize entanglement properties, furthering our understanding of quantum systems.

In our framework, the environment's state consists of a quantum circuit initialized as an identity at the beginning of each episode. The agent incrementally constructs the circuit at every step by selecting a gate from $\mathcal{G}$ according to the policy π encoded in the deep neural network, as depicted in Fig. 2 (a).

We consider a gate set $\mathcal{G}$ that includes single-qubit rotations (along the x, y, or z-directions), Hadamard, and controlled-NOT (CNOT) gates, which together form a universal gate set. Rotation gates are parameterized by continuous variables, while CNOT and Hadamard gates are fixed. All single-qubit rotation gate parameters are initialized to π. It is important to note that the operation space is limited, acting only on the nearest or next-nearest qubits of the target qubit.

At each time step, the reinforcement learning agent is provided with an input vector that encapsulates the architecture of the quantum circuit U_t. This representation employs an array s_t, wherein each element corresponds to a quantum gate belonging to the predetermined gate set (see Fig. 2 (b)).

The agent's reward in each step is given by:

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

Here, $\mathcal{R}$ represents the modified reward function for our learning task. The agent does not receive an immediate reward, but instead at time N. A quantum circuit grants the last step's reward with optimized parameters. After finalizing the quantum circuit's structure, a classical optimizer is employed to adjust the initial parameters of the parameterized rotation gate to those with the minimum entanglement entropy.

We define $\mathcal{R}$ as:

$$\mathcal{R} = \begin{cases} 0, & \text{if } S_{RL} \geq S_0 \\ \dfrac{S_0 - S_{RL}}{S_0}, & \text{otherwise} \end{cases} \tag{3}$$

where S_{RL} denotes the final entanglement entropy achieved by our RL-assisted quantum algorithm, and S_0 represents the ground-state entanglement entropy provided by the physical system.

We employ Deep Q-learning algorithms to train the agents in this work (details of the algorithm are in supplementary materials), relying on the reward function. For the classical optimizer, we utilize the BFGS algorithm [42–45] implemented in SciPy [46], enabling us to determine the parameters for each gate, from U_t to U_t^{opt}.

3 Numerical Results

3.1 Transverse field Ising model

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

$$H = -\sum_{j=0}^{N-1} \left(\lambda \sigma_j^x \sigma_{j+1}^x + \sigma_j^z\right), \tag{4}$$

where σ_j^a is the a component of Pauli matrix ($a = x, y$ or z) at site j, and λ is the inverse strength of the external field. We assume the periodic boundary conditions, i.e., the N-th site is identified with the 0-th 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 change dramatically when crossing this critical point.

When λ approaches zero, the TFIM ground state becomes a product of spins pointing in the positive z direction,

$$|\psi_{\text{TFIM}}\rangle_{\lambda\to 0} = |0\rangle^{\otimes N}, \tag{5}$$

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 is an adiabatic continuation of the GHZ state [47]:

$$|\psi_{\text{TFIM}}\rangle_{\lambda\to\infty} = \frac{1}{\sqrt{2}}(|+\rangle^{\otimes N} + |-\rangle^{\otimes N}), \tag{6}$$

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

$$S_{\lambda\to 0} = 0, \tag{7}$$
$$S_{\lambda\to\infty} = 1. \tag{8}$$

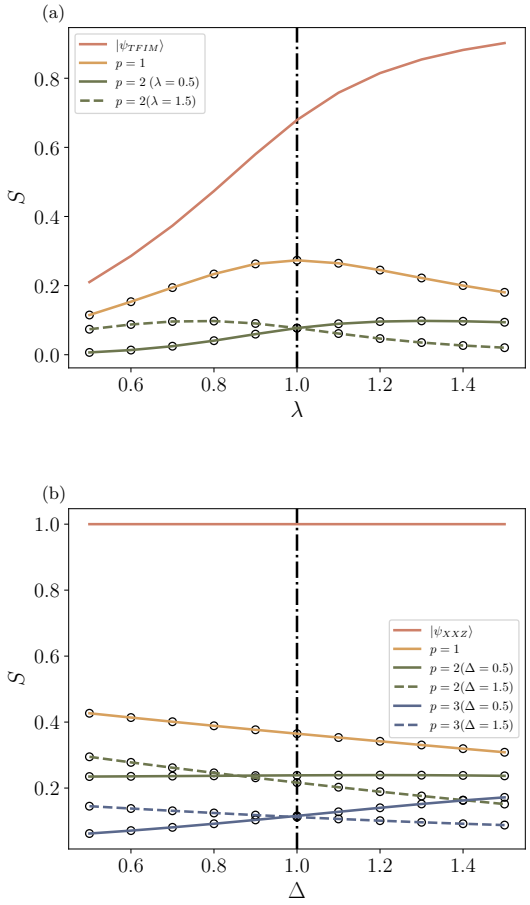


Figure 3: Resulting entanglement entropy of RL-designed disentangling circuits. p represents the number of circuit layers. Solid and dashed lines indicate the disentangling circuits trained from different parameters. The models are for (a) TFIM and (b) XXZ model with 8 sites. The dash-dotted vertical line signifies the critical point.

At the critical point, $\lambda = 1$, there is a phase transition in the ground state in the thermodynamic limit [1]. 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 λ, this decay is exponential.

In Fig. 3 (a), we show the results of the optimal variational algorithm to disentangle the ground state with different values of λ. We take two ground-states from two phases of the system to get the RL-ansatz quantum circuit. The operation space is set to the nearest qubits, i.e., 3 qubits. We also selected two training states asymmetrically from two phases far from the critical

point to exclude the effect of selecting particular training points. (See supplementary material for details on the robustness of the model.)

The performance of the designed algorithm improves as the number circuit layers p increases. As the RL-ansatz disentangling circuit with one layer ($p = 1$), the transformation space is not large enough to cut the correlation with the two subsystems, so different training samples have the same results. However, unlike the raw local entanglement entropy case, the disentangle circuit entropy shows three regions with different values of λ. In both cases, the entanglement entropy rises quickly from a small number from point $\lambda = 0.5$ to point $\lambda = 0.9$ and then slowly rises to the maximum value around the critical point $\lambda = 1$. The situation after the critical point $\lambda = 1$ is precisely the opposite of before. So we can divide the space of λ into "fast-slow-fast" three regions according to the speed of entropy change. We find that the most significant entanglement is present in the parameter region close to the critical point. The slow region corresponds with the quantum critical region [1]. This indicates that the local quantum entanglement has a deep relation with the phase transition region.

However, with the layer number larger than one, the results show different structures of circuits with different phases of λ. As Fig. 3 (a) shows, the designed circuit structures with two layers have two optimal circuits in different phases. Also, in this case, the characteristic is obtained, and we detect the transition at $\lambda = 1$. We confirm that the results with $p = 2$ are indeed convergent (see Fig. S3 in Supplementary materials).

3.1.1 Understanding agent actions

To confirm what agents have learned from disentanglement under TFIM, we first determine the optimal local disentanglement in theory. We proved that the minimum entropy of ρ_P decreased by unitary operation is determined by the initial state ρ_{RP} of the region RP, on which the unitary operation acts. More specifically, we have the following theorem.

Theorem 1 *Suppose the eigen-decomposition of the state ρ_{RP} is $\rho_{RP} = \sum_{i=1}^{d_P d_R} p_i |\psi_i\rangle \langle\psi_i|$, where d_P (d_R) is the dimension of subsystem P (R). Without loss of generality, we assume these eigenvalues are in decreasing order, i.e., $p_i \geq p_{i+1}$. By performing a unitary operation U on the RP, the minimal entanglement entropy of the subsystem P is*

$$\min_U S(\rho_P(U)) = -\sum_m q_m \log_2 q_m, \tag{9}$$

where $\rho_P(U) = \mathrm{Tr}_R\left[U\rho_{RP}U^\dagger\right]$, $S(\rho) = -\mathrm{Tr}\left[\rho\log_2\rho\right]$ is the Von-Neumann entropy of state ρ, and $q_m = \sum_{j=1}^{d_R} p_{(m-1)d_R+j}$.

Proof The proof is given in the given in Supplementary materials, Section S-I. ■

Although **Theorem** 1 gives the optimal local disentanglement theoretically, it does not imply that we can distinguish quantum phases by directly applying the theorem. In fact, the relation between disentanglement and quantum phase is subtle. Because the optimal solution will destroy all entanglement structures equally, the minimal entanglement entropy in our problem is so small in both phases that it can not be used to distinguish the phases. However, the RL agent often finds a local minimum but not a global one, which leads it to destroy the dominant entangled structure under a specific phase. Therefore, even if RL cannot give an optimal strategy in many cases, it may fits the problem.

Interestingly, we found the RL agent finds the optimal solution to disentangle the local entanglement of the nearest two-body in the phase of $\lambda > 1$ (Fig. 4), which is exactly the solution of **Theorem** 1 when RP is restricted to two nearest spins. At first glance, the RL-ansatz disentangling circuit in the phase of $\lambda < 1$ may be a suboptimal disentanglement of a complex three-body local entanglement. However, we realize that there is a well known Kramers-Wannier duality in TFIM [48, 49]. For the model, the self-dual point is the phase transition point $\lambda = 1$. When we perform the duality transformation which defines new Pauli operators μ_i^z and μ_i^x in a dual lattice

$$\mu_j^x = \prod_{i\leq j} \sigma_i^z \tag{10}$$

$$\mu_j^z = \sigma_{j+1}^x \sigma_j^x, \tag{11}$$

and the original Hamiltonian can be written as

$$H = -\sum_{j=0}^{N-1} \left(\mu_j^x \mu_{j+1}^x + \lambda \mu_j^z\right). \tag{12}$$

We found that the optimal disentanglement of the nearest neighbor under the duality transformation coincides with the results of the RL agent in phase $\lambda < 1$ (Fig. 4).

