

Separation between entanglement criteria and entanglement detection protocols

Zhenhuan Liu^{1,*} and Fuchuan Wei^{2,*}

¹Center for Quantum Information, Institute for Interdisciplinary Information Sciences, *Tsinghua University*, Beijing 100084, China

²Yau Mathematical Sciences Center and Department of Mathematics, *Tsinghua University*, Beijing 100084, China



(Received 19 March 2024; accepted 23 July 2025; published 5 August 2025)

Entanglement detection is one of the most fundamental tasks in quantum information science, playing vital roles in theoretical studies and quantum system benchmarking. Researchers have proposed many powerful entanglement criteria with high detection capabilities and small observable numbers. Nonetheless, entanglement criteria only represent mathematical rules deciding the existence of entanglement. The relationship between a good entanglement criterion and an effective experimental entanglement detection protocol (EDP) is poorly understood. In this study, we introduce postulates for EDPs about their detection capabilities and robustness and use them to show the difference between entanglement criteria and EDPs. Specifically, we design an entanglement detection task for unknown pure bipartite states and demonstrate that the sample complexity of any EDP and the number of observables for a good entanglement criterion can have exponential separation. Furthermore, we discover that the optimal EDP with the lowest sample complexity does not necessarily correspond to an entanglement criterion constructed by a small number of observables. Our results can be used to prove the exponential speedups achieved through quantum memory and be generalized to multipartite entanglement detection. By highlighting the significance and independence of EDP design, our work holds practical implications for entanglement detection experiments.

DOI: [10.1103/hm8j-wgqm](https://doi.org/10.1103/hm8j-wgqm)

I. INTRODUCTION

The “spooky action,” entanglement, has long been recognized as a crucial resource that allows quantum physics to exhibit advantages over classical systems in tasks such as quantum communication and simulation [1]. Consequently, the task of entanglement detection holds practical and theoretical significance, leading to the proposal of numerous approaches [2]. Actually, most entanglement detection proposals are entanglement criteria (EC), which are mathematical rules that provide sufficient conditions for the presence of entanglement. Without considering the practical implementation, the main objective in designing a good EC is maximizing the detection capability, i.e., the probability of successfully detecting entanglement. However, experimentally implementing EC necessitates further efforts to devise a practical entanglement detection protocol (EDP), which relies on experimentally accessible quantities. This gives rise to two consequences: Firstly, if the primary focus in designing the EDP is practicality, the detection capability of the original EC may be significantly diminished. Secondly, if the main goal is

to maintain the detection capability, the EDP may necessitate additional resources and consequently become less practical.

The positive map criterion [2] is a typical example to show the gap between EC and EDP, which detects entanglement by telling the semipositivity of the density matrix after a positive but not completely positive map. Including the positive partial transposition (PPT) criterion [3], the positive map criterion finds widespread use in theoretical studies of entanglement due to its concise mathematical form and strong detection capability [4–6]. However, as the positive map is not a physical operation and determining the semipositivity of a given matrix is a task with expensive computational cost, the positive map criterion cannot be directly applied in real experiments. Two main approaches have been developed to experimentally test the positive map criterion, relying on two experimentally accessible quantities. The first approach is measuring entanglement witnesses derived from the corresponding positive map criterion [2], such as $W = |\psi\rangle\langle\psi|^{T_A}$ for PPT criterion, where T_A represents the partial transposition operation on subsystem A. When $|\psi\rangle$ has a simple form, this method is experimentally friendly while exponentially weakening the detection capability of the original positive map criterion [7]. The second approach entails measuring the moments of the density matrices processed by the positive map, such as $\text{Tr}[(\rho^{T_A})^3]$, and using them to infer the semipositivity of the target matrix [8–12]. This method exhibits much higher detection capability compared to the entanglement witness approach [7,11], whereas moment measurement necessitates either exponential experimental times or a significant number of memory qubits [9,10,12,13]. In conclusion, although positive map criteria are theoretically powerful, applying them in

*These authors contributed equally to this work.

[†]Contact author: qubithuan@gmail.com

[‡]Contact author: wfc21@mails.tsinghua.edu.cn

real experiments requires compromising the detection capability or dealing with high experimental costs.

Over the years, researchers gradually realized that there might exist some fundamental limitations in entanglement detection [14–17], especially when they tried to construct a powerful EC using a minimal number of linear observables [18–20]. Moreover, many efforts are paid to explore the relation between observable numbers and the detection capability of EC [21,22]. Actually, the observable number has long been considered as the cost of an EC with a simple example being entanglement witness, which is widely acknowledged to be an experimentally friendly protocol as it only requires one to measure a single observable. Remarkably, the fundamental limitation of EC was derived in Ref. [7], where the authors established a fundamental trade-off between the observable number and the detection capability of EC. However, this result cannot be directly applied to analyze the performances of EDPs as the number of observables does not directly equal the experimental cost.

With the rapid development of quantum hardware, entanglement detection gradually transforms from a theoretical problem to an essential task for quantum simulation and benchmarking [23–32]. At the same time, performances of entanglement detection protocols in practical scenarios attract more and more attention [33–35]. To provide a guide for designing entanglement detection experiments, it is an urgent problem to figure out the complexities and limitations of EDPs. Specifically, some critical questions arise: does a good EC correspond to a good EDP? Does a fundamental limitation of EDP exist? Can we use some quantum resources to reduce the sample complexity of EDP?

In this work, we provide answers to these questions. We first clarify the difference between EC and EDP and propose two natural postulates of a good EDP about its detection capability and error robustness. Then, we design an entanglement detection task that can be solved using an EC constructed with a single observable, while any EDP satisfying our postulates requires exponential repetition times. This example demonstrates that the observable number of EC and sample complexity of EDP can have exponential separation. Furthermore, in this task, we find that the optimal EDP with the lowest sample complexity does not correspond to the optimal EC with minimal observables, further showing that designing EDP and EC are two independent tasks. In addition, our results help to show that quantum memory can bring exponential speedups for EDP. We hope our work can remind entanglement detection experimenters that the practical performance of the corresponding EDP is a nontrivial problem that should be considered independently from EC when designing experiments.

II. EC AND EDP

Entangled states are those that cannot be written in a separable form

$$\rho = \sum_i p_i \rho_A^i \otimes \rho_B^i, \quad (1)$$

where $p_i \geq 0$ represents probability and A and B represent two subsystems. Entanglement criteria are mathematical rules that distinguish some entangled states from separable ones [2].

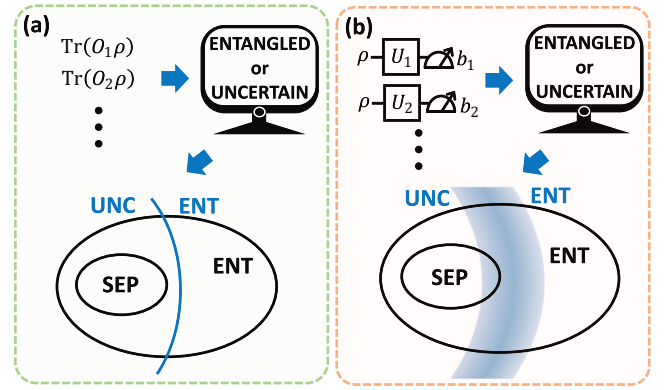


FIG. 1. Difference between EC and EDP. ENT and SEP represent sets of entangled and separable states, respectively. Here, we only consider EC and EDPs constructed by single-copy observable expectation values and quantum experiments. (a) EC represents mathematical rules to decide whether a state is entangled and never makes mistakes. The blue line represents those states with $C(\rho) = 0$, the states at the left of the line satisfy $C(\rho) > 0$, and vice versa. (b) EDP represents rules of physical operations and measurements to infer entanglement with finite rounds of experiments. The blue line for $C(\rho)$ becomes a blue area for $\hat{C}(\rho)$ whose darkness is proportional to the probability for a state to satisfy $\hat{C}(\rho) = 0$. In principle, any state has a nonzero probability of being classified as entangled or uncertain (UNC) by an EDP.

Mathematically speaking, an EC can be represented as a function of the target state $C(\rho)$, which satisfies that $C(\rho) \geq 0$ for all the separable ρ . Thus, $C(\rho) < 0$ is a sufficient condition for entanglement. For example, $C(\rho) = \text{Tr}(W\rho)$ represents the entanglement witness criterion and $C(\rho) = \lambda_{\min}(\rho^{T_A})$ represents the PPT criterion, where W is a witness operator and $\lambda_{\min}$ represents the minimal eigenvalue. As shown in Fig. 1(a), given a target state, a valid EC tells you if the state is entangled when $C(\rho) < 0$, or if it is uncertain when $C(\rho) \geq 0$. As EC are just theoretical rules, we can ask EC to never make mistakes by classifying separable states into entangled ones. Thus, to design a powerful EC, the most important indicator is the ratio of entangled states that can be detected by it, i.e., the detection capability.

In addition, since EC is always the starting point for researchers designing practical entanglement detection protocols, we also need to consider the physical implementation when designing EC. For example, there always exists an optimal yet trivial EC which is defined as $C(\rho) = -1$ when ρ is entangled and $C(\rho) = +1$ when ρ is separable. In principle, one can use infinite computational resources to decide the entanglement using Eq. (1); the function $C(\cdot)$ can be computed and the EC is valid. However, this form of EC does not directly relate to any specific entanglement detection protocol due to its abstract definition. Therefore, we usually consider the EC with the form of $C(\rho) = f[\text{Tr}(O_1\rho), \dots, \text{Tr}(O_M\rho)]$, where $\text{Tr}(O_i\rho)$ represents the expectation value of observable O_i [7]. This kind of EC can be easily transformed into an EDP as $\text{Tr}(O_i\rho)$ can be readily estimated using single-copy measurements. Actually, when $M = 1$, such kind of EC is reduced to the entanglement witness protocols. Therefore, researchers

often use the number of observables, M , to quantify the cost of an entanglement criterion [18–22].

EDP represents rules about operating and measuring the target quantum states and using the measurement results to infer whether the target state is entangled. EDP also decides the entanglement of the target state through a function, $\hat{C}(\rho)$, whose value is calculated using the experiment data. Due to the intrinsic randomness in quantum measurements and the finite experiment times, the value of $\hat{C}(\rho)$ has uncertainty in different rounds of experiments, as shown in Fig. 1(b). Thus, there always exists the possibility for EDP to classify a separable state into an entangled one and make mistakes, which is the most important difference between EC and EDP. We can thus propose two postulates for a good EDP about its detection capability and robustness to errors.

(1) Completeness: the probability of outputting the result of entanglement when the target state is entangled should be large enough, i.e.,

$$\Pr[\hat{C}(\rho) < 0 | \rho \in \text{ENT}] = \Theta(1), \quad (2)$$

where $\Theta(1)$ is a complexity notation meaning that the probability approaches some constant when the qubit number of ρ approaches infinity. We present the formal definitions of complexity notations in Appendix A.

(2) Soundness: the probability of actually being entangled when outputting the result of entanglement should be large enough, i.e.,

$$\Pr[\rho \in \text{ENT} | \hat{C}(\rho) < 0] = p_e + \Theta(1), \quad (3)$$

where p_e is the probability for the random state ρ being an entangled state.

As stated, EC does not make mistakes, which implies $\Pr[\rho \in \text{ENT} | C(\rho) < 0] = 1$. Therefore, completeness is the only postulate for a good EC. Since an EDP is commonly constructed based on some EC, the postulate of completeness requires that an EDP should be based on a good EC with high detection capability. The term of $p_e + \Theta(1)$ in the right-hand side of Eq. (3) ensures that the EDP is remarkably better than the random guess. Thus, the postulate of soundness places a requirement on the experiment sample complexity to suppress the error. Note that all the properties of EC and EDPs are tightly related to probabilities. Therefore, one must first specify the underlying quantum state distribution before evaluating the performance of EC and EDPs.

III. EXPONENTIAL SEPARATION

In this section, we will use a specific entanglement detection task, i.e., a specific state distribution, to show the gap between the observable number of a good EC and the sample complexity of a good EDP. In practice, the state distribution is determined by experiment setups and can be very concentrated around some specific state. For example, the real state distribution might be an ideal state disturbed by a well-calibrated noise. Meanwhile, assessing EDPs without any prior information about the state remains essential, as it reveals the methods' intrinsic, distribution-independent detection power [17]. Intuitively, detecting the entanglement of pure states is an easy task, so we start from the pure state distribution:

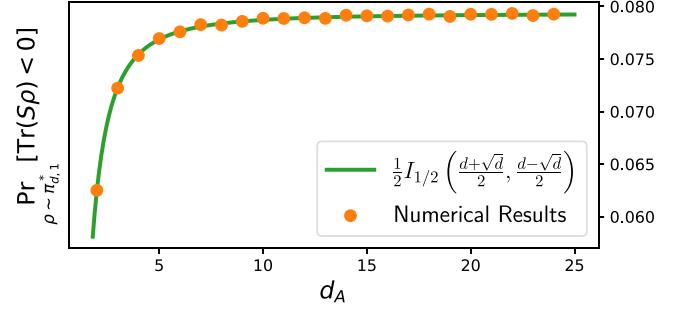


FIG. 2. The probability for detecting the entanglement in $\pi_{d,1}^*$ using the entanglement witness $S = |\Phi^+\rangle\langle\Phi^+|^{T_A}$. The x axis represents the dimension of the subsystem A , which is the square root of the dimension of the whole system. For each orange point, we sample 10^7 ρ 's from $\pi_{d,1}^*$ and calculate the ratio of these states satisfying $\text{Tr}(S\rho) < 0$.

Definition 1. A bipartite state $\rho \in D(\mathcal{H}_d)$ is sampled according to the distribution of $\pi_{d,1}^*$ if and only if with probability 0.5, it is a d -dimensional Haar random pure state; with probability 0.5, it is the tensor product of two independent $\sqrt{d}$ -dimensional Haar random pure states, $\rho = \rho_A \otimes \rho_B$. A Haar random pure state is generated by acting a Haar random unitary on a fixed pure state.

It is easy to see that with probability 0.5, a state sampled according to $\pi_{d,1}^*$ is entangled.

As the only requirement for EC is completeness, we only care about the number of observables needed for detecting the entanglement of $\pi_{d,1}^*$ with a constant probability. Following our intuition that pure state entanglement detection is easy for EC, we find that a single observable is enough to satisfy the completeness postulate. We use the SWAP operator S defined between subsystems A and B , which satisfies $S|\psi_A, \psi_B\rangle = |\psi_B, \psi_A\rangle$ for arbitrary $|\psi_A\rangle$ and $|\psi_B\rangle$, to show this. The SWAP operator is actually an entanglement witness as $S = |\Phi^+\rangle\langle\Phi^+|^{T_A}$ with $|\Phi^+\rangle = \sum_{i=1}^{\sqrt{d}} |ii\rangle$ being the unnormalized maximally entangled state.

Theorem 1. Considering that ρ is sampled according to $\pi_{d,1}^*$, the detection power of S satisfies

$$\Pr_{\rho \sim \pi_{d,1}^*} [\text{Tr}(S\rho) < 0 | \rho \in \text{ENT}] = I_{1/2}\left(\frac{d + \sqrt{d}}{2}, \frac{d - \sqrt{d}}{2}\right), \quad (4)$$

where $I(\cdot, \cdot)$ is the regularized incomplete beta function.

