Towards efficient and generic entanglement detection

Jue Xu* and Qi Zhao†
Department of Computer Science, University of Hong Kong.
(Dated: October 23, 2022)

Detection of entanglement is an indispensable step to practical quantum computation and communication. Compared with existing works, we propose an end-to-end, machine learning assisted entanglement detection protocol that is experimental friendly and robust to different types of noises. In this protocol, an entanglment witness for a generic entangled state is obtained by classical machine learning with a synthetic dataset which consists of classical features of certain states and their labels. The classical features of a state, that is expectation values of a set of Pauli observables, are estimated by the sample-efficient scheme developed recently.

I. INTRODUCTION

Entanglement [1] is the key ingredient of quantum computation [2], quantum communication, and quantum cryptography [3]. However, decoherence and imperfection are inevitable in real-world devices, which means the interaction between a quantum system and classical environment would significantly affect entanglement quality and diminish quantum advantage. So, for practical purpose, it is essential to benchmark (characterize) entanglement structures of certain target states in actual (real) experiments. Though the entanglement detection problem [4] has been widely studied, it is far from being perfectly solved. Even we are given the full density matrix of a general state, it is computationally hard to determine its separability classically [5], even by quantum computation [6]. If we would like to know the separability of an unknown state from experiments, the sample complexity to fully recover a density matrix is prohibitive [7] [8]. So, a more realistic scenario is to determine whether a state from experiments is still entangled, assuming it is a known entangled state subject to noise. This problem for many entangled states of practical interest can be efficiently solved by measuring few observables called entanglement witness [9] [10] [11], but more analysis and measurement settings are required for more general states and noise cases [12] [13] [14].

The goal of this paper is to find an efficient and generic way to detect entanglement around a target state. Machine learning (ML) is a powerful tool for such purpose. As we know, many ML techniques including quantum machine learning models [15] have been proposed for classification tasks in physics, such as classification of phases and prediction of ground states [16] [17]. Entanglement detection as a typical classification problem has been studies by machine learning techniques, such as determining separability by Neural Network [18] [19] and deriving generic entanglement witness by Support Vector Machine [20] [21]. However, these prior machine learning assisted methods only consider the robustness to white noise and do not address the problem how to efficiently extract classical features of quantum states in real experiments.

In contrast, our method exhibits better robustness to white noise than the conventional fidelity witness and also robust to coherent noise which is more realistic in experiments but not widely studied. Specifically, our protocol starts from evaluation of expectations of n-qubit Pauli observables of a target state. The set of expectation values that serves as classical features of the target state, together with its label, consist of a data point in a dataset. Then, a classical ML classifier is obtained by training with this dataset. With the trained classifier at hand, it is expected that brand new samples from real experiments can be classified with high accuracy, where classical features of quantum states are estimated by the classical shadow method [22] with affordable samples complexity.

This paper is organized as follows: in Section II, we briefly present necessary definitions about entanglement structures and mainstream entanglement detection methods; Section III demonstrates our end-to-end protocol including two parts: learning an entanglement witness from synthetic data and estimating classical features of states from experiments; at last, numerical simulation results are discussed in Section IV.

II. PRELIMINARIES

A. Entanglement structures

Large scale entanglement involving multiple particles maybe the main resource for quantum advantages in quantum computation and communication. Roughly, we say a quantum state ρ of n subsystems is *entangled* if it is not fully separable, i.e., the state cannot be written as the tensor product of all subsystems as $\rho = \rho_1 \otimes \cdots \otimes \rho_n$. However, the simple statement 'the state is entangled' would allow that only two of the particles are entangled while the rest is in a product state, which is very weak entanglement. So, the more interesting entanglement property is bipartite separability:

* juexu@cs.umd.edu
† zhaoqi@cs.hku.hk

Definition 1 (bi-separable)**.** A pure state $|\psi\rangle$ is bipartite separable (bi-separable) if and only if it can be written as a tensor product form $|\psi\rangle_{\text{bi}}^{\mathcal{P}} = |\phi_A\rangle \otimes |\phi_B\rangle$ with some bi-partition $\mathcal{P} = \{A, B \equiv \bar{A}\}$. A mixed state ρ is bi-separable if and only if it can be written as a convex combination of pure bi-separable states, i.e., $\rho_{\text{bi}} = \sum_i p_i |\psi_i\rangle\langle\psi_i|_{\text{bi}}^{\mathcal{P}_i}$ ($\mathcal{P}_i$ can be different partitions) with a probability distribution $\{p_i\}$. The set of all bi-separable states is denoted as $\mathcal{S}_{\text{bi}}$.

Definition 2 (GME)**.** On the contrary, if a state ρ is not a convex combination of any (partition) bi-separable states, i.e., $\rho \notin \mathcal{S}_{\text{bi}}$, it possesses genuine multipartite entanglement (GME).

GME implies that all subsystems are indeed entangled with each other, so it is the strongest form of entanglement. Whereas, there is another restricted way for generalizing bi-separability to mixed states: if it is a mixing of pure bi-separable states with the same partition $\mathcal{P}_2$, and we denote the state set as $\mathcal{S}_{\text{bi}}^{\mathcal{P}_2}$. It is practically interesting to study entanglement structure under certain partition, because it naturally indicates the quantum information processing capabilities among a real geometric configuration. We have a definition concerning partitions:

Definition 3 (full entanglement)**.** A state ρ possesses full entanglement if it is outside of the separable state set $\mathcal{S}_{\text{bi}}^{\mathcal{P}_2}$ for any partition, that is, $\forall \mathcal{P}_2 = \{A, \bar{A}\}, \rho \notin \mathcal{S}_{\text{bi}}^{\mathcal{P}_2}$.

For a state with full entanglement, it is possible to prepare it by mixing bi-separable states with different bipartitions, so full entanglement is weaker than GME but still useful.

B. Entanglement detection

After introducing the definitions about entanglement, the next basic question is how to determine entanglement and its computational complexity. Despite its clear definitions, it is a highly non-trivial problem for a general state. For a general review on this subject, we refer readers to [4]. The most widely studied problem in this area maybe bi-separability.

Problem 1 (separability)**.** Given a density matrix [23] ρ, to determine if it is bi-separable (in $\mathcal{S}_{\text{bi}}$).

It is not hard to prove that if a state is bi-separable regarding $\mathcal{P} = \{A, B\}$, then it must have positive partial transpose [24] (PPT), that is, the partially transposed (PT) density matrix $\rho_{AB}^{\Gamma_A}$ is positive, semidefinite [25] [26] [27]. By contrapositive, we have a sufficient condition for (bipartite) entanglement, that is

Theorem 1 (PPT criterion)**.** *If the smallest eigenvalue of partial transpose $\rho_{AB}^{\Gamma_A}$ is negative (NPT), then the state is entangled (cannot be bi-separable with $\mathcal{P} = \{A, B\}$).*

We should mention that PPT criterion is a necessary and sufficient condition for separability only when the system dimension is low ($d_A d_B \leq 6$) [27]. Therefore, no general solution for the separability problem is known. Then, a natural question is whether it is possible to solve separability approximately. By relaxing the defintion (promise a gap between two types of states), a reformulation of separability in the theoretic computer science language is

Problem 2 (Weak membership problem for separability)**.** Given a density matrix ρ with the promise that either (i) $\rho \in \mathcal{S}_{\text{bi}}$ or (ii) $\|\rho - \rho_{\text{bi}}\| \geq \epsilon$ with certain norm, decide which is the case.