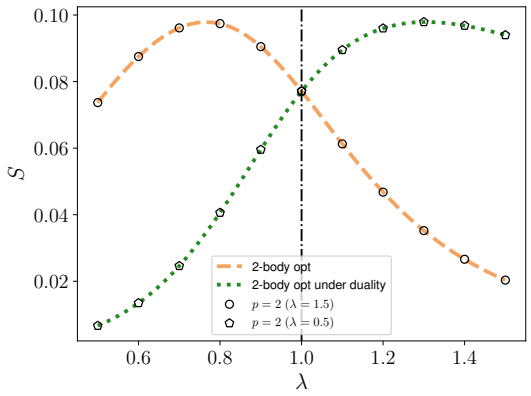


Figure 4: Entanglement entropy versus λ for the transverse field Ising model (TFIM), comparing the reinforcement learning (RL)-ansatz disentangling circuit and the nearest two-body optimal disentanglement. The lines depict the theoretically optimal outcomes, whereas the unfilled markers indicate the numerical results obtained from the quantum circuits designed based on two distinct ground states ($\lambda = 0.5, 1.5$).

The difference between the result of the RL agent and the theoretical optimal solution is negligible. These results suggest that the agent learns the duality of the Hamiltonian, which is an important feature in classifying quantum phases. This further shows that the RL agent has found the main local entanglement structure between different phases.

It is important to recognize that the inherent duality of the model is already known, serving as a fundamental component of our understanding of phase transitions. This established knowledge allows us to apply **Theorem** 1 effectively to determine phase transition points. In contrast, our machine learning approach was not initially programmed with knowledge of this duality, yet it successfully identified the same critical point as established by theoretical analysis. This demonstrates the capability of our method to independently corroborate theoretical insights. In such cases, **Theorem** 1 cannot leverage duality transformations to derive an optimal solution.

It is worth noting that, for small areas, finding the optimal disentangler might be simple, allowing us to use an exact solution instead of a machine learning approximation. However, to do this, one must know the interaction area of the main entanglement structure. Discerning this structure within the context of a quantum system proves to be a nontrivial task, demanding an extensive array of calculations. In practice, the accurate determination of the phase transition point poses a considerable challenge for all currently established methods.

3.2 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

$$H = \sum_{j=0}^{N-1} \left(\sigma_j^x \sigma_{j+1}^x + \sigma_j^y \sigma_{j+1}^y + \Delta \sigma_j^z \sigma_{j+1}^z \right), \quad (13)$$

where Δ is a dimensionless parameter characterizing the anisotropy of the model. The XXZ spin chain has been widely studied [50–54]. 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 result from that the former is gapless while the latter is gapped. The quantum phase transition is located at the isotropic point $\Delta = 1$ which is the Kosterlitz-Thouless quantum phase transition (KT-QPT).

First, we perform our learning scheme, using the same condition for the TFIM case. However, we found the results do not converge. So we set the operation space to include next-nearest qubits, i.e., 5 qubits.

In Fig. 3 (b), we show the results of the optimal circuits to disentangle the ground state with different values of Δ. Similar to the TFIM model mentioned above, the performance of the designed circuits is proportional to the circuit depth p. As the RL-ansatz disentangling circuit with one layer ($p = 1$), this case has a universal circuit structure for different values of Δ, which is similar to the TFIM result.

Nevertheless, unlike the TFIM case, the performance is different in the two-layer case. From Fig. 3 (b), we observe interesting differences for disentangling circuit results at the crossing points. The intersection is far from the critical point because the number of layers of the quantum circuit is just between too shallow and enough layers. So we can see that there are two

optimal solutions, but the intersection of the two is not the critical point. The above situation indicates that the ground-state entanglement structure of the XXZ model might be more complex than the TFIM case.

When $p = 3$, the optimal disentangling circuit is convergent (see Supplementary materials, Fig. S4), and the result is consistent with what we discussed in the former section. Also, we detect the transition at $\Delta = 1$.

4 Comparison with variational quantum classifier

In this section, we present an analytical comparison between our method and the variational quantum classifier, focusing on the perspective of resource requirements.

The variational quantum classifier is a quantum algorithm designed for quantum phase classification. It is worth mentioning that numerous existing quantum classifiers have been proposed, with the majority of them being rooted in supervised learning [55, 56]. This implies that additional information is incorporated during the training process. For the purpose of comparison, we have chosen variational quantum classifiers [41]. Much like our method, the variational quantum classifier utilizes a quantum circuit to identify quantum phase transitions. The training process of this technique involves a supervised learning approach, necessitating the advanced preparation of a substantial quantity of ground states as well as pairwise labels. Ultimately, the input verified quantum states are distinguished by voting on the observations of all qubits to determine the phase to which they belong.

In this analysis, we simplify the variational quantum classifier method; the original version necessitates the use of a variational quantum eigensolver (VQE) [30, 57–60] to simulate the ground state, while our focus lies solely on the classifier's performance. Consequently, we directly employ the system's ground state as input data. We implement this method to the two models previously examined in order to assess whether it demands fewer resources compared to our proposed approach.

As illustrated in Fig. 5, the method demands a minimum of 8 layers in the quantum circuit and 200 and 1600 training ground states for the TFIM and XXZ models, respectively. The y-axis represents the probability of the algorithm predicting that a quantum state belongs to a specific phase in the absence of labels. A reliable algorithm can yield a probability of 0.5 near the critical point. In contrast, our method necessitates only two training ground states and a maximum of 3 layers in the quantum circuit. Furthermore, the method requires measurements of all qubits, whereas our approach mandates local measurements.

5 Scalability and robustness of the RL-ansatz disentangling circuit

In previous discussions, the relationship between the entanglement structure, system size, and robustness was not explored. Further, our method only operates on the local space of the target qubit, which is easy to expand the scale of the system. So we want to study the effect of disentangling circuit training in the small system on a large system.

In Fig. 6 and Fig. 7, we show the results of the disentangling circuit trained in 8-qubit systems on N-qubit problems with $N > 10$. For the RL-designed disentangling circuit, the circuit is obtained by a training process on a problem with qubit number $N = 8$ then applied to problems having a larger number of qubits ($N = 10, 12, 14$). The results show that our solution is not affected by the size of the system. This further implies that the entanglement structure is not affected by the size but is more closely related to the quantum phase.

To verify the robustness of our RL-ansatz disentangling circuit with imperfect input states, we evaluated the performance of the training scheme using imperfect ground states generated by the VQE method. This method is often used in experimental settings where the exact ground state is not accessible. Through this evaluation, we aim to demonstrate the resilience and efficacy of our model, even when faced with suboptimal input data. Since obtaining the ground state of the TFIM is relatively straightforward and we have a theoretically optimal solution for this model, we chose to utilize the ground state of the XXZ model, which poses more significant challenges.

As shown in Fig. 8, our method effectively iden-

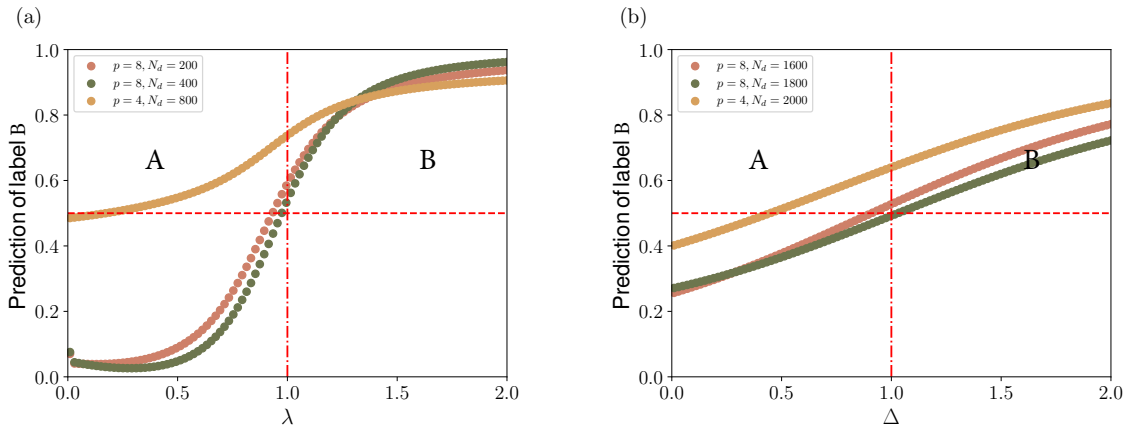


Figure 5: Performance of the variational quantum classifier: (a) TFIM with 8 sites; (b) XXZ model with 8 sites. p represents the number of circuit layers, while N_d denotes the quantity of training ground states. To ensure convergence of the results, we employed 1000 iterations of optimization for each data point. The predicted phase label corresponds to the expected probability that the classifier resides in a fixed quantum phase for a given state. When this probability reaches 0.5, we consider the system to be at a phase transition point.

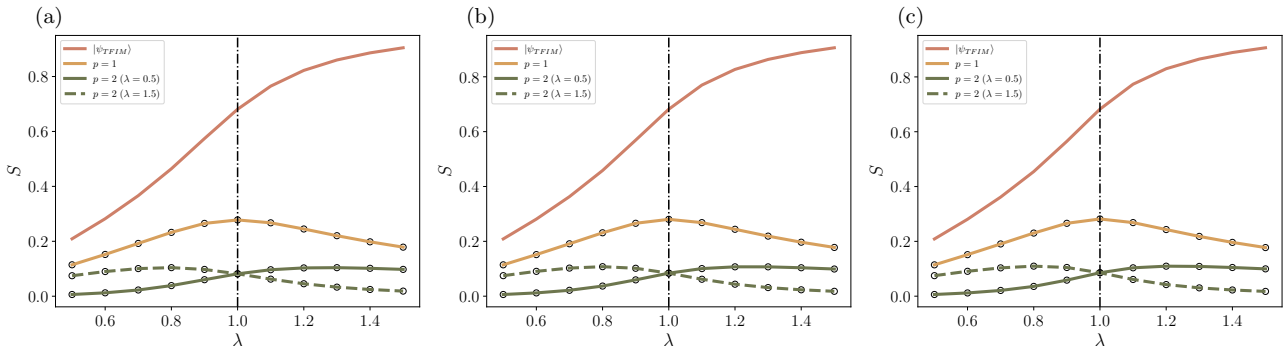


Figure 6: Performance of RL-designed disentangling circuits applied to TFIM with different numbers of qubits. In this plot, the RL-designed circuits are obtained by training on the problem with qubit number $N = 8$ then applied to problems with a larger number of qubits, (a):$N = 10$ (b):$N = 12$ (c):$N = 14$.