We leave the detailed proof in Appendix B and conjecture that the detection power of S is asymptotically a constant, $\lim_{d \rightarrow \infty} I_{1/2}\left(\frac{d + \sqrt{d}}{2}, \frac{d - \sqrt{d}}{2}\right) = \text{const}$. We numerically test this conjecture using Fig. 2, in which the detection power converges to some constant when the dimension increases.

However, considering the physical implementation of this EC, i.e., the corresponding EDP based on single-copy measurement of $\text{Tr}(S\rho)$, the sample complexity is high due to the additional requirement of soundness. The first reason is that if we randomly select a state according to $\pi_{d,1}^*$, the expectation value is normally exponentially small, around $\frac{1}{\sqrt{d}}$. Thus, one should keep the measurement error exponentially small to estimate a negative value with high certainty. Secondly,

considering practical situations, when measuring S , in every experiment time, one randomly gets a result $+1$ or -1 , which are eigenvalues of S . Thus, the variance of a single experiment result is a constant. According to Hoeffding's inequality, to hold the soundness requirement for EDP, the complexity is around $\mathcal{O}(d)$, which is exponential with the qubit number as $d = 2^n$.

From this example, we find that if we directly design an EDP based on a good EC constructed using a small number of observables, the sample complexity might be exponentially high. One may further ask if a better EDP for this entanglement detection task exists or if a fundamental limitation exists. In the following, we will show that a better EDP does exist, and at the same time, an exponential sample complexity lower bound also exists for any EDP, although this is an easy task for EC in terms of the observable number.

Theorem 2. Given multiple copies of a state ρ that is sampled from $\pi_{d,1}^*$, if only single-copy quantum operations are allowed, the minimum number of copies required by an adaptive EDP with completeness, $\Pr_{\rho \sim \pi_{d,1}^*}[\hat{C}(\rho) < 0 | \rho \in \text{ENT}] \geq \frac{1}{2}$, and soundness, $\Pr_{\rho \sim \pi_{d,1}^*}[\rho \in \text{ENT} | \hat{C}(\rho) < 0] \geq \frac{6}{7}$, scales as $\Theta(d^{1/4})$.

We emphasize here that the constants $\frac{1}{2}$ and $\frac{6}{7}$ can be changed into any other constants satisfying postulates of completeness and soundness, which does not affect the validation of the theorem. Moreover, when the subsystem dimensions differ ($d_A \neq d_B$), the sample complexity bound changes from $\Theta(d^{1/4})$ to $\Theta(\min\{d_A, d_B\}^{1/2})$. We leave the detailed proof in Appendix C and sketch core ideas here. To prove the lower bound, we first notice that if the EDP described in this theorem exists, it can be used to solve a state discrimination task. Specifically, given multiple copies of a state ρ , the state discrimination task is to decide whether it is sampled from the set s_1 containing d -dimensional random pure states ρ or s_2 containing product states $\rho_A \otimes \rho_B$ with ρ_A and ρ_B being two independent $\sqrt{d}$ -dimensional random pure states, with equal probabilities. With the EDP given in Theorem 2, one can classify the state to s_1 when the EDP outputs $\hat{C}(\rho) < 0$ and to s_2 otherwise. With the conditions of completeness and soundness, the success probability of this strategy reaches $\frac{2}{3}$. Therefore, if this state discrimination task has a sample complexity lower bound, it is also a lower bound of the EDP.

To find the lower bound for this state discrimination task, we employ a decision tree to describe all the methods, including those depending on adaptive strategies [36,37], as shown in Fig. 3. Every node of the tree has all the experiment data of its child nodes and uses it to decide how to experiment within this step. After many steps, different input states will lead to different probability distributions on the leaves. If the deviation between two probability distributions on the leaves is large enough, we can discriminate two state sets with high probability. We prove that, when only single-copy quantum operations are allowed in this decision tree, the depth of this tree should be at least $\Omega(d^{1/4})$ to make the distance of probability distributions caused by s_1 and s_2 a constant, which is also the sample complexity lower bound for single-copy EDP.

To prove the upper bound, we consider an EDP that satisfies the requirements of Theorem 2 and derive its sample

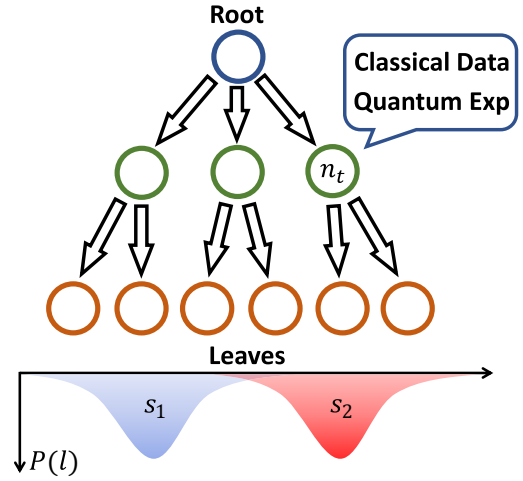


FIG. 3. Every state discrimination protocol can be represented as a decision tree, where n and l stand for the node and leaf, respectively. $P(l)$ stands for the probability distribution on leaves, which is determined by the protocol and input state. The t th layer of the decision tree represents the t th step of the protocol. In our setting, the experimenter has a classical memory that records the experiment data of the former steps and a quantum processor that can perform operations and measurements on quantum states. The protocol has a rule that decides how to perform quantum experiments based on the data recorded in the classical memory. At the end of the experiment, the experimenter will decide which set the target state comes from, depending on the leaf the experimenter reaches.

complexity. This EDP is based on the purity criterion [2], which detects entanglement by measuring $C(\rho) = \text{Tr}(\rho_A^2) - \text{Tr}(\rho^2)$. As states in $\pi_{d,1}^*$ are all pure states, the condition of $\text{Tr}(\rho_A^2) < 1$ can help us to detect the entanglement of ρ . According to the properties of random pure states, if a state is sampled according to $\pi_{d,1}^*$, $\text{Tr}(\rho_A^2)$ is either 1 or exponentially small [38], with probability 0.5 for each scenario. Thus, any protocol that accurately estimates $\text{Tr}(\rho_A^2)$ at least satisfies the requirement for completeness, Eq. (2). Then, we need to find a purity measurement protocol that satisfies soundness with the lowest sample complexity. Due to the restriction of single-copy quantum operations, we employ the randomized measurement protocol [39,40], which is commonly employed for entanglement detection [41,42]. In this protocol, one evolves the subsystem A using many random unitaries, measures it in the computational basis, and computes the purity from these experiment data. In Appendix E, we analytically show that $\mathcal{O}(d^{1/4})$ measurement times is enough to suppress the variance of the randomized measurement protocol to satisfy the soundness requirement, Eq. (3), which concludes our proof for the upper bound.

There are some important consequences from Theorem 2 and its proof. Firstly, the optimal EDP with the lowest sample complexity does not necessarily correspond to a good EC with the smallest number of observables. Here, a good EC refers to those satisfying the postulate of completeness, and the optimal EDP refers to the EDP that satisfies two postulates of completeness and soundness with the lowest sample complexity. As stated before, the optimal single-copy EDP is based on the randomized measurement protocol, which requires

measurements on many bases. If we judge entanglement detection protocols only regarding the number of observables, this randomized measurement EDP is much worse than the entanglement witness EDP we mentioned before, which only requires a single observable. Secondly, a counterintuitive fact from Theorem 2 is that entanglement detection for pure states can be exponentially hard when considering sample complexity. However, such a task is well acknowledged to be easy for EC and theoretical analysis. Note that if considering practical situations with enough prior knowledge of the target state, we can largely avoid the exponential sample complexity.

In addition, Ref. [7] has shown that quantum memory, or equivalently joint operations, brings advantages in constructing EC by exponentially saving the number of observables. Using conclusions in this work, we can further conclude that quantum memory also brings exponential speedups for practical entanglement detection tasks. This is because the SWAP test algorithm, which relies on quantum memory, estimates purity with constant sample complexity. We thus arrive at the following conclusion, which is also proved in Ref. [43] with another quantum algorithm.

Corollary 1. Quantum memory can bring exponential speedups in sample complexities for entanglement detection protocols.

IV. HARDNESS FOR DETECTING MIXED STATES

Until now, we primarily focus on pure bipartite states. Our results, especially the hardness of entanglement detection, can be extended to mixed-state scenarios, as discussed in Appendix D. These results help to understand the influence of unpredictable noises.

We first need to generalize the definition of pure random state $\pi_{d,1}^*$. To consider mixed states, we assume the target state is interacting with a k -dimensional environment. Thus, the global random pure state in the definition of $\pi_{d,1}^*$ is now replaced by a reduced density matrix of a $d \times k$ -dimensional random pure state; and the tensor product of two $\sqrt{d}$ -dimensional pure states is now replaced by the tensor product of two reduced density matrices of $\sqrt{d} \times \sqrt{k}$ -dimensional random pure states. The formal definition is shown below.

Definition 2. A bipartite state $\rho \in D(\mathcal{H}_d)$ is said to be sampled according to the distribution $\pi_{d,k}^*$ if and only if with probability 0.5, it is the reduced density matrix of a $d \times k$ -dimensional random pure state $|\psi\rangle\langle\psi| \in D(\mathcal{H}_{d \times k})$; with probability 0.5, it is the tensor product of two independent random states $\rho = \rho_A \otimes \rho_B$, each of which is a $\sqrt{d}$ -dimensional reduced density matrix of a $\sqrt{d} \times \sqrt{k}$ -dimensional random pure state.

A natural question concerning the distribution of $\pi_{d,k}^*$ is that when k is large, a state sampled according to $\pi_{d,k}^*$ will be separable with a probability close to 1, which would make the discussion of entanglement detection meaningless. According to Ref. [16], when $k = \mathcal{O}(d^{1.5})$, a state sampled according to $\pi_{d,k}^*$ is entangled with probability approximately 0.5, similar with $\pi_{d,1}^*$. Thus, entanglement is actually a robust resource and it is reasonable to analyze the detection capability of EC and EDPs in a wide range of k . Based on the new state distribution, we make the following conclusions.

Theorem 3. Given multiple copies of a state ρ that is sampled from $\pi_{d,k}^*$ with $k = \mathcal{O}(d^{1.5})$, if only single-copy quantum operations are allowed, the minimum number of copies required by an adaptive EDP with completeness, $\Pr_{\rho \sim \pi_{d,k}^*}[\hat{C}(\rho) < 0 | \rho \in \text{ENT}] \geq \frac{1}{2}$, and soundness, $\Pr_{\rho \sim \pi_{d,k}^*}[\rho \in \text{ENT} | \hat{C}(\rho) < 0] \geq \frac{6}{7}$, scales as $\Theta[(dk)^{1/4}]$.

We refer to Appendix D for the detailed proof. From this theorem, it is easy to see the influence caused by the interaction with the environment, as the sample complexity increases from $\Theta(d^{1/4})$ to $\Theta[(dk)^{1/4}]$. Intuitively, the higher sample complexity is caused by the fact that information of the target system is leaked into the environment, which causes the hardness in distinguishing states. In Appendix D, we also discuss how to consider the multipartite entanglement detection and show a reduced sample complexity compared with the bipartite entanglement detection.

V. DISCUSSION

The target for this work is to show that, although the number of observables is widely adopted as the cost of EC [7,18–22], it would result in a resource waste in practical entanglement detection experiments. Specifically, if one designs the practical EDP based on the EC with minimal number of observables, one may not obtain the optimal EDP with the lowest sample complexity. To summarize, some take-home messages of this work are the following. Firstly, in addition to the detection power, the only requirement for an entanglement criterion, an effective EDP must also possess robustness against errors. Secondly, certain entanglement detection tasks exist that are relatively easy for an entanglement criterion, achieving constant detection probability with a small number of observables while proving to be exponentially challenging for any EDP in terms of sample complexity. Thirdly, although an EDP is typically designed based on an entanglement criterion, the optimal EDP with the lowest sample complexity may not necessarily correspond to an entanglement criterion with a small number of observables, as their evaluation criteria differ. Lastly, using quantum memory can significantly reduce the sample complexity of the EDP, demonstrating the great potential of multicopy entanglement detection protocols [44].

Our analysis framework is potentially generalizable to other property estimation tasks. Note that a central object of our framework is the pair of functions $C(\rho)$ and $\hat{C}(\rho)$, used for EC and EDP, respectively. By replacing these two functions with indicators of other properties and adjusting the state distribution, we can analyze the sample complexity of estimating those properties as well. For example, if we set both functions to some entanglement quantifier, such as entanglement negativity [4], we may obtain sample complexity bounds for entanglement quantification.

While our results are derived based on a specific state distribution, they represent an initial step towards systematically considering sample complexities of entanglement detection protocols. Because $\pi_{d,k}^*$ is built from Haar random pure states, the corresponding sample complexity bound has two key interpretations. First, it captures the resources required when the target state is completely unknown and is therefore modeled as being drawn uniformly at random from the entire state space. Second, it serves as a worst-case benchmark: if

the actual state distribution deviates from the Haar prior and concentrates around specific states, this extra structure can be leveraged to design more efficient EDPs, thereby reducing the number of copies required. We hope that our work will inspire further exploration using various practical state distributions, ultimately reducing the cost of real-world entanglement detection experiments.

Note that the optimal EDP explored in this work for detecting states sampled from $\pi_{d,1}^*$ and $\pi_{d,k}^*$ is the combination of randomized measurements and purity criterion, which is an inadaptive single-copy protocol. So, beyond protocols based on quantum memory, it would be valuable to consider other types of protocols and their performance in entanglement detection, such as separable, local, and adaptive protocols [45–47].

ACKNOWLEDGMENTS

The authors thank Satoya Imai, Xiongfeng Ma, Otfried Gühne, Senrui Chen, Pei Zeng, Rui Zhang, Zihao Li, Pengyu Liu, Boyang Chen, Xueying Yang, and Rundi Lu for their insightful discussions and suggestions. Z.L. acknowledges support from the National Natural Science Foundation of China under Grant No. 12174216 and the Innovation Program for Quantum Science and Technology under Grant No. 2021ZD0300804. F.W. is supported by BMSTC and ACZSP under Grant No. Z221100002722017.

DATA AVAILABILITY

All data that support the findings of this study are included within the article.

APPENDIX A: NOTATIONS AND AUXILIARY LEMMAS

We use $\mathcal{H}_d$ to represent the d -dimensional Hilbert space. $D(\mathcal{H}_d)$ represents the set of density matrices on $\mathcal{H}_d$. For convenience, sometimes we omit the Dirac notation and denote $|\psi\rangle\langle\psi|$ as ψ . We will frequently use some complexity notations such as Θ , Ω , and $\mathcal{O}$ in this Appendix. $f(x) = \mathcal{O}[g(x)]$ means that in the limit of $x \rightarrow \infty$, the value of $f(x)$ can be upper bounded by $cg(x)$ where c is some constant. $f(x) = \Omega[g(x)]$ means that in the limit of $x \rightarrow \infty$, the value of $f(x)$ can be lower bounded by $cg(x)$. $f(x) = \Theta[g(x)]$ means that in the limit of $x \rightarrow \infty$, the value of $f(x)$ can be lower bounded by $c_1g(x)$ and upper bounded by $c_2g(x)$.

We first provide the definition of state distribution $\pi_{d,k}$, which will be frequently used.

