Near-Term Efficient Quantum Algorithms for Entanglement Analysis

Ranyiliu Chen⬤,[1,2] Benchi Zhao⬤,[1,3] and Xin Wang⬤[1,3,*]

[1]*Institute for Quantum Computing, Baidu Research, Beijing 100193, China*

[2]*QMATH, Department of Mathematical Sciences, University of Copenhagen, Universitetsparken 5, Copenhagen 2100, Denmark*

[3]*Thrust of Artificial Intelligence, Information Hub, Hong Kong University of Science and Technology (Guangzhou), Nansha, China*

Entanglement plays a crucial role in quantum physics and is the key resource in quantum information processing. However, entanglement detection and quantification are believed to be hard due to the operational impracticality of existing methods. This work proposes three near-term efficient algorithms that exploit the hybrid quantum-classical technique to address this difficulty. The first algorithm finds the Schmidt decomposition—a powerful tool to analyze the properties and structure of entanglement—for bipartite pure states. While the logarithm negativity can be calculated from the Schmidt decomposition, we propose the second algorithm to estimate the logarithm negativity for bipartite pure states, where the width of the parameterized quantum circuits is further reduced. Finally, we generalize our framework for mixed states, leading to our third algorithm that detects entanglement on specific families of states, and determines distillability in general. All three algorithms share a similar framework where the optimizations are accomplished by maximizing a cost function utilizing local parameterized quantum circuits, with better hardware efficiency and practicality compared to existing methods. The experimental implementation on Quantum Leaf using the Institute of Physics, Chinese Academy of Sciences superconducting quantum processor exhibits the validity and practicality of our methods for analyzing and quantifying entanglement on near-term quantum devices.

I. INTRODUCTION

Quantum entanglement [1] is the most nonclassical manifestation of quantum mechanics. As the core ingredient in quantum information, quantum entanglement has found use in a variety of areas, including quantum cryptography [2,3], quantum chemistry [4,5], quantum machine learning [6,7], and quantum communication [8–10].

Detecting the entanglement of a given unknown multiparty state is not simple work, let alone quantifying the entanglement. Positive (but not completely positive) map criteria, such as the well-known positive partial transpose (PPT) criterion [11], can distinguish some entangled states by their nonpositivity after local positive maps. However, this method is less practical compared to its elegant representation from at least two points of view: the positive maps are not physically implementable, and determining positivity relies on the tomography of states [12], which is very costly. Wang *et al.* [13] proposed decomposing the nonphysical positive maps into physical Pauli operations, and estimated the positiveness using hybrid quantum-classical computation [14]. This is an effective way to detect entanglement, but the required resources in the decomposition of positive maps is still worrying.

The Schmidt decomposition [15] is a fundamental and powerful tool for analyzing the entanglement in bipartite systems. According to the Schmidt decomposition, a pure bipartite state is expressed as a simpler combination of the tensor products of two orthonormal bases of subspaces, that is, $|\psi\rangle_{AB} = \sum_j c_j |u_j\rangle_A \otimes |v_j\rangle_B$, where subscripts A and B denote the corresponding subsystems. The positive coefficient in Schmidt decomposition c_j is known as the Schmidt coefficient, and the number of Schmidt coefficients is called the Schmidt rank or Schmidt number, determining whether this pure state is entangled or not: if the Schmidt rank equals one, it is safe to say that there is no entanglement between the two parties. Additionally, for entangled states, entanglement measures such as Von Neumann entropy and logarithm negativity are determined by the Schmidt coefficients [16].

The most straightforward approach to do Schmidt decomposition on a given quantum system is state tomography [12] followed by a singular value decomposition on a classical computer. Nevertheless, it is known that state tomography consumes a huge amount of state copies, and singular value decomposition demands exponential

*felixxinwang@hkust-gz.edu.cn

024071-1

classical storage in the number of qubits. Alternatively, several variational quantum algorithms [17,18] have been proposed to leverage the power of hybrid quantum-classical computing. In Ref. [17], the Schmidt coefficients are read out from the variationally diagonalized marginal states. In Ref. [18], local unitaries are performed and optimized on each party, and when the measurement of each party coincides, the Schmidt coefficients can be evaluated.

In this work, we present variational algorithms for entanglement analysis that have better stability or efficiency than existing approaches. In our first algorithm for Schmidt decomposition for pure states, bilocal parameterized quantum circuits (PQCs) or quantum neural networks act on the two subsystems (inspired by Refs. [14,18]), followed by a depth-2 subcircuit that calculates the fidelity with another elaborated entangled state. Von Neumann's trace inequality [19] guarantees that, when this fidelity is maximized through tuning the circuit parameters, the Schmidt decomposition is accomplished. By exploiting symmetry, we further introduce our second algorithm to estimate the logarithm negativity of bipartite pure states, where the reduction of one side of the PQC improves its efficiency mostly in classical optimization. Taking general mixed states into consideration, we propose the third algorithm for entanglement detection. We show that it detects entanglement for certain families of mixed states of practical interest, and identifies distillability under a particular protocol. Numerical simulation and quantum device implementation in the (up to) eight-qubit and two-qubit cases are undertaken, respectively, illustrating the practicality and validity of our algorithms in noisy intermediate-scale quantum (NISQ) devices. Our algorithms thus notably provide efficient and practical ways to analyze and quantify entanglement on near-term quantum devices.

The rest of the paper is organized as follows. Section II gives notation and preliminaries used in the paper. Section III provides the theoretic foundations of our algorithm design of which the details about cost function evaluation and the adopted optimization are introduced in Sec. IV. In Sec. V we exhibit experiments, including numerical simulation and quantum device implementation. Finally, Sec. VI concludes this work and discusses outlooks.

II. NOTATION AND PRELIMINARIES

A. Notation

In this paper, the standard computational basis vectors of an n-qubit system are denoted by $|j\rangle$, $j = 0, \ldots, 2^n - 1$, with matrix representations $|0\rangle = (1, 0, \ldots, 0)^\mathsf{T}, |1\rangle = (0, 1, \ldots, 0)^\mathsf{T}, \ldots, |2^n - 1\rangle = (0, 0, \ldots, 1)^\mathsf{T}$. For composed systems, we write an operator with a subscript indicating the system that the operator acts on, such as M_{AB}, and write $M_A := \mathrm{Tr}_B M_{AB}$. We mainly focus on the natural partition of the $2n$-qubit system, which corresponds to

a Hilbert space $\mathcal{H}_{AB} = \mathcal{C}^d \otimes \mathcal{C}^d$, where $d = 2^n$ is the local dimension.

B. Schmidt decomposition for pure states

The Schmidt decomposition is a fundamental tool to characterize the entanglement of bipartite pure states. Given a pure state $|\psi\rangle_{AB}$ living in the composed Hilbert space $\mathcal{H}_{AB}$, it always admits the Schmidt decomposition [15]

$$|\psi\rangle_{AB} = \sum_{j=0}^{R-1} c_j |u_j\rangle_A |v_j\rangle_B,$$

where $\{|u_j\rangle_A\}$ and $\{|v_j\rangle_B\}$ are some orthonormal bases in marginal spaces $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. Here the *Schmidt coefficients* $c_j > 0$ are ordered decreasingly, i.e., $c_j \geq c_k$ for all $j < k$. The number of positive coefficients R is called the *Schmidt rank*, satisfying $R \leq \min\{d_A, d_B\}$, where d_A and d_B are the dimensions of $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively.

Once given the Schmidt decomposition, one can deduce entanglement measures from these Schmidt coefficients. For example, the logarithm negativity follows as $E_N(|\psi\rangle_{AB}) = \log_2 N(|\psi\rangle_{AB})$, where $N(|\psi\rangle_{AB}) = 1/2[(\sum_j c_j)^2 - 1]$ is the negativity [16].

C. Reduction map as a separability criterion

The reduction map [20] $\mathcal{R}$ defined by

$$\mathcal{R}(X) = \mathrm{Tr} X \cdot I - X$$

is a positive but not completely positive map. Like every positive but not completely positive map, the reduction map $\mathcal{R}$ can be used as a separability criterion, in the sense that any separable state ρ_{AB} satisfies $(\mathcal{R}_A \otimes \mathrm{id}_B)(\rho_{AB}) = I \otimes \rho_B - \rho_{AB} \geq 0$. The set of bipartite states that are positive under the partial reduction map is called the $\mathcal{R}$ state; then the set of separable states, demoted by SEP, is a subset of the $\mathcal{R}$ state. It is worth mentioning that the PPT state is a subset of the $\mathcal{R}$ state, and when $d_A d_B \leq 6$, they both become necessary conditions for entanglement [20]. Also, any state violating the reduction criterion is distillable [21].

D. Maximally entangled states

In a composed system $\mathcal{H}_{AB} = \mathcal{C}^d \otimes \mathcal{C}^d$, the maximally entangled states are the states with Schmidt rank $R = d$ and Schmidt coefficients $c_j = 1/\sqrt{d}$, i.e., states of the form

$$|\Phi\rangle_{AB} = \frac{1}{\sqrt{d}} \sum_{j=0}^{d-1} |u_j\rangle_A |v_j\rangle_B,$$

where $|u_j\rangle_A$ and $|v_j\rangle_B$ are some orthonormal bases in $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. We denote the set of all maximally

entangled states by MAXE:

$$\mathrm{MAXE} := \left\{ |\Phi\rangle \,\Big|\, |\Phi\rangle = (1/\sqrt{d}) \sum_{j=0}^{d-1} |u_j\rangle_A |v_j\rangle_B \right\}.$$

It is clear that the maximally entangled states of the same dimension are only different in local orthonormal bases. Among states in MAXE we denote by $|\Phi^+\rangle_{AB} = (1/\sqrt{d}) \sum_j |j\rangle_A |j\rangle_B \in$ MAXE the maximally entangled state in the computational basis. A frequently used identity in entanglement theory about $|\Phi^+\rangle$ is the "transpose trick": for any operator M,

$$(M_A \otimes I_B) |\Phi^+\rangle_{AB} = (I_A \otimes M_B^{\mathsf{T}}) |\Phi^+\rangle_{AB}, \qquad (1)$$

where M^{T} denotes the transpose (with respect to the computational basis) of M.

We remark that MAXE can also be expressed as

$$\mathrm{MAXE} = \{|\Phi\rangle \,|\, |\Phi\rangle = U_A \otimes V_B |\Phi^+\rangle\},$$

where $U_A, V_B \in \mathcal{U}_d$ are unitaries on $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. Note that we can further reduce the unitary operator on $\mathcal{H}_B$ via the transpose trick, i.e.,

$$\mathrm{MAXE} = \{|\Phi\rangle \,|\, |\Phi\rangle = V_B^{\mathsf{T}} U_A \otimes I_B |\Phi^+\rangle\}$$
$$= \{|\Phi\rangle \,|\, |\Phi\rangle = U_A \otimes I_B |\Phi^+\rangle\},$$

since $V_B^{\mathsf{T}} U_A \in \mathcal{U}_d$.

