Towards efficient and generic entanglement detection

Jue Xu* and Qi Zhao†
(Dated: October 1, 2022)

Detection of entanglement is an indispensable step for practical quantum computation and communication. In this work, we propose an end-to-end, machine learning assisted entanglement detection protocol. In this protocol, an entanglment witness for a generic entangled state is trained by classical SVM with synthetic data which consist of classical features of quantum state and their labels. In real experiments, classical features of a state, that is expectation values of a set of Pauli measurements, are estimated by sample-efficient methods such as classical shadow.

I. INTRODUCTION

Entanglement [1] is the key ingredient of quantum computation [2], quantum communication, and quantum cryptography [3]. However, decoherence is inevitable in real-world, which means the interaction between a quantum system and classical environment would greatly 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 experiments. Machine techniques have been widely applied to classification tasks in physics [][].

The goal of this paper is to find an efficient and generic way to achieve it. Assume we would like to distinguish an entangled state incluing its 'vicinity' from all separable states, our method derive such a classifier by fitting a dataset sampled from target states with their labels (entangled or not). Specifically, our pipeline starts from evaluation of expectations of a subset of all possible n-qubit pauli operators. The set of expectation values that serves as classical features of a quantum state, together with its label, consist of a data point. A classical machine learning classifier, SVM [4] 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 esitmated by classical shadow method [5] with affordable samples complexity.

II. PRELIMINARIES

Notation 1. If no ambiguity, we omit the tensor products between subsystems and the hats on operators for readability, e.g., $|\psi_A\rangle|\psi_B\rangle \equiv |\psi_A\rangle\otimes|\psi_B\rangle$ and $X^{(1)}Z^{(3)} \equiv \hat{X}\otimes\mathbb{1}\otimes\hat{Z}$.

* juexu@cs.umd.edu
† zhaoqi@cs.hku.hk

A. Entanglement structures

Large scale entanglement involving multiple particles maybe the main resource of quantum advantages in quantum computation and communication. Roughly, we say a quantum state is entangled if it is not fully separable.

Definition 1 (fully separable)**.** An n-particle (qubit) pure state $|\psi_f\rangle$ is fully separable if it can be written as the tensor product of subsystems $\{A_1,\ldots,A_n\}$, i.e., $|\psi_f\rangle = \bigotimes_{i=1}^n |\phi_{A_i}\rangle$. Then, a mixed state ρ_f is fully separable if it can be written as a convex combination of fully separable pure states.

The simple statement 'The state is entangled' would still allow that only two of the qubits are entangled while the rest is in a product state. Firstly, we consider the simplest entanglement structure: bipartite separable case. Consider a bipartite system AB with the Hilbert space $\mathcal{H}_A\otimes\mathcal{H}_B$, where $\mathcal{H}_A$ has dimension d_A and $\mathcal{H}_B$ has dimensional d_B, respectively.

Definition 2 (bipartite separable)**.** A pure state $|\psi\rangle$ is bipartite (bi-)separable if it can be written as a tensor product form $|\psi_{bi}\rangle = |\phi_A\rangle\otimes|\phi_B\rangle$, where $\mathcal{P}_2 = \left\{A, B\equiv\bar{A}\right\}$ is a bipartition of the qubits in the system. A mixed state ρ is separable if and only if it can be written as a convex combination of pure bi-separable states, i.e., $\rho_{\text{bi}} = \sum_i p_i |\psi_i\rangle\langle\psi_i|_{\text{bi}}$ with probability distribution $\{p_i\}$. The set of all bi-separable states is denoted as S_{bi}.

On the other hand, if a state is not a convex combination of any (partition) biseparable states, it means that all qubits in the system are indeed entangled with each other. This is the strongest form of entanglement, called genuine multipartite entanglement, formally defined as

Definition 3 (genuine multipartite entanglement)**.** If a state is not in S_{bi}, it possesses genuine multipartite entanglement (GME).

There is another restricted way for the extension to mixed states. In contrast to S_{bi}, a state is $\mathcal{P}_2$-separable, if it is a mixing of pure separable states with a same partition $\mathcal{P}_2$, and we denote the state set as $S_{\text{bi}}^{\mathcal{P}_2}$.