Definition A1. A state $\rho \in D(\mathcal{H}_d)$ is said to be sampled according to the distribution $\pi_{d,k}$ if and only if it is the reduced density matrix of a $d \times k$ -dimensional random pure state $|\psi\rangle\langle\psi| \in D(\mathcal{H}_{d \times k})$. Here, by random, we mean that $|\psi\rangle = U|0\rangle$ where U is a Haar-measure random unitary matrix and $|0\rangle$ is a fixed pure state.

Let S_T be the symmetric group of order T . For distinct $a_1, \dots, a_l$, we use the notation $(a_1 \cdots a_l)$ to denote a cyclic permutation in S_T , which acts as $a_1 \mapsto a_2, \dots, a_{l-1} \mapsto a_l, a_l \mapsto a_1$. For $\sigma \in S_T$, define a unitary operator W_σ acting on $\mathcal{H}_d^{\otimes T}$ by

$$W_\sigma = \sum_{i_1, \dots, i_T=1}^d |i_{\sigma^{-1}(1)} \cdots i_{\sigma^{-1}(T)}\rangle \langle i_1 \cdots i_T|, \quad (\text{A1})$$

where $\{|i\rangle\}_{i=1}^d$ is an orthonormal basis of $\mathcal{H}_d$.

For two discrete probability distributions $p = \{p_1, \dots, p_N\}$ and $q = \{q_1, \dots, q_N\}$, define the total variation between them by $\text{TV}(p, q) = \frac{1}{2} \sum_{i=1}^N |p_i - q_i|$.

For two systems $\mathcal{H}_1$ and $\mathcal{H}_2$, define the unraveling operation [48] $\mathcal{V}_M : \mathcal{H}_1^{\otimes M} \otimes \mathcal{H}_2^{\otimes M} \rightarrow (\mathcal{H}_1 \otimes \mathcal{H}_2)^{\otimes M}$ by

$$\begin{aligned} \mathcal{V}_M(|x_1\rangle \otimes \cdots \otimes |x_M\rangle) \otimes (|y_1\rangle \otimes \cdots \otimes |y_M\rangle) \\ = (|x_1\rangle \otimes |y_1\rangle) \otimes \cdots \otimes (|x_M\rangle \otimes |y_M\rangle), \end{aligned} \quad (\text{A2})$$

which holds for all $|x_i\rangle \in \mathcal{H}_1$ and $|y_i\rangle \in \mathcal{H}_2$. More generally, for K systems $\mathcal{H}_1, \dots, \mathcal{H}_K$, define $\mathcal{V}_M^{(K)} : \mathcal{H}_1^{\otimes M} \otimes \cdots \otimes \mathcal{H}_K^{\otimes M} \rightarrow (\mathcal{H}_1 \otimes \cdots \otimes \mathcal{H}_K)^{\otimes M}$ by

$$\begin{aligned} \mathcal{V}_M^{(K)}(|x_1^{(1)}\rangle \otimes \cdots \otimes |x_M^{(1)}\rangle) \otimes \cdots \\ \otimes (|x_1^{(K)}\rangle \otimes \cdots \otimes |x_M^{(K)}\rangle) \\ = (|x_1^{(1)}\rangle \otimes \cdots \otimes |x_1^{(K)}\rangle) \otimes \cdots \\ \otimes (|x_M^{(1)}\rangle \otimes \cdots \otimes |x_M^{(K)}\rangle), \end{aligned} \quad (\text{A3})$$

which holds for all $|x_i^{(k)}\rangle \in \mathcal{H}_k$, where $k = 1, \dots, K$ and $i = 1, \dots, M$.

Lemma A1 ([49]). For Haar random pure state $|\psi\rangle \in \mathcal{H}_d$, we have

$$\mathbb{E}_{\psi \sim \text{Haar}} |\psi\rangle\langle\psi|^{\otimes T} = \frac{1}{(d+T-1) \cdots (d+1)d} \sum_{\sigma \in S_T} W_\sigma. \quad (\text{A4})$$

APPENDIX B: DETECTION POWER AND PERFORMANCE OF SWAP OPERATOR

The entanglement witness S is an important example in the main context. In this section, we will first evaluate its detection power as an entanglement criterion and then estimate its sample complexity as an EDP.

Theorem B1. Consider that ρ is sampled according to $\pi_{d,1}^*$; the detection power of S satisfies

$$\Pr_{\rho \sim \pi_{d,1}^*} [\text{Tr}(S\rho) < 0 | \rho \in \text{ENT}] = I_{1/2} \left(\frac{d + \sqrt{d}}{2}, \frac{d - \sqrt{d}}{2} \right), \quad (\text{B1})$$

where $I(\cdot, \cdot)$ is the regularized incomplete beta function.

Proof. We take an orthonormal basis $\{|\alpha_1\rangle, \dots, |\alpha_{(d+\sqrt{d})/2}\rangle\}$ for the $+1$ eigenspace of S , and an orthonormal basis $\{|\alpha_{((d+\sqrt{d})/2)+1}\rangle, \dots, |\alpha_d\rangle\}$ for the -1 eigenspace of S . For a random complex Gaussian vector $v \sim \mathcal{N}_{\mathbb{C}}(0, \mathbb{I}_d)$, take $|\psi\rangle = \sum_{x=1}^d \frac{v_x}{\|v\|} |\alpha_x\rangle \in \mathcal{H}_d$, then $\psi \sim \pi_{d,1}$ [17]. Note that $\text{Tr}(S\psi) = \sum_{x=1}^{(d+\sqrt{d})/2} \frac{|v_x|^2}{\|v\|^2} - \sum_{x=((d+\sqrt{d})/2)+1}^d \frac{|v_x|^2}{\|v\|^2}$. We have

$$\begin{aligned} \Pr_{\rho \sim \pi_{d,1}^*} [\text{Tr}(S\rho) < 0 | \rho \in \text{ENT}] \\ = \Pr_{\psi \sim \pi_{d,1}} [\text{Tr}(S\psi) < 0] \\ = \Pr_{v \sim \mathcal{N}_{\mathbb{C}}(0, \mathbb{I}_d)} \left[\frac{\sum_{x=1}^{(d+\sqrt{d})/2} |v_x|^2}{\sum_{x=((d+\sqrt{d})/2)+1}^d |v_x|^2} < 1 \right]. \end{aligned} \quad (\text{B2})$$

For $v \sim \mathcal{N}_{\mathbb{C}}(0, \mathbb{I}_d)$, we have $\sum_{x=1}^{(d+\sqrt{d})/2} |v_x|^2 \sim \Gamma(\frac{d+\sqrt{d}}{2}, 1)$ and $\sum_{x=((d+\sqrt{d})/2)+1}^d |v_x|^2 \sim \Gamma(\frac{d-\sqrt{d}}{2}, 1)$, thus $\sum_{x=1}^{(d+\sqrt{d})/2} |v_x|^2 / \sum_{x=((d+\sqrt{d})/2)+1}^d |v_x|^2 \sim \beta'(\frac{d+\sqrt{d}}{2}, \frac{d-\sqrt{d}}{2})$, which is the standard beta prime distribution with parameters $\frac{d+\sqrt{d}}{2}$ and $\frac{d-\sqrt{d}}{2}$. Therefore the random variable $\sum_{x=1}^{(d+\sqrt{d})/2} |v_x|^2 / \sum_{x=((d+\sqrt{d})/2)+1}^d |v_x|^2$ has cumulative distribution function

$$\Pr_{v \sim \mathcal{N}_{\mathbb{C}}(0, \mathbb{I}_d)} \left[\frac{\sum_{x=1}^{(d+\sqrt{d})/2} |v_x|^2}{\sum_{x=((d+\sqrt{d})/2)+1}^d |v_x|^2} < y \right] = I_{y/1+y} \left(\frac{d+\sqrt{d}}{2}, \frac{d-\sqrt{d}}{2} \right), \quad (\text{B3})$$

where $I(\cdot, \cdot)$ is the regularized incomplete beta function. Taking $y = 1$, we have

$$\Pr_{\rho \sim \pi_{d,1}^*} [\text{Tr}(S\rho) < 0 | \rho \in \text{ENT}] = I_{1/2} \left(\frac{d+\sqrt{d}}{2}, \frac{d-\sqrt{d}}{2} \right). \quad (\text{B4})$$

Now we briefly analyze the complexity of the EDP designed based on S . Firstly, if we randomly select states from $\pi_{d,1}^*$, the average expectation value is exponentially small as

$$\begin{aligned} \mathbb{E}_{\psi_{AB} \sim \pi_{d,1}^*} \text{Tr}(S\psi_{AB}) &= \mathbb{E}_{\psi_A, \psi_B \sim \pi_{\sqrt{d},1}^*} \text{Tr}(S(\psi_A \otimes \psi_B)) = \text{Tr} \left(S \frac{\mathbb{I}_d}{d} \right) \\ &= \frac{1}{\sqrt{d}}, \end{aligned} \quad (\text{B5})$$

where $\frac{\mathbb{I}_d}{d}$ is the maximally mixed state in $\mathcal{H}_d$. Then, one can prove that if the state ρ is sampled according to the distribution $\pi_{d,1}^*$, the variance of the expectation value is also exponentially small,

$$\text{Var}_{\psi_{AB} \sim \pi_{d,1}^*} [\text{Tr}(S\psi_{AB})] = \mathbb{E}_{\psi_{AB} \sim \pi_{d,1}^*} \text{Tr}(S\psi_{AB})^2 - \frac{1}{d}. \quad (\text{B6})$$

Substituting the result in Lemma A1, we arrive at

$$\text{Var}_{\psi_{AB} \sim \pi_{d,1}^*} [\text{Tr}(S\psi_{AB})] = \frac{2d}{d(d+1)} - \frac{1}{d} \sim \frac{1}{d}. \quad (\text{B7})$$

This means that, when we sample a state ψ_{AB} with negative expectation value $\text{Tr}(S\psi_{AB}) < 0$, the absolute value of $\text{Tr}(S\psi_{AB})$ normally scales as $\frac{1}{\sqrt{d}}$. One can prove that with a constant probability, the value of $\text{Tr}(S\psi_{AB})$ is less than $-\frac{1}{\sqrt{d}}$. Therefore, to satisfy the requirement of soundness, the EDP should have an estimation error approximately $\frac{1}{\sqrt{d}}$. Combining with the property of the SWAP operator mentioned in the main context, this requires the sample complexity of EDP to be approximately $\mathcal{O}(d)$.

APPENDIX C: SAMPLE COMPLEXITY FOR PURE BIPARTITE STATE EDP

Theorem C1. Given multiple copies of a state ρ that is sampled from $\pi_{d,1}^*$, the minimum number of copies required by a single-copy adaptive EDP with com-

pleteness, $\Pr_{\rho \sim \pi_{d,1}^*} [\hat{C}(\rho) < 0 | \rho \in \text{ENT}] \geq \frac{1}{2}$, and soundness, $\Pr_{\rho \sim \pi_{d,1}^*} [\rho \in \text{ENT} | \hat{C}(\rho) < 0] \geq \frac{6}{7}$, scales as $\Theta(d^{1/4})$.

Proof. To prove the lower bound on the sample complexity of EDP, we first find an equivalent state discrimination task and derive the lower bound for that task. Notice that given an EDP satisfying the requirement in Theorem C1, it satisfies

$$\begin{aligned} \Pr_{\rho \sim \pi_{d,1}^*} [\rho \in \text{ENT}, \hat{C}(\rho) < 0] &= \Pr_{\rho \sim \pi_{d,1}^*} [\hat{C}(\rho) < 0 | \rho \in \text{ENT}] \\ &\quad \times \Pr_{\rho \sim \pi_{d,1}^*} [\rho \in \text{ENT}] \geq \frac{1}{4}. \end{aligned} \quad (\text{C1})$$

By $\Pr_{\rho \sim \pi_{d,1}^*} [\rho \in \text{ENT} | \hat{C}(\rho) < 0] \geq \frac{6}{7}$ we know that

$$\begin{aligned} \Pr_{\rho \sim \pi_{d,1}^*} [\rho \in \text{SEP}, \hat{C}(\rho) < 0] &\leq \frac{1}{6} \Pr_{\rho \sim \pi_{d,1}^*} [\rho \in \text{ENT}, \hat{C}(\rho) < 0] \\ &\leq \frac{1}{12}, \end{aligned} \quad (\text{C2})$$

thus

$$\begin{aligned} \Pr_{\rho \sim \pi_{d,1}^*} [\rho \in \text{SEP}, \hat{C}(\rho) \geq 0] &= \Pr_{\rho \sim \pi_{d,1}^*} [\rho \in \text{SEP}] \\ &\quad - \Pr_{\rho \sim \pi_{d,1}^*} [\rho \in \text{SEP}, \hat{C}(\rho) < 0] \\ &\geq \frac{5}{12}. \end{aligned} \quad (\text{C3})$$

Consider the following two pure state sets with distributions defined as follows:

- (1) s_1 : $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle \in \mathcal{H}_d$ is a Haar random pure state.
- (2) s_2 : $\rho = |\psi_A\rangle\langle\psi_A| \otimes |\psi_B\rangle\langle\psi_B|$, where $|\psi_A\rangle, |\psi_B\rangle \in \mathcal{H}_{\sqrt{d}}$ are Haar random pure states.

Suppose a referee tosses a fair coin to select a set, and then chooses a state ρ from the selected set according to the distribution shown above. Then the referee gives T copies of ρ to us. Our task is to decide which set ρ is from, with T copies of ρ . Notice that state sampled from s_1 is entangled with probability 1 and the state sampled from s_2 is entangled with probability 0. Thus, if we can use these T copies of ρ to achieve the EDP with requirements in Theorem C1, we can solve this state discrimination task with a success

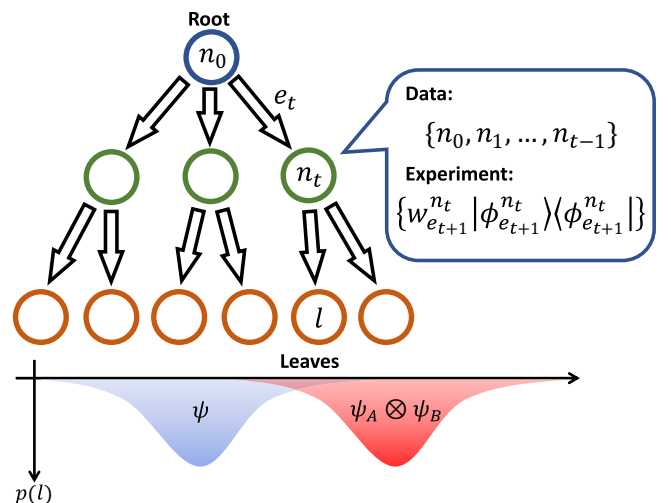


FIG. 4. Decision tree for solving the state discrimination task.

probability of

$$\begin{aligned} & \Pr_{\rho \sim \pi_{d,1}^*} [\rho \in s_1, \hat{C}(\rho) < 0] + \Pr_{\rho \sim \pi_{d,1}^*} [\rho \in s_2, \hat{C}(\rho) \geq 0] \\ & \geq \frac{1}{4} + \frac{5}{12} = \frac{2}{3}. \end{aligned} \quad (C4)$$

So, it suffices to prove that the sample complexity for solving this state discrimination task has a lower bound $\Omega(d^{1/4})$ for all single-copy adaptive strategies.