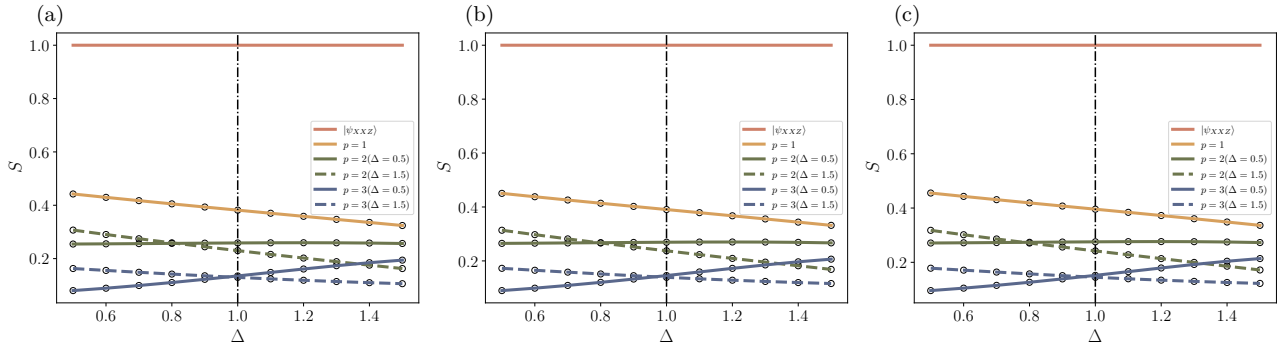


Figure 7: Performance of RL-designed disentangling circuits applied to XXZ model with different numbers of qubits. In this plot, the RL-designed circuits are obtained by training on the problem with qubit number $N = 8$ then applied to problems with a larger number of qubits, (a):$N = 10$ (b):$N = 12$ (c):$N = 14$.

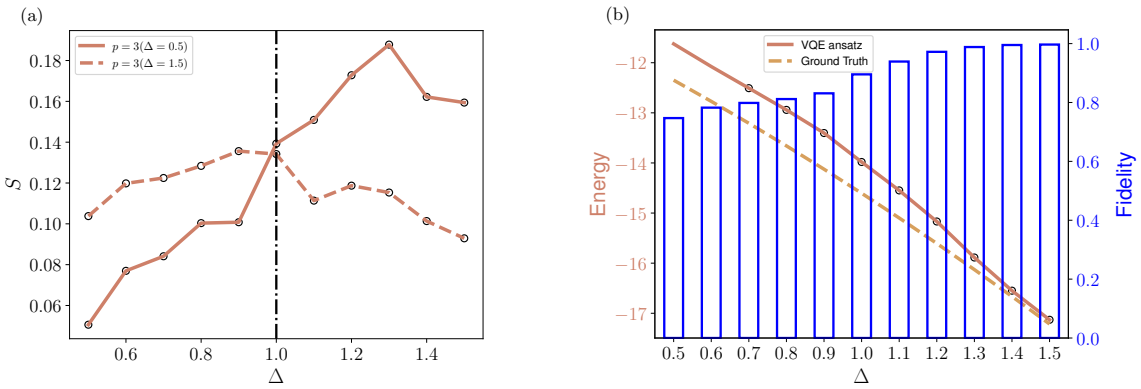


Figure 8: Performance of RL-designed disentangling circuits on Imperfect Ground States in the XXZ Model with $N = 8$. (a) Resulting entanglement entropy of RL-designed disentangling circuits applied to the ground state prepared via the VQE. (b) Comparative analysis of energy and fidelity between the VQE-ansatz ground state and the actual ground state of the XXZ model. The energy of the ground state as predicted by the VQE ansatz (solid line) and the actual ground state (dashed line). The fidelity values between these states are represented as a blue outline bar chart.

tifies phase transition points even at relative low fidelity ground states. This performance highlights the robustness of our methodology, confirming its ability to handle less-than-ideal conditions effectively. In other words, our approach can also be applied to detect phase transitions in scenarios where the ground state is imperfect.

6 Discussion

In this paper, we have presented an algorithm for studying entanglement structures and identifying quantum phases in many-body systems. The approach combines deep reinforcement learning-optimized variational quantum circuits, enabling a systematic investigation of entanglement patterns and offering valuable insights for both theoretical and experimental contexts.

Our method serves distinct purposes compared to theoretically optimal disentanglement methods, such as those described in **Theorem** 1. While **Theorem** 1 provides a mathematical framework to determine the minimal entanglement entropy achievable through unitary transformations, it does not specify the actual quantum state that corresponds to this minimal entropy. Consequently, **Theorem** 1 lacks practical guidance on designing specific unitary operations for achieving this state. Our machine learning approach addresses this gap by not only predicting the minimal entropy but also facilitating the design of the corresponding disentangling unitary operations. This is crucial for practical applications where explicit construction of quantum circuits is necessary. Furthermore, our approach excels in identifying quantum phases and transitions by recognizing and disentangling key entanglement structures, thereby facilitating the detection of quantum phase transitions, theoretical disentanglement methods primarily focus on elucidating the theoretical underpinnings of these disentangling operations. Additionally, the application of RL methods enhances our understanding of entanglement structures and phase transitions across diverse systems. RL approaches are particularly resilient against system imperfections and data noise, capable of learning to identify critical features of phase transitions during the training process. Unlike traditional methods that rely on known order parameters like magnetization or correlation measures, RL can potentially discover novel indicators of phase transitions by directly analyzing data. This capability is especially advantageous in systems with poorly understood transitions or where traditional indicators are difficult to measure.

From a theoretical standpoint, our algorithm sheds light on the understanding and characterization of quantum phases by leveraging the entanglement structures in quantum many-body systems. It not only allows for the identification of essential entanglement structures but also facilitates more efficient classification of quantum phases. Furthermore, the potential connections with measurement-induced phase transitions [61–65] present an exciting avenue for future research. By adapting our method to account for the effects of measurements on quantum many-body

systems, we could uncover a more comprehensive understanding of the interplay between entanglement structures, quantum phase transitions, and measurement-induced phase transitions.

For experimentalists, our algorithm provides a practical and scalable approach to phase identification in quantum many-body systems. It requires only local quantum operations and entanglement measures of local qubits, making it ideally suited for direct application in existing experiments, even without comprehensive knowledge of the quantum states. Utilizing a limited number of operations and measurements, the quantum circuit can be extended to accommodate large-sized systems. When attempting to measure the reduced density matrix in a many-body system, our algorithm can aid in identifying the quantum phase based on entanglement structures. Although performing such measurements can be challenging, indirect methods and experimental techniques are available for extracting information about the reduced density matrix or specific properties of interest, such as entanglement entropy or correlations.

The choice of the region size for disentanglement depends on the entanglement structures' complexity and the specific many-body system under consideration. While our approach demonstrates success in the context of Ising spin chain systems, extending it to higher dimensions, critical systems, and topological systems may necessitate further development and customization. For instance, in topological systems, a more refined approach may be necessary to capture the intrinsic long-range entangled structures that cannot be disentangled via local unitary operations within a specific depth.

Although finding optimal disentanglers analytically may be possible for small regions, the numerical approach using reinforcement learning offers several advantages. As the size of the region or the complexity of the entanglement structures increases, reinforcement learning provides scalability, adaptability, and robustness to imperfections, making it a versatile framework for quantum phase identification across various quantum many-body systems.

7 Conclusion

In conclusion, we have proposed an innovative reinforcement learning approach for classifying quantum phase transitions through disentanglement. Our method employs deep Q-learning to design and optimize variational quantum circuits that effectively disentangle the ground state of a quantum system. By analyzing the performance of these disentangling circuits across a range of system parameters, we can identify the critical point that demarcates distinct quantum phases.

We demonstrated the effectiveness of our approach on two paradigmatic models of quantum magnetism: the transverse field Ising model and the XXZ model. For both models, our method accurately identified the critical point and phase transition. Moreover, we found similarities in the disentangling circuit structures for states within the same phase, shedding light on the relation between entanglement structure and circuit design. Our approach not only classifies quantum phases but also provides insight into their inherent entanglement properties.

Compared to existing methods, our approach requires significantly fewer resources, using local measurements and a few layers of variational quantum circuits. We also showed that disentangling circuits trained on small system sizes can be applied to much larger systems, exhibiting the scalability of our method. 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. It provides a universal way to quantify entanglement structures of quantum phases. Our work highlights the promise of machine learning, and reinforcement learning in particular, for exploring and understanding the properties of complex quantum systems.

8 Acknowledgments

This work is supported by the National Key Research and Development Program of China (Grant No. 2021YFA0718302 and No. 2021YFA1402104), NSF of China (Grants No. 11775300 and No. 12075310), and the Strategic Priority Research Program of Chinese Academy of Sciences (Grant No. XDB28000000).

References

[1] Subir Sachdev. "Quantum phase transitions". Cambridge University Press. (2011). 2 edition.

[2] Xie Chen, Zheng-Cheng Gu, and Xiao-Gang Wen. "Local unitary transformation, long-range quantum entanglement, wave function renormalization, and topological order". Phys. Rev. B **82**, 155138 (2010).

[3] Tobias J. Osborne and Michael A. Nielsen. "Entanglement in a simple quantum phase transition". Phys. Rev. A **66**, 032110 (2002).

[4] G. Vidal, J. I. Latorre, E. Rico, and A. Kitaev. "Entanglement in quantum critical phenomena". Phys. Rev. Lett. **90**, 227902 (2003).