For multipartite quantum systems, it is crucial to identify not only the presence of entanglement but also its detailed structure. An identification of the entanglement structure may thus provide us with a hint about

where imperfections in the setup may occur, as well as where we can identify groups of subsystems that can still exhibit strong quantum-information processing capabilities. Given a n-qubit quantum system and its partition into m subsystems, the *entanglement structure* indicates how the subsystems are entangled with each other. In some specific systems, such as distributed quantum computing[] quantum networks[] or atoms in a lattice, the geometric configuration can naturally determine the system partition. Therefore, it is practically interesting to study entanglement structure under partitions.

Definition 4 (fully entangled)**.** A state ρ is a fully entangled if it is outside of the separable state set $S_{\text{bi}}^{\mathcal{P}_2}$ for any partition, that is, $\forall \mathcal{P}_2 = \{A, \bar{A}\}, \rho \notin S_{\text{bi}}^{\mathcal{P}_2}$.

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.

B. Entanglement detection

Many methods [6] have been developed to determine whether a state is bi-separable.

Problem 1 (separability)**.** Given a state ρ in its density matrix repsentation, to determine if it is bipartite separable.

PPT criterion

Theorem 1 (PPT criterion)**.** *If a state is bipartite separable, then it must have positive partial transpose (PPT), that is, the partially transposed (PT) density matrix $\rho_{AB}^{\intercal_A}$ is PSD [7]. By contrapositive, if the smallest eigenvalue of partial transpose $\rho_{AB}^{\intercal_A}$ is negative, then the state is entangled (cannot be bi-separable) with respect to the partition $\mathcal{P} = \{A, B\}$. Note that PPT is a necessary and sufficient condition for separability when $d_A d_B \leq 6$ [8].*

Restate in theoretic CS language Classically, the hardness of determining the bipartite separability.

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

1. Direct detection and hardness

However, up to now, no general solution for the separability problem is known. Another widely used one is the k-symmetric extension hierarchy [9], which is presently one of the most powerful criteria, but hard to compute in practice due to its exponentially growing complexity with k.[??] In order to apply the PPT criterion (the minimum eigenvalue of the partial-transposed density matrix), the full density matrix must be available. However, quantum state tomography requires an exponential number of measurements.

Similar to the PPT condition, the p_3-PPT condition applies to mixed states and is completely independent of the state in question. This is a key distinction from entanglement witnesses, which can be more powerful, but which **usually require detailed prior information about the state**. From this data set, the PT-moments p_n can be estimated without having to reconstruct the density matrix ρ_{AB}, and with a significantly smaller number of experimental runs M than required for full quantum state tomography.

Theorem 2 ([10])**.** *Weak membership problem for separability is NP-Hard for $\epsilon = 1/\operatorname{poly}(D)$ (with respect to Euclidean norm and trace norm). [11] [12] quasipolynomial-time algorithm ($\|\cdot\|_2$ and $\|\cdot\|_{LOCC}$) [13]*

Even we know the complete information about a general state, it is hard to determine its separability. However, it does not rule out the possibility to solve it efficiently with stronger promise (approximation) or quantum algorithms, even machine learning powered by data.

does not require any a priori knowledge about the quantum state. For $\operatorname{Tr}(\rho_A^m)$, an important application of multivariate trace estimation [14] is to entanglement spectroscopy [15] [16] [17] - deducing the full set of eigenvalues of ρ_A. The smallest eigenvalue of diagnoses whether ψ_{AB} is separable or entangled [18]. The well-known identity (related to the replica trick originating in spin glass theory)

$$\operatorname{Tr}(U^\pi(\rho_1 \otimes \cdots \otimes \rho_m)) = \operatorname{Tr}(\rho_1 \cdots \rho_m) \tag{1}$$

where the RHS is the multivariate trace and U^π is a unitary representation of the cyclic shift permutation. Direct entanglement detections, can be employed as subroutines in quantum computation and realized by quantum circuits because it is naturally quantum data. unknown how to generalize to multipartite.

estimate the spectrum by quantum constant depth circuits, but this method only works for bipartite. solve the quantum problem by quantum, without full tomography.