III. THEORETICAL FRAMEWORK

In this section, we provide the theoretical foundation for our variational algorithms. We start with the pure-state case whose Schmidt decomposition is sufficient to characterize its entanglement properties. Then we extend our method by looking at general mixed states and building the connection with the reduction separability criterion.

A. Schmidt decomposition for bipartite pure states

Consider a bipartite state $|\psi\rangle_{AB}$ operated by *local* unitaries $U_A \otimes V_B$. This unitary operator will not affect any entanglement properties, e.g., the Schmidt coefficients. Our key observation is that the maximal fidelity (maximized over the unitaries) between $U_A \otimes V_B |\psi\rangle_{AB}$ and some elaborated entangled state reflects the weighted sum of the Schmidt coefficients of $|\psi\rangle_{AB}$. It is rigorously stated in the following theorem.

Theorem 1.—Let U_A and V_B be unitaries on systems A and B, respectively. Let $|\Psi\rangle_{AB} = \sum_j p_j |j\rangle_A |j\rangle_B$, where $\{p_j\}$ is positive and ordered *strictly decreasingly*. For any

bipartite state $|\psi\rangle_{AB} = \sum_j c_j |u_j\rangle_A |v_j\rangle_B$ with decreasing coefficients c_j, it holds that, with $|\widetilde{\psi}\rangle_{AB} = U_A \otimes V_B |\psi\rangle_{AB}$,

$$\max_{U_A, V_B} F(|\widetilde{\psi}\rangle_{AB}, |\Psi\rangle_{AB}) = \sum_j p_j c_j, \qquad (2)$$

and when the maximal of Eq. (2) is reached, the Schmidt coefficients can be readout as

$$c_j = F(\widetilde{\rho}_A, |j\rangle_A), \qquad (3)$$

where $\widetilde{\rho}_A = \mathrm{Tr}_B[|\widetilde{\psi}\rangle\langle\widetilde{\psi}|_{AB}]$. Furthermore, if c_j is not degenerate, i.e., $c_{j-1} > c_j > c_{j+1}$, the orthonormal vectors can be prepared (up to global phases $e^{i\theta_{A,B}^{(j)}}$) by

$$e^{i\theta_A^{(j)}} |u_j\rangle_A = U_A^\dagger |j\rangle_A, \qquad e^{i\theta_B^{(j)}} |v_j\rangle_B = V_B^\dagger |j\rangle_B, \qquad (4)$$

for some real $\theta_{A,B}^{(j)}$.

Proof.—Recall the operator-vector correspondence mapping [22] defined by the action on base vectors:

$$\mathrm{vec}(|j\rangle\langle k|) = |j\rangle |k\rangle.$$

Since the vec mapping is bijective and isometric, we define its inverse mapping as

$$\mathrm{mat}(|j\rangle |k\rangle) := \mathrm{vec}^{-1}(|j\rangle |k\rangle) = |j\rangle\langle k|,$$

which is also isometric. Let $X = \mathrm{mat}(|\widetilde{\psi}\rangle_{AB})$ and $Y = \mathrm{mat}(|\Psi\rangle_{AB})$; then $F(|\widetilde{\psi}\rangle_{AB}, |\Psi\rangle_{AB}) = |\langle\Psi|\widetilde{\psi}\rangle_{AB}| = |\mathrm{Tr}\, X^\dagger Y|$ by isometry. What is more, c_j and p_j are singular values of X and Y in decreasing orders, respectively, bringing about $|\mathrm{Tr}\, X^\dagger Y| \le \sum_j c_j p_j$ via Von Neumann's trace inequality [19]. To sum up, we have

$$F(|\widetilde{\psi}\rangle_{AB}, |\Psi\rangle_{AB}) = |\langle\Psi|_{AB} U_A \otimes V_B |\psi\rangle_{AB}|$$
$$= |\mathrm{Tr}\, X^\dagger Y|$$
$$\le \sum_j c_j p_j. \qquad (5)$$

Thus, the fidelity will not exceed $\sum_j c_j p_j$. Evidently, when U_A maps $\{|u_j\rangle_A\}$ to $\{|j\rangle_A\}$ and V_B maps $\{|v_j\rangle_B\}$ to $\{|j\rangle_B\}$, the maximal is reached. Then Eq. (2) holds.

When the maximal of Eq. (2) is reached, by Von Neumann's trace inequality, $U_A |u_j\rangle_A$ (correspondingly, $V_B |v_j\rangle_B$) falls in the subspace spanned by $\{|k\rangle_A |c_k = c_j\}$ (correspondingly, $\{|k\rangle_B |c_k = c_j\}$); thus, Eq. (3) holds; if c_j is further nondegenerate then Eq. (4) holds naturally, which completes the proof. ∎

Remark.—Inequality (5) was also proved in Refs. [23, Eq. (5)] and [24, Eq. (23)], and from von Neumann's trace inequality as well. Beyond the inequality itself, Theorem

1 also shows that, with nondegenerated Schmidt coefficients, the Schmidt vectors can be prepared (up to phases) when the inequality saturates, which is part of the Schmidt decomposition task.

Theorem 1 enables the design of our variational algorithm. The idea is straightforward: we use parameterized quantum circuits to implement U_A, V_B, respectively. Then we set the cost function to be some monotonic function of the fidelity $F(|\widetilde{\psi}\rangle_{AB}, |\Psi\rangle_{AB})$ (in Sec. IV A we choose F^2 to be the cost function for convenience). By tuning the circuits' parameters we can maximize the cost function. Finally, after the optimization we can readout the required Schmidt coefficient. In Sec. IV A the evaluation of the cost function and the readout process will be discussed in detail. It is worth noting that, even though we can prepare each decomposition vector $|e_j\rangle_{A,B}$, we might not be able to reconstruct $|\psi\rangle_{AB}$ via U_A and V_B because the relative phases for each component might not be equal, i.e., $\theta_{A,B}^{(j)}$ are not always the same for different j.

B. Logarithm negativity for bipartite pure states

Entanglement measures quantify the entanglement between quantum systems, being a key figure of merit in entanglement distillation and many other protocols. Known methods, however, are not directly applicable in most cases on near-term devices. In this part, we introduce the variational algorithm to estimate the logarithm negativity of pure bipartite states, which consume fewer resources. After extracting all the Schmidt coefficients from the above approach, entanglement measures such as logarithm negativity can be calculated by (classical) postprocessing [16]. Significantly, in the case where one would like to estimate the logarithm negativity directly and have no interest in each coefficient, we can provide a simpler variational estimation by substituting $|\Psi\rangle_{AB}$ with the bipartite maximally entangled state and using the transpose trick. Specifically, we have the following corollary.

Corollary 1.—Let U_A be a unitary operator on system A. For any bipartite state $|\psi\rangle_{AB} = \sum_j c_j |u_j\rangle_A |v_j\rangle_B$ with decreasing coefficients c_j, it holds that, with $|\widetilde{\psi}\rangle_{AB} = U_A \otimes I_B |\psi\rangle_{AB}$,

$$\max_{U_A} F(|\widetilde{\psi}\rangle_{AB}, |\Phi^+\rangle_{AB}) = \max_{|\Phi\rangle \in \text{MAXE}} F(|\psi\rangle_{AB}, |\Phi\rangle)$$
$$= \frac{\sum_j c_j}{\sqrt{d}}, \qquad (6)$$

where MAXE is the set of all maximally entangled states.

Proof.—In Eq. (2), let $p_j = 1/\sqrt{d}$; we have

$$F(U_A \otimes V_B |\psi\rangle_{AB}, |\Phi\rangle_{AB}) \leq \sum_j c_j p_j = \frac{\sum_j c_j}{\sqrt{d}}.$$

Thus, the maximal fidelity will not exceed $\sum_j c_j/\sqrt{d}$. Evidently, when U_A maps $\{|u_j\rangle_A\}$ to $\{|j\rangle_A\}$ and V_B maps $\{|v_j\rangle_B\}$ to $\{|j\rangle_B\}$, the maximal is reached.

Now note that

$$\max_{U_A, V_B} F(U_A \otimes V_B |\psi\rangle_{AB}, |\Phi^+\rangle_{AB})$$
$$= \max_{U_A, V_B} |\langle\psi|_{AB} U_A^\dagger \otimes V_B^\dagger |\Phi^+\rangle_{AB}|$$
$$= \max_{U_A, V_B} |\langle\psi|_{AB} U_A^\dagger V_B^* \otimes I_B |\Phi^+\rangle_{AB}|$$
$$= \max_{U_A} F(U_A \otimes I_B |\psi\rangle_{AB}, |\Phi^+\rangle_{AB})$$
$$= \max_{|\Phi\rangle \in \text{MAXE}} F(|\psi\rangle_{AB}, |\Phi\rangle_{AB}),$$

where the second equation follows from the transpose trick [Eq. (1)]. Then Eq. (6) holds, completing the proof. ∎

(A similar result is also discussed in Ref. [25].) Corollary 1 gives an estimation of the $L1$ norm (the sum) of the Schmidt coefficients, reflecting entanglement properties of the state. For example, note that the logarithm negativity of $|\psi\rangle_{AB}$ can be written as [16]

$$E_N(|\psi\rangle_{AB}) = \log_2(2N + 1) = \log_2\left(\sum_j c_j\right)^2,$$

where $N(|\psi\rangle_{AB}) = 1/2[(\sum_j c_j)^2 - 1]$ is the negativity of $|\psi\rangle_{AB}$. Specifically, since Eq. (6) is an optimization over unitaries, a variational quantum algorithm implements the optimization via tuning parameterized quantum circuits on near-term devices. The fidelity in Eq. (6) is then obtained by measurement in the computational basis after the subcircuit whose inverse prepares $|\Phi^+\rangle$ from $|0^n\rangle$. In this way, entanglement measures such as logarithm negativity can be inferred without tomography of the state. (In Sec. IV more details on the variational quantum algorithm will be provided.) Notably, Corollary 1 enables us to employ a one-side parameterized quantum circuit (on either the A or B system) in the variational learning. This reduces the parameters required in the parameterized quantum circuit by at least half, saving an appreciable amount of computational resources in classical optimization.

C. Entanglement detection for the general bipartite state

For pure states, entanglement can be easily detected by measuring the purity of the reduced quantum state. However, for mixed states, detecting entanglement becomes more challenging. While the purity and some related entropies (like Tsallis-2 entropy $1 - \text{Tr}[\rho^2]$ and second-order Rényi entropy $-\log\text{Tr}[\rho^2]$) can be computed efficiently, estimating entropies in general is difficult. Moreover, purity and these specific entropies typically do not

provide much operational information, such as distillability of the state. Wang *et al.* [13] realized the positive map criteria by decomposing the maps into Pauli operators, but this decomposition can consume exponentially many resources in multiqubit cases.

In the above logarithm negativity estimation for pure states we considered the maximal fidelity between a pure state and the maximally entangled states. It should not be surprising that this maximal fidelity also reflects some entanglement properties for general mixed states. In fact, the maximal overlap between a mixed state and the family of maximally entangled states is called the *fully entangled fraction* [25].