[5] A. Osterloh, Luigi Amico, G. Falci, and Rosario Fazio. "Scaling of entanglement close to a quantum phase transition". Nature **416**, 608–610 (2002).

[6] Alexei Kitaev and John Preskill. "Topological entanglement entropy". Phys. Rev. Lett. **96**, 110404 (2006).

[7] Michael Levin and Xiao-Gang Wen. "Detecting topological order in a ground state wave function". Phys. Rev. Lett. **96**, 110405 (2006).

[8] Luigi Amico, Rosario Fazio, Andreas Osterloh, and Vlatko Vedral. "Entanglement in many-body systems". Rev. Mod. Phys. **80**, 517–576 (2008).

[9] L.-A. Wu, M. S. Sarandy, and D. A. Lidar. "Quantum phase transitions and bipartite entanglement". Phys. Rev. Lett. **93**, 250404 (2004).

[10] Daniel R. Terno. "Nonlinear operations in quantum-information theory". Phys. Rev. A **59**, 3320–3324 (1999).

[11] Tal Mor and Daniel R. Terno. "Sufficient conditions for a disentanglement". Phys. Rev. A **60**, 4341–4343 (1999).

[12] Tal Mor. "Disentangling quantum states while preserving all local properties". Phys. Rev. Lett. **83**, 1451–1454 (1999).

[13] Sibasish Ghosh, Somshubhro Bandyopadhyay, Anirban Roy, Debasis Sarkar, and Guruprasad Kar. "Optimal universal disentangling machine for two-qubit quantum states". Phys. Rev. A **61**, 052301 (2000).

[14] Zheng-Wei Zhou and Guang-Can Guo. "Disentanglement and inseparability correlation in a two-qubit system". Phys. Rev. A **61**, 032108 (2000).

[15] Ya-fei Yu, Jian Feng, Xiao-qing Zhou, and Ming-sheng Zhan. "Approximate disentanglement". Int. J. Theor. Phys. **43**, 21–26 (2004).

[16] Richard S. Sutton and Andrew G. Barto. "Reinforcement learning: An introduction". The MIT Press. Cambridge, MA, USA (2018). url: http://incompleteideas.net/book/the-book-2nd.html.

[17] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fidjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. "Human-level control through deep reinforcement learning". Nature **518**, 529–533 (2015).

[18] Oriol Vinyals, Igor Babuschkin, Wojciech M Czarnecki, Michaël Mathieu, Andrew Dudzik, Junyoung Chung, David H Choi, Richard Powell, Timo Ewalds, Petko Georgiev, et al. "Grandmaster level in starcraft ii using multi-agent reinforcement learning". Nature **575**, 350–354 (2019).

[19] David Silver, Aja Huang, Chris J. Maddison, Arthur Guez, Laurent Sifre, George van den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, Sander Dieleman, Dominik Grewe, John Nham, Nal Kalchbrenner, Ilya Sutskever, Timothy Lillicrap, Madeleine Leach, Koray Kavukcuoglu, Thore Graepel, Demis Hassabis, David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. "Mastering the game of go with deep neural networks and tree search". Nature **529**, 484–489 (2016).

[20] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. "Mastering the game of go without human knowledge". Nature **550**, 354–359 (2017).

[21] David Silver, Thomas Hubert, Julian Schrit-

twieser, Ioannis Antonoglou, Matthew Lai, Arthur Guez, Marc Lanctot, Laurent Sifre, Dharshan Kumaran, Thore Graepel, Timothy Lillicrap, Karen Simonyan, and Demis Hassabis. "A general reinforcement learning algorithm that masters chess, shogi, and go through self-play". Science **362**, 1140–1144 (2018).

[22] Barret Zoph and Quoc V. Le. "Neural architecture search with reinforcement learning". In 5th International Conference on Learning Representations, ICLR 2017, Toulon, France, April 24-26, 2017, Conference Track Proceedings. (2017). arXiv:1611.01578.

[23] Marin Bukov, Alexandre G. R. Day, Dries Sels, Phillip Weinberg, Anatoli Polkovnikov, and Pankaj Mehta. "Reinforcement learning in different phases of quantum control". Phys. Rev. X **8**, 031086 (2018).

[24] Zheng An and D. L. Zhou. "Deep reinforcement learning for quantum gate control". EPL **126**, 60002 (2019).

[25] Zheng An, Hai-Jing Song, Qi-Kai He, and D. L. Zhou. "Quantum optimal control of multilevel dissipative quantum systems with reinforcement learning". Phys. Rev. A **103**, 012404 (2021).

[26] Thomas Fösel, Petru Tighineanu, Talitha Weiss, and Florian Marquardt. "Reinforcement learning with neural networks for quantum feedback". Phys. Rev. X **8**, 031084 (2018).

[27] Thomas Fösel, Murphy Yuezhen Niu, Florian Marquardt, and Li Li. "Quantum circuit optimization with deep reinforcement learning" (2021). arXiv:2103.07585.

[28] Mateusz Ostaszewski, Lea M. Trenkwalder, Wojciech Masarczyk, Eleanor Scerri, and Vedran Dunjko. "Reinforcement learning for optimization of variational quantum circuit architectures". In M. Ranzato, A. Beygelzimer, Y. Dauphin, P.S. Liang, and J. Wortman Vaughan, editors, Advances in Neural Information Processing Systems. Volume 34, pages 18182–18194. Curran Associates, Inc. (2021). url: https://proceedings.neurips.cc/paper_files/paper/2021/file/9724412729185d53a2e3e7f889d9f057-Paper.pdf.

[29] Chenfeng Cao, Zheng An, Shi-Yao Hou, DL Zhou, and Bei Zeng. "Quantum imaginary time evolution steered by reinforcement learning". Communications Physics **5**, 1–7 (2022).

[30] Alberto Peruzzo, Jarrod McClean, Peter Shadbolt, Man-Hong Yung, Xiao-Qi Zhou, Peter J. Love, Alán Aspuru-Guzik, and Jeremy L. O'Brien. "A variational eigenvalue solver on a photonic quantum processor". Nature Communications **5**, 4213 (2014).

[31] John Preskill. "Quantum Computing in the NISQ era and beyond". Quantum **2**, 79 (2018).

[32] M. Cerezo, Andrew Arrasmith, Ryan Babbush, Simon C. Benjamin, Suguru Endo, Keisuke Fujii, Jarrod R. McClean, Kosuke Mitarai, Xiao Yuan, Lukasz Cincio, and Patrick J. Coles. "Variational quantum algorithms". Nature Reviews Physics **3**, 625–644 (2021).

[33] Kishor Bharti, Alba Cervera-Lierta, Thi Ha Kyaw, Tobias Haug, Sumner Alperin-Lea, Abhinav Anand, Matthias Degroote, Hermanni Heimonen, Jakob S. Kottmann, Tim Menke, Wai-Keong Mok, Sukin Sim, Leong-Chuan Kwek, and Alán Aspuru-Guzik. "Noisy intermediate-scale quantum algorithms". Rev. Mod. Phys. **94**, 015004 (2022).

[34] Chenfeng Cao, Yunlong Yu, Zipeng Wu, Nic Shannon, Bei Zeng, and Robert Joynt. "Mitigating algorithmic errors in quantum optimization through energy extrapolation". Quantum Science and Technology **8**, 015004 (2022).

[35] Chenfeng Cao, Yeqing Zhou, Swamit Tannu, Nic Shannon, and Robert Joynt. "Exploiting many-body localization for scalable variational quantum simulation" (2024). arXiv:2404.17560.

[36] Henrik Dreyer, Mircea Bejan, and Etienne Granet. "Quantum computing critical exponents". Phys. Rev. A **104**, 062614 (2021).

[37] Michael Meth, Viacheslav Kuzmin, Rick van Bijnen, Lukas Postler, Roman Stricker, Rainer Blatt, Martin Ringbauer, Thomas Monz, Pietro Silvi, and Philipp Schindler. "Probing phases of quantum matter with an ion-trap tensor-network quantum eigensolver". Phys. Rev. X **12**, 041035 (2022).

[38] Kevin Lively, Tim Bode, Jochen Szangolies, Jian-Xin Zhu, and Benedikt Fauseweh. "Robust experimental signatures of phase transitions in the variational quantum eigensolver" (2024). arXiv:2402.18953.

[39] Jan Lukas Bosse, Raul A Santos, and Ashley Montanaro. "Sketching phase diagrams using low-depth variational quantum algorithms". Quantum Science and Technology **9**, 035034 (2024).

[40] Chenfeng Cao, Filippo Maria Gambetta, Ashley Montanaro, and Raul A. Santos. "Unveiling quantum phase transitions from traps in variational quantum algorithms" (2024). arXiv:2405.08441.

[41] A. V. Uvarov, A. S. Kardashin, and J. D. Biamonte. "Machine learning phase transitions with a quantum processor". Phys. Rev. A **102**, 012415 (2020).

[42] C. G. BROYDEN. "The Convergence of a Class of Double-rank Minimization Algorithms 1. General Considerations". IMA Journal of Applied Mathematics **6**, 76–90 (1970).

[43] R. Fletcher. "A new approach to variable metric algorithms". The Computer Journal **13**, 317–322 (1970).

[44] Donald Goldfarb. "A family of variable-metric methods derived by variational means". Mathematics of Computation **24**, 23–26 (1970). url: http://www.jstor.org/stable/2004873.

[45] D. F. Shanno. "Conditioning of quasi-newton methods for function minimization". Mathematics of Computation **24**, 647–656 (1970). url: http://www.jstor.org/stable/2004840.

[46] Pauli Virtanen, Ralf Gommers, Travis E Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, et al. "Scipy 1.0: fundamental algorithms for scientific computing in python". Nature methods **17**, 261–272 (2020).

[47] Bei Zeng and Xiao-Gang Wen. "Gapped quantum liquids and topological order, stochastic local transformations and emergence of unitarity". Phys. Rev. B **91**, 125121 (2015).

[48] H. A. Kramers and G. H. Wannier. "Statistics of the two-dimensional ferromagnet. part i". Phys. Rev. **60**, 252–262 (1941).