Unfortunately, even we are given the complete information about a state and promised a gap (error tolerance ϵ), it is still hard to determine separability approximately by classical computation. Weak membership problem for separability is NP-Hard for $\epsilon = 1/\operatorname{poly}(d_A, d_B)$ with respect to Euclidean norm and trace norm [28] [5] [29], while there exists a quasipolynomial-time algorithm with respect to the LOCC norm [30]. A notable example is the widely-used and powerful criteria called k-symmetric extension hierarchy based on SDP [31] [32] [33], which is computationally intractable with growing k. The quantum complexity of a series of related separability testing problems were studied in the framework of quantum interactive proofs [6]. Nevertheless, these hardness results do not rule out the possibility to solve it efficiently with stronger promise (approximation) or by machine learning (heuristic) techniques powered by data.

1. *Entanglement witness based on fidelity*

A related but different problem setting is how to determine bi-separable given copies of an unknown state (from experiments) rather than its density matrix. Since the input to this problem is quantum data (states), directly estimating spectrum of the reduced density matrix $\rho_A := \text{Tr}_B(\rho_{AB})$ by quantum circuits is a good option (without fully recovering density matrix). For example, multivariate trace $\text{Tr}(\rho_A^m)$ encodes the entanglement information (e.g., purity, negativity, and entanglement entropy) of ρ_{AB} with ρ_A being the reduced density matrix [34] [35] [36] . The multivariate trace can be estimated by constant depth quantum circuits [37] [38], but this line of work is still based on PPT criterion (one-side test). The problem we study in this paper is another variant:

Problem 3 (entanglement detection)**.** Given copies of an unknown state ρ (from experiments) that is promised either (i) $\rho \in \mathcal{S}_{\text{bi}}$ or (ii) in proximity of a target $|\psi_{\text{tar}}\rangle$ (i.e., possesses 'useful' entanglement such as GME, full entanglement, depth ...) [39], determine which is the case.

The typical scenario for this problem is one aims to prepare a pure entangled state $|\psi_{\text{tar}}\rangle$ in experiments and

would like to detect (verify) it as true multipartite entangled. While the preparation is not perfect, it is reasonable to assume that the prepared mixed state $\rho_{\rm pre}$ is in the proximity of the target state, that is, $|\psi_{\rm tar}\rangle$ undergoes noise channels restricted to white noise, bit/phase-flip error, or random local unitary.

This problem can be expected to be solved more efficiently, because we have a much stronger promise than the separability problem. The usual method is constructing an observable W called entanglement witness such that

$$\mathrm{Tr}(W\rho_{\rm bi}) \geq 0 \text{ and } \mathrm{Tr}(W|\psi_{\rm tar}\rangle\langle\psi_{\rm tar}|) < 0 \qquad (1)$$

Eq. (1) means that the witness W has a positive expectation value on all separable states, hence a negative expectation value implys the presence of entanglement (GME). Entanglement witness only provides one-side test. For every entangled state, a witness can always be constructed, but no entanglement witness works for all entangled states [40]. For example, the Bell (CHSH) inequalities originally proposed to rule out local hidden variable models, can be regarded as an entanglement witness [41]. A Bell inequality is a linear combination of Pauli observables $W_{\rm Bell} := \mathbf{w}_{\rm Bell}\cdot\mathbf{O}_{\rm Bell}$ such that only entangled states ρ have $|\mathrm{Tr}(\rho W_{\rm Bell})|$ greater than a threshold [42].

While various methods for constructing an entanglement witness exist, the most common one is based on the fidelity between a prepared state $\rho_{\rm pre}$ to the target (pure entangled) state $|\psi_{\rm tar}\rangle$

$$W_\psi = \alpha\mathbb{1} - |\psi_{\rm tar}\rangle\langle\psi_{\rm tar}| \qquad (2)$$

where $\alpha = \max_{\rho_{\rm bi}}\mathrm{Tr}(\rho_{\rm bi}|\psi_{\rm tar}\rangle\langle\psi_{\rm tar}|)$ is the smallest constant such that for every separable state $\mathrm{Tr}(\rho_{\rm bi}W) \geq 0$. For instance, assume the target state is $|\mathrm{GHZ}\rangle := \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} + |1\rangle^{\otimes n})$, the maximal overlap between GHZ and bi-separable states is $1/2$, such that the witness Eq. (2) with $\alpha = 1/2$ certifies tripartite entanglement [43]. We call Eq. (2) as projector-based fidelity witness [9]. In order to effectly measure a witness in an experiment, it is preferrable to decompose the projector term into a sum of locally measurable observables such as [44]. Moreover, for graph states (stabilizer states, i.e., a large class of entanglement states), a witness can be constructed by very few local measurement settings (tradeoff between robustness and meaurement efficiency) [10] [11] [45], ... experiments [46] [47], while for non-stabilizer cases (e.g., W state), more careful analysis is required [14] [20].

III. END-TO-END ENTANGLEMENT DETECTION PROTOCOL

A. Motivation: Beyond fidelity witness

In most studies of fidelity witness, the robustness measure of a fidelity witness is its tolerence (robustness) to white noise:

$$\rho = (1-p_{\rm noise})|\psi_{\rm tar}\rangle\langle\psi_{\rm tar}| + p_{\rm noise}\frac{\mathbb{1}}{2^n} \qquad (3)$$

where the limit of (maximal) $p_{\rm noise}$ indicates the robustness of the witness. For example, the maximally-entangled Bell state can maximally violate the CHSH inequality, but Bell states mixed with white noise don't violate the CHSH inequality when $1-1/\sqrt{2} < p_{\rm noise} < 2/3$, despite they are still entangled in this regime.

For 3-qubit GHZ and W states mixed with white noise, we can analytically compute the white noise threshold for NPT (implies bipartite entanglement): when $p_{\rm noise} < 0.8$ (0.791 W), GHZ states cannot be bi-separable with respect to any partition (that is full entanglement). However, the conventional fidelity witness only detects GME when $p_{\rm noise} < 4/7$ (8/21)[4]. So, it would be practically interesting to have a witness for this white noise regime.