For any single-copy adaptive state discrimination protocol and a given state ρ , consider the decision tree associated with them (Fig. 4). We use l to label the leaves of the tree. Let p_ρ be a probability distribution on the leaves, such that $p_\rho(l)$ equals the probability of arriving at the leaf l . As this decision tree covers all the single-copy adaptive protocols to solve the state discrimination tasks, and the success probability depends on the total variation between two probability distributions on leaves, $p_\psi(l)$ and $p_{\psi_A \otimes \psi_B}(l)$, to find the lower bound of this task, we need to derive the lowest depth that allows the total variation between leaves probability distributions a constant,

$$\text{TV} \left(\mathbb{E}_{\psi \sim \text{Haar}} p_\psi, \mathbb{E}_{\psi_A, \psi_B} p_{\psi_A \otimes \psi_B} \right) \geq \frac{1}{3}. \quad (C5)$$

To benefit our derivation, we will first calculate the total variations between these two state distributions with the distribution caused by the maximally mixed state $p_{\mathbb{I}_d/d}$ and then use the triangle inequality to bound the target total variation.

As discussed in Sec. IV of [37], rank-1 positive operator-valued measure (POVM) can simulate arbitrary POVM if we do not care about the postmeasurement state, thus we can assume that all measurements in the decision tree are rank-1

POVM. For a leaf l in the decision tree, there is a unique path from the root node n_0 to the leaf node $n_T = l$. Denote the nodes on the path by $n_0, \dots, n_T$ and the edges between them by $e_1, \dots, e_T$. For $t = 1, \dots, T$, denote the rank-1 POVM operator associated to e_t by $w_{e_t}^{n_{t-1}} |\phi_{e_t}^{n_{t-1}}\rangle \langle \phi_{e_t}^{n_{t-1}}|$, with $w_{e_t}^{n_{t-1}} > 0$ satisfying the normalization condition $\sum_e w_e^{n_{t-1}} = d$, where the summation is calculated over all the edges connecting node n_{t-1} with its child nodes. To calculate the total variation between the distribution caused by s_1 and the maximally mixed state, we calculate the following quantity:

$$\begin{aligned} \frac{\mathbb{E}_{\psi \sim \text{Haar}} p_\psi(l)}{p_{\mathbb{I}_d/d}(l)} &= \frac{1}{\prod_{t=1}^T w_{e_t}^{n_{t-1}} \frac{1}{d}} \mathbb{E}_{\psi \sim \text{Haar}} \prod_{t=1}^T w_{e_t}^{n_{t-1}} \langle \phi_{e_t}^{n_{t-1}} | \psi \rangle \langle \psi | \phi_{e_t}^{n_{t-1}} \rangle \\ &= d^T \mathbb{E}_{\psi \sim \text{Haar}} \prod_{t=1}^T \langle \phi_{e_t}^{n_{t-1}} | \psi \rangle \langle \psi | \phi_{e_t}^{n_{t-1}} \rangle \\ &= d^T \langle \Phi | \left(\mathbb{E}_{\psi \sim \text{Haar}} |\psi\rangle \langle \psi|^{\otimes T} \right) | \Phi \rangle \\ &= \frac{d^T}{(d+T-1) \cdots (d+1)d} \langle \Phi | \\ &\quad \times \left(\sum_{\sigma \in S_T} W_\sigma \right) | \Phi \rangle, \end{aligned} \quad (C6)$$

where $|\Phi\rangle = \otimes_{t=1}^T |\phi_{e_t}^{n_{t-1}}\rangle$, and the last equality is given by Lemma A1. With the fact $\langle \Phi | (\sum_{\sigma \in S_T} W_\sigma) | \Phi \rangle \geq 1$ (Lemma V.8 in Ref. [37]), we have

$$\frac{\mathbb{E}_{\psi \sim \text{Haar}} p_\psi(l)}{p_{\mathbb{I}_d/d}(l)} \geq \frac{d^T}{(d+T-1) \cdots (d+1)d} = \frac{1}{(1 + \frac{T-1}{d}) \cdots (1 + \frac{1}{d})1} \geq \left(1 + \frac{T-1}{d}\right)^{-(T-1)}. \quad (C7)$$

We can use it to estimate the total variation as

$$\text{TV} \left(\mathbb{E}_{\psi \sim \text{Haar}} p_\psi, p_{\mathbb{I}_d/d} \right) = \sum_l p_{\mathbb{I}_d/d}(l) \max \left\{ 0, 1 - \frac{\mathbb{E}_{\psi \sim \text{Haar}} p_\psi(l)}{p_{\mathbb{I}_d/d}(l)} \right\} \leq 1 - \left(1 + \frac{T-1}{d}\right)^{-(T-1)}. \quad (C8)$$

On the other hand, we have

$$\begin{aligned} \frac{\mathbb{E}_{\psi_A, \psi_B} p_{\psi_A \otimes \psi_B}(l)}{p_{\mathbb{I}_d/d}(l)} &= \frac{1}{\prod_{t=1}^T (w_{e_t}^{n_{t-1}} \frac{1}{d})} \mathbb{E}_{\psi_A, \psi_B} \prod_{t=1}^T w_{e_t}^{n_{t-1}} \langle \phi_{e_t}^{n_{t-1}} | \psi_A \psi_B \rangle \langle \psi_A \psi_B | \phi_{e_t}^{n_{t-1}} \rangle \\ &= d^T \mathbb{E}_{\psi_A, \psi_B} \prod_{i=1}^T \langle \phi_{e_i}^{n_{i-1}} | \psi_A \psi_B \rangle \langle \psi_A \psi_B | \phi_{e_i}^{n_{i-1}} \rangle = d^T \langle \Phi | \left(\mathbb{E}_{\psi_A, \psi_B} (|\psi_A\rangle \langle \psi_A| \otimes |\psi_B\rangle \langle \psi_B|)^{\otimes T} \right) | \Phi \rangle \\ &= d^T \langle \Phi | \mathcal{V}_T \left(\mathbb{E}_{\psi_A} |\psi_A\rangle \langle \psi_A|^{\otimes T} \otimes \mathbb{E}_{\psi_B} |\psi_B\rangle \langle \psi_B|^{\otimes T} \right) \mathcal{V}_T^\dagger | \Phi \rangle \\ &= \left(\frac{(d^{1/2})^T}{(d^{1/2} + T - 1) \cdots (d^{1/2} + 1)d^{1/2}} \right)^2 \langle \Phi | \mathcal{V}_T \left(\sum_{\sigma_1 \in S_T} W_{\sigma_1} \right) \otimes \left(\sum_{\sigma_2 \in S_T} W_{\sigma_2} \right) \mathcal{V}_T^\dagger | \Phi \rangle, \end{aligned} \quad (C9)$$

where the last equality can be obtained by Lemma A1. By Lemma C1 below, we have

$$\frac{\mathbb{E}_{\psi_A, \psi_B} p_{\psi_A \otimes \psi_B}(l)}{p_{\mathbb{I}_d/d}(l)} \geq \left(\frac{(d^{1/2})^T}{(d^{1/2} + T - 1) \cdots (d^{1/2} + 1)d^{1/2}} \right)^2 = \left(\frac{1}{(1 + \frac{T-1}{d^{1/2}}) \cdots (1 + \frac{1}{d^{1/2}})1} \right)^2 \geq \left(1 + \frac{T-1}{d^{1/2}}\right)^{-2(T-1)}. \quad (C10)$$

We can estimate the total variation as

$$\text{TV} \left(\mathbb{E}_{\psi_A, \psi_B} p_{\psi_A \otimes \psi_B}, p_{\mathbb{I}_d/d} \right) = \sum_l p_{\mathbb{I}_d/d}(l) \max \left\{ 0, 1 - \frac{\mathbb{E}_{\psi_A, \psi_B} p_{\psi_A \otimes \psi_B}(l)}{p_{\mathbb{I}_d/d}(l)} \right\} \leq 1 - \left(1 + \frac{T-1}{d^{1/2}} \right)^{-2(T-1)}. \quad (\text{C11})$$

With triangle inequality, we have

$$\begin{aligned} \text{TV} \left(\mathbb{E}_{\psi \sim \text{Haar}} p_{\psi}, \mathbb{E}_{\psi_A, \psi_B} p_{\psi_A \otimes \psi_B} \right) &\leq \text{TV} \left(\mathbb{E}_{\psi \sim \text{Haar}} p_{\psi}, p_{\mathbb{I}_d/d} \right) + \text{TV} \left(\mathbb{E}_{\psi_A, \psi_B} p_{\psi_A \otimes \psi_B}, p_{\mathbb{I}_d/d} \right) \\ &\leq 2 - \left(1 + \frac{T-1}{d} \right)^{-(T-1)} - \left(1 + \frac{T-1}{d^{1/2}} \right)^{-2(T-1)}. \end{aligned} \quad (\text{C12})$$

By Le Cam's two-point method [36,50], the success probability for any procedure to distinguish between the two sets is upper bounded by

$$\frac{1}{2} + \frac{1}{2} \text{TV} \left(\mathbb{E}_{\psi \sim \text{Haar}} p_{\psi}, \mathbb{E}_{\psi_A, \psi_B} p_{\psi_A \otimes \psi_B} \right). \quad (\text{C13})$$

Combined with Eq. (C12), to achieve a success probability of at least 2/3, it must hold that

$$\begin{aligned} \frac{2}{3} &\leq \frac{1}{2} + \frac{1}{2} \text{TV} \left(\mathbb{E}_{\psi \sim \text{Haar}} p_{\psi}, \mathbb{E}_{\psi_A, \psi_B} p_{\psi_A \otimes \psi_B} \right) \leq \frac{3}{2} - \frac{1}{2} \left(1 + \frac{T-1}{d} \right)^{-(T-1)} - \frac{1}{2} \left(1 + \frac{T-1}{d^{1/2}} \right)^{-2(T-1)} \\ &\leq \frac{3}{2} - \left(1 + \frac{T-1}{d^{1/2}} \right)^{-2(T-1)}, \end{aligned} \quad (\text{C14})$$

which implies

$$\left(1 + \frac{T-1}{d^{1/2}} \right)^{-2(T-1)} \leq \frac{5}{6}. \quad (\text{C15})$$

Thus we must have

$$T \geq \sqrt{\frac{1}{2} \ln \left(\frac{6}{5} \right)} d^{1/4} + 1 = \Omega(d^{1/4}). \quad (\text{C16})$$

We leave the proof of the upper bound in Appendix E. ■

Lemma C1. For any product pure state $|\Phi\rangle = |\phi_1\rangle \otimes \cdots \otimes |\phi_T\rangle$ in $\mathcal{H}_d^{\otimes T}$, we have

$$\langle \Phi | \mathcal{V}_T \left(\sum_{\sigma_1 \in S_T} W_{\sigma_1} \right) \otimes \left(\sum_{\sigma_2 \in S_T} W_{\sigma_2} \right) \mathcal{V}_T^\dagger | \Phi \rangle \geq 1, \quad (\text{C17})$$

where W_{σ_1} and W_{σ_2} both act on the space of $\mathcal{H}_{\sqrt{d}}^{\otimes T}$.

Proof. We prove by induction on T . When $T = 1$, the first-order permutation group S_1 only has one element, id, thus $W_{\text{id}} = \mathbb{I}_{d^{1/2}}$ and $\mathcal{V}_1 = \mathbb{I}_d$, therefore

$$\langle \phi_1 | \mathcal{V}_1 (\mathbb{I}_{d^{1/2}} \otimes \mathbb{I}_{d^{1/2}}) \mathcal{V}_1^\dagger | \phi_1 \rangle = \langle \phi_1 | \phi_1 \rangle = 1. \quad (\text{C18})$$

Now suppose the statement holds for $T-1$. Denote $\Pi = \frac{1}{T!} \sum_{\sigma \in S_T} W_\sigma$ as the projector to the symmetric subspace of $\mathcal{H}_{\sqrt{d}}^{\otimes T}$ and $\tilde{\Pi} = \frac{1}{(T-1)!} \sum_{\tilde{\sigma} \in S_{T-1}} W_{\tilde{\sigma}}$ as the projector to the symmetric subspace of $\mathcal{H}_{\sqrt{d}}^{\otimes (T-1)}$. We have $\Pi(\mathbb{I} \otimes \tilde{\Pi}) = (\mathbb{I} \otimes \tilde{\Pi})\Pi = \Pi$ and $\tilde{\Pi}W_{\tilde{\sigma}} = W_{\tilde{\sigma}}\tilde{\Pi} = \tilde{\Pi}, \forall \tilde{\sigma} \in S_{T-1}$. Note that

$$\begin{aligned} \langle \Phi | \mathcal{V}_T \left(\sum_{\sigma_1 \in S_T} W_{\sigma_1} \right) \otimes \left(\sum_{\sigma_2 \in S_T} W_{\sigma_2} \right) \mathcal{V}_T^\dagger | \Phi \rangle &= \langle \Phi | \mathcal{V}_T \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_1 \in S_T} W_{\sigma_1} (\mathbb{I} \otimes \tilde{\Pi}) \right) \otimes \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_2 \in S_T} W_{\sigma_2} (\mathbb{I} \otimes \tilde{\Pi}) \right) \mathcal{V}_T^\dagger | \Phi \rangle \\ &= \langle \Phi | \mathcal{V}_T \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_1(1)=1} W_{\sigma_1} (\mathbb{I} \otimes \tilde{\Pi}) \right) \otimes \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_2(1)=1} W_{\sigma_2} (\mathbb{I} \otimes \tilde{\Pi}) \right) \mathcal{V}_T^\dagger | \Phi \rangle \\ &\quad + \langle \Phi | \mathcal{V}_T \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_1(1) \neq 1} W_{\sigma_1} (\mathbb{I} \otimes \tilde{\Pi}) \right) \otimes \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_2(1) \neq 1} W_{\sigma_2} (\mathbb{I} \otimes \tilde{\Pi}) \right) \mathcal{V}_T^\dagger | \Phi \rangle \end{aligned}$$

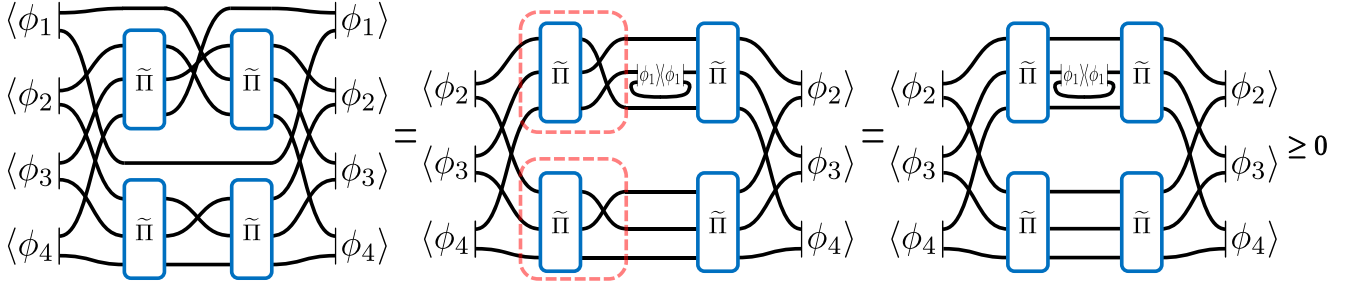


FIG. 5. We take $T = 4$, $W_{\sigma_1} = (1423)$, and $W_{\sigma_2} = (23)$ as an example to illustrate the correctness of Eq. (C22). The general situation is similar to this example. The product of two terms in the red dashed box equals $\tilde{\Pi}$. The last tensor network can be seen as an unnormalized state doing inner product with itself, and thus is non-negative.