Definition 1.—The *fully entangled fraction* χ of any bipartite state ρ_{AB} is defined as the maximal state overlap to MAXE:

$$\chi(\rho_{AB}) := \max_{|\Phi\rangle \in \text{MAXE}} F^2(\rho_{AB}, |\Phi\rangle_{AB})$$

$$= \max_{|\Phi\rangle \in \text{MAXE}} \text{Tr}[\rho_{AB}|\Phi\rangle\langle\Phi|]$$

$$= \max_{U_A \in \mathcal{U}_d} \text{Tr}[\rho_{AB}U_A \otimes I_B|\Phi^+\rangle\langle\Phi^+|U_A^\dagger \otimes I_B]$$

with MAXE the set of all maximally entangled states.

A χ *state* is defined as the set of all bipartite states ρ_{AB} satisfying $\chi(\rho_{AB}) \leq 1/d$, where d is the local dimension.

The fully entangled fraction χ was previously studied in the qubit-qubit case in Refs. [26,27], and, more generally, in Refs. [25,28,29]. Specifically, Ganguly *et al.* [28] showed that the χ-state set is convex and compact.

We point out that $\chi \leq 1/d$ is a valid separability criterion: recall that the $\mathcal{R}$ state is the set of states that remain positive under the partial $\mathcal{R}$ map. We illustrate that the χ state contains the $\mathcal{R}$ state.

Lemma 1.—The $\mathcal{R}$ state is a subset of the χ state.

Proof.—By definition, we need to show that $\chi(\rho_{AB}) \leq 1/d$ for any $\rho_{AB} \in \mathcal{R}$ state.

Given any $\mathcal{R}$ state ρ_{AB}, it holds that

$$\mathcal{R}_A \otimes \text{id}_B(\rho_{AB}) \geq 0$$

$$\iff U_A \circ \mathcal{R}_A \otimes \text{id}_B(\rho_{AB}) \geq 0$$

$$\implies d\,\text{Tr}[U_A \circ \mathcal{R}_A \otimes \text{id}_B(\rho_{AB})|\Phi^+\rangle\langle\Phi^+|] \geq 0$$

for any unitary map U_A. The right-hand side satisfies

$$d\,\text{Tr}[U_A \circ \mathcal{R}_A \otimes \text{id}_B(\rho_{AB})|\Phi^+\rangle\langle\Phi^+|]$$

$$= d\,\text{Tr}[\rho_{AB}\mathcal{R}_A^\dagger \circ U_A^\dagger \otimes \text{id}_B(|\Phi^+\rangle\langle\Phi^+|)]$$

$$= d\,\text{Tr}[\rho_{AB}\mathcal{R}_A \circ U_A^\dagger \otimes \text{id}_B(|\Phi^+\rangle\langle\Phi^+|)]$$

$$= 1 - \text{Tr}[\rho_{AB}U_A^\dagger \otimes I_B|\Phi^+\rangle\langle\Phi^+|U_A \otimes I_B]$$
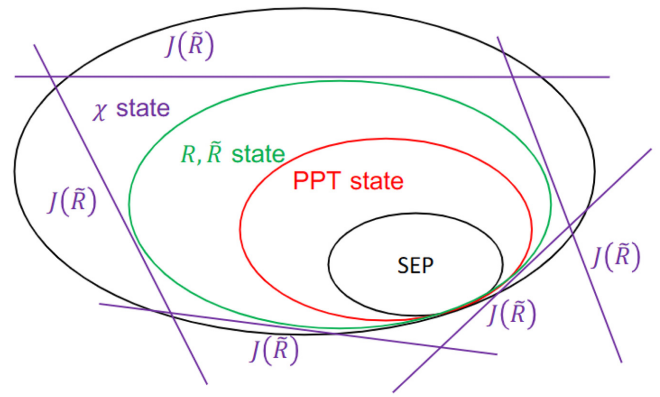
$$\geq 0,$$

FIG. 1. The geometrical explanation of the χ state. The black, red, and green circles correspond to the separable, PPT, and $\mathcal{R}$ state, respectively. The purple lines are hyperplanes defined by $\text{Tr}[\mathcal{J}(\tilde{R})\rho] = 0$, where $\mathcal{J}(\tilde{R})$ is the Choi operator of the local $\tilde{\mathcal{R}}$ map. Note that each line corresponds to a different $\tilde{\mathcal{R}}$ map defined by different U_A.

where in the first equivalence $\mathcal{R}_A^\dagger$ is the adjoint map of $\mathcal{R}_A$, and in the second equivalence we used $\mathcal{R}_A^\dagger = \mathcal{R}_A$. Then

$$\text{Tr}[\rho_{AB}U_A \otimes I_B|\Phi^+\rangle\langle\Phi^+|U_A^\dagger \otimes I_B] \leq 1/d \quad \text{for all } U_A,$$

leading to

$$\chi(\rho_{AB}) = \max_{U_A \in \mathcal{U}_d} \text{Tr}[\rho_{AB}U_A \otimes I_B|\Phi^+\rangle\langle\Phi^+|U_A^\dagger \otimes I_B]$$

$$\leq 1/d,$$

which completes the proof. ∎

With Lemma 1 one can think of χ as a separability criterion that is no stronger than the reduction criterion. That is, all states with fully entangled fraction $\chi(\rho) > 1/d$ will also violate the reduction criterion. It is known that states with fully entangled fraction $\chi(\rho) > 1/d$ can be distilled by the protocol consisting of twirling and the generalized XOR operation [21]. Therefore, our criterion determines distillability for this protocol.

Geometrically, the $\mathcal{R}$ state is a convex set that contains SEP and PPT. Consider the map $U_A \circ \mathcal{R}_A \otimes \text{id}_B$ for some fixed unitary U_A, and let the $\tilde{\mathcal{R}}$ state be the set of states that is positive under this map. Then the $\tilde{\mathcal{R}}$ state is equal to the $\mathcal{R}$ state since unitary operators preserve positivity. Then $\text{Tr}[W\rho] = 0$ corresponds to the hyperplane "outside" of the $\tilde{\mathcal{R}}$ state, where $W = \mathcal{J}(U_A \circ \mathcal{R}_A \otimes \text{id}_B) = \mathcal{R}_A \circ U_A^\dagger \otimes \text{id}_B(|\Phi^+\rangle\langle\Phi^+|)$ is the Choi operator of the partial $\tilde{R}$ map. Take U_A over the unitary group; those hyperplanes envelop a convex set, which is essentially the χ state by our definition (see Fig. 1).

We study the value of χ for certain families of entangled mixed states that are of practical interest. We have the following result.

Proposition 1.—The $\chi(\rho) > 1$ criterion can detect entanglement on the following families of states:

(i) isotropic states $\rho = p|\Phi^+\rangle\langle\Phi^+| + (1-p)I/d^2, p \in [1/(d^2-1), 1]$;

(ii) S states $\rho = p|\Phi^+\rangle\langle\Phi^+| + (1-p)|00\rangle\langle00|, p \in [0, 1]$;

(iii) Werner states (2×2) $\rho = 1/(d^3 - d)((d - \alpha)I + (1 - d\alpha)F)$, where $d = 2$, $\alpha \in [-1, 1]$, and F is the flip operator;

(iv) Bit-and-phase-flipped Bell states $\rho = p\rho_1 + (1-p)X \otimes I\rho_1 X \otimes I$, $\rho_1 = q|\Phi_2^+\rangle\langle\Phi_2^+| + (1-q)Z \otimes I$ $|\Phi_2^+\rangle\langle\Phi_2^+|Z \otimes I$.

Proof.—We calculate χ for each case individually.

(i) For isotropic states,

$$\chi(\rho) = p\chi(|\Phi^+\rangle\langle\Phi^+|) + (1-p)/d^2 = p + (1-p)/d^2.$$

So $\chi(\rho) > 1/d$ if and only if $p > 1/(d+1)$. On the other hand, it is known [21] that isotropic states are entangled if and only if $p > 1/(d+1)$, and the reduction criterion detects all entangled isotropic states. So χ also detects all entangled isotropic states.

(ii) For *S* states,

$$\chi(\rho) \leq p\chi(|\Phi^+\rangle\langle\Phi^+|)$$
$$+ (1-p)\chi(|00\rangle\langle00|) = p + (1-p)/d.$$

This bound is achievable by taking $U = I$. So $\chi(\rho) = p + (1-p)/d$. For all $p > 0$, $\chi(\rho) > p/d + (1-p)/d = 1/d$. On the other hand, all *S* states are entangled unless $p = 0$, as they all violate the PPT criterion.

(iii) For $d > 2$, it is known that the reduction criterion does not detect any entanglement for Werner states [21]. So the best hope is the two-dimensional case. For $d = 2$, Werner states are equivalent to isotropic states up to local unitary operators, and χ is local-unitary invariant. So χ detects all entangled two-dimensional Werner states.

(iv) For bit-and-phase-flipped Bell states, decompose the optimizing unitary operators as $U = R_z(z_1/2)R_y(y/2)$ $R_z(z_2/2)$; then direct calculation shows that

$$\chi(\rho) = \max_{z_1, y, z_2} \cos^2(y)(1 + (1 - 2p)(1 - 2q)\cos(z_1 + z_2))/2$$
$$= 1/2 + |(1 - 2p)(1 - 2q)|/2$$
$$\geq 1/2.$$

It is known that these states are entangled unless $p = q = 1/2$. So χ detects all such entangled states.

This completes the proof. ∎

Proposition 1 shows that our criterion detects entanglement for states that are relevant in practical applications.

One might also hope that χ forms some entanglement measure [30]. Unfortunately, there are counterexamples against it.

Property 1. *It holds that χ is local-unitary invariant, but not local-operation-and-classical-communication (LOCC) nonincreasing.*

Proof.—It is evident that χ is invariant under local unitary operators since the optimization is over all local unitaries.

We show that χ is not LOCC nonincreasing by contradiction. Suppose that χ is LOCC nonincreasing; then χ must be constant on separable states since separable states are mutually LOCC convertible. Then consider $\rho_0 = I_{d^2}/d^2$ the maximally mixed state and $\rho_1 = |00\rangle\langle00|$ the pure separable state; we have

$$\chi(\rho_0) = \frac{1}{d^2} \mathrm{Tr}[|\Phi^+\rangle\langle\Phi^+|] = \frac{1}{d^2},$$

while $\chi(\rho_1) = (2N(|00\rangle) + 1)/d = 1/d \neq \chi(\rho_0)$ from Property 1. This leads to a contradiction, so χ is not LOCC nonincreasing. ∎

Also, an example of the χ state not equaling the *R* state is provided in Appendix A 4.

IV. VARIATIONAL QUANTUM ALGORITHMS

With the theoretical framework in hand, we develop the variational quantum algorithms for Schmidt decomposition, logarithm negativity estimation, and entanglement detection. Here, the variational quantum algorithm (VQA) [31–33] is a popular paradigm for near-term quantum applications, which uses a classical optimizer to train parameterized quantum circuits to achieve certain tasks. VQA is applied to solve problems in many areas, including ground and excited state preparations [34–36], quantum data compression [37–39], combinatorial optimization [40], quantum classifier [41–43], and quantum classification [44,45].