Other than white noise, more realistic noise happened in (photonic) experiments is coherent noise, e.g., local rotations. Take GHZ state as an example, unconscious phase accumulation and rotation on the first control qubit can be modeled as [13]

$$|\mathrm{GHZ}(\phi,\theta)\rangle = \cos\theta\,|0\rangle^{\otimes n} + e^{\mathrm{i}\phi}\sin\theta\,|1\rangle^{\otimes n}. \qquad (4)$$

In certain noise regime (see Fig. 3 in [13]), $|\mathrm{GHZ}(\phi,\theta)\rangle$ cannot be detected by conventional fidelity witness because coherent noises diminish the fidelity but not change entanglement property.

To formally characterize the cases beyond fidelity witness, Weilenmann et. al [48] [49] coined the term *unfaithful states* which systematically analyze 2-qudit entangled state mixed with white noise that cannot be detected by fidelity witness. They found that for $d \geq 3$ that almost all states in the Hilbert space are unfaithful. Subsequently, Güthe et. al [50] [51] give a formal definition: A 2-qudit state ρ_{AB} is faithful if and only if there are local unitary transformations U_A and U_B such that $\langle\phi^+|U_A\otimes U_B\rho_{AB}U_A^\dagger\otimes U_B^\dagger|\phi^+\rangle > \frac{1}{d}$. Consequently, they found a necessary and sufficient condition for 2-qubit unfaithfulness: a 2-qubit state ρ_{AB} is faithful if and only if the maximal eigenvalue of

$$\mathcal{X}_2(\rho_{AB}) = \rho_{AB} - \frac{1}{2}(\rho_A\otimes I + I\otimes\rho_B) + \frac{1}{2}I\otimes I \qquad (5)$$

is larger than $1/2$. We can see in Fig. 1, even for 2-qubit states, nonnegligible portion of randomly sampled states are unfaithful but still entangled (NPT).

Although there are variants of witness [13], such as nonlinear witness [12] and post-processing [52], designed to remedy the shortcomings of conventional fidelity witness respectively, it would be meaningful in practice to find a generic method to construct witnesses for entanglement detection. Machine learning techniques satisfy the needs well because supervised learning can be regarded as a powerful nonlinear post-processing tool.

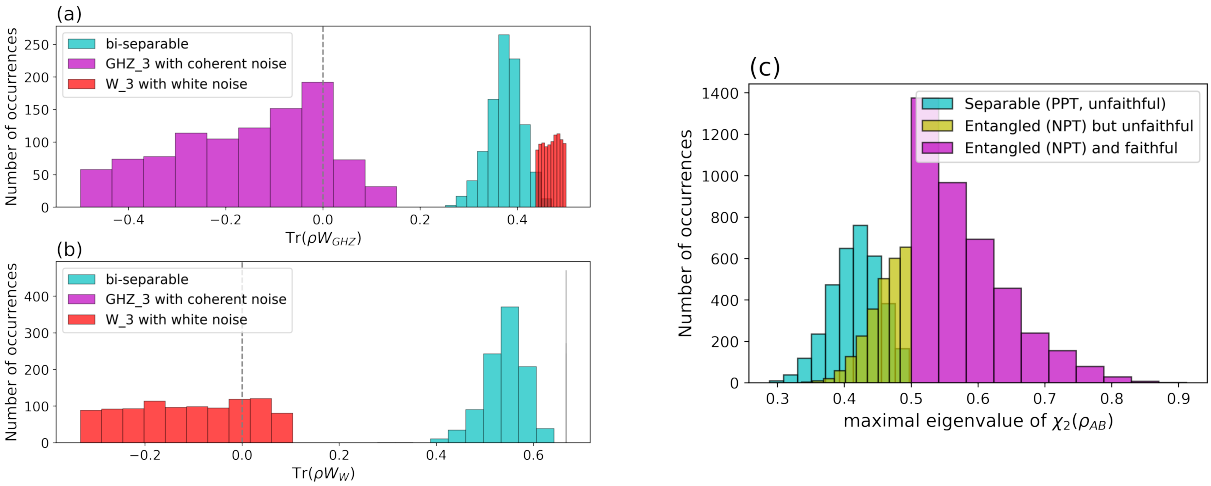

FIG. 1: The entanglement states cannot be detected by conventional fidelity witnesses. Entangled states should be on the left of the dashed vertical line, i.e., have negative expectation value of the witnesses $\mathrm{Tr}(\rho W)$. (a) GHZ states with coherent noise sampled $\theta \in [0, 0.5]$ and $\phi \in [0, 0.6]$ cannot be completely detected by the GHZ projector fidelity witness W_{GHZ}, see Eq. (2); (b) W states with large white noise $p_{\mathrm{noise}} \in [0, 0.5]$ cannot be completely detected by W_{w}. And we can see W states with white noise has $\mathrm{Tr}(\rho_{\mathrm{W}} W_{\mathrm{GHZ}}) > 0$, vice versa. (c) Unfaithfulness of 2-qubit states: 10^4 randomly sampled 2-qubit states are categorized according to the minimal eigenvalue of partial transpose $\rho_{AB}^{\mathsf{T}_A}$ and the maximal eigenvalue of $\chi_2(\rho_{AB})$.

B. Training a generic witness via kernel SVM

One of basic tasks in classical machine learning (ML) is binary classification, such as cat/dog images classification. In this case, the input to a ML algorithm is a (training) dataset $\left\{(\mathbf{x}^{(i)}, y^{(i)})\right\}_{i=1}^{m}$ consists of m data points, where each data point is a pair of feature vector $\mathbf{x} \in \mathbb{R}^d$ of d features and its label (scalar) y (either -1 or 1). For example, the feature $\mathbf{x}$ of an image is a flatten vector of all pixel values and the label $y = -1$ for cat images (1 for dog). It is clear that separability or entanglement detection problem are exactly such binary classification problems where each quantum state has a binary label, such as either entangled or separable. The features $\mathbf{x}$ of a quantum state ρ can be the entries of its density matrix, or more realistically, the expectation values of selected observables.

With the surge of ML research, ML algorithms have been proposed for classification tasks related to entanglement. Lu et. al [18] trained a (universal) separability classifier by classical neural network where features $\mathbf{x}$ are the entries of density matrices. For the similar purpose, Ma and Yung [19] generalized Bell inequalities to a Bell-like ansatz $W_{\mathrm{ml}} := \mathbf{w}_{\mathrm{ml}} \cdot \mathbf{O}_{\mathrm{Bell}}$ where the optimal weights $\mathbf{w}_{\mathrm{ml}}$ are obtained via a neural network. And they found the tomographic ansatz

$$\langle W_{\mathrm{ml}} \rangle := \mathbf{w}_{\mathrm{ml}} \cdot \langle \mathbf{O}_\sigma \rangle \ , \ \forall \sigma \in \{I, X, Y, Z\}^n \qquad (6)$$

where the feature $\mathbf{x} := \langle \mathbf{O}_\sigma \rangle$ is the vector of expectations of all 4^n Pauli observables [53]. The tomographic ansatz not only has better performance, also required [54] for a universal separability classifier. It is worth noting that training such a universal classifier for high-dimensional systems is hard if the gap between two state sets is small (weak promise).

