Efficient bipartite entanglement detection scheme with a quantum adversarial solver

Xu-Fei Yin,1,2,3 Yuxuan Du,4,5 Yue-Yang Fei,1,2,3 Rui Zhang,1,2,3 Li-Zheng Liu,1,2,3 Yingqiu Mao,1,2,3 Tongliang Liu,5 Min-Hsiu Hsieh,6 Li Li,1,2,3 Nai-Le Liu,1,2,3 Dacheng Tao,4,5 Yu-Ao Chen,1,2,3 and Jian-Wei Pan1,2,3

1Hefei National Laboratory for Physical Sciences at the Microscale and Department of Modern Physics, University of Science and Technology of China, Hefei 230026, China
2Shanghai Branch, CAS Center for Excellence in Quantum Information and Quantum Physics, University of Science and Technology of China, Shanghai 201315, China
3Shanghai Research Center for Quantum Sciences, Shanghai 201315, China
4JD Explore Academy, Beijing 101111, China
5School of Computer Science, Faculty of Engineering, University of Sydney, Sydney 2006, Australia
6Hon Hai Quantum Computing Research Center, Taipei 114, Taiwan

The recognition of entanglement states is a notoriously difficult problem when no prior information is available. Here, we propose an efficient quantum adversarial bipartite entanglement detection scheme to address this issue. Our proposal reformulates the bipartite entanglement detection as a two-player zero-sum game completed by parameterized quantum circuits, where a two-outcome measurement can be used to query a classical binary result about whether the input state is bipartite entangled or not. In principle, for an $N$-qubit quantum state, the runtime complexity of our proposal is $O(poly(N)T)$ with $T$ being the number of iterations. We experimentally implement our protocol on a linear optical network and exhibit its effectiveness to accomplish the bipartite entanglement detection for 5-qubit quantum pure states and 2-qubit quantum mixed states. Our work paves the way for using near-term quantum machines to tackle entanglement detection on multipartite entangled quantum systems.

Introduction.—Quantum entanglement, the gem of quantum computation and information processing, allows quantum computers to tackle certain problems beyond the traditionally possible. The representative examples include quantum phase estimation for factoring large integers, quantum cryptography for secure communication, and quantum secret sharing. In the last decade, tremendous progress has been achieved to fabricate various physical systems to manipulate entanglement. Existing entanglement detection methods can be divided into three main groups. In the first group, the positive partial transposition method is generally efficient but only provides sufficient conditions for up to $2 \times 3$ dimensional cases; the entanglement witness approach only distinguishes specific entangled states from separable ones; and the method based on self-testing only effectively identifies product states. These methods can therefore only be regarded as “incomplete.” Conversely, methods in the second group, such as symmetric extension and the improved semidefinite programming hierarchy, are complete in the sense that they can correctly identify any entangled state at any dimension, but the runtime cost is extremely expensive and exponentially scales with the qubits count. In addition to these deterministic methods, classical machine learning techniques have also provided novel insights into the entanglement detection. For example, entanglement classifiers based on deep neural network can approximately recognize all entangled states with a high accuracy. Current learning algorithms oriented to entanglement tests belong to the supervised and static learning paradigm, where a labeled dataset containing thousands of quantum states is required to train classifiers. The adopted supervised learning techniques prohibit their scalability, since labeling and processing quantum states involving a large number of qubits is impractical on classical devices. Moreover, it has been recently shown that static algorithms are susceptible to adversarial attack and consequently false classifications, limiting their applicability. The above observations imply that there is currently no efficient solution to (approximately) identify the entanglement of hundreds of qubit states due to the intrinsic computational hardness of the problem and the exponentially large Hilbert space. Hence, the certification of high-dimensional entanglement is a major challenge in quantum information processing.

Here we devise an efficient scheme to tackle entanglement detection with scalability. Our proposal, the quantum adversarial bipartite entanglement detection scheme, is versatile and only requires modest resources, such that it can be efficiently built on noisy intermediate-scale quantum (NISQ) machines. Furthermore, the proposed scheme is scalable and can be used to approximately recognize all quantum entangled states represented by any arbitrarily large qubit count, since the required computational complexity polynomially scales with the number of qubits. Mathematically, given an $N$-qubit state, the runtime complexity of our proposal is $O(poly(N)T)$, where $T$ refers to the total number of iterations. To our best knowledge, our theoretical study for the first time accomplishes high-dimensional entanglement testing of thousands of qubits with an acceptable computational overhead. Furthermore, our proposal benefits from robustness to adversarial attack, ensuring its practical applicability and reliability. Our framework represents a new approach that uses NISQ devices to tackle...
practical quantum information processing problems with exponential advantages.