In this section, we present the details of cost function evaluation and optimization methods in the algorithm design. The diagram of our algorithm is shown in Fig. 2 and the algorithm boxes are given in Appendix A 3.

A. Cost functions

We first look at the cost function in the Schmidt decomposition task. Recall that $\max F(|\widetilde{\psi}\rangle_{AB}, |\Psi\rangle_{AB}) = \sum_j p_j c_j$ according to Theorem 1. In the variational algorithm, we set our cost function to be

$$C^{\mathrm{SD}} := F^2(|\widetilde{\psi}\rangle_{AB}, |\Psi\rangle_{AB}) = \mathrm{Tr}[\widetilde{\rho}_{AB} |\Psi\rangle\langle\Psi|_{AB}]. \quad (7)$$

Recall that $\widetilde{\rho}_{AB} = |\widetilde{\psi}\rangle\langle\widetilde{\psi}|_{AB}$ and $|\widetilde{\psi}\rangle_{AB} = U_A \otimes V_B |\psi\rangle_{AB}$. Evidently, F reaches its maximal if and only if C^{SD} reaches
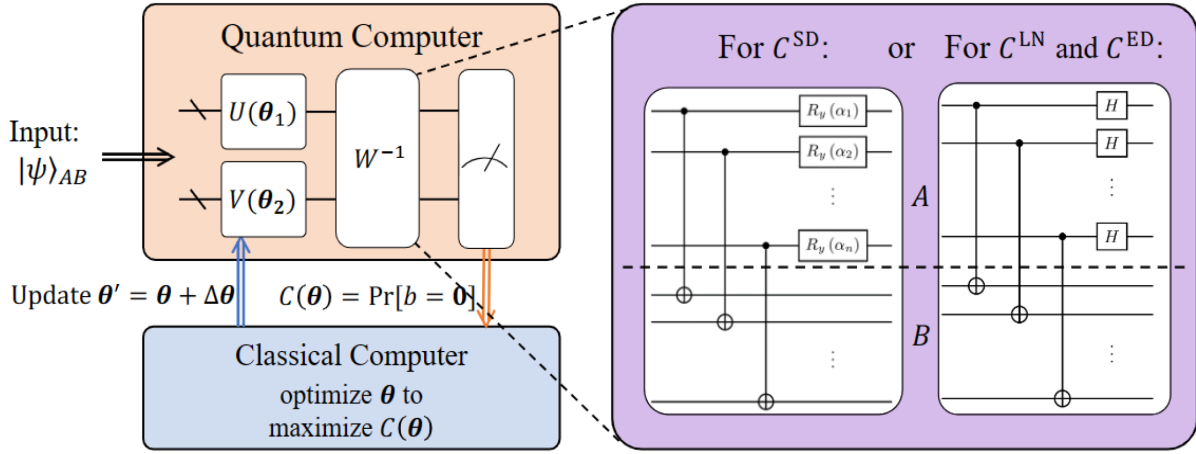
FIG. 2. Diagram of the variational quantum algorithms for Schmidt decomposition and logarithm negativity. Apply parameterized quantum circuits $U(\boldsymbol{\theta}_1)$ and $V(\boldsymbol{\theta}_2)$ onto the two parts A and B, respectively [for logarithm negativity estimation and entanglement detection, I_B can replace $V(\boldsymbol{\theta}_2)$]. Then operate an inverse of the state-preparing circuit (W for Schmidt decomposition and W' for logarithm negativity estimation and entanglement detection). After measurement, we take the probability of the all-zero measurement outcome as the cost function, and the classical optimizer is used to maximize the cost function by updating the parameters in the PQC iteratively.

the maximal. Thus, after the cost function C is optimized the Schmidt coefficients can be read out through measurement:

$$c_j = F(|\widetilde{\psi}\rangle_A, |j\rangle_A) = \sqrt{\mathrm{Tr}[\widetilde{\rho}_A |j\rangle\langle j|_A]}.$$

We then show that the cost function can be evaluated efficiently on a NISQ device. Suppose that circuit W prepares state $|\Psi\rangle_{AB} = W|0\rangle_{AB}$. By the cyclic property of trace we have

$$\begin{aligned} C^{\mathrm{SD}} &= \mathrm{Tr}[\widetilde{\rho}_{AB} W|0\rangle\langle 0|_{AB} W^\dagger] \\ &= \mathrm{Tr}[W^\dagger \widetilde{\rho}_{AB} W|0\rangle\langle 0|_{AB}]. \end{aligned} \quad (8)$$

Thus, the cost function C^{SD} is equal to the probability of all-zero measurement outcomes of $|\widetilde{\psi}\rangle_{AB}$ acted on by $W^\dagger = W^{-1}$. So the key to evaluating C is to implement W (W^{-1}) efficiently.

Note that W can be constructed by W_A, which prepares the superposition $\sum_j p_j |j\rangle_A$ in system A and n controlled-NOT (CNOT) gates connecting all qubit pairs across systems A and B. If we preset the coefficients p_j then constructing the circuit for W_A is essentially the amplitude encoding [46] task, for which the best-known construction [47] requires $O(2^n)$ CNOT count and $O(2^n)$ depth. Nevertheless, in our algorithm, we only require coefficients $\{p_j\}$ to be decreasing. We provide a circuit construction for this "weak" state preparation task that uses n single-qubit gates in parallel. In fact, the circuit composed of n paralleled

y-axis rotation gates,

$$\left.\begin{array}{l} |0\rangle - \boxed{R_y(\alpha_1)} - \\ |0\rangle - \boxed{R_y(\alpha_2)} - \\ \qquad \cdots \\ |0\rangle - \boxed{R_y(\alpha_n)} - \end{array}\right\} \sum_j p_j |j\rangle_A$$

with carefully chosen parameters $\{\alpha_j\}$, satisfies the condition that p_j decreases for arbitrary n. When n trivially equals 1, the parameter α_1 can be set to 0. We refer the reader to Appendix A 1 for the selection of α_j for $n \geq 2$. Note that the total depth in this construction of W is 2 with CNOT count n, leading to an efficient implementation of W (thus W^{-1}) in the cost function evaluation.

We next turn to the logarithm negativity estimation. In this case, since $\max F(|\widetilde{\psi}\rangle_{AB}, |\Phi^+\rangle_{AB}) = \sum_j c_j/\sqrt{d}$ according to Corollary 1, the cost function is set as

$$C^{\mathrm{LN}} := F^2(|\widetilde{\psi}\rangle_{AB}, |\Phi^+\rangle_{AB}) = \mathrm{Tr}[\widetilde{\rho}_{AB}|\Phi^+\rangle\langle\Phi^+|_{AB}].$$

Evidently, $C^{\mathrm{LN}}_{\max} = (\sum_j c_j)^2/d$. We use the variational method to maximize the cost function C^{LN}, and then the logarithm negativity is estimated as

$$E_N(|\psi\rangle_{AB}) = \log_2(C^{\mathrm{LN}}_{\max} d) = \log_2(C^{\mathrm{LN}}_{\max}) + n. \quad (9)$$

The cost function C^{LN} also has an efficient evaluation using a similar technique to that discussed above. Note that the maximally entangled state $|\Phi\rangle_{AB}$ can be prepared by a depth-2 circuit W' consisting of n Hadamard gates and n

CNOT gates. With the inverse circuit $W'^\dagger$ in hand, the cost function C^{LN} is evaluated as

$$
\begin{aligned}
C^{\mathrm{LN}} &= \mathrm{Tr}[\widetilde{\rho}_{AB}|\Phi^+\rangle\langle\Phi^+|_{AB}] \\
&= \mathrm{Tr}[W'^\dagger \widetilde{\rho}_{AB} W' \, |0\rangle\langle0|_{AB}], \quad\quad (10)
\end{aligned}
$$

which is the probability of the all-zero measurement outcome of $|\widetilde{\psi}\rangle_{AB}$ acted on by $W'^\dagger$.

Finally, for the entanglement detection algorithm, we use the same cost function as we used when estimating logarithm negativity:

$$
\begin{aligned}
C^{\mathrm{ED}} &= C^{\mathrm{LN}} \\
&= \mathrm{Tr}[\widetilde{\rho}_{AB}|\Phi^+\rangle\langle\Phi^+|_{AB}] \\
&= \mathrm{Tr}[W'^\dagger \widetilde{\rho}_{AB} W' \, |0\rangle\langle0|_{AB}]. \quad\quad (11)
\end{aligned}
$$

Thus, the depth-2 circuit W' is also employed. We remark that in entanglement detection we may not need the cost function to converge; the algorithm can halt when the cost function $C^{\mathrm{ED}} \geq 1$ and the output is a positive detection result.

B. Parameterized quantum circuit

As discussed in Sec. III, we adopt the bilocal PQC $U(\boldsymbol{\theta}_1) \otimes V(\boldsymbol{\theta}_2)$ for variational Schmidt decomposition and the one-side PQC $U(\boldsymbol{\theta}_1) \otimes I$ for logarithm negativity estimation. Since our algorithm makes no assumption about input states, we recommend the class of hardware-efficient PQCs [32] as the tunable unitary to further enhance the experimental realizability of our algorithms. In the experiments, the PQCs consist of single-qubit rotations and CNOT gates or controlled-Z gates as the entangled gates. We refer the reader to Sec. V for more details of the circuit implementation in each experiment.

C. Optimization

Both gradient-based [48,49] and gradient-free [50–52] optimization methods can be applied to our variational quantum algorithms. For gradient-based optimization, the analytic gradients of C^{SD}, C^{LN}, and C^{ED} are supported by the parameter shift rule [53], which provides an unbiased estimate compared to the finite difference method [54]. We have adopted various optimization methods in both the numerical simulation and the real-device implementation of our algorithms (see Sec. V for more details about the optimization adopted in our experiments).

Here we also discuss the possible solutions to the gradient vanishing issue for our algorithms. According to Ref. [55], our algorithms could suffer from "barren plateaus" throughout training. Additionally, this gradient vanishing issue is unlikely to be solved by a local cost function [56] due to the global connectivity brought by subcircuit W^{-1}.

On the other hand, proposed strategies including layerwise learning [57], parameter correlations [58], and a quantum convolutional neural network ansatz [59] have been shown to be helpful in certain training tasks. We leave the adaptation of these and other training techniques to our algorithms for future study.

D. Summary of algorithms and comparisons

Before moving to the experiments, we summarize our algorithms and explain improvements over known methods. From the previous discussion, one can observe that our algorithms share a similar framework: the target state is operated by local parameterized quantum circuits, and the computational basis measurement after a depth-2 subcircuit estimates the fidelity with some elaborated entangled state as the optimization function. The VQA approach provides practicality on near-term quantum devices, and our framework guarantees their efficiency due to the constant extra consumption of computational resources. Exploiting symmetry, our framework can further reduce the number of parameters in the optimization of the VQA. This idea of "computing the state overlap with some entangled state after operating local unitaries" lies at the heart of our algorithms' framework, and turns out to reflect entanglement properties for near-term devices. As the algorithms advance over known variational approaches, our framework technically and conceptually contributes to entanglement analysis of near-term quantum devices, and is worth further study, development, and extension, e.g., in multipartite cases.

