Variational Quantum Algorithm for Schmidt Decomposition

Ranyiliu Chen,¹ Benchi Zhao,¹ and Xin Wang¹,∗

¹Institute for Quantum Computing, Baidu Research, Beijing 100193, China
(Dated: September 23, 2021)

Entanglement plays a crucial role in quantum physics and is the key resource in quantum information processing. In entanglement theory, Schmidt decomposition is a powerful tool to analyze the fundamental properties and structure of quantum entanglement. This work introduces a hybrid quantum-classical algorithm for Schmidt decomposition of bipartite pure states on near-term quantum devices. First, we show that the Schmidt decomposition task could be accomplished by maximizing a cost function utilizing bi-local quantum neural networks. Based on this, we propose a variational quantum algorithm for Schmidt decomposition (named VQASD) of which the cost function evaluation notably requires only one estimate of expectation with no extra copies of the input state. In this sense, VQASD outperforms existent approaches in resource cost and hardware efficiency. Second, by further exploring VQASD, we introduce a variational quantum algorithm to estimate the logarithm negativity, which can be applied to efficiently quantify entanglement of bipartite pure states. Third, we experimentally implement our algorithm on Quantum Leaf using the IoP CAS superconducting quantum processor. Both experimental implementations and numerical simulations exhibit 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 been found use in a variety of areas including quantum cryptography [2, 3], quantum chemistry [4, 5], quantum machine leaning [6, 7], and quantum communication [8–10].

Detecting the entanglement of a given unknown multiparty state is not a simple work, let alone quantifying the entanglement. Positive partial transpose (PPT) criterion [11] is a well-known way to detect whether a bipartite quantum system is entangled. However, this method is less practical comparing to its elegant representation, because the partial transpose is not physically implementable (not completely positive) while tomography of the state is very costly. Entanglement witness [12, 13] is another important methods of entanglement detection, while it requires prior knowledge of the state. Recently, Ref. [14] proposed to decompose the nonphysical operation partial transpose into physical Pauli operations, and estimate the positiveness using hybrid quantum-classical computation. This is an effective way to detect entanglement, but the required resources in the decomposition of transpose operation is still non-negligible. Moreover, the quantification of quantum entanglement is ever more difficult than entanglement detection.

The Schmidt decomposition [15] is a fundamental and powerful tool for analysing the entanglement between bipartite system. According to the Schmidt decomposition, a pure bipartite state is expressed as a simpler combination of tensor product of two orthonormal bases of sub-spaces, that is, \( |\psi\rangle = \sum_j c_j |u_j\rangle \otimes |v_j\rangle \). The positive coefficient of Schmidt decomposition \( c_j \) is known as Schmidt coefficient, and the number of Schmidt coefficients is called Schmidt rank or Schmidt number, determining whether this pure state is entangled or not. Fewer Schmidt rank means less entanglement, if the Schmidt rank of the given bipartite state equals to one, it is safe to say there is no entanglement between the two parties. Additionally, for entangled states, entanglement measures such as Von Neumann entropy and logarithm negativity can be inferred from the Schmidt coefficients [16].

The most straightforward approach to do Schmidt decomposition on a given quantum system is the state tomography [17] followed by a singular value decomposition on a classical computer. Nevertheless, it is known that state tomography has to consume a huge amount of state copies, and singular value decomposition demands exponential many classical storage in the number of qubit. Alternatively, several variational quantum algorithms [18, 19] have been proposed to leverage the power of hybrid quantum-classical computing. In [18], the Schmidt coefficients are read out from the variationally diagonalized reduce states. In [19], local unitaries are performed and optimized on each party, and when the measurement of each party coincides the Schmidt coefficients can be evaluated. However, these algorithms requires either extra qubit resource [18] or multiple expectation estimates [19] in each iteration. All these high costs make the algorithms less practical.

In this work, we present a variational Schmidt decomposition algorithm which, comparing to [18] and [19], demands no extra qubit and estimates only one probability as the cost function for each optimization iteration. In our approach, bi-local parameterized quantum circuits (PQCs) or quantum neural networks are acted on the two subsystems, followed by a depth-2 subcircuit which calculates the fidelity with another elaborated entangled state. The von Neumann’s trace inequality [20] guarantees that when this fidelity is maximized through tuning the circuit parameters, the Schmidt decomposition is accomplished. By exploiting symmetry, we further introduce a variational quantum algorithm to estimate the logarithm negativity of bipartite pure states. Numerical simulation and quantum device implementation in the (up to) 8-qubit and 2-qubit cases are undertaken, respectively, illustrating the practicality and validity of our algorithm in NISQ de-
vices. Our algorithms for Schmidt decomposition and logarithm negativity estimation notably provides efficient and practical ways to analyze and quantify entanglement on near-term quantum devices.

The rest of this paper is organized as follows: Section II gives preliminaries and notations used in this paper. Section III provides the theoretical foundations of our algorithm design of which the details about cost function evaluation and the adopted optimization is introduced in Section IV. In Section V we exhibit experiments including numerical simulation and quantum device implementation. Then, we compare our work with existing works in Section VI. At last, Section VII concludes this work and discusses outlooks.