[49] Eduardo Fradkin and Leonard Susskind. "Order and disorder in gauge systems and magnets". Phys. Rev. D **17**, 2637–2658 (1978).

[50] C. N. Yang and C. P. Yang. "One-dimensional chain of anisotropic spin-spin interactions. i. proof of bethe's hypothesis for ground state in a finite system". Phys. Rev. **150**, 321–327 (1966).

[51] C. N. Yang and C. P. Yang. "One-dimensional chain of anisotropic spin-spin interactions. ii. properties of the ground-state energy per lattice site for an infinite system". Phys. Rev. **150**, 327–339 (1966).

[52] C. N. Yang and C. P. Yang. "One-dimensional chain of anisotropic spin-spin interactions. iii. applications". Phys. Rev. **151**, 258–264 (1966).

[53] Eytan Barouch, Barry M. McCoy, and Max Dresden. "Statistical mechanics of the XY model. i". Phys. Rev. A **2**, 1075–1092 (1970).

[54] Eytan Barouch and Barry M. McCoy. "Statistical mechanics of the xy model. ii. spin-correlation functions". Phys. Rev. A **3**, 786–804 (1971).

[55] Maria Schuld, Alex Bocharov, Krysta M. Svore, and Nathan Wiebe. "Circuit-centric quantum classifiers". Phys. Rev. A **101**, 032308 (2020).

[56] Iris Cong, Soonwon Choi, and Mikhail D. Lukin. "Quantum convolutional neural networks". Nat. Phys. **15**, 1273–1278 (2019).

[57] P. J. J. O'Malley, R. Babbush, I. D. Kivlichan, J. Romero, J. R. McClean, R. Barends, J. Kelly, P. Roushan, A. Tranter, N. Ding, B. Campbell, Y. Chen, Z. Chen, B. Chiaro, A. Dunsworth, A. G. Fowler, E. Jeffrey, E. Lucero, A. Megrant, J. Y. Mutus, M. Neeley, C. Neill, C. Quintana, D. Sank, A. Vainsencher, J. Wenner, T. C. White, P. V. Coveney, P. J. Love, H. Neven, A. Aspuru-Guzik, and J. M. Martinis. "Scalable quantum simulation of molecular energies". Phys. Rev. X **6**, 031007 (2016).

[58] J. I. Colless, V. V. Ramasesh, D. Dahlen, M. S. Blok, M. E. Kimchi-Schwartz, J. R. McClean, J. Carter, W. A. de Jong, and

I. Siddiqi. “Computation of molecular spectra on a quantum processor with an error-resilient algorithm”. Phys. Rev. X **8**, 011021 (2018).

[59] Chenfeng Cao, Hiroshi Yano, and Yuya O. Nakagawa. “Accelerated variational quantum eigensolver with joint bell measurement”. Phys. Rev. Res. **6**, 013205 (2024).

[60] Yihao Liu, Min-Quan He, and Z. D. Wang. “Variational quantum eigensolvers with quantum gaussian filters for solving ground-state problems in quantum many-body systems” (2024). arXiv:2401.13459.

[61] Yaodong Li, Xiao Chen, and Matthew P. A. Fisher. “Quantum zeno effect and the many-body entanglement transition”. Phys. Rev. B **98**, 205136 (2018).

[62] Yaodong Li, Xiao Chen, and Matthew P. A. Fisher. “Measurement-driven entanglement transition in hybrid quantum circuits”. Phys. Rev. B **100**, 134306 (2019).

[63] Brian Skinner, Jonathan Ruhman, and Adam Nahum. “Measurement-induced phase transitions in the dynamics of entanglement”. Phys. Rev. X **9**, 031009 (2019).

[64] Michael J. Gullans and David A. Huse. “Dynamical purification phase transition induced by quantum measurements”. Phys. Rev. X **10**, 041020 (2020).

[65] Amos Chan, Rahul M. Nandkishore, Michael Pretko, and Graeme Smith. “Unitary-projective entanglement dynamics”. Phys. Rev. B **99**, 224307 (2019).

[66] R. Bhatia. “Matrix analysis”. Graduate Texts in Mathematics. Springer New York. (2013).

[67] Christopher Watkins. “Learning from delayed rewards”. PhD thesis. University of Cambridge. (1989).

[68] Tom Schaul, John Quan, Ioannis Antonoglou, and David Silver. “Prioritized experience replay” (2015). arXiv:1511.05952.

[69] Shi-Jian Gu, Hai-Qing Lin, and You-Quan Li. “Entanglement, quantum phase transition, and scaling in the XXZ chain”. Phys. Rev. A **68**, 042330 (2003).

[70] William K. Wootters. “Entanglement of formation of an arbitrary state of two qubits”. Phys. Rev. Lett. **80**, 2245–2248 (1998).

Supplemental Materials: Learning quantum phases via single-qubit disentanglement

S-I Optimal local disentanglement

In this part, we aim to prove Theorem 1 in the main text. To this end, let's review the theory of majorization first, which plays an important role in quantum information theory.

Suppose we have two n-dimensional vectors $p = (p_1, p_2, \cdots, p_n)$ and $q = (q_1, q_2, \cdots, q_n)$. Without loss of generality, we assume that $p_i \geq p_{i+1}$ and $q_i \geq q_{i+1}$ for $i = 1, 2, \cdots, n-1$. The theory of majorization gives a way to compare these two vectors:

Definition S1. Vector p is majorized by q, denoted as $p \prec q$, if

$$\sum_{i=1}^{k} p_i \leq \sum_{i=1}^{k} q_i \tag{S1}$$

holds for every $k = 1, 2, \cdots, n$.

From the Definition S1, we can see that if $p \prec q$ then vector p is more disordered in the distribution of components than vector q. Indeed, we have a theorem [66]:

Theorem S1. $p \prec q$ if and only if $p = Dq$, where D is doubly stochastic matrix, i.e., D has non-negative entries and each row and column sums to 1.

Now we can prove Theorem 1 in the main text. Suppose state $|\Phi\rangle$ is a pure state of system RPQ. Then the reduced density matrices of subsystem RP is defined as

$$\rho_{RP} = \mathrm{Tr}_Q \left[|\Phi\rangle \langle\Phi| \right], \tag{S2}$$

When we perform a local unitary operation U acting on subsystem RP, the reduced state of subsystem P becomes

$$\rho_P(U) = \mathrm{Tr}_R \left[U \rho_{RP} U^\dagger \right] \tag{S3}$$

By choosing a proper unitary transformation we aim to minimize the entropy

$$\min_U S(\rho_P(U)) \tag{S4}$$

where $S(\rho) = -\mathrm{Tr}\left[\rho \log_2 \rho\right]$, and $S(\rho_P)$ characterizes the entanglement between subsystem P and subsystem RQ.

Let the dimension of the Hilbert space of subsystem P be d_P, and let the dimension of the Hilbert space of subsystem R be d_R. We start with the eigen-decomposition of the state ρ_{RP}

$$\rho_{RP} = \sum_{i=1}^{d_P d_R} p_i \, |\psi_i\rangle \langle\psi_i| \,, \tag{S5}$$

where p_i is the i-th eigenvalue, and $|\psi_i\rangle$ is the corresponding eigenstate. Without loss of generality, we assume these eigenvalues are in decreasing order, i.e., $p_i \geq p_{i+1}$. Then, for $1 \leq m \leq d_P$ let

$$q_m = \sum_{j=1}^{d_R} p_{(m-1)d_R + j}. \tag{S6}$$

Now we are ready to give the following theorem on optimal local disentanglement.

Theorem S2: The minimal entanglement entropy

$$\min_U S(\rho_P(U)) = H(\{q_m\}) \equiv -\sum_m q_m \log_2 q_m. \tag{S7}$$

Proof: Let us take a basis of $\mathcal{H}_P$ as $\{|m\rangle, 1 \leq m \leq d_P\}$, and a basis of $\mathcal{H}_R$ as $\{|\alpha\rangle, 1 \leq \alpha \leq d_R\}$. Then we construct a unitary transformation

$$V |\psi_{(m-1)d_R + \alpha}\rangle = |m, \alpha\rangle \tag{S8}$$

Then

$$\begin{aligned}\rho_P(V) &= \sum_{m,\alpha} p_{(m-1)d_R+\alpha} \operatorname{Tr}_R(|m,\alpha\rangle\langle m,\alpha|) \\ &= \sum_m q_m |m\rangle\langle m| \end{aligned} \tag{S9}$$

which implies that

$$S(\rho_P(V)) = H(\{q_m\}) \tag{S10}$$

For any local unitary U, the eigen-decomposition of $\rho_P(U)$ is

$$\rho_P(U) = \sum_m q_m^{(U)} |m^{(U)}\rangle\langle m^{(U)}| \tag{S11}$$

Then

$$\begin{aligned}\rho_P(U) &= \rho_P(WV) \\ &= \sum_{m,\alpha} p_{(m-1)d_R+\alpha} \operatorname{Tr}_R(W|m,\alpha\rangle\langle m,\alpha|W^\dagger) \\ &= \sum_{n,\beta}\sum_{m,\alpha} p_{(m-1)d_R+\alpha} \langle n^{(U)},\beta|W|m,\alpha\rangle\langle m,\alpha|W^\dagger|n^{(U)},\beta\rangle |n^{(U)}\rangle\langle n^{(U)}| \\ &= \sum_{n,\beta} p_{n,\beta}^{(U)} |n^{(U)}\rangle\langle n^{(U)}| \end{aligned} \tag{S12}$$

where $W = UV^{-1}$ and

$$p_{n,\beta}^{(U)} = \sum_{m,\alpha} B_{n,\beta;m,\alpha} p_{(m-1)d_R+\alpha} \tag{S13}$$

with

$$B_{n,\beta;m,\alpha} = \langle n^{(U)},\beta|W|m,\alpha\rangle\langle m,\alpha|W^\dagger|n^{(U)},\beta\rangle \tag{S14}$$

Combining Eq. S11 with Eq. S12, we arrives at

$$q_m^{(U)} = \sum_\beta p_{m,\beta}^{(U)} \tag{S15}$$

Notice that $B_{n,\beta;m,\alpha} \geq 0$ and

$$\sum_{n,\beta} B_{n,\beta;m,\alpha} = \sum_{m,\alpha} B_{n,\beta;m,\alpha} = 1 \tag{S16}$$

Namely, B is a doubly stochastic matrix. According to the Theorem S1 above or Theorem II.1.9 in Bhatia's book [66], we obtain

$$p^{(U)} \prec p \tag{S17}$$

Now we denote $p_\downarrow^{(U)}$ as $p^{(U)}$ in the decreasing order. Similarly, we introduce

$$q_{\downarrow,m}^{(U)} = \sum_\alpha p_{\downarrow,(m-1)d_R+\alpha}^{(U)} \tag{S18}$$

Then according to the definition of majorization,

$$q_\downarrow^{(U)} \prec q \tag{S19}$$

$$q^{(U)} \prec q_\downarrow^{(U)} \tag{S20}$$

which imples that $q^{(U)} \prec q$. Thus according to Theorem II.3.1 in Bhatia's book [66],

$$S(\rho_P(V)) \leq S(\rho_P(U)). \tag{S21}$$

This completes our proof.