The quantum adversarial bipartite entanglement detection scheme.—Before describing the implementation details, we first introduce how to reformulate the bipartite entanglement detection as a two-player zero-sum game [22]. This reformulation represents a precondition for devising efficient quantum algorithms for bipartite entanglement detections. Recall that all quantum states form a convex set and all separable states form a convex subset [23]. The convex property implies that given any separable-entangled state pair, there always exists a hyperplane that partitions these two states on opposite sides. This geometric observation allows us to set up a two-player zero-sum game [23].

We have a convex subset, while the second player, i.e., the discriminator $D$, is trainable two-outcome positive operator valued measure $U$. The aim of $D$ is to discriminate $\sigma_G$ from $\rho$ by finding a hyperplane that places $\sigma_G$ and $\rho$ on opposite sides, while the aim of $G$ is to produce $\sigma_G$, which approximates $\rho$ to fool $D$ in the sense that $\sigma_G$ and $\rho$ stay on the same side. Remarkably, the restricted power of $G$ means that the Nash equilibrium, the status in which no player can improve their individual gain by choosing a different strategy, can be conditionally achieved if and only if $\rho$ is separable [22]. This result converts the bipartite entanglement detection into a binary outcome optimization task, i.e., the achievable or unachievable equilibrium correspond to $\rho$ being separable or entangled.

We next elaborate on the quantum adversarial bipartite entanglement detection scheme, which accomplishes the zero-sum game under the NISQ setting. Specifically, denoted $\rho \in \mathbb{C}^{2^N \times 2^N}$ as a bipartite state that is composed of two subsystems $A$ and $B$. The task is to identify whether there exists entanglement between these two subsystems of $\rho$. To achieve this goal, the generator $G$ and the discriminator $D$ used in our proposal refer to two trainable unitaries $U_G(\theta)$ and $U_D(\gamma)$, which are implemented by parameterized quantum circuits [27, 28]. Mathematically, we have $U_G(\theta) = \prod_{l=1}^{L_1} U_l(\theta)$ and $U_D(\gamma) = \prod_{l=1}^{L_2} U_l(\gamma)$, where $L_1$ ($L_2$) refers to the number of blocks in $U_G(\theta)$ ($U_D(\gamma)$) and each block $U_l(\theta)$ ($U_l(\gamma)$) has an identical arrangement of quantum gates. The generated state is $\sigma_G = \text{Tr}_E(U_G(\theta)(|0\rangle\langle 0|)\otimes^N U_G(\theta)\dagger)$, where $N = N + N_A$, the subscript “$A$” represents the ancillary system that is formed by $N_A$ qubits, and $\sigma_G$ is obtained by partial tracing this ancillary system. When $\rho$ is a pure state, it is sufficient to set $N_A = 0$. Note that to ensure that $U_G(\theta)$ can only prepare the states that are always separable between subsystems $A$ and $B$, we impose a restriction on $U_l(\theta)$, where no two-qubit gate is allowed to span subsystems $A$ and $B$. The discriminator $D$ refers to a trainable two-outcome positive operator valued measure $M_D$, i.e., $M_D = U_D(\gamma)(I \otimes E)U_D(\gamma)^\dagger$, where $I \otimes E$ is the measurement operator with $E = |0\rangle\langle 0|$. In the above notation, the loss function of the zero-sum game yields

$$\max_{\theta} \min_{\gamma} \mathcal{L}(\rho, \sigma_G) = \frac{\text{Tr}(M_D \rho)}{2} + \frac{\text{Tr}(M_D^2 \sigma_G)}{2},$$

where $M_D^2 = I - M_D$. The optimization of $\mathcal{L}$ amounts to iteratively updating $\theta$ and $\gamma$ with $T$ iterations. This updating process can be effectively completed by a classical optimizer (see Supplemental Material A and B in [29]). The theoretical foundation of our proposal is guaranteed by the following theorem, whose proof is provided in Supplemental Material C [29].

**Theorem 1** The zero-sum formulated in Eq. (1) has the equilibrium value $1/2$ if and only if the given state $\rho$ is separable. In this case, the generated state is identical to the given state, i.e., $\sigma_G = \rho$.