II. PRELIMINARIES AND NOTATIONS

The Schmidt decomposition for a bipartite quantum state |φ⟩_{AB} living in composed Hilbert space $\mathcal{H}_{AB}$ is defined as [15]

$$|\psi⟩_{AB} = \sum_{j=0}^{R-1} c_j |u_j⟩_A |v_j⟩_B,$$  

where \{u_j⟩_A\} and \{v_j⟩_B\} are some orthonormal basis in subspaces $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. $c_j$ are positive numbers called the Schmidt coefficients. Here we order the coefficients 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 dimension of $\mathcal{H}_A$ and $\mathcal{H}_B$, respectively. Without loss of generality, in the rest of this paper we consider natural partition of 2n-qubit system, i.e., $d = d_A = d_B = 2^n$.

A frequently used identity in entanglement theory is the “transpose trick”: for any operator $M$, 

$$(M_A \otimes I_B) |\psi⟩_{AB} = (I_A \otimes M_B^T) |\psi⟩_{AB},$$

where $M_B^T$ denotes the transpose of $M_B$, and $|\psi⟩_{AB} = 1/d \sum_j |j⟩_A |j⟩_B$ is the maximally entangled state on the composed system AB.

In this paper, the standard computational basis elements of an $n$-qubit system are denoted by $|j⟩$, $j = 0, \ldots, 2^n - 1$, i.e., $|0⟩ = (1, 0, \ldots, 0)^T$, $|1⟩ = (0, 1, \ldots, 0)^T$, $\ldots$, $|2^n - 1⟩ = (0, 0, \ldots, 1)^T$. For composed systems, we write an operator with a subscript indicating the system that the operator acts on, such as $M_{AB}$, and write $M_A := Tr_B M_{AB}$.

III. THEORETICAL FRAMEWORK

Our algorithm bases on the observation that the Schmidt coefficients could be achieved by optimizing the parameters in local PQCs. In this section, we provide the theorem which converts the Schmidt decomposition into an optimization over unitaries, then we derive a corollary to calculate logarithm negativity from the proved theorem.

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

Theorem 1 Let $U_A$ and $V_B$ be unitaries on system $A$ and $B$, respectively. Let $|Ψ⟩_{AB} = \sum_j p_j |j⟩_A |j⟩_B$ where \{p_j\} is positive and ordered strictly decreasingly. For any bipartite state $|ψ⟩_{AB} = \sum_j c_j |u_j⟩_A |v_j⟩_B$ with decreasing coefficients $c_j$, denote $|ψ⟩_{AB} = U_A \otimes V_B |ψ⟩_{AB}$, it holds that

$$\max_{U_A, V_B} F(|ψ⟩_{AB}, |Ψ⟩_{AB}) = \sum_j p_j c_j,$$

and when the maximal of Eq. (3) is reached, the Schmidt coefficients can be readout by

$$c_j = F(\tilde{ρ}_A, |j⟩_A),$$

where $\tilde{ρ}_A = Tr_B(|ψ⟩⟨ψ|_{AB})$. Further 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 $e^{iθ_{AB}}$) by

$$e^{iθ_{AB}} ^j |u_j⟩_A = U_A^j |j⟩_A,$$

$$e^{iθ_{AB}} ^j |v_j⟩_B = V_B^j |j⟩_B,$$

for some real $θ_{AB}$.

Proof Recall the operator-vector correspondence mapping [21] defined by the action:

$$\text{vec}(|j⟩⟨k|) = |j⟩⟨k|.$$  

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

$$\text{mat}(|j⟩⟨k|) = \text{vec}^{-1}(|j⟩⟨k|) = |j⟩⟨k|,$$

which is also isometric. Let $X = \text{mat}(|ψ⟩_{AB})$ and $Y = \text{mat}(|Ψ⟩_{AB})$, then $F(|ψ⟩_{AB}, |Ψ⟩_{AB}) = |⟨Ψ|_{AB} |ψ⟩_{AB}| = |\text{Tr} X Y|$, by isometry. What’s more, $c_j$ and $p_j$ are singular values of $X$ and $Y$ in decreasing orders, respectively, bringing about $|\text{Tr} X Y| \leq \sum_j c_j p_j$ via the Von Neumann’s trace inequality [20]. To sum up, we have

$$F(|ψ⟩_{AB}, |Ψ⟩_{AB}) = |⟨Ψ|_{AB} U_A \otimes V_B |ψ⟩_{AB}| = |\text{Tr} X Y| \leq \sum_j c_j p_j.$$  

(8)

Thus the fidelity will not exceed $\sum_j c_j p_j$. Evidently, when $U_A$ maps \{u_j⟩_A\} to \{j⟩_A\} and $V_B$ maps \{v_j⟩_B\} to \{j⟩_B\}, the maximal is reached. Then Eq. (3) holds.
When the maximal of Eq. (3) is reached, by the Von Neumann’s trace inequality we have that $U_A |u_j⟩_A$ (correspondingly, $V_B |v_j⟩_B$) falls in the subspace spanned by $|k⟩_A |c_k⟩$ (correspondingly, $|k⟩_B |c_k⟩ = |c_j⟩$), thus Eq. (4) holds; if $c_j$ is further non-degenerate, then Eq. (5) holds naturally, which complete the proof of Theorem 1.

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 $F(⟨\tilde{\psi}⟩_{AB}, |\Psi⟩_{AB})$ (in Sect. IV A we choose $F^2$ to be the cost function for convenience). By adjusting circuits’ parameters we can maximize the cost function. Finally after the optimization we can readout the required Schmidt coefficient. In Section IV A the evaluation of the cost function and the readout process will be discussed in detail. It is worth noting that, even we can prepare each decomposition vectors $|e_j⟩_{A,B}$, we might not be able to reconstruct $|ψ⟩_{AB}$ via $U_A$ and $V_B$ because the relative phases for each component might not be equal, i.e., $|ψ⟩_{A,B}$ are not always the same for different $j$.