In our context, $|\Phi\rangle$ is the ground state of the considered Hamiltonian H, P is the target qubit that we aim to disentangle from the ground state, and RP is the set of qubits that the unitary transformations U for disentangling P act on.

S-II Reinforcement learning: Q-learning method

The reinforcement learning process is a finite Markov decision process [16]. As shown in Fig. S1, 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

$$G_t = \sum_{k=0}^{\infty} \gamma^k R_{t+k+1}, \tag{S22}$$

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

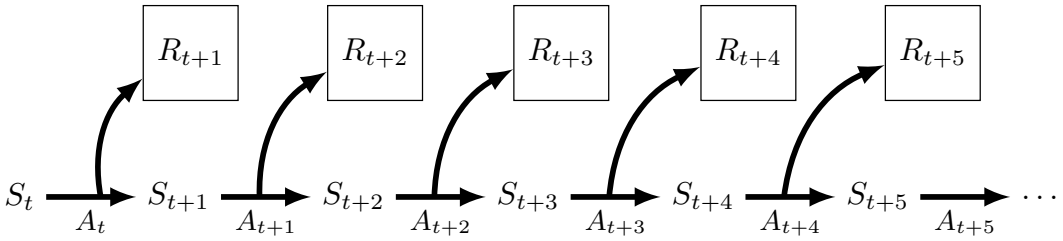


Figure S1: A schematic diagram of Markov decision process.

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

In the Q-learning algorithm [67], the objective is to maximize the value of state-action function, the conditional discount return

$$Q_\pi(s, a) = E_\pi(G_t|S_t = s, A_t = a). \tag{S23}$$

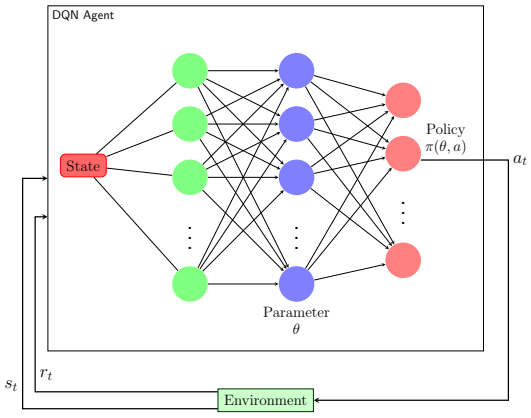


Figure S2: A diagram of deep Q-network architecture.

If we have a policy π, 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 π'. 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 π^* we are looking for.

The update of Q-learning algorithm is defined as

$$Q(S_t, A_t) \leftarrow Q(S_t, A_t) + \Delta Q \tag{S24}$$

with

$$\Delta Q = \alpha[R_{t+1} + \gamma \max_a Q(S_{t+1}, a) - Q(S_t, A_t)], \tag{S25}$$

where α is the step size parameter.

As shown in Fig. S2, 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 agent's 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 [68]. 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. (S26). At the end of each step of training, the evaluation network is updated through the back-propagation by minimizing the loss. Based on Eq. (S24), the loss is the mean square error (MSE) of the difference between the evaluation $Q(S_t, A_t; \theta)$ and the target $\max_{a'} Q(S_{t+1}, a'; \theta^-)$

$$\text{loss} = \text{MSE}((\text{R}_{\text{t+1}} + \gamma \max_{\text{a}'} \text{Q}(\text{S}_{\text{t+1}}, \text{a}'; \theta^-)) - \text{Q}(\text{S}_\text{t}, \text{A}_\text{t}; \theta)). \tag{S26}$$

During the learning episodes, the agent updates the parameters of the target network $\theta^- \to \theta$ to make better decisions.

S-III The robustness of the training

To confirm that the training scheme indeed extracts non-trivial structures from the input state, we have checked the performance from the training scheme when the quantum circuit is trained on different initial values and states. To show the robustness of the RL-ansatz quantum circuit, we examined our method with two example cases: (i) the initial parameters of rotation gate extracted from the uniform distribution with mean value $\{0, \pi/2, \pi\}$ and (ii) the disentangling circuit training from different ground states of parameters (λ or Δ).

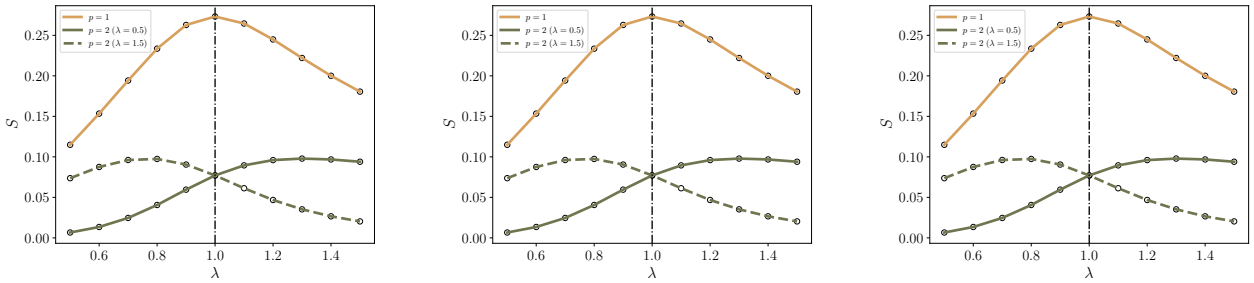
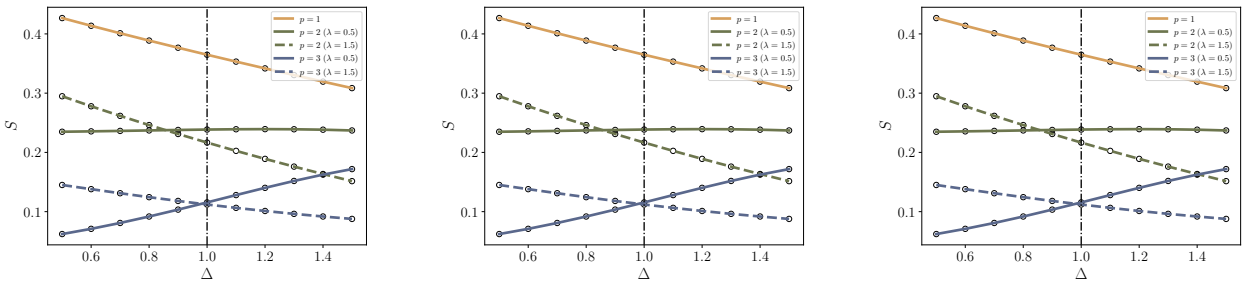


Figure S3: Entanglement entropy versus λ with the RL-ansatz disentangling circuit for TFIM with $N = 8$ of (i) under different initial values. (A): uniform distribution with mean value 0 (B): uniform distribution with mean value $\pi/2$ (C): uniform distribution with mean value π. The solid line denotes the disentangling circuit training from $\lambda = 0.5$; the dashed line denotes the disentangling circuit training from $\lambda = 1.5$.

As Fig. S3 and Fig. S5 show, the algorithm's performance with the TFIM is robust under different conditions. The same conclusion can be drawn in Fig. S4 and Fig. S6 with the XXZ model. In other words, we may also look for the phase transition at which the training is most independent of chosen conditions.

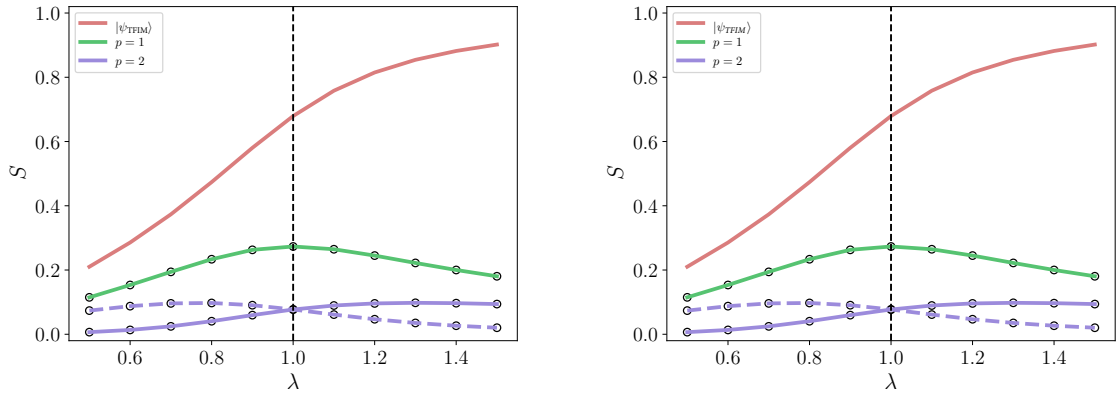
S-IV Concurrence in quantum phase transition

In previous studies, a partial relationship between entanglement and quantum phase transitions has been found [3, 5, 9, 69]. Here we show the relationship between concurrence and quantum phase



(a) (b) (c)

Figure S4: Entanglement entropy versus Δ with the RL-ansatz disentangling circuit for XXZ model with $N = 8$ of (i) under different initial values. (A): uniform distribution with mean value 0 (B): uniform distribution with mean value $\pi/2$ (C): uniform distribution with mean value π. The solid line denotes the disentangling circuit training from $\Delta = 0.5$; the dashed line denotes the disentangling circuit training from $\Delta = 1.5$.