As highlighted by Theorem 1, the criterion for identifying the entangled states on NISQ devices shows that: during $T$ iterations, if the loss $\mathcal{L}$ converges to $1/2$ within a tolerable error, the state $\rho$ will be labeled as separable; otherwise, the state $\rho$ will be labeled as entangled. The runtime complexity of our scheme is $O((L_1 + L_2)N^3)$. This computational efficiency has two main origins. First, our proposal does not encounter read-in and read-out bottlenecks. Moreover, the gradient information for each parameter and the evaluation of $\mathcal{L}$ can be achieved in $O(1)$ runtime, which indicates that the optimization of each iteration takes $O((L_1 + L_2)N^3)$ runtime (see Supplemental Material A [29]). Second, unlike prior supervised learning based entanglement classifiers, our proposal does not require a preprocessing procedure to prepare labeled quantum states that may take $O(2^N)$ runtime. The low runtime cost assures the scalability of our proposal.

**Experimental setup.—**To benchmark the performance of the proposed scheme, we use linear optical circuits and entangled photon pairs to demonstrate the bipartite entanglement detection for two 5-qubit pure states and two 2-qubit mixed states, respectively.

The construction of these states is as follows. The first pure quantum state to be recognized is partially separable $\rho_{G_{23}} = |G_2\rangle\langle G_2| \otimes |G_3\rangle\langle G_3|$, where $|G_N\rangle = \frac{1}{\sqrt{2}}(|H\rangle \otimes^N + |V\rangle \otimes^N)$ denotes the $N$-qubits Greenberger-Horne-Zeilinger state. The second pure quantum state to be recognized is fully entangled $\rho_{G_5} = |G_5\rangle\langle G_5|$. Note that following the convention of linear optics, the qubit state $\{0, 1\}$ is represented by the polarization degree of freedom $\{H, V\}$, where $H$ and $V$ refer to the horizontal and vertical polarizations, respectively. Moreover, the first mixed quantum state to be recognized is separable $\rho_S = E^{(S)}_p(|\psi_S\rangle\langle \psi_S|)$, where $|\psi_S\rangle = \frac{1}{\sqrt{2}}(|HH\rangle + |VV\rangle)$ is the separable pure state and $E^{(S)}_p : |\psi_S\rangle\langle \psi_S| \rightarrow |\psi_S\rangle\langle \psi_S| - (1 - p)(|H\rangle\langle V| + |V\rangle\langle H|)$.
FIG. 1: The implementation of the quantum adversarial bipartite entanglement detection scheme based on quantum circuit models. (a) The bipartite entanglement detection between pure state \( \rho_{G_3} \) and \( \rho_{G_{23}} \) is shown in the left and right panels, respectively. In particular, the pink and green regions separately represent the state preparation of \( \rho_{G_3} \) and \( \rho_{G_{23}} \). The purple region represents generator \( U_G \), which is used to prepare \( \sigma_G \) or \( \sigma_D \), the blue region represents discriminator \( U_D \) which is used to create a hyperplane that lies in the pure state space. (b) bipartite entanglement detection between pure state \( |\psi\rangle \) and \( |\psi\rangle \). The classical optimizer exploits the measured results to update trainable parameters in \( U_G \) and \( U_D \). The classical optimizer exploits the measured results to update trainable parameters in \( U_G \) and \( U_D \).

- For the mixed state case, the block number for \( U_G(\theta) \) and \( U_D(\gamma) \) is set as \( L_1 = 2 \) and \( L_2 = 2 \). \( U_G \) includes 15 parameterized single-qubit gates, two controlled-Z (CZ) gates, and a 6 parameterized deterministic controlled unitary gate. For the mixed state case, the block number for \( U_G(\theta) \) and \( U_D(\gamma) \) is set as \( L_1 = 1 \) and \( L_2 = 2 \). \( U_G \) includes 6 parameterized single-qubit gates. To ensure that the generator can only produce separable states, for the pure state case, there is no two-qubit gate between the first two and the last three qubits in \( U_G \); for the mixed state case, \( U_G \) only includes single-qubit gates without two-qubit gates. In addition, for both cases, the discriminator \( U_D \) consists of 12 parameterized single-qubit gates and a CZ gate. Furthermore, in the mixed state case, to guarantee the same noise environment between the generated state and the input state, the quantum channel \( \mathcal{E}_p^{(S)} \) is also employed in the generator with \( p = 0.8 \).}