After extracting all the Schmidt coefficients from the above approach, entanglement measures such as logarithm negativity can be calculated by (classical) post-processing [16]. Significantly, in the case that 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 $|Ψ⟩_{AB}$ with the bipartite maximally entangled state and the transpose trick. Specifically, we have the following corollary:

**Corollary 2** Let $U_A$ be a unitary on system $A$. For any bipartite state $|ψ⟩_{AB} = \sum_j c_j |u_j⟩_A |v_j⟩_B$ with decreasing coefficients $c_j$, denote $|\tilde{ψ}⟩_{AB} = U_A \otimes I_B |ψ⟩_{AB}$, it holds that

$$\max_{U_A} F(⟨\tilde{ψ}⟩_{AB}, |Φ⟩_{AB}) = \frac{\sum_j c_j}{\sqrt{d}},$$

where $|Φ⟩_{AB} = 1/\sqrt{d} \sum_j |j⟩_A |j⟩_B$ is the bipartite maximally entangled state.

**Proof** In Eq. (3), let $p_j = 1/\sqrt{d}$, we have that:

$$F(U_A \otimes V_B |ψ⟩_{AB}, |Φ⟩_{AB}) \leq \sum_j c_j p_j = \frac{\sum_j c_j}{\sqrt{d}}. \tag{10}$$

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

Now note that

$$\max_{U_A, V_B} F(U_A \otimes V_B |ψ⟩_{AB}, |Φ⟩_{AB}) = \max_{U_A, V_B} |⟨ψ⟩_{AB} U_A \otimes V_B |Φ⟩_{AB}|$$

$$= \max_{U_A, V_B} |⟨ψ⟩_{AB} U_A V_B^† \otimes I_B |Φ⟩_{AB}|$$

$$= \max_{U_A, V_B} F(V_B U_A \otimes I_B |ψ⟩_{AB}, |Φ⟩_{AB})$$

$$= \max_{U_A} F(⟨\tilde{ψ}⟩_{AB}, |Φ⟩_{AB}),$$

where the second equation comes from the transpose trick (Eq. (2)). Then Eq. (9) holds, completing the proof of Lemma 2.

Corollary 2 can lead to a simple estimation of the logarithm negativity for pure state $|ψ⟩_{AB}$. Note that the logarithm negativity of $|ψ⟩_{AB}$ can be written as [16]

$$E_N(|ψ⟩_{AB}) = \log_2(2\mathcal{N} + 1) = \log_2 \left( \sum_j c_j^2 \right), \tag{12}$$

where $\mathcal{N}(|ψ⟩_{AB}) = 1/2[(\sum_j c_j^2) - 1]$ is the negativity of $|ψ⟩_{AB}$. Thus, if $\max F(⟨\tilde{ψ}⟩_{AB}, |Φ⟩_{AB})$ is estimated by our variational approach, the logarithm negativity follows immediately. Notably, Corollary 2 enables us to employ a one-side PQC (on either $A$ or $B$ system) in the variational learning, saving appreciable amount of computational resource in classical optimization.

**IV. VARIATIONAL QUANTUM ALGORITHMS**

With the theoretical framework in hand we develop the variational quantum algorithms for Schmidt decomposition and logarithm negativity. Here, the variational quantum algorithm (VQA) [22–24] 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 states preparation [25–27], quantum data compression [28–30], combinatorial optimization [31], quantum classifier [32–34], and quantum metrology [35, 36].

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. 1 and the algorithm boxes are given in SM III.

**A. Cost functions**

We first look at the cost function in the Schmidt decomposition task. Recall that $\max F(⟨\tilde{ψ}⟩_{AB}, |Φ⟩_{AB}) = \sum_j p_j c_j$ according to Theorem 1. In the variational algorithm, we set our cost function to be:

$$C^{SD} := F^2(⟨\tilde{ψ}⟩_{AB}, |Ψ⟩_{AB}) = \text{Tr} [\bar{ρ}_{AB} |Ψ⟩⟨Ψ|_{AB}]. \tag{13}$$

Recall that $\bar{ρ}_{AB} = |Ψ⟩⟨ψ⟩_{AB}$ and $|ψ⟩_{AB} = U_A \otimes V_B |ψ⟩_{AB}$. Evidently $F$ reaches its maximal iff $C^{SD}$ reaches maximal. Thus after the cost function $C$ is optimized the Schmidt coefficients can be read out through measurement:

$$c_j = F(⟨\tilde{ψ}⟩_{A}, |j⟩_A) = \sqrt{\text{Tr} [\bar{ρ}_A |j⟩⟨j|_A]}. \tag{14}$$

We then show that the cost function can be evaluated efficiently on NISQ device. Suppose that circuit $W$ prepares the state $|Ψ⟩_{AB} = W |0⟩_{AB}$. By the cyclic property of trace we
have
\[ C^{SD} = \text{Tr}[\tilde{\rho}_{AB} W |0\rangle\langle 0|_A W^\dagger] \]
\[ = \text{Tr}[W^\dagger \tilde{\rho}_{AB} W |0\rangle\langle 0|_A]. \] (15)