(a) (b)

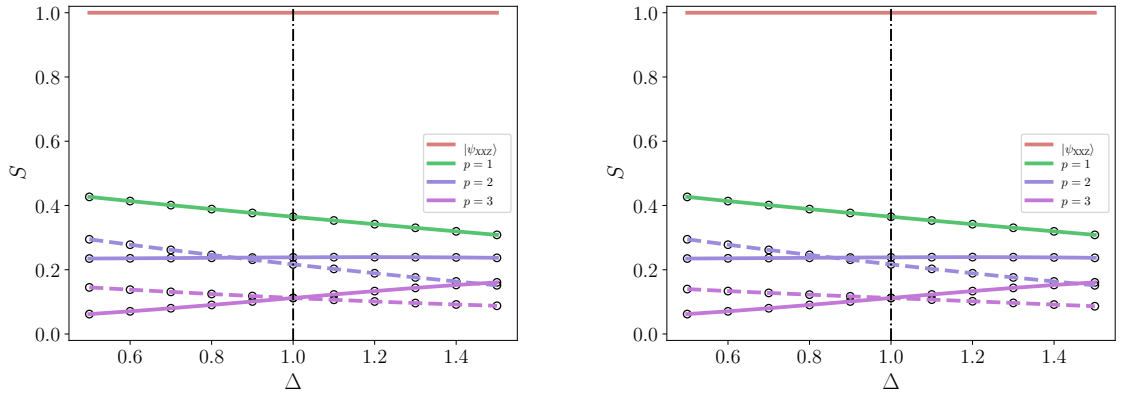
Figure S5: Entanglement entropy versus λ with the RL-ansatz disentangling circuit for TFIM with $N = 8$ of (ii) under different training states. (A): ground state with $\lambda = 0.6, 1.3$ (B): ground state with $\lambda = 0.7, 1.4$. The solid line denotes the disentangling circuit training from $\lambda = 0.6, 0.7$; the dashed line denotes the disentangling circuit training from $\lambda = 1.3, 1.4$.

transition, which is used to compare with our method. The concurrence [70] between sites i and j is defined as

$$C(i,j) = \max\left\{r_1(i,j) - r_2(i,j) - r_3(i,j) - r_4(i,j), 0\right\}, \tag{S27}$$

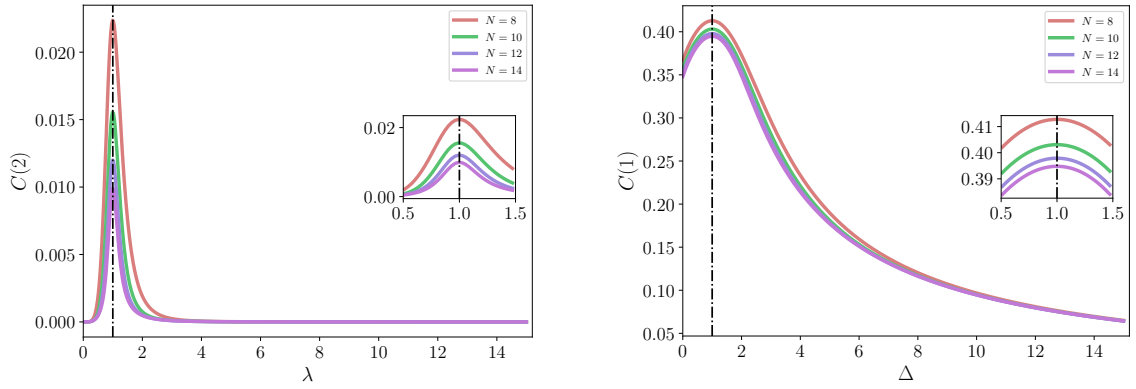
where $r_\alpha(i,j)$ are the square roots of the eigenvalues of the product matrix $\rho(i,j)\tilde{\rho}(i,j)$ in descending order; the spin flipped matrix is defined as $\tilde{\rho} = \sigma^y \bigotimes \sigma^y \rho^* \sigma^y \bigotimes \sigma^y$.

Based on previous studies, we know that the next nearest neighbor concurrence $C(2)$ is related to the quantum phase transition of the TFIM model. The relationship is shown in Fig. S7 (a). In the XXZ model, the nearest-neighbor concurrence $C(1)$ is related to its quantum phase transition, and the relationship is shown in Fig. S7 (b). From Fig. S7 (a), it can be seen that as the spin number N of the system increases, the mutation region of concurrence becomes narrower near the critical point. However, this method can only give a phase transition range in a finite size of systems and cannot precisely determine the critical point. As can be seen in Fig. S7 (b), the maximum value of concurrence is at the critical point of the XXZ model. However, unlike TFIM, it is independent of the system size. We can conclude that there is no general method to observe only the change in the quantity of entanglement to determine the quantum phase transition.



(a) (b)

Figure S6: Entanglement entropy versus Δ with the RL-ansatz disentangling circuit for XXZ model with $N = 8$ of (ii) under different training states. (A): ground state with $\Delta = 0.6, 1.3$ (B): ground state with $\Delta = 0.7, 1.4$. The solid line denotes the disentangling circuit training from $\Delta = 0.6, 0.7$; the dashed line denotes the disentangling circuit training from $\Delta = 1.3, 1.4$.



(a) (b)

Figure S7: Concurrence as function of parameters. (a): next nearest neighbor concurrence $C(2)$ in TFIM with N sites (b): nearest-neighbor concurrence $C(1)$ in XXZ model with N sites.

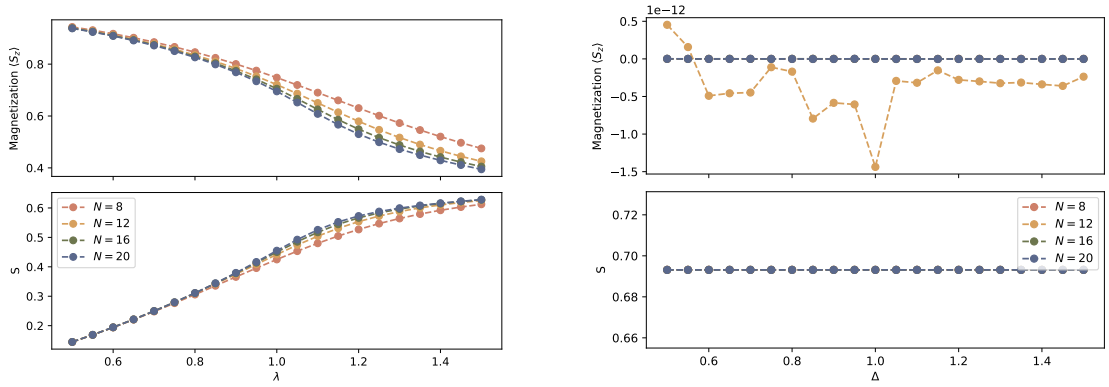
S-V DMRG in phase identification

Fig. S8 presented illustrates the results obtained from the Density Matrix Renormalization Group (DMRG) method in determining phase transitions in the Transverse Field Ising Model (TFIM) and the XXZ model. We evaluated the performance of two critical physical quantities in the DMRG ground state. The first quantity is the magnetization, which corresponds to the average of the pauli operator observations in the z-direction over all sites. The second quantity is the entanglement entropy, calculated by considering the partition between half of the chain and the remaining half.

From the analysis of the figure, it becomes apparent that pinpointing the precise phase transition point using the DMRG method is a challenging task. Nevertheless, the primary advantage of employing DMRG lies in its ability to calculate the ground state of larger systems with remarkable accuracy.

S-VI Local Order Parameters and Phase Transition Indicators

For the finite 1D transverse-field Ising model (TFIM) presented in our example, the limited system size introduces significant boundary effects. As illustrated in Fig. S9(a), magnetization changes smoothly as a function of the transverse field λ, contrasting with the sharp transitions observed in infinite

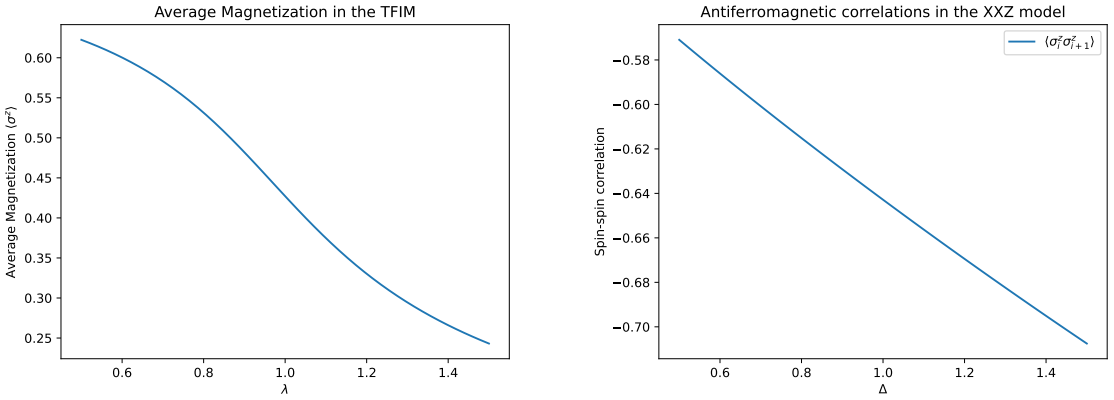


(a) (b)

Figure S8: Ground state magnetization and entanglement entropy for (a) Transverse Field Ising Model (TFIM) and (b) XXZ model with varying system sizes. The top subplots display the magnetization $\langle S_z \rangle$ as a function of the system parameter, while the bottom subplots show the entanglement entropy S as a function of the system parameter.

systems. This smoothness arises from the lack of a true phase transition in finite systems.