In our paper, we focus on the problem entanglement detection with training data. In other words, we derive the entanglement witness for certain target states with desired entannglement structure by fitting a synthetic dataset.

Problem 4 (learning an entanglement witness)**.**

- **Input**: a dataset $\left\{\left(\rho^{(i)}, y^{(i)}\right)\right\}$ consist of randomly sampled entangled states ρ around $|\psi_{\mathrm{tar}}\rangle$ with label $y = -1$ and randomly sampled separable states with label 1.
- **Output**: a learned classifier $f_{\boldsymbol{\sigma}}(\mathbf{x})$ with high accuracy where $\boldsymbol{\sigma}$ is a subset of all Pauli observables and $\mathbf{x}$ is a vector of corresponding expectation values.

The learning an entanglement witness problem has also been studied by classical ML [20] [21], but by a technique different from Neural Network, called Support Vector Machine (SVM) [55]. A classification task performed by SVM can be formulated as a convex optimization problem: find a hyperplane $(\mathbf{w}, b)$ in a feature space (a linear function $f_{\mathbf{w},b}$) that maximizes the margin between two sets subject to the constraint that two set of data points are separated (on the two sides of the hyperplane see Fig. 2)

$$\max_{\mathbf{w}} \|\mathbf{w}\|_2^2 \text{ s.t. } \forall i,\ y^{(i)} \cdot (\mathbf{w} \cdot \mathbf{x}^{(i)} + b) \geq 1. \qquad (7)$$

where $\mathbf{w}$ is the (not necessarily normalized) normal vector to the hyperplane and b is a bias term sim-

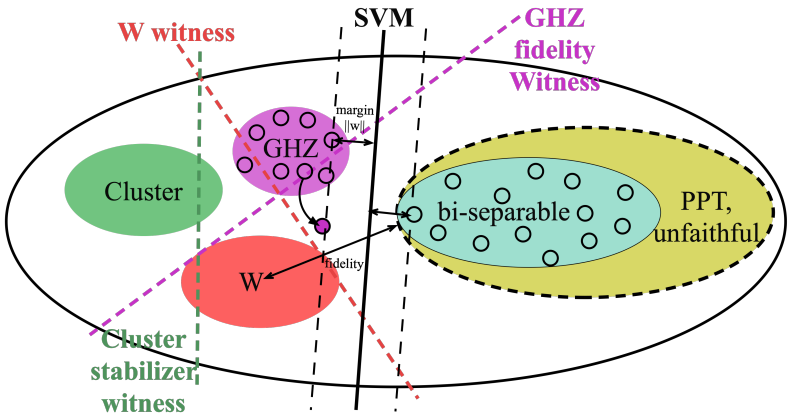

FIG. 2: Schematic diagram for different entanglement detection methods: the colored ellipses indicate the vicinity (white noise) of certain entangled state such as GHZ, W, cluster states. Conventional fidelity witnesses for different states are depicted by colored dash lines (hyperplanes in feature space). Entangled state with large white noise or coherent (local rotation) noise cannot be detected by conventional fideilty witnesses. SVM without kernel is a hyperplane separating two sets of colored dots (synthetic dataset). The data points on the boundary (dashed black lines) are called support vectors. The distance between the SVM hyperplane and boundary is the margin to be minimized via optimization. PPT criterion is a nonlinear but one-side classifier without prior knowledge.

ilar to α in Eq. (2). Therefore, the predicted label is given by the sign of the inner product (projection) of the hyperplane and the feature vector $\mathbf{x}$, i.e., $y = f_{\mathbf{w},b}(\mathbf{x}) = \text{sign}(\mathbf{w}\cdot\mathbf{x}+b)$. Geometrically, both SVM witness and conventional fidelity witness, i.e., $\text{Tr}(W\rho) \equiv \langle W\rangle = \mathbf{w}\cdot\mathbf{x}$, are hyperplanes in feature spaces, but the SVM witness is more flexible because the classifier $(\mathbf{w}, b)$ are numerically derived by optimization for any generic target state. This method only requires local (Pauli) measurements $\mathbf{x} := \text{Tr}(\rho O_\sigma)$ even when the target state is a non-stabilizer state, such as W state, which normally need nonlocal measurements.

Algorithm III.1: train a witness via kernel SVM

input : dataset $\left\{\left(\rho^{(i)}, y^{(i)}\right)\right\}_{i=1}^{m}$, minimal number of features M, and tolerance ϵ
output: a classifier $f_{\boldsymbol{\sigma}}(\mathbf{x})$

Evaluate Pauli observables $\mathbf{x}^{(i)} := \text{Tr}\left(\rho^{(i)} O_{\boldsymbol{\sigma}}\right), \forall i$
while accuracy $< \epsilon$ or $len(\mathbf{x}) > m$ **do**
  for j in range$(len(\mathbf{x}))$ **do**
    /* eliminate j-th feature */
    $\forall i$, let $\tilde{\mathbf{x}}^{(i)}$ be $\mathbf{x}^{(i)}$ without the j-th feature
    /* Train SVM with the new feature vectors */
    accuracy, classifier = SVM$(\left\{\left(\tilde{\mathbf{x}}^{(i)}, y^{(i)}\right)\right\}_i^m)$
    if accuracy $> \epsilon$ **then**
      $\mathbf{x}^{(i)} := \tilde{\mathbf{x}}^{(i)}$ and then **break**
    else if accuracy <= ϵ and $j = len(\mathbf{x})$ **then**
      /* If cannot find a classifier with less features, then output the last classifier with high accuracy */
      return a classifier $f_{\boldsymbol{\sigma}}(\mathbf{x})$

A key drawback of conventional witnesses is its linearity because many real-world datasets are not linearly-separable in a low-dimensional feature space. Despite the nonlinear witness [12] proposed, its experimental implementation is more challenging than linear ones. The good news is, within the framework of SVM, nonlinearity can be easily achieved by the so-called kernel method [56]. The main idea is mapping the features $\mathbf{x}$ to a higher dimensional space via a feature map $\phi(\mathbf{x})$ such that they can be linearly separated in the high dimensional feature space. The kernel function

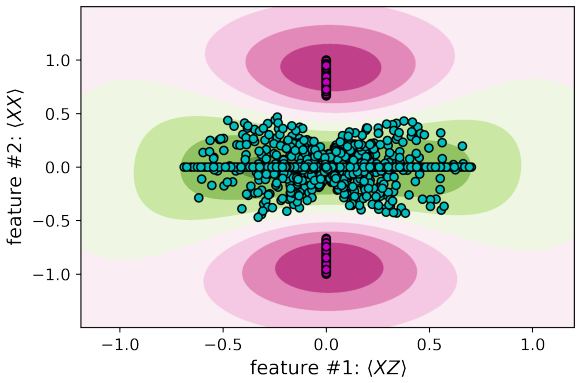

FIG. 3: The two-dimensional embedding (a low-dimensinoal feature space $\langle XZ\rangle$ VS $\langle XX\rangle$) of 2-qubit states: green dots represent randomly sampled separable states, while pink ones represent entangled Bell states mixed with white noise in the range [0,1/3]. The colored shade indicates the nonlinear decision boundary of the kernel SVM classifier.