Thus, the cost function \( C^{SD} \) is equal to the probability of all zero measurement outcome of \( |\tilde{\psi}\rangle_{AB} \) acted by \( W^\dagger = W^{-1} \). So the key to evaluate \( C \) 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 \( n \) CNOT gates connecting all qubit pairs across system \( A \) and \( B \). If we preset the coefficients \( p_j \), then constructing the circuit for \( W_A \) is essentially the amplitude encoding [37] task, for which the best known construction[38] requires \( O(2^n) \) CNOT-count and \( O(2^n) \) depth. Nevertheless, in our algorithm we only require the coefficients \( \{p_j\} \) to be decreasing. We provide a circuit construction for this ‘weak’ state preparation task which uses \( n \) single qubit gates in parallel. In fact, the circuit composed by \( n \) paralleled \( y \)-axis rotation gate

\[
\begin{bmatrix}
|0\rangle \\
|0\rangle \\
\vdots \\
|0\rangle
\end{bmatrix} \xrightarrow{R_y(\alpha_1)}
\begin{bmatrix}
R_y(\alpha_1) |0\rangle \\
R_y(\alpha_2) |0\rangle \\
\vdots \\
R_y(\alpha_n) |0\rangle
\end{bmatrix}
\]

with carefully chosen parameters \( \{\alpha_j\} \) satisfies the condition that \( p_j \) decreases for arbitrary \( n \). When \( n \) trivially equals 1 the parameter \( \alpha_1 \) can be set to 0. We refer the selection of \( \alpha_j \) for \( n \geq 2 \) in SM 1. Note that the total depth in this construction of \( W \) is 2 with CNOT-count \( \log_2(C_{\text{max}}^n) \), leading to a efficient implementation of \( W \) (thus \( W^{-1} \)) in the cost function evaluation.

We next discuss about the logarithm negativity estimation. In this case, since \( \max F(|\tilde{\psi}\rangle_{AB}, |\Phi\rangle_{AB}) = \sum_j |c_j|^2 \) according to Corollary 2, the cost function is set as:

\[ C^{LN} := F^2(|\tilde{\psi}\rangle_{AB}, |\Phi\rangle_{AB}) = \text{Tr} [\tilde{\rho}_{AB} |\Phi\rangle\langle \Phi|_AB] . \] (16)

Evidently, \( C_{\text{max}}^n = (\sum_j |c_j|^2)^2 \). Use the variational method to maximize the cost function \( C^{LN} \), then the logarithm negativity is estimated as

\[ E_N(|\tilde{\psi}\rangle_{AB}) = \log_2(C_{\text{max}}^n + n) . \] (17)

The cost function \( C^{LN} \) also has an efficient evaluation with similar technique 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

\[ C^{LN} = \text{Tr} [\tilde{\rho}_{AB} |\Phi\rangle\langle \Phi|_AB] \]
\[ = \text{Tr} [W^\dagger \tilde{\rho}_{AB} W^\dagger |0\rangle\langle 0|_AB] , \] (18)

which is the probability of all zero measurement outcome of \( |\tilde{\psi}\rangle_{AB} \) acted by \( W^\dagger \).

B. Parameterized quantum circuit

As discussed in Section III, we adopt the bi-local PQC \( U(\theta_1) \otimes V(\theta_2) \) for variational Schmidt decomposition and the one-side PQC \( U(\theta_1) \otimes I \) for logarithm negativity estimation. Since our algorithm makes no assumption of input states, we recommend the class of hardware-efficient PQCs [39] as the tunable unitary to further enhance the experimental realizability of our algorithms. In the experiments, the PQCs consists of single qubit rotations and CNOTs or CZs as the entangled gates. We refer the readers to subsections of Section V for more details of the circuit implementation in each experiments.
C. Optimization

Both gradient-based [40, 41] and gradient-free [42–44] optimization methods can be applied to our variational quantum algorithms. For gradient-based optimization, the analytic gradients of both $C_{SD}$ and $C_{LN}$ are supported by the parameter shift rule [45], which provides unbiased estimate comparing to the finite difference method [46]. We have adopted various optimization method in both the numerical simulation and the real-device implementation of our algorithms (see Section V for more details about the optimization adopted in our experiments).

Here we also discuss the possible solutions to the gradient vanishing issue for VQASD. According to [47], VQASD could suffer from the ‘Barren Plateaus’ throughout training. Additionally, this gradient vanishing issue is unlikely to be solved by a local cost function [48] due to the global connectivity brought by subcircuit $W^{-1}$. On the other hand, proposed strategies including layerwise learning [49], parameter correlations [50], and a quantum convolutional neural network ansatz [51] have been shown to be helpful in certain training tasks. We leave the adaptation of these and other training techniques to our VQASD for a future study.

V. NUMERICAL SIMULATION AND IMPLEMENTATION ON QUANTUM DEVICE

In this section, we will show the effectiveness of variational quantum algorithm in Schmidt decomposition through numerical simulation and implementation on quantum device. The simulation experiments are operated on the Paddle quantum platform and the Quantum Leaf platform. We also realize our algorithm on the superconducting quantum device of Institute of Physics, CAS through the Quantum Leaf platform.

A. Simulating on circuits with different depth

Here, we conduct numerical experiments to decompose a bipartite quantum system with 8 qubits to investigate the impact of different numbers of circuit layers. At the beginning, we randomly generate a 8-qubit bipartite entangled state $|\psi\rangle_{AB}$, both parties have 4 qubits. Then, they apply PQCs $U_A(\theta_A)$ and $V_B(\theta_B)$ on two parties respectively. The structure of the PQCs used in this task are shown in Fig. 2. As discussed above we adopt Eq. (15) as the cost function. This simulation is performed on the Paddle Quantum platform.