$$\begin{aligned}
 & + \langle \Phi | \mathcal{V}_T \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_1(1)=1} W_{\sigma_1} (\mathbb{I} \otimes \tilde{\Pi}) \right) \otimes \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_2(1) \neq 1} W_{\sigma_2} (\mathbb{I} \otimes \tilde{\Pi}) \right) \mathcal{V}_T^\dagger | \Phi \rangle \\
 & + \langle \Phi | \mathcal{V}_T \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_1(1) \neq 1} W_{\sigma_1} (\mathbb{I} \otimes \tilde{\Pi}) \right) \otimes \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_2(1) \neq 1} W_{\sigma_2} (\mathbb{I} \otimes \tilde{\Pi}) \right) \mathcal{V}_T^\dagger | \Phi \rangle.
 \end{aligned} \tag{C19}$$

Denoting $|\tilde{\Phi}\rangle = |\phi_2\rangle \otimes \cdots \otimes |\phi_T\rangle$, the first term of the last equation in Eq. (C19) satisfies

$$\begin{aligned}
 & \langle \Phi | \mathcal{V}_T \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_1(1)=1} W_{\sigma_1} (\mathbb{I} \otimes \tilde{\Pi}) \right) \otimes \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_2(1)=1} W_{\sigma_2} (\mathbb{I} \otimes \tilde{\Pi}) \right) \mathcal{V}_T^\dagger | \Phi \rangle \\
 & = \langle \phi_1 | \phi_1 \langle \tilde{\Phi} | \mathcal{V}_{T-1} \left(\tilde{\Pi} \sum_{\tilde{\sigma}_1 \in S_{T-1}} W_{\tilde{\sigma}_1} \tilde{\Pi} \right) \otimes \left(\tilde{\Pi} \sum_{\tilde{\sigma}_2 \in S_{T-1}} W_{\tilde{\sigma}_2} \tilde{\Pi} \right) \mathcal{V}_{T-1}^\dagger | \tilde{\Phi} \rangle \\
 & = \sum_{\tilde{\sigma}_1, \tilde{\sigma}_2 \in S_{T-1}} \langle \tilde{\Phi} | \mathcal{V}_{T-1} (\tilde{\Pi} W_{\tilde{\sigma}_1} \tilde{\Pi}) \otimes (\tilde{\Pi} W_{\tilde{\sigma}_2} \tilde{\Pi}) \mathcal{V}_{T-1}^\dagger | \tilde{\Phi} \rangle \\
 & = (T-1)!(T-1)! \langle \tilde{\Phi} | \mathcal{V}_{T-1} \tilde{\Pi} \otimes \tilde{\Pi} \mathcal{V}_{T-1}^\dagger | \tilde{\Phi} \rangle \\
 & = \langle \tilde{\Phi} | \mathcal{V}_{T-1} \left(\sum_{\tilde{\sigma}_1 \in S_{T-1}} W_{\tilde{\sigma}_1} \right) \otimes \left(\sum_{\tilde{\sigma}_2 \in S_{T-1}} W_{\tilde{\sigma}_2} \right) \mathcal{V}_{T-1}^\dagger | \tilde{\Phi} \rangle \geq 1,
 \end{aligned} \tag{C20}$$

where the last inequality is given by the induction hypothesis. The second term satisfies

$$\begin{aligned}
 & \langle \Phi | \mathcal{V}_T \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_1(1) \neq 1} W_{\sigma_1} (\mathbb{I} \otimes \tilde{\Pi}) \right) \otimes \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_2(1)=1} W_{\sigma_2} (\mathbb{I} \otimes \tilde{\Pi}) \right) \mathcal{V}_T^\dagger | \Phi \rangle \\
 & = \sum_{\sigma_1(1) \neq 1} \sum_{\sigma_2(1)=1} \langle \Phi | \mathcal{V}_T [(\mathbb{I} \otimes \tilde{\Pi}) W_{\sigma_1} (\mathbb{I} \otimes \tilde{\Pi})] \otimes [(\mathbb{I} \otimes \tilde{\Pi}) W_{\sigma_2} (\mathbb{I} \otimes \tilde{\Pi})] \mathcal{V}_T^\dagger | \Phi \rangle.
 \end{aligned} \tag{C21}$$

We have

$$\langle \Phi | \mathcal{V}_T [(\mathbb{I} \otimes \tilde{\Pi}) W_{\sigma_1} (\mathbb{I} \otimes \tilde{\Pi})] \otimes [(\mathbb{I} \otimes \tilde{\Pi}) W_{\sigma_2} (\mathbb{I} \otimes \tilde{\Pi})] \mathcal{V}_T^\dagger | \Phi \rangle \geq 0 \tag{C22}$$

for all σ_1 and σ_2 satisfying $\sigma_1(1) \neq 1$ and $\sigma_2(1) = 1$, whose proof is shown graphically in Fig. 5 using tensor network diagrams.

Following the same logic, the third term of the last equation in Eq. (C19) is also non-negative. The fourth term of the last equation in Eq. (C19) satisfies

$$\begin{aligned}
 & \langle \Phi | \mathcal{V}_T \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_1(1) \neq 1} W_{\sigma_1} (\mathbb{I} \otimes \tilde{\Pi}) \right) \otimes \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_2(1) \neq 1} W_{\sigma_2} (\mathbb{I} \otimes \tilde{\Pi}) \right) \mathcal{V}_T^\dagger | \Phi \rangle \\
 & = \sum_{\sigma_1(1) \neq 1} \sum_{\sigma_2(1) \neq 1} \langle \Phi | \mathcal{V}_T [(\mathbb{I} \otimes \tilde{\Pi}) W_{\sigma_1} (\mathbb{I} \otimes \tilde{\Pi})] \otimes [(\mathbb{I} \otimes \tilde{\Pi}) W_{\sigma_2} (\mathbb{I} \otimes \tilde{\Pi})] \mathcal{V}_T^\dagger | \Phi \rangle,
 \end{aligned} \tag{C23}$$

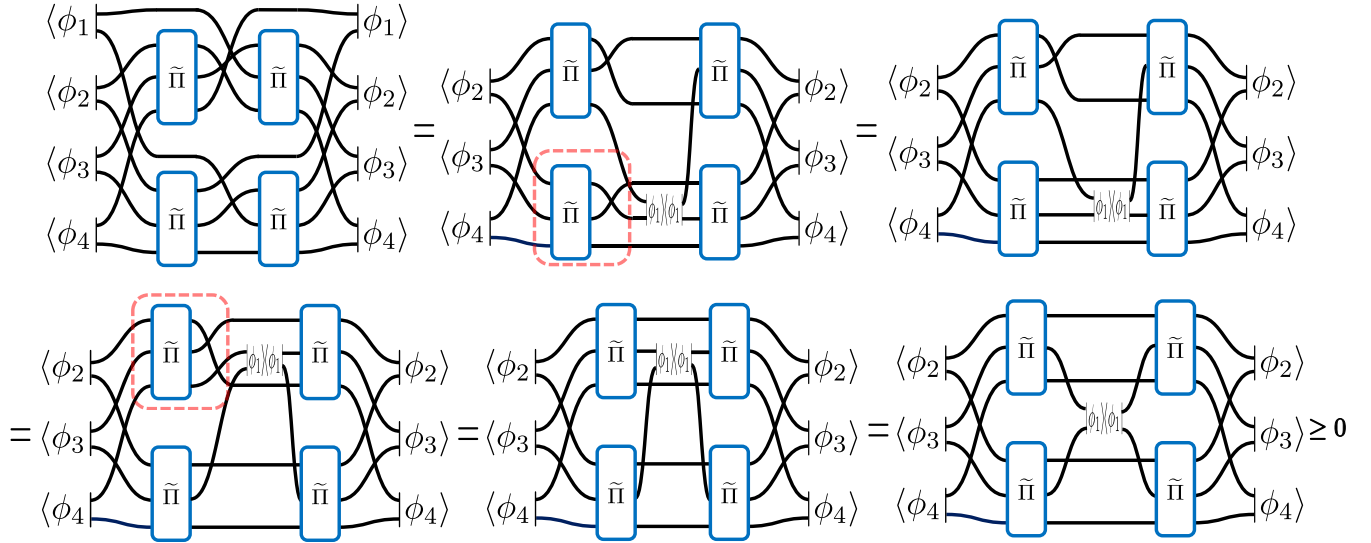


FIG. 6. We take $T = 4$, $W_{\sigma_1} = (1423)$, and $W_{\sigma_2} = (123)$ as an example to illustrate the positivity of Eq. (C24). The product of two terms in the red dashed box equals $\tilde{\Pi}$. The general situation is similar to this example. The last tensor network can be seen as an unnormalized state doing inner product with itself, and thus is non-negative.

where all the terms in the summation are again all non-negative:

$$\langle \Phi | \mathcal{V}_T [(\mathbb{I} \otimes \tilde{\Pi}) W_{\sigma_1} (\mathbb{I} \otimes \tilde{\Pi})] \otimes [(\mathbb{I} \otimes \tilde{\Pi}) W_{\sigma_2} (\mathbb{I} \otimes \tilde{\Pi})] \mathcal{V}_T^\dagger | \Phi \rangle \geq 0. \quad (\text{C24})$$

We provide graphical proof as shown in Fig. 6. ■

APPENDIX D: GENERALIZATION

Until now, we mainly focus on the pure bipartite states. Our result, especially the lower bounds, can be generalized to mixed-state and multipartite cases. Exponential lower bounds of sample complexities also exist in these cases.

For mixed-state entanglement detection, we have the following result.

Theorem D1. Given multiple copies of a state ρ that is sampled from $\pi_{d,k}^*$ with $k = \mathcal{O}(d^{3/2})$, all single-copy adaptive EDPs with completeness, $\Pr_{\rho \sim \pi_{d,k}^*} [\hat{C}(\rho) < 0 | \rho \in \text{ENT}] \geq \frac{1}{2}$, and soundness, $\Pr_{\rho \sim \pi_{d,k}^*} [\rho \in \text{ENT} | \hat{C}(\rho) < 0] \geq \frac{6}{7}$, require $\Theta[(dk)^{1/4}]$ copies of ρ .

For multipartite states, we first define a new state distribution.

Definition D1. A state $\rho \in D(\mathcal{H}_d)$ is said to be sampled according to the distribution of $\pi_{d,1}^{(K)*}$ if and only if with probability 0.5, it is a Haar random pure state $\psi \in \mathcal{H}_d$; with probability 0.5, it is the tensor product of K Haar random pure state $\psi = \bigotimes_{i=1}^K \psi_i$ with $\psi_i \in D(\mathcal{H}_{d^{1/K}})$.

A K -partite state ρ is said to have K -partite entanglement if and only if it cannot be written as

$$\rho = \sum_i p_i \bigotimes_{N=1}^K \rho_N^i, \quad (\text{D1})$$

where p_i is probability satisfying $\sum_i p_i = 1$. Then, the definition of EDP can also be naturally generalized from a bipartite to a multipartite case. We obtain the following result.

Theorem D2. Given multiple copies of a state ρ that is sampled from $\pi_{d,k}^{(K)*}$, all single-copy adaptive multipartite EDPs with completeness, $\Pr_{\rho \sim \pi_{d,k}^{(K)*}} [\hat{C}(\rho) < 0 | \rho \in \text{ENT}] \geq \frac{1}{2}$, and soundness, $\Pr_{\rho \sim \pi_{d,k}^{(K)*}} [\rho \in \text{ENT} | \hat{C}(\rho) < 0] \geq \frac{6}{7}$, require $\Theta(d^{1/2K})$ copies of ρ .

1. Proof of Theorem D1

Here we consider the case of $k = \mathcal{O}(d^{3/2})$ as in this range, all states sampled according to $\pi_{d,k}$ are entangled with probability asymptotically 1 [51]. Thus, the analysis of EDP is also equivalent to the analysis of a specific state discrimination task. Consider the state discrimination task $\{\pi_{d,k}\}$ vs $\{\pi_{\sqrt{d},\sqrt{k}} \otimes \pi_{\sqrt{d},\sqrt{k}}\}$. More precisely, consider the following two state sets with distributions defined as follows:

- (1) $\{\rho = \text{Tr}_k(|\psi\rangle\langle\psi|)\}$, where $|\psi\rangle \in \mathcal{H}_{d \times k}$ is a Haar random pure state.
- (2) $\{\rho = \text{Tr}_{\sqrt{k}}(|\psi_A\rangle\langle\psi_A|) \otimes \text{Tr}_{\sqrt{k}}(|\psi_B\rangle\langle\psi_B|)\}$, where $|\psi_A\rangle, |\psi_B\rangle \in \mathcal{H}_{\sqrt{d} \times \sqrt{k}}$ are Haar random pure states.

Suppose a referee tosses a fair coin to select a set, and then chooses a state ρ from the selected set according to the distribution shown above. Then the referee gives T copies of ρ to us. Our task is to decide which set ρ is from, with T copies of ρ .

Given an EDP satisfying the requirements in Theorem D1, we can solve this state discrimination task with a success probability of $2/3$. Thus, it suffices to prove the following.

Proposition D1. Using single-copy adaptive measurement, we need at least $T = \Omega[(dk)^{1/4}]$ copies of ρ to find out which set ρ is taken from, with a success probability of at least $2/3$.