$k(\mathbf{x}, \mathbf{x}') : \mathcal{X}\times\mathcal{X} \to \mathbb{R}$ measures the similarity between two input data points in a high-dimensional feature space because a kernel can be written as an inner product

$\langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle$. The commonly used kernel is the radial basis function (RBF) kernel which a Gaussian function $k_{\text{rbf}}(\mathbf{x}, \mathbf{x}') := \exp\left(-\gamma \|\mathbf{x} - \mathbf{x}'\|_2^2\right)$ with l_2 Euclidean norm $\|\cdot\|_2$ a parameter γ. Fig. 3 exhibits that two kinds of data points are clearly classified by a nonlinear (RBF kernel) SVM classifier, though it is not linearly separable in this 2-dimensional space.

	# observables	weights	promise
fidelity witness	few local	fixed	strongest
Bell (CHSH) inequality	constant	fixed	weak
tomographic classifier	$4^n - 1$	trained	weakest
SVM (kernel) witness	$\ll 4^n - 1$	trained	strong

TABLE I: Comparison of fidelity witness, CHSH inequality, tomographic classifier, and SVM witness.

However, these prior ML witnesses only consider the robustness to white noise and cannot be directly applied to experiments. In numerical simulation, we can efficiently evaluate classical features by direct calculation, but, in actual experiments, entries of a density matrix are not explicitly known. Instead, we need to estimate the observables by repeat measurements, which we are going to discuss in next section.

C. Sample-efficient expectation estimation methods

The brute force approach to fully characterize a state in an experiment is quantum state tomography [57] [58]. With a recovered densitry matrix, we can directly calculate classcial features or separability measures, but full tomography is experimentally and computationally demanding. Even if adaptive or collective measurements (and post-processing) allowed [59], rigorous analysis [7] [8] showed that $\Omega(D^2/\epsilon^2)$ measurements (copies) are required for recovering a $D \times D$ density matrix with error tolerence ϵ measured by trace distance. Now that full tomography is intractable for large systems, a workaround is to extract (partial) information about a state without fully recovering it:

Problem 5 (shadow tomography)**.** Given m copies of an unknown D-dimensional state and M known 2-outcome measurements $\{E_1, \ldots, E_M\}$, to estimate $\forall i, \text{Tr}(\rho E_i)$ within additive error ϵ with success probability at least $1 - \delta$.

Since shadow tomography can be implemented with $\tilde{\mathcal{O}}(\log^4 M \cdot \log D \cdot \log 1/\delta \cdot \epsilon^{-4})$ copies [60] [61], we can evaluate classical features $x_{\rho,\sigma} = \text{Tr}(\rho O_\sigma)$ for entanglement witness in a samples-efficient (copies) manner. However, Aaronson's shadow tomography procedure is very demanding in terms of quantum hardware (in the collective preparation and measurement on $\rho^{\otimes m}$). To be more friendly to experiments, Huang et. al [22] introduced classical shadow (CS) scheme which we use in our protocol. A classical shadow is a succinct classical description of a quantum state, which can be extracted by performing reasonably simple single-copy measurements (i.e., each measurement measures all qubits in some Pauli X, Y, or Z- basis) on a reasonably small number of copies of the state. The classical shadow attempts to approximate this expectation value by an empirical average over R independent samples, i.e., $o_i = \text{Tr}(O_i \rho_{\text{cs}})$ obeys $\mathbb{E}[o_i] = \text{Tr}(O_i \rho)$.

Algorithm III.2: estimate features by CS

input : samples of ρ and $O_{\boldsymbol{\sigma}_{\text{ml}}}$
output: estimation of $\mathbf{x}_{\rho,\boldsymbol{\sigma}_{\text{ml}}} := \text{Tr}(\rho O_{\boldsymbol{\sigma}_{\text{ml}}})$

for $i = 1, 2, \ldots, R$ **do**
  $\rho \mapsto U \rho U^\dagger$ // apply a random unitary
  $|b\rangle \in \{0,1\}^n$ // measurement outcome
  $\rho_{\text{cs}} = \mathcal{M}^{-1}\left(U^\dagger |b\rangle\langle b| U\right)$ // $\mathcal{M}$ quantum channel
$\text{CS}(\rho, R) = \{\rho_{\text{cs}_1}, \ldots, \rho_{\text{cs}_R}\}$ // classical shadow
// estimate features for SVM from classical shadow
return $\mathbf{x}_{\rho,\boldsymbol{\sigma}_{\text{ml}}} = \text{MedianOfMeans}(\text{CS}(\rho, R)\boldsymbol{\sigma}_{\text{ml}})$

Given a quantum state ρ, a classical shadow is created by repeatedly performing a simple procedure: Apply a unitary transformation $\rho \mapsto U \rho U^\dagger$, and then measure all the qubits in the computational basis $|\mathbf{b}\rangle \in \{|0\rangle, |1\rangle\}^{\otimes n}$. Its classical shadow (snapshots) ρ_{cs} (a density matrix) can be reconstructed

$$\rho_{\text{cs}} := \mathcal{M}^{-1}\left(U^\dagger |\mathbf{b}\rangle\langle\mathbf{b}| U\right), \tag{8}$$

where $\mathcal{M}$ is a quantum channel that depends on the ensemble of random unitary transformation... . The algorithm is summarized in Algorithm. III.2. The number of times this procedure is repeated is called the size of the classical shadow. Classical shadows with size of order $\log(M)$ suffice to predict M target functions $\{O_1, \ldots, O_M\}$. The classical shadow size required to accurately approximate all reduced k-body density matrices scales exponentially in subsystem size k, $\Omega(\log(M) 3^k/\epsilon^2)$ [22], but is independent of the total number of qubits n.

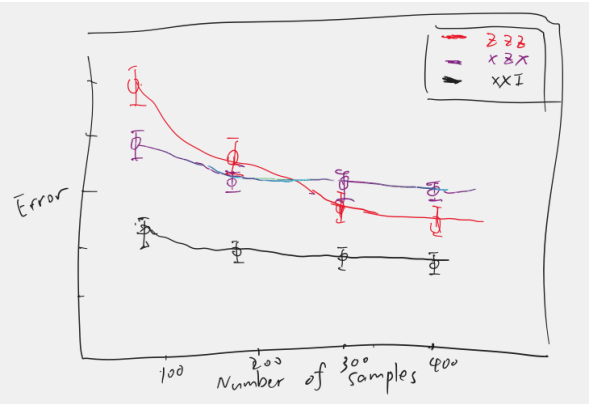

FIG. 4: [TODO] Error of estimating expectation values VS classical shadow sizes (number of samples)

There are several variants of classical shadow method [62] [63]. The derandomized version is the refinement of the original randomized protocol which provides better performance for k-local observables, but not guarantees better performance for global observables (involving all subsystems). The detect entanglement by estimating p_3-PPT with classical shadow and comparison with related methods have been done experimentally [64] [65]. Interestingly, the task of estimating expectation values can also be achieved efficiently by both classical [66] [67] [68] and quantum machine learning [69] [17]. Huang et. al rigorously show that, for achieving accurate prediction on all Pauli observables $\mathrm{Tr}(\rho O_\sigma), \forall \sigma \in \{I, X, Y, Z\}^n$, the exponential quantum advantage over classical ML is possible [70]. Training a more powerful classifier with all Pauli observables as features might be one of our future research direction.