In this experiment, we set the DEPTH of layers of PQCs to be 1, 2, 4, and 8, which means apply the structure in Fig. 2 recurrently for DEPTH times, to clarify the impact of the number of circuit layers. The results are shown in Fig. 3. We use ADAM optimizer (a gradient-based optimizer) to maximize our cost function.

The simulation results are shown in Fig. 3. 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 $c'_j$ are the estimated Schmidt coefficients and $c_j$ are the actual values. It is obvious that circuits with larger DEPTH have better performance in accuracy, satisfying the intuition that deeper circuit has better expressibility [52].

B. Decomposing noisy state

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

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

$$N_{dep}^{\rho}(\rho) := (1 - p) \rho + p \text{Tr}(\rho) \frac{I}{2},$$

respectively, where $I$ stands for identity, and

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

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

$$= \begin{pmatrix}
0.455 & 0.459 & 0.136 & -0.137 \\
0.459 & 0.463 & 0.137 & -0.138 \\
0.136 & 0.137 & 0.041 & -0.041 \\
-0.137 & -0.138 & -0.041 & 0.041
\end{pmatrix}. \quad (20)$$

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 $U(\theta_1)$ and $V(\theta_2)$ which are single-qubit universal ansatz. This simulation is performed on the Paddle quantum platform.

![FIG. 4: Decompose quantum state affected by noisy state. The blue circles and triangles are the Schmidt coefficients of amplitude damping noised state. The red circles and triangles are the Schmidt coefficients of depolarizing noised state. The black dashed line are the theoretical Schmidt coefficients of the state without being noised.](image)

The experiment results are displayed in Fig. 5. As demonstrated, the cost raises dramatically in first two iterations, and the simulator converges to a higher value than that of quantum device, which may caused by the quantum device noises. Correspondingly, the achieved Schmidt values by simulator ($c_1 = 0.958 \pm 0.001$ and $c_2 = 0.286 \pm 0.002$) is closer to the theoretical values ($c_1 = 0.958$ and $c_2 = 0.286$) than that of quantum device ($c_1 = 0.919 \pm 0.006$ and $c_2 = 0.311 \pm 0.003$).

![FIG. 5: Learning curves respect to simulator and quantum device. The shadowed area spans from 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 $\pm$ 0.001 and 0.286 $\pm$ 0.002, while the results from CAS quantum device are 0.919 $\pm$ 0.006 and 0.311 $\pm$ 0.003.](image)

**C. Implementations on superconducting quantum processor**

We also apply our algorithm to decompose a 2-qubits bipartite quantum system on quantum device, and compare the result with the values achieved from simulation. The experiments are performed on the Quantum Leaf platform, loading the quantum device from Institution of Physics, Academy of Science (IoP CAS). This quantum device containing 10 direct coupling transmon qubits, whose topology is neighbor coupling one-dimensional chain-like.

The input state $|\psi\rangle_{AB}$ we use here is exactly the same as the one in Sec. V B. On the quantum device, the input state could be prepared by applying two $R_y$ gates on the first and second qubits with parameters $0.58$ and $1.58$ respectively, followed by a Control-$Z$ gate. Next, we operate parameterized quantum circuit on qubits, and use sequential minimal optimization [43] (gradient-free) to optimize the parameters until the cost converge to its maximum. We repeat 20 independent experiments with same input states and randomly initialized parameters.

**D. Estimating logarithm negativity**

From Corollary 2, 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 with theoretical values.

In this simulation, $|\psi\rangle_{AB}$ is a bipartite with 3 qubits in each party. In order to make the results comparable, the am-
Quantum Leaf

... with the theoretical values. This simulation is performed on the Quantum Leaf platform. The results are shown in Fig. 7. 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.

![Parameterized Circuit ansatz for logarithm negativity estimation](image)

**FIG. 6**: Parameterized Circuit ansatz for logarithm negativity estimation. The circuits in the dashed box should be repeated three times.

![Logarithm negativity estimation with respect to different Schmidt ranks](image)

**FIG. 7**: Logarithm negativity estimation with respect to different Schmidt ranks. The red bar is the median logarithm negativity over 10 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.

VI. COMPARISONS

We here make a detailed comparison between our VQASD and existing Schmidt decomposition methods proposed in [18] and [19], which are also based on the hybrid quantum-classical approach. Essentially, all three methods performs local unitaries on each party. And after the optimization, all three methods should diagonalize the reduce states. In this sense, it is reasonable to assume that three methods need same expressibility [52] for ansatz. However, comparing to [18] and [19], VQASD demands no extra qubit and estimates one probability as an observable expectation for each optimization iteration, therefore is more efficient in terms of quantum resources consumption. We explicitly compare the consumed resources (quantum registers, copies of states in each circuit run, estimates of expectation in each optimization iteration) in term of \( n \), as well as quantum device implementation in Table I.

Recall that in [18] the authors estimate the Schmidt coefficients via diagonalizing the reduce state. In their approach (with global cost function), two copies of the state is consumed in each optimization iteration. Thus, the resource of state preparation as well as the quantum register is doubled. In [19], the Hamming weight of bipartite measurement outcome is adopted by the authors as their cost function. According to its evaluation, this 2-local cost function is calculated from \( n \) measurement outcomes, requiring \( n \) expectation estimates in each optimization iteration, while our QVQSD estimates one probability (although our circuit is 2-layer deeper than what in [19]). Comparing to the above hybrid approaches, our algorithm is therefore more hardware-efficient.