Proof. Denote p_ψ and p_{ψ_A, ψ_B} as probability distributions on the leaves induced by $\rho = \text{Tr}_2(|\psi\rangle\langle\psi|)$ and $\rho = \text{Tr}_2(|\psi_A\rangle\langle\psi_A|) \otimes \text{Tr}_2(|\psi_B\rangle\langle\psi_B|)$. Consider the following quantity:

$$\begin{aligned} \frac{\mathbb{E}_{\psi \sim \text{Haar}} p_\psi(l)}{p_{\mathbb{I}_d/d}(l)} &= \frac{1}{\prod_{t=1}^T w_{e_t}^{n_t-1} \frac{1}{d}} \mathbb{E}_{\psi \sim \text{Haar}} \prod_{t=1}^T w_{e_t}^{n_t-1} \langle \phi_{e_t}^{n_t-1} | \text{Tr}_k(|\psi\rangle\langle\psi|) | \phi_{e_t}^{n_t-1} \rangle = d^T \mathbb{E}_{\psi \sim \text{Haar}} \prod_{t=1}^T \langle \phi_{e_t}^{n_t-1} | \text{Tr}_k(|\psi\rangle\langle\psi|) | \phi_{e_t}^{n_t-1} \rangle \\ &= d^T \langle \Phi | \left(\mathbb{E}_{\psi \sim \text{Haar}} \text{Tr}_k(|\psi\rangle\langle\psi|)^{\otimes T} \right) | \Phi \rangle = d^T \langle \Phi | \left[\text{Tr}_{\text{even}} \left(\mathbb{E}_{\psi \sim \text{Haar}} |\psi\rangle\langle\psi|^{\otimes T} \right) \right] | \Phi \rangle \\ &= \frac{d^T}{(dk + T - 1) \cdots (dk + 1) dk} \langle \Phi | \text{Tr}_{\text{even}} \left(\sum_{\sigma \in S_T} W_\sigma \right) | \Phi \rangle, \end{aligned} \quad (\text{D2})$$

where W_σ are permutation operators acting on $\mathcal{H}_{d \times k}^{\otimes T}$, and Tr_{even} denotes tracing out systems $2, 4, 6, \dots, 2T$. By Lemma D1, we have

$$\frac{\mathbb{E}_{\psi \sim \text{Haar}} p_\psi(l)}{p_{\mathbb{I}_d/d}(l)} \geq \frac{(dk)^T}{(dk + T - 1) \cdots (dk + 1) dk} = \frac{1}{\left(1 + \frac{T-1}{dk}\right) \cdots \left(1 + \frac{1}{dk}\right)} \geq \left(1 + \frac{T-1}{dk}\right)^{-(T-1)}. \quad (\text{D3})$$

On the other hand, we have

$$\begin{aligned} \frac{\mathbb{E}_{\psi_A, \psi_B} p_{\psi_A, \psi_B}(l)}{p_{\mathbb{I}_d/d}(l)} &= \frac{1}{\prod_{t=1}^T (w_{e_t}^{n_t-1} \frac{1}{d})} \mathbb{E}_{\psi_A, \psi_B} \prod_{t=1}^T w_{e_t}^{n_t-1} \langle \phi_{e_t}^{n_t-1} | \text{Tr}_{\sqrt{k}}(|\psi_A\rangle\langle\psi_A|) \otimes \text{Tr}_{\sqrt{k}}(|\psi_B\rangle\langle\psi_B|) | \phi_{e_t}^{n_t-1} \rangle \\ &= d^T \mathbb{E}_{\psi_A, \psi_B} \prod_{t=1}^T \langle \phi_{e_t}^{n_t-1} | \text{Tr}_{\sqrt{k}}(|\psi_A\rangle\langle\psi_A|) \otimes \text{Tr}_{\sqrt{k}}(|\psi_B\rangle\langle\psi_B|) | \phi_{e_t}^{n_t-1} \rangle \\ &= d^T \langle \Phi | \left(\mathbb{E}_{\psi_A, \psi_B} [\text{Tr}_{\sqrt{k}}(|\psi_A\rangle\langle\psi_A|) \otimes \text{Tr}_{\sqrt{k}}(|\psi_B\rangle\langle\psi_B|)]^{\otimes T} \right) | \Phi \rangle \\ &= d^T \langle \Phi | \mathcal{V}_T \left(\mathbb{E}_{\psi_A} \text{Tr}_{\sqrt{k}}(|\psi_A\rangle\langle\psi_A|)^{\otimes T} \otimes \mathbb{E}_{\psi_B} \text{Tr}_{\sqrt{k}}(|\psi_B\rangle\langle\psi_B|)^{\otimes T} \right) \mathcal{V}_T^\dagger | \Phi \rangle \\ &= d^T \langle \Phi | \mathcal{V}_T \left[\text{Tr}_{\text{even}} \left(\mathbb{E}_{\psi_A} |\psi_A\rangle\langle\psi_A|^{\otimes T} \right) \otimes \text{Tr}_{\text{even}} \left(\mathbb{E}_{\psi_B} |\psi_B\rangle\langle\psi_B|^{\otimes T} \right) \right] \mathcal{V}_T^\dagger | \Phi \rangle \\ &= d^T \left(\frac{1}{(\sqrt{dk} + T - 1) \cdots (\sqrt{dk} + 1) \sqrt{dk}} \right)^2 \langle \Phi | \mathcal{V}_T \text{Tr}_{\text{even}} \left(\sum_{\sigma_1 \in S_T} W_{\sigma_1} \right) \otimes \text{Tr}_{\text{even}} \left(\sum_{\sigma_2 \in S_T} W_{\sigma_2} \right) \mathcal{V}_T^\dagger | \Phi \rangle, \end{aligned} \quad (\text{D4})$$

where $W_{\sigma_1}, W_{\sigma_2}$ are permutation operators acting on $(\mathcal{H}_{\sqrt{dk}})^{\otimes T}$. By Lemma D2, we have

$$\frac{\mathbb{E}_{\psi_A, \psi_B} p_{\psi_A, \psi_B}(l)}{p_{\mathbb{I}_d/d}(l)} \geq \left(\frac{\sqrt{dk}^T}{(\sqrt{dk} + T - 1) \cdots (\sqrt{dk} + 1) \sqrt{dk}} \right)^2 = \left(\frac{1}{\left(1 + \frac{T-1}{\sqrt{dk}}\right) \cdots \left(1 + \frac{1}{\sqrt{dk}}\right)} \right)^2 \geq \left(1 + \frac{T-1}{\sqrt{dk}}\right)^{-2(T-1)}. \quad (\text{D5})$$

Putting everything together, we have

$$\begin{aligned} \text{TV} \left(\mathbb{E}_{\psi \sim \text{Haar}} P_{\psi}, \mathbb{E}_{\psi_A, \psi_B} P_{\psi_A, \psi_B} \right) &\leq \text{TV} \left(\mathbb{E}_{\psi \sim \text{Haar}} P_{\psi}, P_{\mathbb{I}_d/d} \right) + \text{TV} \left(\mathbb{E}_{\psi_A, \psi_B} P_{\psi_A, \psi_B}, P_{\mathbb{I}_d/d} \right) \\ &\leq 2 - \left(1 + \frac{T-1}{dk} \right)^{-(T-1)} - \left(1 + \frac{T-1}{\sqrt{dk}} \right)^{-2(T-1)} \leq 2 - 2 \left(1 + \frac{T-1}{\sqrt{dk}} \right)^{-2(T-1)}. \end{aligned} \quad (\text{D6})$$

To achieve a success probability of at least $2/3$, it must hold that

$$\frac{2}{3} \leq \frac{1}{2} + \frac{1}{2} \text{TV} \left(\mathbb{E}_{\psi \sim \text{Haar}} P_{\psi}, \mathbb{E}_{\psi_A, \psi_B} P_{\psi_A, \psi_B} \right) \leq \frac{3}{2} - \left(1 + \frac{T-1}{\sqrt{dk}} \right)^{-2(T-1)}, \quad (\text{D7})$$

which implies $(1 + \frac{T-1}{\sqrt{dk}})^{-2(T-1)} \leq \frac{5}{6}$, thus $T \geq \Omega[(dk)^{1/4}]$. $\blacksquare$

Lemma D1. For any product state $|\Phi\rangle = |\phi_1\rangle \otimes \cdots \otimes |\phi_T\rangle \in \mathcal{H}_d^{\otimes T}$ we have

$$\langle \Phi | \text{Tr}_{\text{even}} \left(\sum_{\sigma \in S_T} W_{\sigma} \right) | \Phi \rangle \geq k^T. \quad (\text{D8})$$

Proof. We prove by induction on T . When $T = 1$, we have

$$\langle \phi_1 | \text{Tr}_k(\mathbb{I}_d \otimes \mathbb{I}_k) | \phi_1 \rangle = k \langle \phi_1 | \phi_1 \rangle = k. \quad (\text{D9})$$

Now suppose the statement holds for $T - 1$. Denote $\Pi = \frac{1}{T!} \sum_{\sigma \in S_T} W_{\sigma}$ as the projector to the symmetric subspace of $\mathcal{H}_{d \times k}^{\otimes T}$ and $\tilde{\Pi} = \frac{1}{(T-1)!} \sum_{\sigma \in S_{T-1}} W_{\sigma}$ as the projector to the symmetric subspace in of $\mathcal{H}_{d \times k}^{\otimes (T-1)}$. We have

$$\begin{aligned} \langle \Phi | \text{Tr}_{\text{even}} \left(\sum_{\sigma \in S_T} W_{\sigma} \right) | \Phi \rangle &= \langle \Phi | \text{Tr}_{\text{even}} \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma \in S_T} W_{\sigma} (\mathbb{I} \otimes \tilde{\Pi}) \right) | \Phi \rangle \\ &= \langle \Phi | \text{Tr}_{\text{even}} \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma(1)=1} W_{\sigma} (\mathbb{I} \otimes \tilde{\Pi}) \right) | \Phi \rangle + \langle \Phi | \text{Tr}_{\text{even}} \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma(1) \neq 1} W_{\sigma} (\mathbb{I} \otimes \tilde{\Pi}) \right) | \Phi \rangle. \end{aligned} \quad (\text{D10})$$

Denote $|\tilde{\Phi}\rangle = |\phi_2\rangle \otimes \cdots \otimes |\phi_T\rangle$. The first term of the last equation in Eq. (D10) satisfies

$$\begin{aligned} \langle \Phi | \text{Tr}_{\text{even}} \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma(1)=1} W_{\sigma} (\mathbb{I} \otimes \tilde{\Pi}) \right) | \Phi \rangle &= k \langle \phi_1 | \phi_1 \rangle \langle \tilde{\Phi} | \text{Tr}_{\text{even}} \left(\tilde{\Pi} \sum_{\tilde{\sigma} \in S_{T-1}} W_{\tilde{\sigma}} \tilde{\Pi} \right) | \tilde{\Phi} \rangle \\ &= k \sum_{\tilde{\sigma} \in S_{T-1}} \langle \tilde{\Phi} | \text{Tr}_{\text{even}} (\tilde{\Pi} W_{\tilde{\sigma}} \tilde{\Pi}) | \tilde{\Phi} \rangle = k(T-1)! \langle \tilde{\Phi} | \text{Tr}_{\text{even}} (\tilde{\Pi}) | \tilde{\Phi} \rangle \\ &= k \langle \tilde{\Phi} | \text{Tr}_{\text{even}} \left(\sum_{\tilde{\sigma} \in S_{T-1}} W_{\tilde{\sigma}} \right) | \tilde{\Phi} \rangle \geq k^T, \end{aligned} \quad (\text{D11})$$

where the last inequality is given by the induction hypothesis. The second term of the last equation in Eq. (D10) satisfies

$$\langle \Phi | \text{Tr}_{\text{even}} \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma(1) \neq 1} W_{\sigma} (\mathbb{I} \otimes \tilde{\Pi}) \right) | \Phi \rangle = \sum_{\sigma(1) \neq 1} \langle \Phi | \text{Tr}_{\text{even}} [(\mathbb{I} \otimes \tilde{\Pi}) W_{\sigma} (\mathbb{I} \otimes \tilde{\Pi})] | \Phi \rangle. \quad (\text{D12})$$

We have

$$\langle \Phi | \text{Tr}_{\text{even}} [(\mathbb{I} \otimes \tilde{\Pi}) W_{\sigma} (\mathbb{I} \otimes \tilde{\Pi})] | \Phi \rangle \geq 0 \quad (\text{D13})$$

for all W_{σ} satisfying $\sigma(1) \neq 1$, which has a graphical proof as shown in Fig. 7. $\blacksquare$

Lemma D2. For any product pure state $|\Phi\rangle = |\phi_1\rangle \otimes \cdots \otimes |\phi_T\rangle$ in $\mathcal{H}_d^{\otimes T}$, we have

$$\langle \Phi | \mathcal{V}_T \text{Tr}_{\text{even}} \left(\sum_{\sigma_1 \in S_T} W_{\sigma_1} \right) \otimes \text{Tr}_{\text{even}} \left(\sum_{\sigma_2 \in S_T} W_{\sigma_2} \right) \mathcal{V}_T^{\dagger} | \Phi \rangle \geq k^T. \quad (\text{D14})$$

Proof. We prove by induction on T . When $T = 1$, we have

$$\langle \phi_1 | \text{Tr}_{2,4}(\mathbb{I}_{\sqrt{d}} \otimes \mathbb{I}_{\sqrt{k}} \otimes \mathbb{I}_{\sqrt{d}} \otimes \mathbb{I}_{\sqrt{k}}) | \phi_1 \rangle = k \langle \phi_1 | \phi_1 \rangle = k. \quad (\text{D15})$$

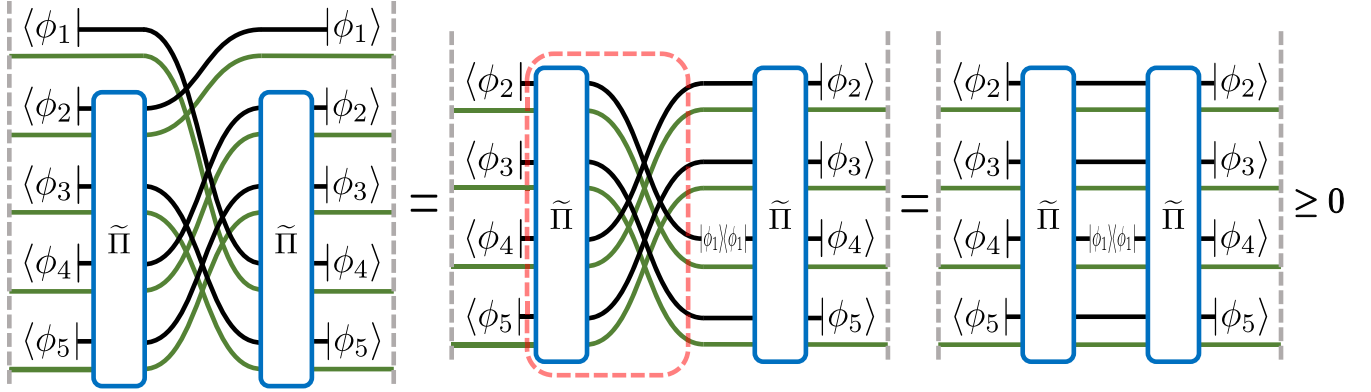


FIG. 7. We take $W_\sigma = (124)(35)$ as an example to illustrate the correctness of Eq. (D13). The general situation is similar to this example. We use black and green legs to represent the indices of subsystems $\mathcal{H}_d$ and $\mathcal{H}_k$, respectively. The gray dashed lines denote the trace function, which is represented by connecting legs in tensor network diagrams. The product of two terms in the red dashed box equals $\tilde{\Pi}$. The last tensor network can be seen as an unnormalized state doing inner product with itself, and thus is non-negative.