IV. NUMERICAL SIMULATION AND DISCUSSION

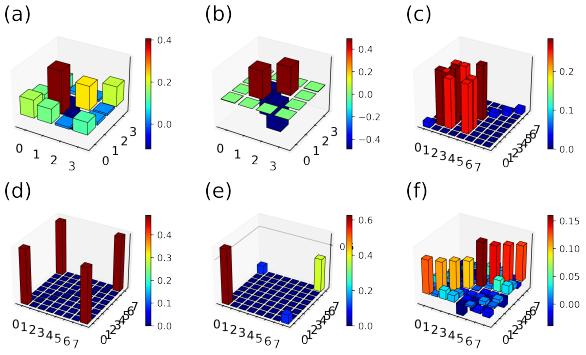

FIG. 5: The real part of sampled density matrices: (a) a random 2-qubit state; (b) Bell state (singlet) with white noise; (c) 3-qubit W state with white noise; (d) 3-qubit GHZ state with white noise; (e) 3-qubit GHZ state with coherent noise; (f) a random 3-qubit bi-separable state

We generate quantum state samples, construct quantum circuits, and manipulate quantum objects numerically by QuTiP Python library [71] [72]. We generate multi-partite entangled states (synthetic data) including: Bell states, 3-qubit GHZ and W states, see Fig. 5 for examples. For different noise channels: white noise according to Eq. (3), coherent noise according to Eq. (4). the noise is randomly (uniform) sampled from $p_{\text{noise}} \in [0, 0.5]$, $\theta \in [0.0.5]$, and $\phi \in [0, 0.6]$. In contrast to entangled states, we generate random separable states for different number of qubits by tensoring random (sampled by Haar measure) density matrices of subsystems. For example, there are three different partitions $\rho_A \otimes \rho_{BC}$, $\rho_{AB} \otimes \rho_C$, and $\rho_B \otimes \rho_{AC}$ for 3-qubit separable states. It is not necessary to prepare the mixed separable states as convex combination of separable states with different partitions because SVM can correctly classifiy a mixture if it can classify each case. For training a 3-qubit SVM classifier with accuracy 0.999, we generate 10^3 states for each cases.

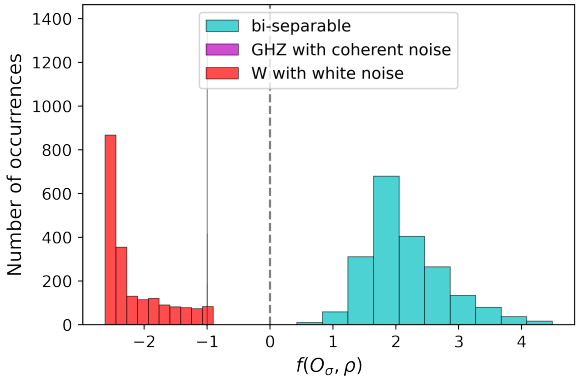

FIG. 6: The states beyond detection by fidelity witnesses (GHZ state with coherence noise, and W state with large white noise) can be classified by the kernel SVM classifier with high accuracy.

For the machine learning part, we make use of scikit-learning Python package [73] to train SVM with RBF kernel. Fig. 1 and Fig. 6 show that conventional fidelity witnesses cannot correctly classify when GHZ states with coherent noises $\theta = 0.5, \phi > 0.5$ and W states mixed with white noise $p_{\text{noise}} > 8/21$, while the SVM classifier can classify them with hight accuracy. One set of features found by the kernel SVM are $\mathbf{x} = (\langle ZZZ \rangle, \langle XZX \rangle, \langle ZZX \rangle, \langle ZIZ \rangle, \langle IZZ \rangle)$. The error of estimating these features VS the size of shadow is shown in Fig. 4 [TODO].

In conclusion, our protocol is flexible and sample-efficient in detecting entanglement around certain entangled state by training a kernel SVM classifier with a synthetic dataset. We test our protocol for 3-qubit GHZ states with coherent noise and W states with large white noise. We also show that the features for training such machine learned classifier can be estimated by a sample-efficient scheme. The reason for choosing SVM as our machine learning technique is many-fold: (1) its clear geometric interpretation analoguous to entanglement witness; (2) the kernel SVM is powerful for classification and theoretically equivalent to Neural Network (nonlinear) in terms of Neural Tangent Kernel [74]; (3) the training of an SVM is convex: if a solution exists for the given target state and ansatz, the optimal SVM will be found; (4) this SVM formalism allows for the programmatic elimination of features [75], i.e., reducing the cost of experimental measurements and copies (samples).

There are also several potential directions for future research: (1) rigorous proof for dataset size and number of features (required for high training accuracy) scaling with the system size (more than 3 qubits); (2) find better kernels, such as graph kernel, quantum kernel, shadow kernel, and neural tangent kernel, for SVM with higher

accuracy and less number of features; (3) quantum machine learning for efficiently estimating all classical features allows for training a tomographic SVM classifier for wider range of states.

ACKNOWLEDGMENTS

[1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009), arXiv:quant-ph/0702225.

[2] H. J. Briegel, D. E. Browne, W. Dür, R. Raussendorf, and M. V. den Nest, Nature Phys **5**, 19 (2009), arXiv:0910.1116.

[3] F. Xu, X. Ma, Q. Zhang, H.-K. Lo, and J.-W. Pan, Rev. Mod. Phys. **92**, 025002 (2020), arXiv:1903.09051.

[4] O. Gühne and G. Toth, Physics Reports **474**, 1 (2009), arXiv:0811.2803 [cond-mat, physics:physics, physics:quant-ph].

[5] L. Gurvits, Classical deterministic complexity of Edmonds' problem and Quantum Entanglement (2003), arXiv:quant-ph/0303055.

[6] G. Gutoski, P. Hayden, K. Milner, and M. M. Wilde, Theory of Comput. **11**, 59 (2015), arXiv:1308.5788 [quant-ph].

[7] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, IEEE Trans. Inform. Theory , 1 (2017).

[8] R. O'Donnell and J. Wright, in *Proc. Forty-Eighth Annu. ACM Symp. Theory Comput.* (ACM, Cambridge MA USA, 2016) pp. 899–912.