VII. CONCLUSION AND OUTLOOK

We proposed a variational quantum algorithm to realize Schmidt decomposition (VQASD) on near-term quantum devices. Comparing to previous methods, our VQASD is more efficient in terms of resource consumption. Simultaneously, a new method to compute the logarithm negativity on near-term quantum devices was derived from VQASD. We have shown the validity and practicality of our methods via both numerical simulations and experimental implementations. Numerical simulations shows that VQASD is resilient to amplitude damping and depolarizing noises. And 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 for Schmidt decomposition and logarithm negativity estimation notably provides efficient and practical tools for entanglement quantification and analysis in the NISQ era.

Beyond quantifying entanglement for bipartite pure states, our VQASD could have a wide range of extensions and applications. One direct extension of VQASD is the to consider the multiparty case. We briefly argue that, if the Schmidt decomposition of a multiparty entangled state exists [53, 54], VQASD can be adapted to decompose these states. We leave the adaptation of VQASD’s techniques to investigate multi-
explore the power of bi-local PQC, which is a key ingredient VQASD training. Moreover, it will be interesting to further reduce its scale after acting the unitary obtained through (or numbers of qubits), it is always possible for the larger party example, if the two party of the state have distinct dimensions includes application such as quantum data compression. For states for future study. Future direction of our work also in- party pure states without Schmidt decomposition or mixed states for future study. Future direction of our work also includes application such as quantum data compression. For example, if the two party of the state have distinct dimensions (or numbers of qubits), it is always possible for the larger party to reduce its scale after acting the unitary obtained through VQASD training. Moreover, it will be interesting to further explore the power of bi-local PQC, which is a key ingredient in our method.

TABLE I: Resource comparison for $2n$-qubit state Schmidt decomposition

| methods                  | Quantum Register Copies of $|\psi\rangle$ | Estimates of Expectations | Implementation                        |
|--------------------------|-----------------------------|-----------------------------|---------------------------------------|
| State Diagonization [18] | $4n$                        | $2n$                        | 1 (global cost)                       | Rigetti’s 8Q-Agave                     |
| Hamming Weight [19]      | $2n$                        | $2n$                        | 1                                     | -                                     |
| This work                | $2n$                        | $2n$                        | 1 (using the IoP CAS quantum processor)| Quantum Leaf                           |

Acknowledgements.

We thank Zelin Meng for helpful discussions. R. C. and B. Z. contributed equally to this work. This work was done when R. C. and B. Z. were research interns at Baidu Research.

[1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Reviews of modern physics 81, 865 (2009).
[2] D. Naik, C. Peterson, A. White, A. Berglund, and P. G. Kwiat, Physical Review Letters 84, 4733 (2000).
[3] A. K. Ekert, Physical review letters 67, 661 (1991).
[4] S. Mcardle, S. Endo, A. Aspuru-Guzik, S. C. Benjamin, and X. Yuan, Reviews of Modern Physics 92, 015003 (2020).
[5] F. Arute, K. Arya, R. Babbush, D. Bacon, J. C. Bardin, R. Barends, S. Boixo, M. Broughton, B. B. Buckley, D. A. Buell, et al., arXiv preprint arXiv:2004.04174 (2020).
[6] J. Biamonte, P. Wittek, N. Pancotti, P. Rebentrost, N. Wiebe, and S. Lloyd, 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, L. Cincio, et al., arXiv preprint arXiv:2012.09265 (2020).
[8] C. H. Bennett and G. Brassard, arXiv preprint arXiv:2003.06557 (2020).
[9] D. Bouwmeester, J. W. Pan, K. Mattle, M. Eibl, H. Weinfurter, and A. Zeilinger, Nature 390, 575 (1997).
[10] N. Gisin and R. Thew, Nature photonics 1, 165 (2007).
[11] R. Simon, Physical Review Letters 84, 2726 (2000).
[12] O. Gühne and N. Lütkenhaus, Phys. Rev. Lett. 96, 170502 (2006).
[13] O. Gühne and G. Toth, Physics Reports 474, 1 (2009).
[14] K. Wang, Z. Song, X. Zhao, Z. Wang, and X. Wang, arXiv preprint arXiv:2012.14311 (2020).
[15] M. A. Nielsen and I. L. Chuang, Quantum computation and quantum information (Cambridge university press, 2010).
[16] G. Vidal and R. F. Werner, Phys. Rev. A 65, 032314 (2002).
[17] G. M. D’Ariano, M. G. Paris, and M. F. Sacchi, “Quantum tomographic methods,” in Quantum State Estimation (Springer, 2004) Chap. 2, pp. 7–58.
[18] R. LaRose, A. Tikku, É. O’Neel-Judy, L. Cincio, and P. J. Coles, npj Quantum Information 5, 1 (2019).
[19] C. Bravo-Prieto, D. Garcia-Martin, and J. I. Latorre, Phys. Rev. A 101, 062310 (2020).
[20] R. A. Horn and C. R. Johnson, Matrix analysis (Cambridge university press, 2012).
[21] J. Watrous, The Theory of Quantum Information (Cambridge University Press, 2018).
[22] 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, Nature Reviews Physics, 1 (2021), arXiv:2012.09265.
[23] 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, arXiv:2101.08448, 1 (2021), arXiv:2101.08448.
[24] S. Endo, Z. Cai, S. C. Benjamin, and X. Yuan, arXiv:2011.01382, 1 (2020), arXiv:2011.01382.
[25] 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, Chemical reviews 119, 10856 (2019).
[26] K. M. Nakanishi, K. Mitarai, and K. Fujii, Physical Review Research 1, 033062 (2019).
[27] O. Higgott, D. Wang, and S. Brierley, Quantum 3, 156 (2019), arXiv:1805.08138.
[28] J. Romero, J. P. Olson, and A. Aspuru-Guzik, Quantum Science and Technology 2, 045001 (2017), arXiv:1612.02806.
[29] C. Cao and X. Wang, Physical Review Applied 15, 054012 (2021), arXiv:2012.08331.
[30] X. Wang, Z. Song, and Y. Wang, Quantum 5, 483 (2021), arXiv:2006.02336.
[31] E. Farhi, J. Goldstone, and S. Gutmann, arXiv:1411.4028 (2014), arXiv:1411.4028.
[32] W. Li and D.-L. Deng, arXiv:2108.13421 (2021), arXiv:2108.13421.
[33] M. Schuld, A. Bocharov, K. M. Svore, and N. Wiebe, Physical Review A 101, 032308 (2020), arXiv:1804.00633.
[34] G. Li, Z. Song, and X. Wang, Proceedings of the AAAI Conference on Artificial Intelligence, 35(9), 8357–8365 (2021).
[35] J. J. Meyer, J. Borregaard, and J. Eisert, npj Quantum Information 7, 89 (2021), arXiv:2006.06303.
[36] B. Koczor, S. Endo, T. Jones, Y. Matsuzaki, and S. C. Benjamin, New Journal of Physics, 35(9), 8357-8365 (2021).
[37] M. Schuld, “Supervised quantum machine learning models are kernel methods,” arXiv:2101.11020 [quant-ph].
[38] M. Plesch and i. c. v. Brukner, Phys. Rev. A 83, 032302 (2011).
[39] 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,” (2021), arXiv:2101.08448 [quant-ph].