Here we make a detailed comparison between our algorithms and known methods. For the Schmidt decomposition algorithm, methods proposed in Refs. [17,18] are also based on the hybrid quantum-classical approach. Essentially, all three methods perform local unitaries on each party, and should diagonalize the marginal states after optimization over unitaries. In this sense, it is reasonable to assume that the three methods need the same expressibility [60] for the ansatz.

Compared to Ref. [17], our algorithm demands no extra qubit, while in their approach (with a global cost function), two copies of the state are consumed in each optimization iteration. Thus, the resource of state preparation as well as the quantum register is doubled. The method in Ref. [18] seems to require a slightly shallower circuit than our algorithm due to the two-layer depth subcircuit $W^\dagger$. Nevertheless, our algorithm can detect the case where the input state is polluted into a separable but classically correlated state. With a similar argument as in the proof of Lemma 1, one can show that, for a separable state, $\max C^{\mathrm{SD}} \leq \max p_i^2$. So, if, for example, the input pure state $0.6\,|00\rangle + 0.8\,|11\rangle$ decoheres to mixed state $0.36|00\rangle\langle00| + 0.64|11\rangle\langle11|$, our algorithm will detect this decoheres, while the method in Ref. [18] will not be able to distinguish them.

For the entanglement detection algorithm, recent work in Ref. [13] realized various positive (but not completely positive) map criteria, including the reduction criterion in the variational approaches. Their method employs the quasiprobability sampling, which suffers from an exponential sampling overhead in the number of qubits. Although our algorithm, as a separability criterion, is no stronger than what Wang *et al.* [13] implemented, we estimate only one expectation value in each optimization iteration that saves us from exponential resource consumption for large quantum systems.

V. NUMERICAL SIMULATION AND IMPLEMENTATION ON A QUANTUM DEVICE

In this section, we show the effectiveness of variational quantum algorithms in Schmidt decomposition and logarithm negativity estimation through numerical simulation and implementation on quantum devices. The simulation experiments are operated on the *Paddle Quantum* [61] platform and the *Quantum Leaf* [62] platform. We also realize our algorithm on the superconducting quantum device at the Institute of Physics, Chinese Academy of Sciences through the *Quantum Leaf* [62] platform.

A. Simulating on circuits with different depths

Here, we conduct numerical experiments to decompose a bipartite quantum system with eight qubits to investigate the impact of different numbers of circuit layers. In the beginning, we randomly generate an eight-qubit bipartite entangled state $|\psi\rangle_{AB}$, and both parties have four qubits. Then, we apply PQCs $U_A(\vec{\theta}_A)$ and $V_B(\vec{\theta}_B)$ on the two parties, respectively. The structure of the PQCs used in this task is shown in Fig. 3. As discussed above, we adopt Eq. (8) as the cost function. This simulation is performed on the *Paddle Quantum* [61] platform.

In this experiment, we set the depth of the PQC layers to be $1, 2, 4, 8$, which means applying the structure in Fig. 3 recurrently for depth times, to clarify the impact of the
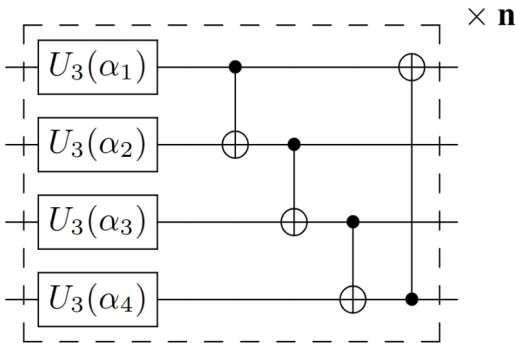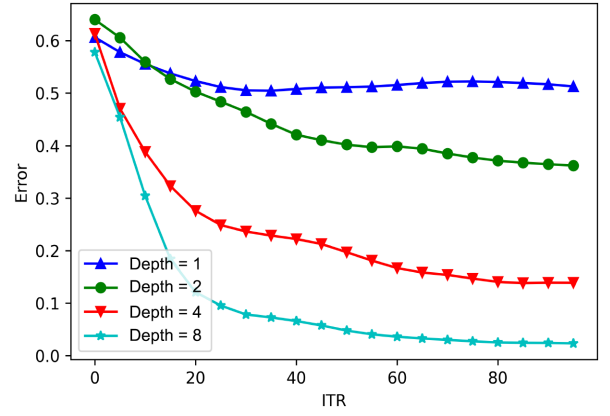

FIG. 4. Simulating an eight-qubit Schmidt decomposition on PQCs of depths 1, 2, 4, and 8.

number of circuit layers. The results are shown in Fig. 4. We use the ADAM optimizer (a gradient-based optimizer) to maximize our cost function.

The simulation results are shown in Fig. 4. The error (vertical axis) is defined as the squared $L2$ distance between the real and estimated Schmidt coefficient vectors, i.e., $\sum_i |c_j - c'_j|^2$, where the c'_j are the estimated Schmidt coefficients and the c_j are the actual values. It is obvious that circuits with larger depth have better performance in accuracy, satisfying the intuition that deeper circuits have better expressibility [60].

B. Decomposing the noisy state

Next, we investigate the performance of our algorithm in decomposing quantum states affected by noisy channels. We take the amplitude damping channel $N_p^{\mathrm{amp}}(\rho)$ and depolarizing channel $N_p^{\mathrm{depl}}(\rho)$ into consideration, which are defined as

$$N_p^{\mathrm{amp}}(\rho) := E_0 \rho E_0^\dagger + E_1 \rho E_1^\dagger,$$

$$N_p^{\mathrm{depl}}(\rho) := (1 - p)\rho + p \operatorname{Tr}(\rho)\frac{I}{2},$$

respectively, and

$$E_0 = \begin{bmatrix} 1 & 0 \\ 0 & \sqrt{1-p} \end{bmatrix}, \qquad E_1 = \begin{bmatrix} 0 & \sqrt{p} \\ 0 & 0 \end{bmatrix},$$

with the noise level $p \in [0, 1)$. In this simulation, we decompose a bipartite system with one qubit in each party, and the input state is

$$\rho_{AB} = |\psi\rangle\langle\psi|_{AB}$$

$$= \begin{pmatrix} 0.455 & 0.459 & 0.136 & -0.137 \\ 0.459 & 0.463 & 0.137 & -0.138 \\ 0.136 & 0.137 & 0.041 & -0.041 \\ -0.137 & -0.138 & -0.041 & 0.041 \end{pmatrix}. \quad (12)$$

FIG. 3. Structure of one entangled layer. Here U_3 is the universal single-qubit template (*ZYZ* rotation gate), α_j is the parameter of the gate, and **n** is the depth of PQC.
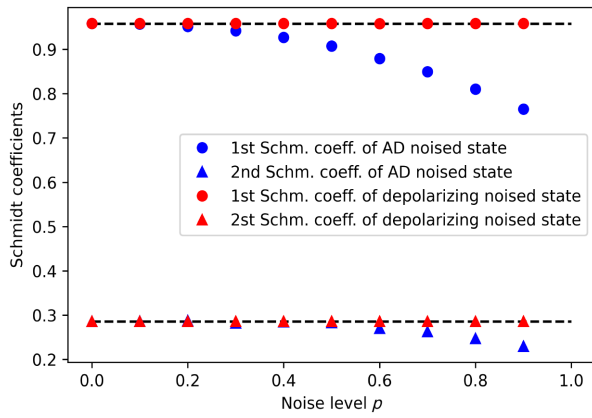
FIG. 5.　Schmidt coefficients of the quantum states affected by noise. The blue circles and triangles are the Schmidt coefficients of the amplitude damping (AD) noised state. The red circles and triangles are the Schmidt coefficients of the depolarizing noised state. The black dashed line is the theoretical Schmidt coefficients of the state without noise.
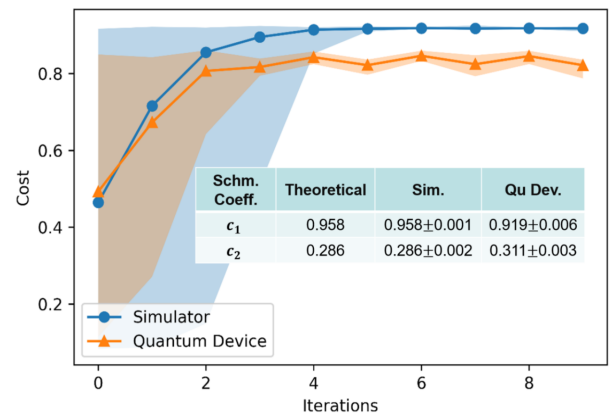

FIG. 6.　Learning curves with respect to the simulator and quantum device. The shadowed area spans from the minimal to maximal value of 20 experiments. Theoretically, the Schmidt coefficients are 0.958 and 0.286, respectively. The simulator-estimated coefficients are 0.958 ± 0.001 and 0.286 ± 0.002, while the results from the CAS quantum device are 0.919 ± 0.006 and 0.311 ± 0.003.

The Schmidt coefficients of the state are $c_1 = 0.958, c_2 = 0.286$. We employ noisy channels to each party before applying PQCs $U(\boldsymbol{\theta}_1)$ and $V(\boldsymbol{\theta}_2)$ which are single-qubit universal ansatz. This simulation is performed on the *Paddle Quantum* [61] platform.

Figure 5 displays the Schmidt coefficients of the states affected by both noisy channels with different noise levels p. For the depolarizing channel, the estimated Schmidt coefficients equal the theoretical values (black dashed line) all the time; for the amplitude damping channel, when the noise level is low (smaller than 0.5), the estimated Schmidt coefficients are close to the theoretical values but diverge as the noise level increases. We can conclude that our algorithm is robust to depolarizing channels, and can achieve relatively accurate results if the state is noised by an amplitude damping channel (low noise level).

C. Implementations on a superconducting quantum processor

We also apply our algorithm to decompose a two-qubit bipartite quantum system on quantum devices and compare the results with the values achieved from the simulation. The experiments are performed on the *Quantum Leaf* [62] platform, loading the quantum device from the Institute of Physics, Chinese Academy of Sciences (IoP CAS). This quantum device contains ten direct coupling transmon qubits, whose topology is neighbor coupling one-dimensional chainlike.

The input state $|\psi\rangle_{AB}$ we use here is exactly the same as that in Sec. V B. On the quantum device, the input state could be prepared by applying two Ry gates on the first and second qubits with parameters 0.58 and 1.58, respectively, followed by a controlled-Z gate. Next, we operate the

parameterized quantum circuit on qubits and use sequential minimal optimization [51] (gradient-free) to optimize the parameters until the cost converges to its maximum. We repeat 20 independent experiments with the same input states and randomly initialized parameters.