Now suppose the statement holds for $T - 1$. We have

$$\begin{aligned}
 & \langle \Phi | \mathcal{V}_T \text{Tr}_{\text{even}} \left(\sum_{\sigma_1 \in S_T} W_{\sigma_1} \right) \otimes \text{Tr}_{\text{even}} \left(\sum_{\sigma_2 \in S_T} W_{\sigma_2} \right) \mathcal{V}_T^\dagger | \Phi \rangle \\
 &= \langle \Phi | \mathcal{V}_T \text{Tr}_{\text{even}} \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_1 \in S_T} W_{\sigma_1} (\mathbb{I} \otimes \tilde{\Pi}) \right) \otimes \text{Tr}_{\text{even}} \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_2 \in S_T} W_{\sigma_2} (\mathbb{I} \otimes \tilde{\Pi}) \right) \mathcal{V}_T^\dagger | \Phi \rangle \\
 &= \langle \Phi | \mathcal{V}_T \text{Tr}_{\text{even}} \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_1(1)=1} W_{\sigma_1} (\mathbb{I} \otimes \tilde{\Pi}) \right) \otimes \text{Tr}_{\text{even}} \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_2(1)=1} W_{\sigma_2} (\mathbb{I} \otimes \tilde{\Pi}) \right) \mathcal{V}_T^\dagger | \Phi \rangle \\
 &\quad + \langle \Phi | \mathcal{V}_T \text{Tr}_{\text{even}} \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_1(1) \neq 1} W_{\sigma_1} (\mathbb{I} \otimes \tilde{\Pi}) \right) \otimes \text{Tr}_{\text{even}} \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_2(1)=1} W_{\sigma_2} (\mathbb{I} \otimes \tilde{\Pi}) \right) \mathcal{V}_T^\dagger | \Phi \rangle \\
 &\quad + \langle \Phi | \mathcal{V}_T \text{Tr}_{\text{even}} \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_1(1)=1} W_{\sigma_1} (\mathbb{I} \otimes \tilde{\Pi}) \right) \otimes \text{Tr}_{\text{even}} \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_2(1) \neq 1} W_{\sigma_2} (\mathbb{I} \otimes \tilde{\Pi}) \right) \mathcal{V}_T^\dagger | \Phi \rangle \\
 &\quad + \langle \Phi | \mathcal{V}_T \text{Tr}_{\text{even}} \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_1(1) \neq 1} W_{\sigma_1} (\mathbb{I} \otimes \tilde{\Pi}) \right) \otimes \text{Tr}_{\text{even}} \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_2(1) \neq 1} W_{\sigma_2} (\mathbb{I} \otimes \tilde{\Pi}) \right) \mathcal{V}_T^\dagger | \Phi \rangle. \tag{D16}
 \end{aligned}$$

The first term of the last equation in Eq. (D16) satisfies

$$\begin{aligned}
 & \langle \Phi | \mathcal{V}_T \text{Tr}_{\text{even}} \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_1(1)=1} W_{\sigma_1} (\mathbb{I} \otimes \tilde{\Pi}) \right) \otimes \text{Tr}_{\text{even}} \left((\mathbb{I} \otimes \tilde{\Pi}) \sum_{\sigma_2(1)=1} W_{\sigma_2} (\mathbb{I} \otimes \tilde{\Pi}) \right) \mathcal{V}_T^\dagger | \Phi \rangle \\
 &= k \langle \phi_1 | \phi_1 \langle \tilde{\Phi} | \mathcal{V}_{T-1} \text{Tr}_{\text{even}} \left(\tilde{\Pi} \sum_{\tilde{\sigma}_1 \in S_{T-1}} w_{\tilde{\sigma}_1} \tilde{\Pi} \right) \otimes \text{Tr}_{\text{even}} \left(\tilde{\Pi} \sum_{\tilde{\sigma}_2 \in S_{T-1}} w_{\tilde{\sigma}_2} \tilde{\Pi} \right) \mathcal{V}_{T-1}^\dagger | \tilde{\Phi} \rangle \\
 &= k \sum_{\tilde{\sigma}_1, \tilde{\sigma}_2 \in S_{T-1}} \langle \tilde{\Phi} | \mathcal{V}_{T-1} \text{Tr}_{\text{even}} (\tilde{\Pi} w_{\tilde{\sigma}_1} \tilde{\Pi}) \otimes \text{Tr}_{\text{even}} (\tilde{\Pi} w_{\tilde{\sigma}_2} \tilde{\Pi}) \mathcal{V}_{T-1}^\dagger | \tilde{\Phi} \rangle \\
 &= k(T-1)!(T-1)! \langle \tilde{\Phi} | \mathcal{V}_{T-1} \text{Tr}_{\text{even}} (\tilde{\Pi}) \otimes \text{Tr}_{\text{even}} (\tilde{\Pi}) \mathcal{V}_{T-1}^\dagger | \tilde{\Phi} \rangle \\
 &= k \langle \tilde{\Phi} | \mathcal{V}_{T-1} \text{Tr}_{\text{even}} \left(\sum_{\tilde{\sigma}_1 \in S_{T-1}} W_{\tilde{\sigma}_1} \right) \otimes \text{Tr}_{\text{even}} \left(\sum_{\tilde{\sigma}_2 \in S_{T-1}} W_{\tilde{\sigma}_2} \right) \mathcal{V}_{T-1}^\dagger | \tilde{\Phi} \rangle \geq k^T, \tag{D17}
 \end{aligned}$$

where $|\tilde{\Phi}\rangle = |\phi_2\rangle \otimes \cdots \otimes |\phi_T\rangle$ and the last inequality is given by the induction hypothesis.

Using proof techniques in Lemmas C1 and D1, we can show that the summands in the second, third, and fourth terms of the last equation in Eq. (D16) are all non-negative.

We also leave the proof for the upper bound in Appendix E. ■

2. Proof of Theorem D2

For $K \geq 2$, consider the following two pure state sets with distributions defined as follows:

(1) s_1 : $\rho = |\psi\rangle\langle\psi|$, where $|\psi\rangle \in \mathcal{H}_d$ is a Haar random pure state.

(2) s_2 : $\rho = |\psi_1\rangle\langle\psi_1| \otimes \cdots \otimes |\psi_K\rangle\langle\psi_K|$, where $|\psi_1\rangle, \dots, |\psi_K\rangle \in \mathcal{H}_{d^{1/K}}$ are Haar random pure states.

Suppose a referee tosses a fair coin to select a set, and then chooses a state ρ from the selected set according to the distribution shown above. Then the referee gives T copies of ρ to us. Our task is to decide which set ρ is from, with T copies of ρ .

It is easy to find that the state sampled in s_1 is K -partite entangled with probabilities 1 and 0 for s_2 . Thus, given an EDP satisfying the requirements in Theorem D2, we can solve this state discrimination task with a success probability of $2/3$. It suffices to prove the following.

Proposition D2. Using single-copy adaptive measurement, we need at least $T = \Theta(d^{1/2K})$ copies of ρ to find out which set ρ is taken from, to achieve a success probability of at least $2/3$.

Proof. Similar to the proof in the last section,

$$\begin{aligned} \frac{\mathbb{E}_{\psi_1, \dots, \psi_K} p_{\psi_1 \otimes \dots \otimes \psi_K(l)}}{p_{\mathbb{I}_d/d}(l)} &= \frac{1}{\prod_{t=1}^T (w_{e_t}^{n_t-1} \frac{1}{d})} \mathbb{E}_{\psi_1, \dots, \psi_K} \prod_{t=1}^T w_{e_t}^{n_t-1} \langle \phi_{e_t}^{n_t-1} | \psi_1 \cdots \psi_K \rangle \langle \psi_1 \cdots \psi_K | \phi_{e_t}^{n_t-1} \rangle \\ &= d^T \mathbb{E}_{\psi_1, \dots, \psi_K} \prod_{i=1}^T \langle \phi_{e_i}^{n_i-1} | \psi_1 \cdots \psi_K \rangle \langle \psi_1 \cdots \psi_K | \phi_{e_i}^{n_i-1} \rangle \\ &= d^T \langle \Phi | \left(\mathbb{E}_{\psi_1, \dots, \psi_K} (|\psi_1\rangle\langle\psi_1| \otimes \cdots \otimes |\psi_K\rangle\langle\psi_K|)^{\otimes T} \right) | \Phi \rangle \\ &= d^T \langle \Phi | \mathcal{V}_T^{(K)} \left(\mathbb{E}_{\psi_1} |\psi_1\rangle\langle\psi_1|^{\otimes T} \otimes \cdots \otimes \mathbb{E}_{\psi_K} |\psi_K\rangle\langle\psi_K|^{\otimes T} \right) \mathcal{V}_T^{(K)\dagger} | \Phi \rangle \\ &= \left(\frac{(d^{1/K})^T}{(d^{1/K} + T - 1) \cdots (d^{1/K} + 1) d^{1/K}} \right)^K \langle \Phi | \mathcal{V}_T^{(K)} \left(\sum_{\sigma_1 \in S_T} W_{\sigma_1} \right) \otimes \cdots \otimes \left(\sum_{\sigma_K \in S_T} W_{\sigma_K} \right) \mathcal{V}_T^{(K)\dagger} | \Phi \rangle, \quad (\text{D18}) \end{aligned}$$

Similar to the proof of Lemma C1, we have

$$\langle \Phi | \mathcal{V}_T^{(K)} \left(\sum_{\sigma_1 \in S_T} W_{\sigma_1} \right) \otimes \cdots \otimes \left(\sum_{\sigma_K \in S_T} W_{\sigma_K} \right) \mathcal{V}_T^{(K)\dagger} | \Phi \rangle \geq 1, \quad (\text{D19})$$

which implies

$$\frac{\mathbb{E}_{\psi_1, \dots, \psi_K} p_{\psi_1 \otimes \dots \otimes \psi_K(l)}}{p_{\mathbb{I}_d/d}(l)} \geq \left(\frac{(d^{1/K})^T}{(d^{1/K} + T - 1) \cdots (d^{1/K} + 1) d^{1/K}} \right)^K = \left(\frac{1}{(1 + \frac{T-1}{d^{1/K}}) \cdots (1 + \frac{1}{d^{1/K}}) 1} \right)^K \geq \left(1 + \frac{T-1}{d^{1/K}} \right)^{-K(T-1)}. \quad (\text{D20})$$

Now we have

$$\text{TV} \left(\mathbb{E}_{\psi_1, \dots, \psi_K} p_{\psi_1 \otimes \dots \otimes \psi_K}, p_{\mathbb{I}_d/d} \right) = \sum_l p_{\mathbb{I}_d/d}(l) \max \left\{ 0, 1 - \frac{\mathbb{E}_{\psi_1, \dots, \psi_K} p_{\psi_1 \otimes \dots \otimes \psi_K(l)}}{p_{\mathbb{I}_d/d}(l)} \right\} \leq 1 - \left(1 + \frac{T-1}{d^{1/K}} \right)^{-K(T-1)}. \quad (\text{D21})$$

By triangle inequality and Eq. (C8), we have

$$\begin{aligned} \text{TV} \left(\mathbb{E}_{\psi \sim \text{Haar}} p_{\psi}, \mathbb{E}_{\psi_1, \dots, \psi_K} p_{\psi_1 \otimes \dots \otimes \psi_K} \right) &\leq \text{TV} \left(\mathbb{E}_{\psi \sim \text{Haar}} p_{\psi}, p_{\mathbb{I}_d/d} \right) + \text{TV} \left(\mathbb{E}_{\psi_1, \dots, \psi_K} p_{\psi_1 \otimes \dots \otimes \psi_K}, p_{\mathbb{I}_d/d} \right) \\ &\leq 2 - \left(1 + \frac{T-1}{d} \right)^{-(T-1)} - \left(1 + \frac{T-1}{d^{1/K}} \right)^{-K(T-1)} \leq 2 - 2 \left(1 + \frac{T-1}{d^{1/K}} \right)^{-K(T-1)}. \quad (\text{D22}) \end{aligned}$$

To achieve a success probability of at least $2/3$, it must hold that

$$\frac{2}{3} \leq \frac{1}{2} + \frac{1}{2} \text{TV} \left(\mathbb{E}_{\psi \sim \text{Haar}} p_{\psi}, \mathbb{E}_{\psi_1, \dots, \psi_K} p_{\psi_1 \otimes \dots \otimes \psi_K} \right) \leq \frac{3}{2} - \left(1 + \frac{T-1}{d^{1/K}} \right)^{-K(T-1)}, \quad (\text{D23})$$

which implies $(1 + \frac{T-1}{d^{1/K}})^{-K(T-1)} \leq \frac{5}{6}$, thus we must have

$$T \geq \sqrt{\frac{1}{K} \ln \left(\frac{6}{5} \right)} d^{1/2K} + 1 = \Omega(d^{1/2K}). \quad (\text{D24})$$

We also leave the proof of the upper bound in Appendix E. ■

APPENDIX E: SAMPLE COMPLEXITY UPPER BOUND

To derive the sample complexity upper bound of the pure state discrimination task stated in Appendix C, we need to design a protocol and analyze its complexity. The protocol we choose is to measure the purity of subsystem A using randomized measurements. This protocol works because for pure product state $\psi_A \otimes \psi_B$, the purity of subsystem A is always $\text{Tr}(\rho_A^2) = 1$. While for global pure state ψ_{AB} , the purity can be exponentially small,

$$\mathbb{E}_{\psi_{AB}}[\text{Tr}(\rho_A^2)] = \frac{2\sqrt{d}}{d+1}. \quad (\text{E1})$$

Therefore, the purity estimation protocol needs to measure purity with constant accuracy to solve this state discrimination task. The randomized measurement protocol works as follows.

- (1) Pick N_U different unitaries U_A from the Haar measure distribution.
- (2) Act each unitary on subsystem A and discard subsystem B to get $U_A \rho_A U_A^\dagger$.
- (3) Measure each $U_A \rho_A U_A^\dagger$ in computational basis for N_M times to get the data $\{b_{U_A}^i\}_{i=1}^{N_M}$.

After acquiring these measurement data, we perform the following calculation to derive the purity of ρ_A :

$$\hat{P}_A = \frac{1}{N_U} \frac{1}{N_M(N_M-1)} \sum_{U_A} \sum_{i \neq j} X(b_{U_A}^i, b_{U_A}^j), \quad (\text{E2})$$

where $X(b, b') = -(-d_A)^{\delta_{b,b'}}$. The unbiasedness of this estimator can be verified using the second-order integral over Haar measure unitaries,

$$\begin{aligned} \mathbb{E}_{U_A \sim \text{Haar}} \mathbb{E}_{\{b_{U_A}^i\}_i}(\hat{P}_A) &= \mathbb{E}_{U_A \sim \text{Haar}} \sum_{b, b'} X(b, b') \text{Pr}(b|U_A) \text{Pr}(b'|U_A) \\ &= \mathbb{E}_{U_A \sim \text{Haar}} \sum_{b, b'} X(b, b') \langle b| U_A \rho_A U_A^\dagger |b \rangle \langle b'| U_A \rho_A U_A^\dagger |b' \rangle \\ &= \mathbb{E}_{U_A \sim \text{Haar}} \text{Tr}[X(U_A \rho_A U_A^\dagger)^{\otimes 2}] = \mathbb{E}_{U_A \sim \text{Haar}} \text{Tr}[U_A^{\otimes 2} X U_A^{\otimes 2} \rho_A^{\otimes 2}] = \text{Tr}(S_A \rho_A^{\otimes 2}) = \text{Tr}(\rho_A^2), \end{aligned} \quad (\text{E3})$$

where $X = \sum_{b, b'} X(b, b') |bb'\rangle \langle bb'|$ is a diagonal matrix satisfying $\mathbb{E}_{U \sim \text{Haar}} U^{\otimes 2} X U^{\dagger \otimes 2} = S$ with S being the SWAP operator.