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

[41] J. Stokes, J. Izaac, N. Killoran, and G. Carleo, Quantum 4, 269 (2020).

[42] R. M. Parrish, J. T. Iosue, A. Ozaeta, and P. L. McMahon, arXiv, 1904.03206 (2019).

[43] K. M. Nakanishi, K. Fujii, and S. Todo, Phys. Rev. Research 2, 043158 (2020).

[44] M. Ostaszewski, E. Grant, and M. Benedetti, Quantum 5, 391 (2021).

[45] M. Schuld, V. Bergholm, C. Gogolin, J. Izaac, and N. Killoran, Physical Review A 99 (2019), 10.1103/physreva.99.032331.

[46] M. Benedetti, E. Lloyd, S. Sack, and M. Fiorentini, Quantum Science and Technology 4, 043001 (2019).

[47] J. R. McClean, S. Boixo, V. N. Smelyanskiy, R. Babbush, and H. Neven, Nature Communications 9 (2018), 10.1038/s41467-018-07090-4.

[48] M. Cerezo, A. Sone, T. Volkoff, L. Cincio, and P. J. Coles, Nature Communications 12 (2021), 10.1038/s41467-021-21728-w.

[49] A. Skolik, J. R. McClean, M. Mohseni, P. van der Smagt, and M. Leib, Quantum Machine Intelligence 3, 1 (2021).

[50] T. Volkoff and P. J. Coles, Quantum Science and Technology 6, 025008 (2021).

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

[52] S. Sim, P. D. Johnson, and A. Aspuru-Guzik, Advanced Quantum Technologies 2, 1900070 (2019).

[53] S. Das and G. Paul, arXiv preprint arXiv:1805.12274 (2018).

[54] A. K. Pati, Physics Letters A 278, 118 (2000).
I. SELECTION OF PARAMETERS IN \( W_A \)

Here we show that how to choose parameters \( \alpha_j \) for \( j \in [1, n] \) in the circuit \( W_A \) defined in Section IV A when \( n \geq 2 \).

We first choose three real numbers \( \beta_n, \beta_{n-1} \), 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 \( O(n) \).

To show that \( W_A \) prepares the state \( |\Phi\rangle_A = \sum_j p_j |j\rangle_A \) such that \( p_j \) strictly decreases, let’s look at the coefficients \( \{p_j\} \) determined by \( \{\alpha_j\} \). The output state can be written as

\[
|\Psi\rangle_A = \bigotimes_{j=1}^{n} \left( \cos \frac{\alpha_j}{2} |0\rangle_j + \sin \frac{\alpha_j}{2} |1\rangle_j \right) = \bigotimes_{j=1}^{n} \left( \sqrt{\gamma_j} |0\rangle_j + \sqrt{1 - \gamma_j} |1\rangle_j \right). \tag{S1}
\]

Note that \( \beta_j > 1 \) implies \( \gamma_j > 1 - \gamma_j \). Let the binary representation of \( j \) be \( j = (j_1 j_2 \ldots j_n)_{2} \) where \( j_i, l \in [1, n] \) are binary bits. Then the coefficient \( p_j \) is

\[
p_j = \prod_{l=1}^{n} \sqrt{\gamma_l^{1 - j_l} (1 - \gamma_l)^{j_l}}. \tag{S2}
\]

(For example, when \( n = 4 \) and \( j = 9 = (1001)_{2} \), we have \( p_j = p_9 = \sqrt{\gamma_1^{2} \gamma_2^{1} \gamma_3^{1} \gamma_4^{1} (1 - \gamma_1)^{1} (1 - \gamma_2)^{1} (1 - \gamma_3)^{1} (1 - \gamma_4)^{1}} = \sqrt{(1 - \gamma_1) (1 - \gamma_2) (1 - \gamma_3) (1 - \gamma_4)} \).