[9] M. Bourennane, M. Eibl, C. Kurtsiefer, S. Gaertner, H. Weinfurter, O. Guehne, P. Hyllus, D. Bruss, M. Lewenstein, and A. Sanpera, Phys. Rev. Lett. **92**, 087902 (2004), arXiv:quant-ph/0309043.

[10] G. Toth and O. Guehne, Phys. Rev. Lett. **94**, 060501 (2005), arXiv:quant-ph/0405165.

[11] G. Tóth and O. Gühne, Phys. Rev. A **72**, 022340 (2005).

[12] O. Gühne and N. Lütkenhaus, Phys. Rev. Lett. **96**, 170502 (2006).

[13] Y. Zhou, Phys. Rev. A **101**, 012301 (2020), arXiv:1907.11495 [quant-ph].

[14] Y. Zhang, Y. Tang, Y. Zhou, and X. Ma, Phys. Rev. A **103**, 052426 (2021), arXiv:2012.07606 [quant-ph].

[15] I. Cong, S. Choi, and M. D. Lukin, Nat. Phys. **15**, 1273 (2019), arXiv:1810.03787 [cond-mat, physics:quant-ph].

[16] J. Carrasquilla and R. G. Melko, Nature Phys **13**, 431 (2017), arXiv:1605.01735.

[17] H.-Y. Huang, R. Kueng, G. Torlai, V. V. Albert, and J. Preskill, Science **377**, eabk3333 (2022), arXiv:2106.12627.

[18] S. Lu, S. Huang, K. Li, J. Li, J. Chen, D. Lu, Z. Ji, Y. Shen, D. Zhou, and B. Zeng, Phys. Rev. A **98**, 012315 (2018), arXiv:1705.01523 [quant-ph].

[19] Y.-C. Ma and M.-H. Yung, npj Quantum Inf **4**, 34 (2018), arXiv:1705.00813 [quant-ph].

[20] E. Y. Zhu, L. T. H. Wu, O. Levi, and L. Qian, Machine Learning-Derived Entanglement Witnesses (2021), arXiv:2107.02301 [quant-ph].

[21] S. V. Vintskevich, N. Bao, A. Nomerotski, P. Stankus, and D. A. Grigoriev, Classification of four-qubit entangled states via Machine Learning (2022), arXiv:2205.11512 [quant-ph].

[22] H.-Y. Huang, R. Kueng, and J. Preskill, Nat. Phys. **16**, 1050 (2020), arXiv:2002.08953 [quant-ph].

[23] A quantum (mixed) state ρ can be represented by a density matrix which is a Hermitian, positive semidefinite operator (matrix) of trace one. If the rank of ρ is 1, then the state is a pure state.

[24] The partial transpose (PT) operation acting on subsystem A is defined as $|k_A, k_B\rangle\langle l_A, l_B|^{\intercal_A} := |l_A, k_B\rangle\langle k_A, l_B|$ where $\{|k_A, k_B\rangle\}$ is a product basis of the joint system $\mathcal{H}_{AB}$.

[25] A matrix (operator) is positive, semidefinite (PSD) if all its eigenvalues are non-negative.

[26] A. Peres, Phys. Rev. Lett. **77**, 1413 (1996), arXiv:quant-ph/9604005.

[27] M. Horodecki, P. Horodecki, and R. Horodecki, Physics Letters A **223**, 1 (1996), arXiv:quant-ph/9605038.

[28] The Euclidean norm of a matrix A is defined as $\|A\|_2 := \sqrt{\mathrm{Tr}(A^\dagger A)}$. The trace norm of A is defined as $\|A\|_{\mathrm{Tr}} \equiv \|A\|_1 := \mathrm{Tr}(|A|) \equiv \mathrm{Tr}\left(\sqrt{A^\dagger A}\right)$. Correspondingly, trace distance between two density matrices is $d_{\mathrm{tr}}(\rho, \rho') := \frac{1}{2}\|\rho - \rho'\|_1$.

[29] S. Gharibian, Strong NP-Hardness of the Quantum Separability Problem (2009), arXiv:0810.4507 [quant-ph].

[30] F. G. Brandão, M. Christandl, and J. Yard, in *Proc. 43rd Annu. ACM Symp. Theory Comput. - STOC 11* (ACM Press, San Jose, California, USA, 2011) p. 343, arXiv:1011.2751 [quant-ph].

[31] A. C. Doherty, P. A. Parrilo, and F. M. Spedalieri, Phys. Rev. A **69**, 022308 (2004), arXiv:quant-ph/0308032.

[32] L. M. Ioannou, Quantum Inf. Comput. **7**, 335 (2007), arXiv:quant-ph/0603199.

[33] M. Navascues, M. Owari, and M. B. Plenio, Phys. Rev. A **80**, 052306 (2009), arXiv:0906.2731 [quant-ph].

[34] A. K. Ekert, C. M. Alves, D. K. L. Oi, M. Horodecki, P. Horodecki, and L. C. Kwek, Phys. Rev. Lett. **88**, 217901 (2002), arXiv:quant-ph/0203016.

[35] P. Horodecki and A. Ekert, Phys. Rev. Lett. **89**, 127902 (2002), arXiv:quant-ph/0111064.

[36] Y. Wang, Y. Li, Z.-q. Yin, and B. Zeng, npj Quantum Inf **4**, 46 (2018), arXiv:1801.03782.

[37] S. Johri, D. S. Steiger, and M. Troyer, Phys. Rev. B **96**, 195136 (2017), arXiv:1707.07658.

[38] Y. Quek, M. M. Wilde, and E. Kaur, Multivariate trace estimation in constant quantum depth (2022), arXiv:2206.15405 [hep-th, physics:quant-ph].

[39] For fidelity witness, promise that the state is either (1) fidelity $\|\rho_{\mathrm{bi}} - |\psi_{\mathrm{tar}}\rangle\langle\psi_{\mathrm{tar}}|\| \geq \alpha$; (2) fidelity $< \alpha$ implies $\rho \in \mathcal{S}_{\mathrm{bi}}$.

[40] T. Heinosaari and M. Ziman, *The Mathematical Language of Quantum Theory: From Uncertainty to Entanglement*, 1st ed. (Cambridge University Press, 2011).

[41] B. M. Terhal, Physics Letters A **271**, 319 (2000), arXiv:quant-ph/9911057.

[42] The CHSH inequality (witness): $\mathbf{O}_{\text{CHSH}} = (\mathbb{1}, ab, ab', a'b, a'b')$ with $a = Z, a' = X, b = (X - Z)/\sqrt{2}, b = (X + Z)/\sqrt{2}$ and $\mathbf{w}_{\text{CHSH}} = (\pm 2, 1, -1, 1, 1)$.

[43] A. Acin, D. Bruss, M. Lewenstein, and A. Sanpera, Phys. Rev. Lett. **87**, 040401 (2001), arXiv:quant-ph/0103025.