The experimental results are displayed in Fig. 6. As demonstrated, the cost raises dramatically in the first two iterations, and the simulator converges to a higher value than that of the quantum device, which may be caused by the quantum device noise. Correspondingly, the achieved Schmidt values by the simulator ($c_1 = 0.958 \pm 0.001$ and $c_2 = 0.286 \pm 0.002$) are closer to the theoretical values ($c_1 = 0.958$ and $c_2 = 0.286$) than those of the quantum device ($c_1 = 0.919 \pm 0.006$ and $c_2 = 0.311 \pm 0.003$).

D. Estimating the logarithm negativity

From Corollary 1, we could derive the logarithm negativity estimation, which quantifies the entanglement of the input state $|\psi\rangle_{AB}$. Here, we investigate the accuracy of our method in estimating the logarithm negativity at different input state ranks by comparing it with theoretical values.

In this simulation, $|\psi\rangle_{AB}$ is bipartite with three qubits in each party. In order to make the results comparable, the amplitude of the input state is designed to be the identity, i.e., $|\psi\rangle = (1/\sqrt{r}) \sum_{i=1}^{r} |i\rangle$, where r denotes the Schmidt rank ranging from 1 to 8. The circuit for the state preparation is displayed in Fig. 9 in Appendix A 2. Then, we apply the parameterized quantum circuit, which is shown in Fig. 7, followed by the subcircuit W^{-1} as in Fig. 2. The sequential minimal optimization [51] is utilized to maximize the cost function, which is set as Eq. (10), until it
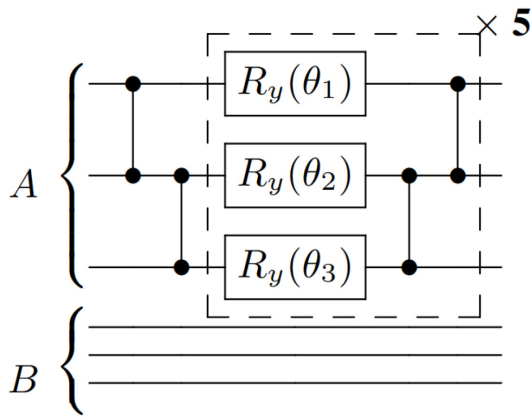
FIG. 7. Parameterized circuit ansatz for logarithm negativity estimation. The circuits in the dashed box should be repeated five times.

converges to its maximum value. At this stage, the logarithm negativity of the input state $|\psi\rangle$ can be calculated by Eq. (9). We repeat the calculation for logarithm negativity ten times for the input state with different Schmidt ranks, i.e., $r = 1, 2, 3, 4, 5, 6, 7, 8$, and compare the calculated values with the theoretical values. This simulation is performed on the *Quantum Leaf* [62] platform. The results are shown in Fig. 8. From the results, we can tell that the estimated values from our method are close to the ideal values even for input states with a high Schmidt rank. The difference between the theoretical and estimated values comes from the statistical error due to finite sampling. In this numerical simulation, the results are calculated by sampling each circuit execution 1024 times. It should be
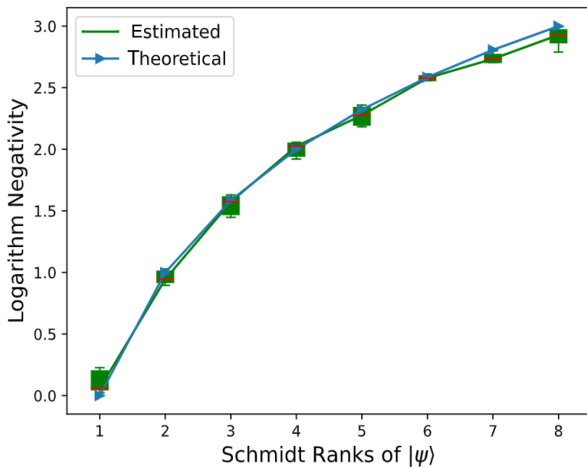

FIG. 8. Logarithm negativity estimation with respect to different Schmidt ranks. The red bar is the median logarithm negativity calculation repeated ten times, and the green rectangles and bar stand for the standard deviation and error, respectively. This experiment includes the no-entanglement (rank $r = 1$) case and the maximally entangled ($r = 8$) case.

noted that if a shallow circuit is adopted, the experimental results may exhibit a degradation in performance.

VI. CONCLUSION AND OUTLOOK

We proposed variational quantum algorithms to realize Schmidt decomposition and estimated logarithm negativity for pure states, and detected entanglement for general mixed states on near-term quantum devices. Compared to previous methods, our Schmidt decomposition algorithms are more efficient in terms of resource consumption or more stable for polluted inputs. For general mixed states, a new method to detect entanglement on near-term quantum devices was derived, which efficiently detects entanglement for specific families of states and detects distillability in general. We have shown the validity and practicality of our methods via both numerical simulations and experimental implementations. Numerical simulations show that the variational Schmidt decomposition algorithm is resilient to amplitude damping and depolarizing noises. Experiments on superconducting quantum devices show relatively accurate results, illustrating that our algorithms are executable and valid on near-term quantum devices. Our variational quantum algorithms notably provide efficient and practical tools for entanglement quantification and analysis in the NISQ era.

Beyond quantifying entanglement for bipartite pure states, our algorithms could have a wide range of applications and extensions. Direct applications of our work include quantum data compression. For example, if the two parties of the state have distinct dimensions (or numbers of qubits), it is always possible for the larger party to reduce its scale by applying the unitary transformation obtained through training. Moreover, it would be of independent interest to further explore the power of local PQC, which is a key ingredient in our method. One more extension of our framework is to consider the multiparty case. If the Schmidt decomposition of a multiparty entangled state exists [63,64], our method can be adapted to decompose these states. We leave the adaptation of our techniques to investigate multiparty pure or mixed states for future study, as well as its implications in many-body quantum systems that exist in today's various quantum computation platforms [65,66].

The simulation and experimental codes are available on Github [67].

APPENDIX: VARIATIONAL QUANTUM ALGORITHMS FOR SCHMIDT DECOMPOSITION

1. Selection of parameters in subcircuit W_A

Here we show how to choose parameters α_j for $j \in [1, n]$ in circuit W_A defined in Sec. IV A when $n \geq 2$.

We first choose three real numbers β_n, β_{n-1}, and $\Delta\beta$ satisfying $1 < \beta_n < \beta_{n-1}$ and $\Delta\beta > 0$. If $n > 2$, we recursively set $\beta_j = \prod_{k=j+1}^{n} \beta_k + \Delta\beta$ for $j = n - 2, \ldots, 1$. Then we let $\gamma_j = \beta_j / (\beta_j + 1)$ and let $\alpha_j = 2 \arccos \sqrt{\gamma_j}$. We remark that the parameters are evaluated on the classical computer once for each algorithm run, with running time $O(n)$.
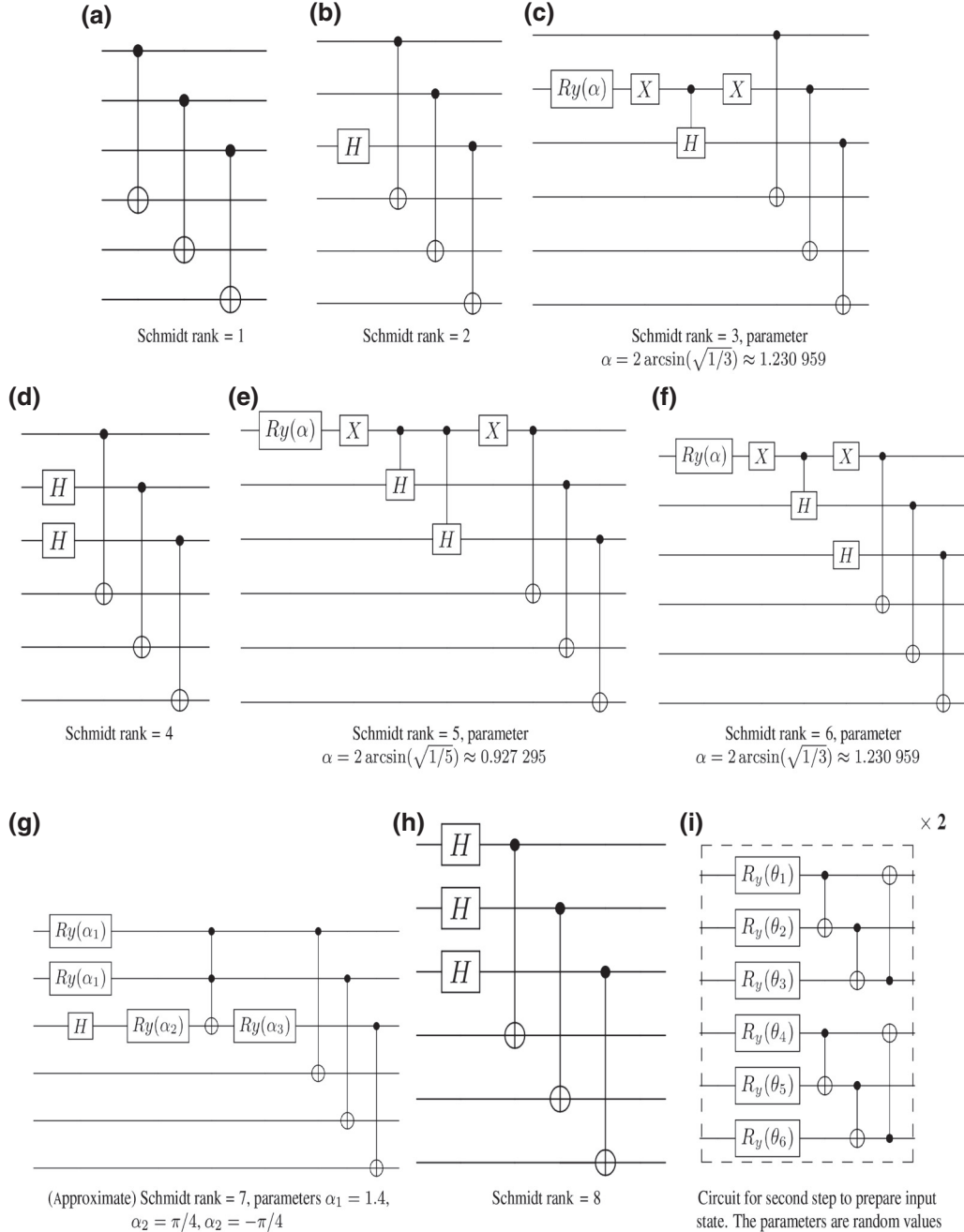

FIG. 9. Quantum circuits for preparing input states with different Schmidt ranks. Circuits (a)–(h) are the first steps to prepare the input state. (a) Schmidt rank of 1. (b) Schmidt rank of 2. (c) Schmidt rank of 3, parameter $\alpha = 2 \arcsin(\sqrt{1/3}) \approx 1.230\,959$. (d) Schmidt rank of 4. (e) Schmidt rank of 5, parameter $\alpha = 2 \arcsin(\sqrt{1/5}) \approx 0.927\,295$. (f) Schmidt rank of 6, parameter $\alpha = 2 \arcsin(\sqrt{1/3}) \approx 1.230\,959$. (g) (Approximate) Schmidt rank of 7, parameters $\alpha_1 = 1.4$, $\alpha_2 = \pi/4$, $\alpha_2 = -\pi/4$. (h) Schmidt rank of 8. (i) Circuit for the second step to prepare input state. The parameters are random values.