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 \( j_+ = (j_1 \ldots j_{n_0} 11 \ldots 1)_{2} \) and \( k_- = (k_1 \ldots k_{n_0} 00 \ldots 0)_{2} \), so \( j \leq j_+ \) and \( k \geq k_- \). Also we have that

\[
\frac{p_j}{p_{j_+}} = \prod_{l=n_0+1}^{n} \frac{1}{1 - \gamma_l^{1 - j_l}} \geq 1, \quad \frac{p_k}{p_{k_-}} = \prod_{l=n_0+1}^{n} \frac{1 - \gamma_l^{j_l}}{1 - \gamma_l} \leq 1, \tag{S3}
\]

so \( p_j \geq p_{j_+} \) and \( p_k \leq p_{k_-} \). Finally, we compare \( p_{j_+} \) and \( p_{k_-} \):

\[
\frac{p_{j_+}}{p_{k_-}} = \frac{\gamma_{n_0} \prod_{l=n_0+1}^{n} 1 - \gamma_l}{\gamma_l} = \sqrt{\frac{\gamma_{n_0} \prod_{l=n_0+1}^{n} 1 - \gamma_l}{\gamma_l}}. \tag{S4}
\]

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

\[
\frac{p_{j_+}}{p_{k_-}} = \sqrt{\frac{\gamma_{n_0} \prod_{l=n_0+1}^{n} 1 - \gamma_l}{\gamma_l}} = \sqrt{\beta_{n_0} \prod_{l=n_0+1}^{n} \beta_l} > 1. \tag{S5}
\]

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

II. CIRCUITS FOR PREPARING STATE WITH DIFFERENT SCHMIDT RANKS

Here we display the circuits to generate a 6-qubit bipartite entangled state with desired Schmidt rank. The circuits are shown below. It is notable that the first three qubits belongs to one party and left three qubits belongs to the other. For example, if we wish to prepare a input state with Schmidt rank equals to 5, we should firstly apply the circuit in Fig. 8(e), then operate the second circuit in Fig. 8(i).
(a) Schmidt rank=1.

(b) Schmidt rank=2.

(c) Schmidt rank=3, parameter $\alpha = 1.23959$.

(d) Schmidt rank=4.

(e) Schmidt rank=5, parameter $\alpha = 0.927295$.

(f) Schmidt rank=6, parameter $\alpha = 1.230959$.

(g) Schmidt rank=7, parameter $\alpha_1 = 1.4, \alpha_2 = \pi/4, \alpha_2 = -\pi/4$.

(h) Schmidt rank=8.

(i) Circuit for second step to prepare input state. The parameters are random values.

FIG. 8: Quantum circuits for preparing input state with different Schmidt ranks. (a)–(h) are the first steps to prepare the input state, (i) is the second step.
III. ALGORITHM BOXES

Algorithm 1 Variational Quantum Algorithm for Schmidt Decomposition (VQASD)

Input: bi-partite $d$-dimension quantum state $\rho_{AB}$, parameterized quantum circuit (PQC) $U(\theta_1)$ and $V(\theta_2)$, inverse of state preparing circuit $W^{-1}$ (Fig. 1), number of iterations ITR;

Output: Schmidt coefficients $c_j$ and Schmidt states $u_j, v_j$.

Initialize parameters $\theta$.

# Training iteration
for $itr = 1, \ldots, ITR$

Apply $U(\theta_1)$ to party A and $V(\theta_2)$ to party B in state $\rho_{AB}$, and the state $\rho_{AB}$ becomes $\tilde{\rho}_{AB}$.

Apply the inverse of state preparing circuit $W^{-1}$ to the state $\tilde{\rho}_{AB}$.

Compute the cost function $C_{SD} = \Pr[b = 0] = \text{Tr}[W^{-1}\rho_{AB} W |0\rangle\langle 0|_{AB}]$.

Maximize the cost function $C_{SD}$ and update parameters $\theta$.

end for

# Coefficient readout
Denote the trained PQCs as $U(\theta^{opt})$.

Apply the trained PQCs $U(\theta^{opt})$ to the party A in target state $\rho_{AB}$, then achieve $\tilde{\rho}'_{AB}$.

Measure the reduced state $\tilde{\rho}'_A$ on the computational basis $\{|j\rangle_A\}$. The $j$-th Schmidt coefficient is calculated by $c_j = \sqrt{\Pr[b = j]} = \sqrt{\langle j | \tilde{\rho}'_A | j \rangle}$.

Algorithm 2 Variational Quantum Algorithm for Logarithm Negativity

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

Output: logarithm negativity of the bi-partite quantum state $E_N(\rho_{AB})$.

Initialize parameters $\theta$.

# Training iteration
for $itr = 1, \ldots, ITR$

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

Apply the inverse of state preparing circuit $W'^{-1}$ to the state $\tilde{\rho}_{AB}$.

Calculate the loss function $C_{LN} = \Pr[b = 0] = \text{Tr}[W'^{-1}\tilde{\rho}_{AB} W' |0\rangle\langle 0|_{AB}]$.

Maximize the loss function $C_{LN}$, and update parameters $\theta$.

end for

# Logarithm negativity readout
Denote the maximum loss function as $C_{LN}^{max}$, and the logarithm negativity is calculated by $E_N(\rho_{AB}) = \log_2(C_{LN}^{max}) + n$.