Experiment results.—The experimental identification of the entanglement of pure state \( \rho_{G_{23}} \) is shown in Fig. 2(a). The loss \( \mathcal{L}(\theta^{(1)}, \gamma^{(1)}) \) at the initial step is only \( 0.3784 \pm 0.00173 \), which is far away from the Nash equilibrium of 0.5. With an increased number of iterations, the loss quickly converges to the Nash equilibrium, i.e., \( \mathcal{L}(\theta^{(10)}, \gamma^{(10)}) \) reaches \( 0.4820 \pm 0.0180 \) at the 10th iteration. Moreover, the two terms \( \text{Tr}(M_D \sigma_G^{(10)}) \) and \( \text{Tr}(M_D \sigma_{G_{23}}) \) tend to be equivalent, i.e., the state distance \( D(\sigma_G^{(10)}, \sigma_{G_{23}}) = \frac{1}{2} |\text{Tr}(M_D \sigma_G^{(10)}) - \text{Tr}(M_D \sigma_{G_{23}})| \) equals to \( 0.0360 \pm 0.0284 \). Given access to the reconstructed state, we obtain the fidelity \( F(\rho_{G_{23}}, \sigma_G^{(1)}) \) is only \( 0.5299 \pm 0.0256 \), while \( F(\rho_{G_{23}}, \sigma_G^{(10)}) \) increases to \( 0.9202 \pm 0.0204 \). These results echo with Theorem 4 for the separable case.

The experimental results for identifying the entanglement of \( \rho_{G_3} \) are illustrated in Fig. 2(b). In contrast to the partial separable case, the loss \( \mathcal{L} \) for the state \( \rho_{G_3} \), highlighted by the purple dashed line, converges to a point that is far away from the Nash equilibrium after 10 iterations. Particularly, although the loss \( \mathcal{L}(\theta^{(1)}, \gamma^{(1)}) \) at the initial iteration is \( 0.4468 \pm 0.0230 \), it decreases to \( 0.2908 \pm 0.0250 \) at the 10th iteration. Moreover, the distance \( D(\sigma_G^{(10)}, \sigma_{G_3}) \) is \( 0.1560 \pm 0.0280 \). We apply full tomographic measurements to reconstruct the states \( \rho_{G_3} \) and \( \sigma_G^{(1)} \) with \( t = 1 \) and \( t = 10 \). The collected results show that the reconstructed state density matrix between \( \rho_{G_3} \) and \( \sigma_G^{(10)} \) is evidently disparate. By lever-
The experimental results for identifying the entanglement of \( \rho_{G23}, \rho_{G1}, \rho_S \) and \( \rho_E \). The results between (a) \( \sigma_G \) and \( \rho_{G23} \), (b) \( \sigma_G \) and \( \rho_{G1} \), (c) \( \sigma_G \) and \( \rho_S \), (d) \( \sigma_G \) and \( \rho_E \). The labels ‘L’ and ‘P\( \rho_{G} \)’ refer to the loss \( \mathcal{L}(\theta^{(i)}, \gamma^{(i)}) \) and the terms \( \text{Tr}(M_D \sigma_G^{(i)}) \) in Eq. (1). The labels ‘P\( \rho_G \)’, ‘P\( \rho_{G1} \)’, ‘P\( \rho_S \)’ and ‘P\( \rho_E \)’ refer to the terms \( \text{Tr}(M_D \rho_{G23}), \text{Tr}(M_D \rho_{G1}), \text{Tr}(M_D \rho_S) \) and \( \text{Tr}(M_D \rho_E) \) in Eq. (1), respectively. The red dash line indicates when loss \( \mathcal{L}(\theta^{(i)}, \gamma^{(i)}) \) converges to 0, ‘Nash equilibrium’ is reached.

We next apply our proposal to recognize the entanglement of the state \( \rho_S \) and \( \rho_E \). The experimental identification of the entanglement of \( \rho_S \) is shown in Fig. 2(c). Specifically, at the initial step, the loss \( \mathcal{L}(\theta^{(1)}, \gamma^{(1)}) \) is only 0.2299 \( \pm \) 0.0030, which is far away from the Nash equilibrium of 0.5. Analogous to the pure state case, the loss quickly converges to the Nash equilibrium, i.e., \( \mathcal{L}(\theta^{(20)}, \gamma^{(20)}) \) reaches 0.4969 \( \pm \) 0.0033 at the 20th iteration. Moreover, the two terms \( \text{Tr}(M_D \sigma_G^{(20)}) \) and \( \text{Tr}(M_D \rho_S) \) tend to be equivalent, i.e., the experimental results lead to the distance \( D(\sigma_G^{(20)}, \rho_S) = 0.0031 \pm 0.0032 \).