Turning to the finite 1D XXZ spin chain, we agree that the antiferromagnetic (AFM) correlation function $\langle S_i^z S_{i+1}^z \rangle$ conveys important information about the system's phase. It's crucial to recognize that the transition at $\Delta = 1$, as discussed in our manuscript, is of the Kosterlitz-Thouless (KT) type. This infinite order phase transition does not feature a discontinuous change in any local order parameter, which can render the transition's signatures less pronounced in finite-size systems. While Fig. S9(b) demonstrates that the AFM correlation function for pinpointing the transition at $\Delta = 1$ remains a challenge.



(a) (b)

Figure S9: compare the reduced ground states with 2 qubits via magnetization or correlation function with the models of size $N = 8$ in our manuscript.

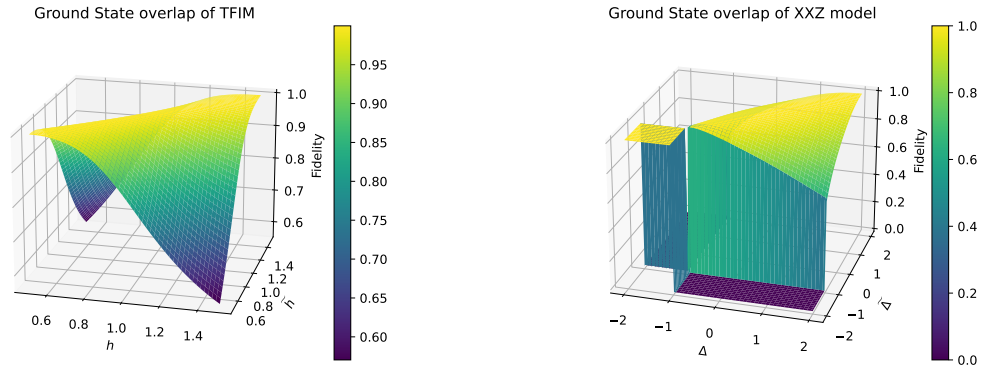
The findings discussed above align with the Density Matrix Renormalization Group (DMRG) results we presented in our previous analysis (refer to Fig. S8). Although the order parameter in a finite system provides valuable insights into the system's state and its proximity to phase transitions, it alone cannot precisely pinpoint the critical point due to the inherent limitations posed by finite size effects. A more accurate determination of the critical point can be achieved through the application of scaling laws and by examining the behavior of the system across various sizes.

Furthermore, our approach serves as an exemplary case of how our methodology can be differentiated from traditional methods used in identifying quantum phase transitions. This demonstrates that our technique not only complements existing phase transition identification methods but also enhances

the robustness and depth of analysis in studying critical phenomena in finite systems.

S-VII Ground state overlap measures

Additionally, we evaluated the ground state overlap for the two models under investigation. As depicted in Fig. S-VII, similar to the expectation or correlation methods, this approach does not yield precise information about the phase transition point in the models we analyzed. Concerning alternative methods that employ entanglement to identify phase transitions, such as the concurrence method outlined in the supplementary material, we suggest that relying exclusively on changes in entanglement measures is not a universally dependable indicator of quantum phase transitions. This observation underscores the necessity for a multifaceted approach when analyzing and identifying critical points in quantum systems.

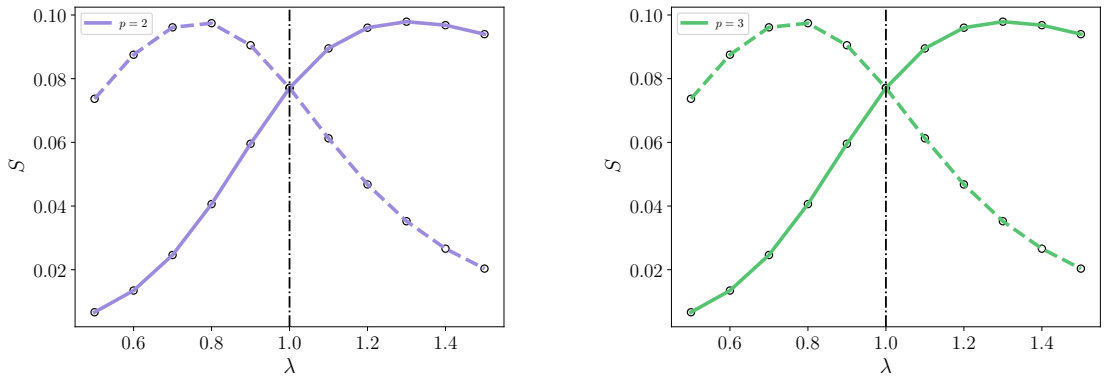


(a) (b)

Figure S10: Compare the ground states via Ground state overlap with the models in our manuscript.

S-VIII The convergency of RL-ansatz disentangling circuit

As Fig. S11 and Fig. S12 shown, the performance of the RL-ansatz disentangling circuit is convergence with the increase of circuit layers.



(a) (b)

Figure S11: Entanglement entropy versus λ for 8-site TFIM. P is the number of circuit layers. (a): circuit layers of $p = 2$ (b): circuit layers of $p = 3$. The solid line denotes the disentangling circuit training from $\lambda = 0.5$; the dashed line denotes the disentangling circuit training from $\lambda = 1.5$.

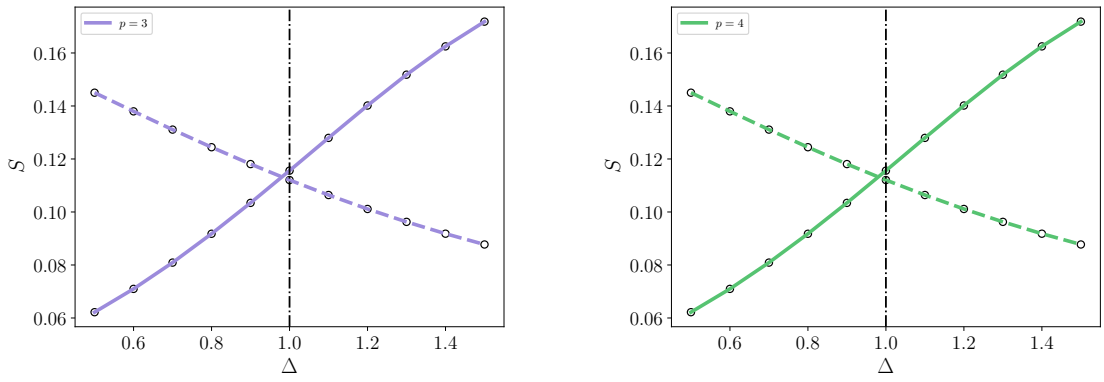


Figure S12: Entanglement entropy versus Δ for 8-site XXZ model. P is the number of circuit layers (a): circuit layers of $p = 3$ (b): circuit layers of $p = 4$. The solid line denotes the disentangling circuit training from $\Delta = 0.5$; the dashed line denotes the disentangling circuit training from $\Delta = 1.5$.

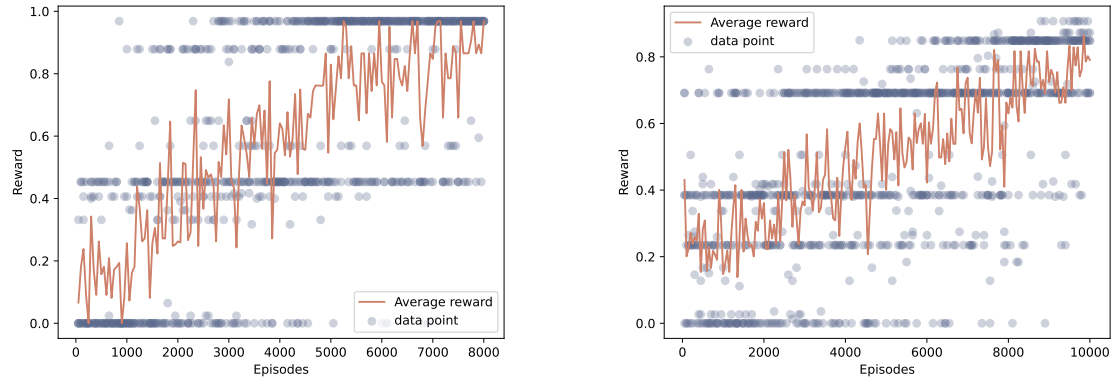


Figure S13: Rewards for a reinforcement learning (RL) agent over a series of episodes in five different runs, applied to (a) Transverse Field Ising Model (TFIM) with 8 sites and a quantum circuit layer $p = 2$, and (b) XXZ model with 8 sites and a quantum circuit layer $p = 3$. Individual data points are depicted as semi-transparent blue dots, while the average reward curve is represented by a solid red line.

Fig. S13 presents the rewards monitored over episodes for 5 agents applied to the TFIM and XXZ models. We observe that the average reward across the 5 agents increases, indicating that our algorithm successfully learns to design disentanglement quantum circuits. However, it is crucial to note that the convergence of a DQN is not solely determined by the rewards. Due to the off-line learning and epsilon-greedy exploration properties in DQN, rewards are influenced by both the current policy and the experiences stored in the memory buffer. During the learning process, a random action selection weight typically starts with a high value and gradually decays over time. This enables the agent to explore the environment more during the initial stages and rely more on its learned Q-values as it gains experience. Consequently, the rewards might not exhibit smooth behavior.

S-IX Hyper-parameters 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 6 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 S1.

Table S1: Training Hyper-Parameters

Hyper-parameter	Values
Neurons in neural network	$\{512, 512, 512, 512, 512, 512\}$
Minibatch size	120
Replay memory size	100000
Learning rate	0.0001
Update period	100
Reward decay γ	0.99