[44] $W_{\text{GHZ}_3} = \frac{1}{8}(3 * III - XXX - \text{Perm}(IZZ) + \text{Perm}(XYY))$ where $ZZI \equiv Z \otimes Z \otimes I$ and $\text{Perm}(IZZ) \equiv ZZI + ZIZ + IZZ$ for readability.

[45] Y. Zhou, Q. Zhao, X. Yuan, and X. Ma, npj Quantum Inf **5**, 83 (2019).

[46] H. Lu, Q. Zhao, Z.-D. Li, X.-F. Yin, X. Yuan, J.-C. Hung, L.-K. Chen, L. Li, N.-L. Liu, C.-Z. Peng, Y.-C. Liang, X. Ma, Y.-A. Chen, and J.-W. Pan, Phys. Rev. X **8**, 021072 (2018).

[47] Y. Zhou, B. Xiao, M.-D. Li, Q. Zhao, Z.-S. Yuan, X. Ma, and J.-W. Pan, npj Quantum Inf **8**, 1 (2022).

[48] M. Weilenmann, B. Dive, D. Trillo, E. A. Aguilar, and M. Navascués, Phys. Rev. Lett. **124**, 200502 (2020), arXiv:1912.10056 [quant-ph].

[49] X.-M. Hu, W.-B. Xing, Y. Guo, M. Weilenmann, E. A. Aguilar, X. Gao, B.-H. Liu, Y.-F. Huang, C.-F. Li, G.-C. Guo, Z. Wang, and M. Navascués, Phys. Rev. Lett. **127**, 220501 (2021).

[50] O. Gühne, Y. Mao, and X.-D. Yu, Phys. Rev. Lett. **126**, 140503 (2021), arXiv:2008.05961 [quant-ph].

[51] G. Riccardi, D. E. Jones, X.-D. Yu, O. Gühne, and B. T. Kirby, Exploring the relationship between the faithfulness and entanglement of two qubits (2021), arXiv:2102.10121 [quant-ph].

[52] Y. Zhan and H.-K. Lo, Detecting Entanglement in Unfaithful States (2021), arXiv:2010.06054 [quant-ph].

[53] Denote $O_\sigma \in \{I, X, Y, Z\}^{\otimes n}$ for a Pauli observable. Denote $\mathbf{x}_{\rho,\boldsymbol{\sigma}} := (\text{Tr}(\rho O_{\sigma_1}), \ldots, \text{Tr}(\rho O_{\sigma_M}))$ for expectations of M Pauli observables with respect to the state ρ where $\boldsymbol{\sigma} \subseteq \{I, X, Y, Z\}^n$.

[54] D. Lu, T. Xin, N. Yu, Z. Ji, J. Chen, G. Long, J. Baugh, X. Peng, B. Zeng, and R. Laflamme, Phys. Rev. Lett. **116**, 230501 (2016), arXiv:1511.00581 [quant-ph].

[55] C. Cortes and V. Vapnik, Mach Learn **20**, 273 (1995).

[56] T. Hofmann, B. Schölkopf, and A. J. Smola, Ann. Statist. **36**, 10.1214/009053607000000677 (2008).

[57] J. Altepeter, E. Jeffrey, and P. Kwiat, in *Advances In Atomic, Molecular, and Optical Physics*, Vol. 52 (Elsevier, 2005) pp. 105–159.

[58] Quantum state tomography refers to the task of recovering the density matrix of an unknown D-dimensional state ρ within error tolerence ϵ, given the ability to prepare and measure copies of ρ.

[59] Adaptive measurements are the intermediate between independent measurements and collective (entangled) measurements, in which the copies of ρ are measured individually, but the choice of measurement basis can change in response to earlier measurements.

[60] $\tilde{O}$ hides a polylog factor. A full tomography require estimate D^2 measuremments (observables) with additive error $\epsilon \ll 1/D$ for all E_i, so the sample complexity of shadow tomography is compatiable with lower bounds of full quantum state tomography.

[61] S. Aaronson, in *Proc. 50th Annu. ACM SIGACT Symp. Theory Comput.*, STOC 2018 (Association for Computing Machinery, New York, NY, USA, 2018) pp. 325–338, arXiv:1711.01053.

[62] H.-Y. Huang, R. Kueng, and J. Preskill, Phys. Rev. Lett. **127**, 030503 (2021), arXiv:2103.07510 [quant-ph].

[63] S. Chen, W. Yu, P. Zeng, and S. T. Flammia, PRX Quantum **2**, 030348 (2021), arXiv:2011.09636 [quant-ph].

[64] A. Elben, R. Kueng, H.-Y. Huang, R. van Bijnen, C. Kokail, M. Dalmonte, P. Calabrese, B. Kraus, J. Preskill, P. Zoller, and B. Vermersch, Phys. Rev. Lett. **125**, 200501 (2020), arXiv:2007.06305 [cond-mat, physics:quant-ph].

[65] T. Zhang, J. Sun, X.-X. Fang, X.-M. Zhang, X. Yuan, and H. Lu, Experimental quantum state measurement with classical shadows (2021), arXiv:2106.10190 [physics, physics:quant-ph].

[66] X. Gao and L.-M. Duan, Nat Commun **8**, 662 (2017), arXiv:1701.05039 [cond-mat, physics:quant-ph].

[67] G. Torlai, G. Mazzola, J. Carrasquilla, M. Troyer, R. Melko, and G. Carleo, Nature Phys **14**, 447 (2018), arXiv:1703.05334.

[68] Y. Zhu, Y.-D. Wu, G. Bai, D.-S. Wang, Y. Wang, and G. Chiribella, Flexible learning of quantum states with generative query neural networks (2022), arXiv:2202.06804 [quant-ph].

[69] H.-Y. Huang, M. Broughton, M. Mohseni, R. Babbush, S. Boixo, H. Neven, and J. R. McClean, Nat Commun **12**, 2631 (2021), arXiv:2011.01938 [quant-ph].

[70] H.-Y. Huang, R. Kueng, and J. Preskill, Phys. Rev. Lett. **126**, 190505 (2021), arXiv:2101.02464 [quant-ph].

[71] J. R. Johansson, P. D. Nation, and F. Nori, Computer Physics Communications **184**, 1234 (2013), arXiv:1110.0573.

[72] B. Li, S. Ahmed, S. Saraogi, N. Lambert, F. Nori, A. Pitchford, and N. Shammah, Quantum **6**, 630 (2022), arXiv:2105.09902 [quant-ph].

[73] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, J. Mach. Learn. Res. **12**, 2825 (2011).

[74] A. Jacot, F. Gabriel, and C. Hongler, Neural Tangent Kernel: Convergence and Generalization in Neural Networks (2020), arXiv:1806.07572 [cs, math, stat].

[75] I. Guyon, J. Weston, S. Barnhill, and V. Vapnik, Machine Learning **46**, 389 (2002).