Through applying quantum state tomography [35, 36], we further obtain the fidelity \( F(\rho_S, \sigma_G^{(1)}) \) is only 0.5546 \( \pm \) 0.0013, while \( F(\rho_S, \sigma_G^{(20)}) \) increases to 0.9802 \( \pm \) 0.0004. The experimental results for identifying the entanglement of \( \rho_E \) are illustrated in Fig. 2(d). In contrast to the separable case, the loss \( \mathcal{L} \) for the state \( \rho_E \) converges to a point that is far away from the Nash equilibrium after 20 iterations. In particular, although the loss \( \mathcal{L}(\theta^{(i)}, \gamma^{(i)}) \) at the initial iteration is 0.4578 \( \pm \) 0.0037, it decreases to 0.2326 \( \pm \) 0.0028 at the 20th iteration with \( D(\sigma_G^{(20)}, \rho_E) = 0.2675 \pm 0.0028 \). By leveraging the reconstructed states, we obtain the fidelity at \( t = 1 \) and \( t = 20 \) iterations are \( F(\rho_S, \sigma_G^{(1)}) = 0.4994 \pm 0.0011 \) and \( F(\rho_S, \sigma_G^{(20)}) = 0.7094 \pm 0.0020 \), respectively. The above observations confirm the feasibility of our proposal in the mixed state scenario. Refer to Supplementary D [29] for the omitted experimental results.

To exhibit that our proposal outperforms entanglement witness method [3] when no prior knowledge is available, we defer the comparison of our proposal with entanglement witness method to identify \( \rho_S \) and \( \rho_E \) in mixed scenario. Concretely, let \( W_N = \frac{1}{N} \mathbb{1}_{2^N} - |G_N \rangle \langle G_N| \) be the employed multipartite entanglement witness operator, where \( \mathbb{1}_{2^N} \) is the \( 2^N \)-dimensional identity matrix. The criterion of entanglement witness methods is assigning the input state as entangled if \( \text{Tr}(W_N \rho) < 0 \). The experimental results of applying \( W_N \) to identify \( \rho_S \) and \( \rho_E \) are shown in Fig. 3. Specifically, the measured result of \( \langle W_E \rangle \) and \( \langle W_S \rangle \) for the state \( \rho_E \) and \( \rho_S \) is 0.4010 \( \pm \) 0.0007 and 0.2532 \( \pm \) 0.0015, which are both much larger than 0. Consequently, the entanglement witness operator \( W_E \) provides a wrong prediction for \( \rho_E \).

Remark. We heuristically estimate when the quantum adversarial bipartite entanglement detection scheme advances other complete methods and hence demonstrates potential quantum advantages. Also, our scheme can be effectively adapted to the fault-tolerant setting with provable convergence [35, 36]. Concisely, for an \( N \)-qubit mixed state \( \rho \), by setting the training rounds as \( T \), our scheme completes the bipartite entanglement detection with the error \( |\mathcal{L}(\sigma_G, \rho) - 1/2| \leq 3 \sqrt{N/T} \). Refer to Supplemental Material H and Theorem 2 in Supplemental Material 1 [29] for details, respectively. Moreover, we conduct extensive experiments to analyze the robustness of our protocol. Given a set of 2-qubit mixed states, our protocol achieves 83.36% accuracy for identifying the bipartite entanglement in the measurement of the confusion matrix (See Supplemental Material E [29] for details).
The most exciting future work is extending our proposal to efficiently quantify unknown entanglement following these measures, which enables us to seek diverse potential applications of quantum computers with evident advantages.

This work was supported by the National Natural Science Foundation of China (Grant No. 11975222), Shanghai Municipal Science and Technology Major Project (Grant No. 2019SHZDZX01), and Chinese Academy of Sciences and the Shanghai Science and Technology Development Funds (Grant No. 18JC1414700).

X.-F Y, Y.D., and Y.-Y. F. contributed to this work equally.
programmable superconducting processor, npj Quantum Inf. (2020).

[49] H. R. Grimsley, S. E. Economou, E. Barnes and N. J. Mayhall, An adaptive variational algorithm for exact molecular simulations on a quantum computer, Nat. Commun. 10, 3007 (2019).

[50] Y. Du, T. Huang, S. You, M.-H. Hsieh, D. Tao, Quantum circuit architecture search: Error mitigation and trainability enhancement for variational quantum solvers, arXiv:2010.10217 (2020).

[51] J. Yao, L. Lin, and M. Bukov, Reinforcement Learning for Many-Body Ground-State Preparation Inspired by Counterdiabatic Driving, Physical Review X. 11, 031070 (2021).