Input: bipartite d-dimension pure state ρ_{AB}, parameterized quantum circuit (PQC) $U(\boldsymbol{\theta}_1)$ and $V(\boldsymbol{\theta}_2)$, inverse of state preparing circuit W^{-1} (Fig. 2), number of iterations ITR;
Output: Schmidt coefficients $\{c_j\}$ and Schmidt bases $\{u_j\}, \{v_j\}$.
    Initialize parameters $\boldsymbol{\theta}$.
    # Training iteration
    for itr $= 1, \ldots,$ ITR **do**
        Apply $U(\boldsymbol{\theta}_1)$ to party A and $V(\boldsymbol{\theta}_2)$ to party B in state ρ_{AB}, and the state ρ_{AB} becomes $\tilde{\rho}_{AB}$.
        Apply the inverse of state preparing circuit $W^\dagger$ to the state $\tilde{\rho}_{AB}$.
        Compute the cost function $C^{SD} = \Pr[\boldsymbol{b} = \boldsymbol{0}] = \mathrm{Tr}[W^\dagger \tilde{\rho}_{AB} W \left|0\right\rangle\!\left\langle0\right|_{AB}]$.
        Maximize the cost function C^{SD} and update parameters $\boldsymbol{\theta}$.
    end for
    # Coefficient readout
    Denote the trained PQCs as $U(\boldsymbol{\theta}^{opt})$ and $V(\boldsymbol{\theta}^{opt})$.
    Apply the trained PQCs $U(\boldsymbol{\theta}^{opt})$ to the party A, and $V(\boldsymbol{\theta}^{opt})$ to the party B in target state ρ_{AB}, then achieve $\tilde{\rho}'_{AB}$.
    Measure the reduced state $\tilde{\rho}'_A$ on the computational basis $\{\left|j\right\rangle_A\}$. The j-th Schmidt coefficient is calculated by $c_j = \sqrt{\Pr[\boldsymbol{b} = j]} = \sqrt{\left\langle j\right| \tilde{\rho}'_A \left|j\right\rangle}$. The corresponding Schmidt bases are $u_j = U^\dagger(\boldsymbol{\theta}^{opt}) \left|j\right\rangle_A$ and $v_j = V^\dagger(\boldsymbol{\theta}^{opt}) \left|j\right\rangle_B$, respectively for $j \in [0, d-1]$.

Algorithm 1.   Variational quantum algorithm for Schmidt decomposition.

To show that W_A prepares state $\left|\Phi\right\rangle_A = \sum_j p_j \left|j\right\rangle_A$ such that p_j strictly decreases; let us look at the coefficients $\{p_j\}$ determined by $\{\alpha_j\}$. The output state can be written as

$$|\Psi\rangle_A = \bigotimes_{j=1}^{n} \left(\cos\frac{\alpha_j}{2} |0\rangle_j + \sin\frac{\alpha_j}{2} |1\rangle_j \right)$$

$$= \bigotimes_{j=1}^{n} (\sqrt{\gamma_j} |0\rangle_j + \sqrt{1-\gamma_j} |1\rangle_j).$$

Note that $\beta_j > 1$ implies that $\gamma_j > 1 - \gamma_j$. Let the binary representation of j be $j = (j_1 j_2 \cdots j_n)_2$, where j_l, $l \in [1, n]$, are binary bits. Then coefficient p_j is

$$p_j = \prod_{l=1}^{n} \sqrt{\gamma_l^{1-j_l}(1-\gamma_l)^{j_l}}.$$

[For example, when $n = 4$ and $j = 9 = (1001)_2$, we have $p_j = p_9 = \sqrt{\gamma_1^0 \gamma_2^1 \gamma_3^1 \gamma_4^0 (1-\gamma_1)^1 (1-\gamma_2)^0 (1-\gamma_3)^0 (1-\gamma_4)^1} = \sqrt{(1-\gamma_1)\gamma_2\gamma_3(1-\gamma_4)}.$]

For any $j < k$, denote by n_0 the highest bit such that j_{n_0} and k_{n_0} are different (hence $j_{n_0} = 0$, $k_{n_0} = 1$, and $j_l = k_l$ for $l < n_0$). Consider the index $j_+ = (j_1 \cdots j_{n_0} 1 1 \cdots 1)_2$ and $k_- = (k_1 \cdots k_{n_0} 0 0 \cdots 0)_2$, so $j \le j_+$ and $k \ge k_-$. Also, we have

$$\frac{p_j}{p_{j_+}} = \prod_{l=n_0, j_l \neq 1}^{n} \sqrt{\frac{\gamma_l}{1-\gamma_l}} \ge 1, \qquad \frac{p_k}{p_{k_-}}$$

$$= \prod_{l=n_0, k_l \neq 0}^{n} \sqrt{\frac{1-\gamma_l}{\gamma_l}} \le 1,$$

so $p_j \ge p_{j_+}$ and $p_k \le p_{k_-}$. Finally, we compare p_{j_+} and p_{k_-}:

$$\frac{p_{j_+}}{p_{k_-}} = \sqrt{\frac{\gamma_{n_0}}{1-\gamma_{n_0}} \prod_{l=n_0+1}^{n} \frac{1-\gamma_l}{\gamma_l}}.$$

Input: bipartite d-dimension pure state ρ_{AB}, parameterized quantum circuit (PQC) $U(\boldsymbol{\theta})$, inverse of state preparing circuit W'^{-1} (Fig. 2), number of iterations ITR;
Output: logarithm negativity of the bipartite pure state $E_N(\rho_{AB})$.
    Initialize parameters $\boldsymbol{\theta}$.
    # Training iteration
    for itr $= 1, \ldots,$ ITR **do**
        Apply $U(\boldsymbol{\theta})$ to party A in state ρ_{AB}, and the state ρ_{AB} becomes $\tilde{\rho}_{AB}$.
        Apply the inverse of state preparing circuit $W'^\dagger$ to the state $\tilde{\rho}_{AB}$.
        Calculate the loss function $C^{LN} = \Pr[\boldsymbol{b} = \boldsymbol{0}] = \mathrm{Tr}[W'^\dagger \tilde{\rho}_{AB} W' \left|0\right\rangle\!\left\langle0\right|_{AB}]$.
        Maximize the loss function C^{LN}, and update parameters $\boldsymbol{\theta}$.
    end for
    # Logarithm negativity readout
    Denote the maximum loss function as C^{LN}_{max}, and the logarithm negativity is calculated by $E_N(\rho_{AB}) = \log_2(C^{LN}_{max}) + n$

Algorithm 2.   Variational quantum algorithm for logarithm negativity estimation.

Input: bipartite d-dimension quantum state ρ_{AB}, parameterized quantum circuit (PQC) $U(\boldsymbol{\theta})$, inverse of state preparing circuit W'^{-1} (Fig. 2), number of iterations ITR;

Output: entanglement detection of the bipartite quantum state.

    Initialize parameters $\boldsymbol{\theta}$.

    # Training iteration

    for itr $= 1, \dots,$ ITR **do**

        Apply $U(\boldsymbol{\theta})$ to party A in state ρ_{AB}, and the state ρ_{AB} becomes $\tilde{\rho}_{AB}$.

        Apply the inverse of state preparing circuit $W'^{\dagger}$ to the state $\tilde{\rho}_{AB}$.

        Calculate the loss function $C^{ED} = \Pr[\boldsymbol{b} = \boldsymbol{0}] = \mathrm{Tr}[W'^{\dagger}\tilde{\rho}_{AB}W'\,|0\rangle\langle0|_{AB}]$.

        Maximize the loss function C^{ED}, and update parameters $\boldsymbol{\theta}$.

        If loss function $C^{ED} > 1/d$:

        Entanglement is detected.

        break

    end for

    If loss function $C^{ED} > 1/d$:

    Entanglement is detected.

Algorithm 3.   Variational quantum algorithm for entanglement detection.

By the definition of β, we have $\beta_{n_0} > \prod_{l=n_0+1}^{n}\beta_l$. Together with $\beta_j = \gamma_j/(1-\gamma_j)$, it holds that

$$\frac{p_{j_+}}{p_{k_-}} = \sqrt{\frac{\gamma_{n_0}}{1-\gamma_{n_0}}\prod_{l=n_0+1}^{n}\frac{1-\gamma_l}{\gamma_l}} = \sqrt{\beta_{n_0}\prod_{l=n_0+1}^{n}\frac{1}{\beta_l}} > 1.$$

So $p_j \geq p_{j_+} > p_{k_-} \geq p_k$, guaranteeing a strictly decreasing sequence $\{p_j\}$.

2. Circuits for preparing states with different Schmidt ranks

Here we display the circuits to generate a six-qubit bipartite entangled state with desired Schmidt rank. The circuits are shown in Fig. 9. It is notable that the first three qubits belong to one party and remaining three qubits belongs to the other. For example, if we wish to prepare an input state with Schmidt rank 5, we should firstly apply the circuit in Fig. 9(e), then operate the second circuit in Fig. 9(i).

3. Algorithm boxes

4. Examples that a χ state violates the reduction criterion

We show that certain AD-noisy Bell states (Bell state under one-sided amplitude damping noise) have $\chi < 1/d = 1/2$, but violates the reduction criterion.

Let $E_0 = |0\rangle\langle0| + \sqrt{1-\gamma}\,|1\rangle\langle1|, E_1 = \sqrt{\gamma}\,|0\rangle\langle1|, \mathcal{N}_{AP}(\cdot) = E_0 \cdot E_0^{\dagger} + E_1 \cdot E_1^{\dagger}$; then the AD-noisy Bell states are expressed as $\rho = \mathcal{N}_{AP} \otimes \mathrm{id}(|\Phi_2^+\rangle\langle\Phi_2^+|)$.

Let the PQC $U = R_z(z_1/2)R_y(y/2)R_z(z_2/2)$; then direct calculation shows that

$$\chi(\rho) = \frac{1}{2}\max_{z_1,y,z_2} 1 + \sqrt{1-\gamma}\cos(z_1+z_2) + \cos(y)(1-\gamma$$

$$+ \sqrt{1-\gamma}\cos(z_1+z_2)) = \frac{2+2\sqrt{1-\gamma}-\gamma}{4}.$$

So $\chi(\rho) > 1/2$ if and only if $\gamma < 2\sqrt{2} - 2 \approx 0.828$. On the other hand, all AD-noisy Bell states are entangled unless $\gamma = 1$, as they all violate the PPT criterion. So χ fails to detect entanglement when $\gamma \in (2\sqrt{2-2}, 1)$. But the reduction criterion can detect entanglement for those states, as it is a necessary and sufficient condition in the 2×2 case.