To derive the sample complexity of this protocol, we need to calculate the variance of this estimator. By definition,

$$\text{Var}(\hat{P}_A) = \mathbb{E}(\hat{P}_A^2) - \mathbb{E}(\hat{P}_A)^2. \quad (\text{E4})$$

As experiments and the estimators for different choices of unitary U_A are independent, the variance can be simplified into

$$\text{Var}(\hat{P}_A) = \frac{1}{N_U} \mathbb{E}_{U_A \sim \text{Haar}} \mathbb{E}_{\{b_{U_A}^i\}_i} \left[\left(\frac{1}{N_M(N_M-1)} \sum_{i \neq j} X(b_{U_A}^i, b_{U_A}^j) \right)^2 - \text{Tr}(\rho_A^2)^2 \right]. \quad (\text{E5})$$

By expanding the summation term, we have

$$\begin{aligned} \left(\frac{1}{N_M(N_M-1)} \sum_{i \neq j} X(b_{U_A}^i, b_{U_A}^j) \right)^2 &= \frac{1}{N_M^2(N_M-1)^2} \sum_{i \neq j} X(b_{U_A}^i, b_{U_A}^j)^2 + \frac{1}{N_M^2(N_M-1)^2} \sum_{i \neq j, k \neq i} X(b_{U_A}^i, b_{U_A}^j) X(b_{U_A}^j, b_{U_A}^k) \\ &\quad + \frac{1}{N_M^2(N_M-1)^2} \sum_{i \neq j \neq k \neq i} X(b_{U_A}^i, b_{U_A}^j) X(b_{U_A}^k, b_{U_A}^i). \end{aligned} \quad (\text{E6})$$

Defining $X_3 = \sum_{b_1, b_2, b_3} X(b_1, b_2) X(b_2, b_3) |b_1 b_2 b_3\rangle \langle b_1 b_2 b_3|$, the variance can be written as

$$\begin{aligned} \text{Var}(\hat{P}_A) &= \frac{1}{N_U} \mathbb{E}_{U_A \sim \text{Haar}} \left[\frac{2}{N_M(N_M-1)} \text{Tr}(U_A^{\dagger \otimes 2} X^2 U_A^{\otimes 2} \rho_A^{\otimes 2}) + \frac{4(N_M-2)}{N_M(N_M-1)} \text{Tr}(U_A^{\dagger \otimes 3} X_3 U_A^{\otimes 3} \rho_A^{\otimes 3}) \right. \\ &\quad \left. + \frac{(N_M-2)(N_M-3)}{N_M(N_M-1)} \text{Tr}(U_A^{\dagger \otimes 4} X^{\otimes 2} U_A^{\otimes 4} \rho_A^{\otimes 4}) - \text{Tr}(\rho_A^2)^2 \right]. \end{aligned} \quad (\text{E7})$$

Using our conclusion in Ref. [13], the expectation terms in this equation are

$$\begin{aligned}\mathbb{E}_{U_A \sim \text{Haar}} \text{Tr}(U_A^{\dagger \otimes 2} X^2 U_A^{\otimes 2} \rho_A^{\otimes 2}) &= d_A + (d_A - 1) \text{Tr}(\rho_A^2), \\ \mathbb{E}_{U_A \sim \text{Haar}} \text{Tr}(U_A^{\dagger \otimes 3} X_3 U_A^{\otimes 3} \rho_A^{\otimes 3}) &= -\frac{1}{d_A + 2} [1 + 2 \text{Tr}(\rho_A^2)] + \frac{3(d_A + 1)}{d_A + 2} \text{Tr}(\rho_A^3) \sim 3 \text{Tr}(\rho_A^3), \\ \mathbb{E}_{U_A \sim \text{Haar}} \text{Tr}(U_A^{\dagger \otimes 4} X^{\otimes 2} U_A^{\otimes 4} \rho_A^{\otimes 4}) &\sim \text{Tr}(\rho_A^2)^2.\end{aligned}\tag{E8}$$

Thus, the variance can be bounded by

$$\text{Var}(\hat{P}_A) \leq \frac{1}{N_U} \left[\frac{C_1 d_A}{N_M^2} + \frac{C_2 \text{Tr}(\rho_A^3)}{N_M} \right],\tag{E9}$$

where C_1 and C_2 are constants that are independent of system size. Therefore, to make sure the variance is a constant independent of the system size, it is sufficient for N_U to be a constant and $N_M = \mathcal{O}(d_A^{1/2})$. Thus the total sample complexity scales as $\mathcal{O}(d^{1/4})$.

The conclusion is easy to generalize to multipartite scenarios, where $d_A = d^{1/K}$ and the sample complexity is $\mathcal{O}(d^{1/2K})$, which proves the conclusion made in Theorem D2. For the mixed-state discrimination task, i.e., $\pi_{d,k}^*$, we need to first decide the accuracy for estimating the purity. Similar to Eq. (E1), if a state is sampled according to $\pi_{d,k}$, the average purity of state of system A , a $\sqrt{d}$ -dimensional subsystem, is approximately $\frac{1}{\sqrt{d}}$, while if the state is sampled according to $\pi_{\sqrt{d}, \sqrt{k}} \otimes \pi_{\sqrt{d}, \sqrt{k}}$, the average purity of subsystem A is approximately $\frac{1}{\sqrt{d}} + \frac{1}{\sqrt{k}}$. Thus, to distinguish these two distributions, the variance in estimating purity should be suppressed to $\frac{1}{k}$. In this case, the sample complexity scales as $\mathcal{O}[(dk)^{1/4}]$.

-
- [1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Quantum entanglement, *Rev. Mod. Phys.* **81**, 865 (2009).
 - [2] O. Gühne and G. Tóth, Entanglement detection, *Phys. Rep.* **474**, 1 (2009).
 - [3] A. Peres, Separability criterion for density matrices, *Phys. Rev. Lett.* **77**, 1413 (1996).
 - [4] G. Vidal and R. F. Werner, Computable measure of entanglement, *Phys. Rev. A* **65**, 032314 (2002).
 - [5] P. Calabrese, J. Cardy, and E. Tonni, Entanglement negativity in quantum field theory, *Phys. Rev. Lett.* **109**, 130502 (2012).
 - [6] H. Shapourian, S. Liu, J. Kudler-Flam, and A. Vishwanath, Entanglement negativity spectrum of random mixed states: A diagrammatic approach, *PRX Quantum* **2**, 030347 (2021).
 - [7] P. Liu, Z. Liu, S. Chen, and X. Ma, Fundamental limitation on the detectability of entanglement, *Phys. Rev. Lett.* **129**, 230503 (2022).
 - [8] J. Gray, L. Banchi, A. Bayat, and S. Bose, Machine-learning-assisted many-body entanglement measurement, *Phys. Rev. Lett.* **121**, 150503 (2018).
 - [9] A. Elben, R. Kueng, H.-Y. R. Huang, R. van Bijnen, C. Kokail, M. Dalmonte, P. Calabrese, B. Kraus, J. Preskill, P. Zoller, and B. Vermersch, Mixed-state entanglement from local randomized measurements, *Phys. Rev. Lett.* **125**, 200501 (2020).
 - [10] Y. Zhou, P. Zeng, and Z. Liu, Single-copies estimation of entanglement negativity, *Phys. Rev. Lett.* **125**, 200502 (2020).
 - [11] X.-D. Yu, S. Imai, and O. Gühne, Optimal entanglement certification from moments of the partial transpose, *Phys. Rev. Lett.* **127**, 060504 (2021).
 - [12] A. Neven, J. Carrasco, V. Vitale, C. Kokail, A. Elben, M. Dalmonte, P. Calabrese, P. Zoller, B. Vermersch, R. Kueng, and B. Kraus, Symmetry-resolved entanglement detection using partial transpose moments, *npj Quantum Inf.* **7**, 152 (2021).
 - [13] Z. Liu, Y. Tang, H. Dai, P. Liu, S. Chen, and X. Ma, Detecting entanglement in quantum many-body systems via permutation moments, *Phys. Rev. Lett.* **129**, 260501 (2022).
 - [14] M. Žnidarič, T. Prosen, G. Benenti, and G. Casati, Detecting entanglement of random states with an entanglement witness, *J. Phys. A: Math. Theor.* **40**, 13787 (2007).
 - [15] U. T. Bhosale, S. Tomsovic, and A. Lakshminarayanan, Entanglement between two subsystems, the Wigner semicircle and extreme-value statistics, *Phys. Rev. A* **85**, 062331 (2012).
 - [16] G. Aubrun, S. J. Szarek, and D. Ye, Entanglement thresholds for random induced states, *Commun. Pure Appl. Math.* **67**, 129 (2014).
 - [17] B. Collins and I. Nechita, Random matrix techniques in quantum information theory, *J. Math. Phys.* **57**, 015215 (2016).
 - [18] M. Weilenmann, B. Dive, D. Trillo, E. A. Aguilar, and M. Navascués, Entanglement detection beyond measuring fidelities, *Phys. Rev. Lett.* **124**, 200502 (2020).
 - [19] O. Gühne, Y. Mao, and X.-D. Yu, Geometry of faithful entanglement, *Phys. Rev. Lett.* **126**, 140503 (2021).
 - [20] Y. Zhou, C. Guo, and X. Ma, Decomposition of a symmetric multipartite observable, *Phys. Rev. A* **99**, 052324 (2019).
 - [21] J. Bae, A. Bera, D. Chruściński, B. C. Hiesmayr, and D. McNulty, How many mutually unbiased bases are needed to detect bound entangled states? *J. Phys. A: Math. Theor.* **55**, 505303 (2022).
 - [22] S. Morelli, M. Huber, and A. Tavakoli, Resource-efficient high-dimensional entanglement detection via symmetric projections, *Phys. Rev. Lett.* **131**, 170201 (2023).
 - [23] H. Häffner, W. Hänsel, C. F. Roos, J. Benhelm, D. Chek-al kar, M. Chwalla, T. Körber, U. D. Rapol, M. Riebe, P. O. Schmidt, C. Becher, O. Gühne, W. Dür, and R. Blatt, Scalable multiparticle entanglement of trapped ions, *Nature (London)* **438**, 643 (2005).
 - [24] D. Leibfried, E. Knill, S. Seidelin, J. Britton, R. B. Blakestad, J. Chiaverini, D. B. Hume, W. M. Itano, J. D. Jost, C. Langer,

- R. Ozeri, R. Reichle, and D. J. Wineland, Creation of a six-atom ‘Schrödinger cat’ state, *Nature (London)* **438**, 639 (2005).
- [25] T. Monz, P. Schindler, J. T. Barreiro, M. Chwalla, D. Nigg, W. A. Coish, M. Harlander, W. Hänsel, M. Hennrich, and R. Blatt, 14-qubit entanglement: Creation and coherence, *Phys. Rev. Lett.* **106**, 130506 (2011).
- [26] J.-W. Pan, Z.-B. Chen, C.-Y. Lu, H. Weinfurter, A. Zeilinger, and M. Żukowski, Multiphoton entanglement and interferometry, *Rev. Mod. Phys.* **84**, 777 (2012).
- [27] A. M. Kaufman, M. E. Tai, A. Lukin, M. Rispoli, R. Schittko, P. M. Preiss, and M. Greiner, Quantum thermalization through entanglement in an isolated many-body system, *Science* **353**, 794 (2016).
- [28] A. Omran, H. Levine, A. Keesling, G. Semeghini, T. T. Wang, S. Ebadi, H. Bernien, A. S. Zibrov, H. Pichler, S. Choi, J. Cui, M. Rossignolo, P. Rembold, S. Montangero, T. Calarco, M. Endres, M. Greiner, V. Vuletić, and M. D. Lukin, Generation and manipulation of Schrödinger cat states in Rydberg atom arrays, *Science* **365**, 570 (2019).
- [29] T. Brydges, A. Elben, P. Jurcevic, B. Vermersch, C. Maier, B. P. Lanyon, P. Zoller, R. Blatt, and C. F. Roos, Probing Rényi entanglement entropy via randomized measurements, *Science* **364**, 260 (2019).
- [30] C. Song, K. Xu, H. Li, Y.-R. Zhang, X. Zhang, W. Liu, Q. Guo, Z. Wang, W. Ren, J. Hao, H. Feng, H. Fan, D. Zheng, D.-W. Wang, H. Wang, and S.-Y. Zhu, Generation of multicomponent atomic Schrödinger cat states of up to 20 qubits, *Science* **365**, 574 (2019).
- [31] X.-L. Wang, Y.-H. Luo, H.-L. Huang, M.-C. Chen, Z.-E. Su, C. Liu, C. Chen, W. Li, Y.-Q. Fang, X. Jiang, J. Zhang, L. Li, N.-L. Liu, C.-Y. Lu, and J.-W. Pan, 18-qubit entanglement with six photons’ three degrees of freedom, *Phys. Rev. Lett.* **120**, 260502 (2018).
- [32] S. Cao *et al.*, Generation of genuine entanglement up to 51 superconducting qubits, *Nature (London)* **619**, 738 (2023).
- [33] Y. Zhou, Entanglement detection under coherent noise: Greenberger-Horne-Zeilinger-like states, *Phys. Rev. A* **101**, 012301 (2020).
- [34] S. Morelli, H. Yamasaki, M. Huber, and A. Tavakoli, Entanglement detection with imprecise measurements, *Phys. Rev. Lett.* **128**, 250501 (2022).
- [35] D. Miller, D. Loss, I. Tavernelli, H. Kampermann, D. Bruß, and N. Wyderka, Shor–Laflamme distributions of graph states and noise robustness of entanglement, *J. Phys. A: Math. Theor.* **56**, 335303 (2023).
- [36] H.-Y. Huang, M. Broughton, J. Cotler, S. Chen, J. Li, M. Mohseni, H. Neven, R. Babbush, R. Kueng, J. Preskill, and J. R. McClean, Quantum advantage in learning from experiments, *Science* **376**, 1182 (2022).
- [37] S. Chen, J. Cotler, H.-Y. Huang, and J. Li, Exponential separations between learning with and without quantum memory, in *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)* (IEEE, Piscataway, NJ, 2022), pp. 574–585.
- [38] D. N. Page, Information in black hole radiation, *Phys. Rev. Lett.* **71**, 3743 (1993).
- [39] A. Elben, B. Vermersch, C. F. Roos, and P. Zoller, Statistical correlations between locally randomized measurements: A toolbox for probing entanglement in many-body quantum states, *Phys. Rev. A* **99**, 052323 (2019).
- [40] A. Elben, S. T. Flammia, H.-Y. Huang, R. Kueng, J. Preskill, B. Vermersch, and P. Zoller, The randomized measurement toolbox, *Nat. Rev. Phys.* **5**, 9 (2023).
- [41] P. Cieřliński, S. Imai, J. Dziewior, O. Gühne, L. Knips, W. Laskowski, J. Meinecke, T. Paterek, and T. Vértesi, Analysing quantum systems with randomised measurements, *Phys. Rep.* **1095**, 1 (2024).
- [42] S. Imai, Randomized measurements as a tool in quantum information processing, Ph.D. thesis, Universität Siegen, 2023.
- [43] F. Wei, Z. Liu, G. Liu, Z. Han, D.-L. Deng, and Z. Liu, Realizing non-physical actions through Hermitian-preserving map exponentiation, *npj Quantum Inf.* **10**, 134 (2024).
- [44] A. Rico and F. Huber, Entanglement detection with trace polynomials, *Phys. Rev. Lett.* **132**, 070202 (2024).
- [45] A. W. Harrow, Approximate orthogonality of permutation operators, with application to quantum information, *Lett. Math. Phys.* **114**, 1 (2023).
- [46] X.-D. Yu, J. Shang, and O. Gühne, Optimal verification of general bipartite pure states, *npj Quantum Inf.* **5**, 112 (2019).
- [47] M. Weilenmann, E. A. Aguilar, and M. Navascués, Analysis and optimization of quantum adaptive measurement protocols with the framework of preparation games, *Nat. Commun.* **12**, 4553 (2021).
- [48] C. J. Wood, J. D. Biamonte, and D. G. Cory, Tensor networks and graphical calculus for open quantum systems, *Quant. Inf. Comp.* **15**, 0759 (2015).
- [49] A. W. Harrow, The church of the symmetric subspace, *arXiv:1308.6595*.
- [50] B. Yu, Assouad, fano, and le cam, in *Festschrift for Lucien Le Cam: Research Papers in Probability and Statistics*, edited by D. Pollard, E. Torgersen, and G. L. Yang (Springer, New York, NY, 1997), pp. 423–435.
- [51] G. Aubrun, S. J. Szarek, and D. Ye, Phase transitions for random states and a semicircle law for the partial transpose, *Phys. Rev. A* **85**, 030302 (2012).