[1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, Rev. Mod. Phys. **81**, 865 (2009).

[2] D. Naik, C. Peterson, A. White, A. Berglund, and P. G. Kwiat, Entangled state Quantum Cryptography: Eavesdropping on the Ekert Protocol, Phys. Rev. Lett. **84**, 4733 (2000).

[3] A. K. Ekert, Quantum Cryptography Based on Bell's Theorem, Phys. Rev. Lett. **67**, 661 (1991).

[4] S. McArdle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Quantum computational chemistry, Rev. Mod. Phys. **92**, 015003 (2020).

[5] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, S. Boixo, M. Broughton, B. B. Buckley, and D. A. Buell *et al.*, Hartree-Fock on a superconducting qubit quantum computer, Science **369**, 1084 (2020).

[6] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, Quantum machine learning, Nature **549**, 195 (2017).

[7] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, and

L. Cincio *et al.*, Variational quantum algorithms, Nat. Rev. Phys. **3,** 625 (2021).

[8] C. H. Bennett and G. Brassard, Quantum cryptography: Public key distribution and coin tossing, arXiv preprint arXiv:2003.06557 (2020).

[9] D. Bouwmeester, J.-W. Pan, K. Mattle, M. Eibl, H. Weinfurter, and A. Zeilinger, Experimental quantum teleportation, Nature **390,** 575 (1997).

[10] N. Gisin and R. Thew, Quantum communication, Nat. Photonics **1,** 165 (2007).

[11] R. Simon, Peres-Horodecki Separability Criterion for Continuous Variable Systems, Phys. Rev. Lett. **84,** 2726 (2000).

[12] G. M. D'Ariano, M. G. Paris, and M. F. Sacchi, in *Quantum State Estimation* (Springer Berlin, Heidelberg, 2004), Chap. 2, p. 7.

[13] K. Wang, Z. Song, X. Zhao, Z. Wang, and X. Wang, Detecting and quantifying entanglement on near-term quantum devices, Npj Quantum Inf. **8,** 1 (2022).

[14] R. Chen, Z. Song, X. Zhao, and X. Wang, Variational quantum algorithms for trace distance and fidelity estimation, Quantum Sci. Technol. **7,** 015019 (2021).

[15] M. A. Nielsen and I. L. Chuang, *Quantum Computation and Quantum Information* (Cambridge University Press, Cambridge, 2010).

[16] G. Vidal and R. F. Werner, Computable measure of entanglement, Phys. Rev. A **65,** 032314 (2002).

[17] R. LaRose, A. Tikku, É. O'Neel-Judy, L. Cincio, and P. J. Coles, Variational quantum state diagonalization, Npj Quantum Inf. **5,** 1 (2019).

[18] C. Bravo-Prieto, D. Garcia-Martin, and J. I. Latorre, Quantum singular value decomposer, Phys. Rev. A **101,** 062310 (2020).

[19] R. A. Horn and C. R. Johnson, *Matrix Analysis* (Cambridge University Press, Cambridge, 2012).

[20] N. J. Cerf and C. Adami, Quantum extension of conditional probability, Phys. Rev. A **60,** 893 (1999).

[21] M. Horodecki and P. Horodecki, Reduction criterion of separability and limits for a class of protocols of entanglement distillation, arXiv preprint, arXiv:quant-ph/9708015 (1998).

[22] J. Watrous, *The Theory of Quantum Information* (Cambridge University Press, Cambridge, 2018).

[23] C. Zhang, S. Yu, Q. Chen, H. Yuan, and C. Oh, Evaluation of entanglement measures by a single observable, Phys. Rev. A **94,** 042325 (2016).

[24] K. Antipin, Lower bounds on concurrence and negativity from a trace inequality, Mod. Phys. Lett. A **35,** 2050254 (2020).

[25] M.-J. Zhao, Z.-G. Li, S.-M. Fei, and Z.-X. Wang, A note on fully entangled fraction, J. Phys. A: Math. Theor. **43,** 275203 (2010).

[26] P. Badziag, M. Horodecki, P. Horodecki, and R. Horodecki, Local environment can enhance fidelity of quantum teleportation, Phys. Rev. A **62,** 012311 (2000).

[27] F. Verstraete and H. Verschelde, Fidelity of mixed states of two qubits, Phys. Rev. A **66,** 022307 (2002).

[28] N. Ganguly, S. Adhikari, A. S. Majumdar, and J. Chatterjee, Entanglement Witness Operator for Quantum Teleportation, Phys. Rev. Lett. **107,** 270501 (2011).

[29] G. Vidal, D. Jonathan, and M. A. Nielsen, Approximate transformations and robust manipulation of bipartite pure-state entanglement, Phys. Rev. A **62,** 012304 (2000).

[30] I. Bengtsson and K. Życzkowski, *Geometry of Quantum States: An Introduction to Quantum Entanglement* (Cambridge University Press, Cambridge, 2017).

[31] M. Cerezo, A. Arrasmith, R. Babbush, S. C. Benjamin, S. Endo, K. Fujii, J. R. McClean, K. Mitarai, X. Yuan, L. Cincio, and P. J. Coles, Variational quantum algorithms, Nat. Rev. Phys. **3,** 625 (2021).

[32] K. Bharti, A. Cervera-Lierta, T. H. Kyaw, T. Haug, S. Alperin-Lea, A. Anand, M. Degroote, H. Heimonen, J. S. Kottmann, T. Menke, W.-K. Mok, S. Sim, L.-C. Kwek, and A. Aspuru-Guzik, Noisy intermediate-scale quantum (NISQ) algorithms, Rev. Mod. Phys. **94,** 015004 (2022).

[33] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, Hybrid quantum-classical algorithms and quantum error mitigation, arXiv:2011.01382, 1 (2020).

[34] Y. Cao, J. Romero, J. P. Olson, M. Degroote, P. D. Johnson, M. Kieferová, I. D. Kivlichan, T. Menke, B. Peropadre, and N. P. D. Sawaya, Quantum chemistry in the age of quantum computing, Chem. Rev. **119,** 10856 (2019).

[35] K. M. Nakanishi, K. Mitarai, and K. Fujii, Subspace-search variational quantum eigensolver for excited states, Phys. Rev. Res. **1,** 033062 (2019).

[36] O. Higgott, D. Wang, and S. Brierley, Variational quantum computation of excited states, Quantum **3,** 156 (2019).

[37] J. Romero, J. P. Olson, and A. Aspuru-Guzik, Quantum autoencoders for efficient compression of quantum data, Quantum Sci. Technol. **2,** 045001 (2017).

[38] C. Cao and X. Wang, Noise-Assisted Quantum Autoencoder, Phys. Rev. Appl. **15,** 054012 (2021).

[39] X. Wang, Z. Song, and Y. Wang, Variational quantum singular value decomposition, Quantum **5,** 483 (2021).

[40] E. Farhi, J. Goldstone, and S. Gutmann, A quantum approximate optimization algorithm, arXiv:1411.4028 (2014).

[41] W. Li and D.-L. Deng, Recent advances for quantum classifiers, arXiv:2108.13421 (2021).

[42] M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, Circuit-centric quantum classifiers, Phys. Rev. A **101,** 032308 (2020).

[43] G. Li, Z. Song, and X. Wang, VSQL: Variational shadow quantum learning for classification, Proceedings of the AAAI Conference on Artificial Intelligence **35,** 8357 (2021).

[44] J. J. Meyer, J. Borregaard, and J. Eisert, A variational toolbox for quantum multi-parameter estimation, Npj Quantum Inf. **7,** 89 (2021).

[45] B. Koczor, S. Endo, T. Jones, Y. Matsuzaki, and S. C. Benjamin, Variational-state quantum metrology, New J. Phys. **22,** 083038 (2020).

[46] M. Schuld, Supervised quantum machine learning models are kernel methods, arXiv:2101.11020 [quant-ph] (2021).

[47] M. Plesch and i. c. v. Brukner, Quantum-state preparation with universal gate decompositions, Phys. Rev. A **83,** 032302 (2011).

[48] K. Mitarai, M. Negoro, M. Kitagawa, and K. Fujii, Quantum circuit learning, Phys. Rev. A **98,** 032309 (2018).

[49] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, Quantum natural gradient, Quantum **4,** 269 (2020).

[50] R. M. Parrish, J. T. Iosue, A. Ozaeta, and P. L. McMahon, A Jacobi diagonalization and anderson acceleration algorithm for variational quantum algorithm parameter optimization, arXiv:1904.03206 (2019).

[51] K. M. Nakanishi, K. Fujii, and S. Todo, Sequential minimal optimization for quantum-classical hybrid algorithms, Phys. Rev. Res. **2**, 043158 (2020).

[52] M. Ostaszewski, E. Grant, and M. Benedetti, Structure optimization for parameterized quantum circuits, Quantum **5**, 391 (2021).

[53] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Evaluating analytic gradients on quantum hardware, Phys. Rev. A **99**, 032331 (2019).

[54] M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini, Parameterized quantum circuits as machine learning models, Quantum Sci. Technol. **4**, 043001 (2019).

[55] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Barren plateaus in quantum neural network training landscapes, Nat. Commun. **9**, 4812 (2018).

[56] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Cost function dependent barren plateaus in shallow parametrized quantum circuits, Nat. Commun. **12**, 1791 (2021).

[57] A. Skolik, J. R. McClean, M. Mohseni, P. van der Smagt, and M. Leib, Layerwise learning for quantum neural networks, Quantum Mach. Intell. **3**, 1 (2021).

[58] T. Volkoff and P. J. Coles, Large gradients via correlation in random parameterized quantum circuits, Quantum Sci. Technol. **6**, 025008 (2021).

[59] A. Pesah, M. Cerezo, S. Wang, T. Volkoff, A. T. Sornborger, and P. J. Coles, Absence of barren plateaus in quantum convolutional neural networks, arXiv:2011.02966 [quant-ph] (2020).

[60] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, Expressibility and entangling capability of parameterized quantum circuits for hybrid quantum-classical algorithms, Adv. Quantum Technol. **2**, 1900070 (2019).

[61] https://qml.baidu.com.

[62] https://quantum-hub.baidu.com/.

[63] S. Das and G. Paul, Necessary and sufficient condition for the existence of Schmidt decomposition in multipartite Hilbert spaces, arXiv preprint arXiv:1805.12274 (2018).

[64] A. K. Pati, Existence of the Schmidt decomposition for tripartite systems, Phys. Lett. A **278**, 118 (2000).

[65] C. D. Bruzewicz, J. Chiaverini, R. McConnell, and J. M. Sage, Trapped-ion quantum computing: Progress and challenges, Appl. Phys. Rev. **6**, 021314 (2019).

[66] M. Kjaergaard, M. E. Schwartz, J. Braumüller, P. Krantz, J. I.-J. Wang, S. Gustavsson, and W. D. Oliver, Superconducting qubits: Current state of play, Annu. Rev. Condens. Matter Phys. **11**, 369 (2020).

[67] https://github.com/benchizhao/near-term-efficient-quantum-algorithms-or-entanglement-analysis.