Problem 3 (Entanglement detection)**.** Different from separability problem, we assume prior knowledge of the target state $|\psi\rangle$

- **Input**: an unknown state ρ (from experiments) is promised in 'vicinity' (undergo noise channels: white noise, bit/phase-flip error, or random local unitary) of a target $|\psi\rangle_{\text{tar}}$
- **Output**: ρ possesses 'useful' entanglement (GME, full entanglement, $S_b^{\mathcal{P}}$, intactness, depth ...) or not

difficulty: multi(n)-partite, high-dimensional (qudit) [19], pure/mixed state, with/out prior knowledge, universal?, non-stabilizer [6], certain partition

Bell inequalities as the oldest tool to detect entanglement [20]. Originally, Bell inequalities were designed to rule out local hidden variable (LHV) models. CHSH inequality $\hat{\mathbf{p}} = (\mathbb{1}, ab, ab', a'b, a'b')$

$$a = Z, a' = X, b = (X - Z)/\sqrt{2}, b = (X + Z)/\sqrt{2}, \quad (2)$$

features as the input. CHSH operator $W_{\text{CHSH}} := \hat{\mathbf{p}} \cdot \mathbf{w}_{\text{CHSH}}$ with $\mathbf{w}_{\text{CHSH}} = \{\pm 2, 1, -1, 1, 1\}$. However, even for two-qubit systems there exist entangled states which do not violate any Bell inequality. [21] For example, the maximally-entangled Bell state can maximally violate the CHSH inequality, but this state mixed with certain extent white noise (2/3) don't violate the CHSH inequality despite it is still entangled.

2. *Entanglement witness based on fidelity*

Another approach for detecting multipartite entanglement is using entanglement witnesses. While various methods for constructing an entanglement witness exist, one of the most common is based on the fidelity of a state to a pure entangled state. Different Bell inequalities can be regarded as entanglement witness for different types of entanglement in a multi-party entangled state.

see Fig. 2 for relations. entanglement detection [22].

Definition 5 (entanglement witness)**.** Given a speific entangled state $|\psi\rangle_{\text{tar}}$, its entanglement witness W is an obseverable such that

$$\text{Tr}(W\rho_{bi}) \geq 0 \text{ and } \text{Tr}(W|\psi_{\text{tar}}\rangle\langle\psi_{\text{tar}}|) < 0 \quad (3)$$

In general, an entanglement witness is an observable which has a positive expectation value on all separable states, hence a negative mean value implys the presence of entanglement.

There is no entanglement witness that detects all entangled states [23].

In a typical experiment one aims to prepare a pure state, $|\psi\rangle$, and would like to detect it as true multipartite entangled. While the preparation is never perfect, it is still expected that the prepared mixed state is in the proximity of $|\psi\rangle$. The usual way to construct entanglement witnesses using the knowledge of this state is

$$W_\psi = \alpha\mathbb{1} - |\psi\rangle\langle\psi| \quad (4)$$

where α is the smallest constant such that for every product state $\text{Tr}(\rho W) \geq 0$. This kind of witness is projector-based witness [24]. However, it is generally difficult to evaluate the quantity $\text{Tr}(\rho_{\text{pre}}|\psi_{\text{tar}}\rangle\langle\psi_{\text{tar}}|)$ by the direct projection, because the target state is an entangled state. In order to measure the witness in an experiment, it is supposed be decomposed into a sum of locally measurable operators. The number of local measurements in these decompositions seems to increase exponentially with the number of qubits.[??] local measurements [25] stabilizer [6] Detecting Genuine Multipartite Entanglement with Two Local Measurements [25] graph state [26]

It is natural to ask how nonlinear entanglement witness [27] and the kernel method (nonlinear boundary) in machine learning can be applied.

3. *Beyond fidelity and stabilizer witness (robustness)*

coherent noise [28]. local unitary/rotation Unconscious phase accumulation Rotation on the first control qubit

$$|\psi(\phi, \theta)\rangle = \cos\theta\,|000\rangle + e^{i\phi}\sin\theta\,|111\rangle \quad (5)$$

the authors coined the term faithful. Weilenmann et. al [29] proposed the idea of unfaithful states which systematically (2-qudit) analyze entangled state with noise cannot be detected by fidelity witness. faithful states are useful for quantum teleportation. [30] [31] [32]

Definition 6 (unfaithful state)**.** a state ρ_{AB} is faithful if and only if there are local unitary transformations U_A and U_B such that

$$\langle\phi^+|U_A \otimes U_B \rho_{AB} U_A^\dagger \otimes U_B^\dagger|\phi^+\rangle > \frac{1}{d} \quad (6)$$

They found that for $d \geq 3$ that almost all states in the Hilbert space are unfaithful. Although there are nonlinear witnesses which also can detect entanglement in unfaithful states, they usually require more measurements [] [33]. Moreover, they can only be applied to bipartite systems, which means they cannot be generalized to detect genuine entanglement in multipartite states.

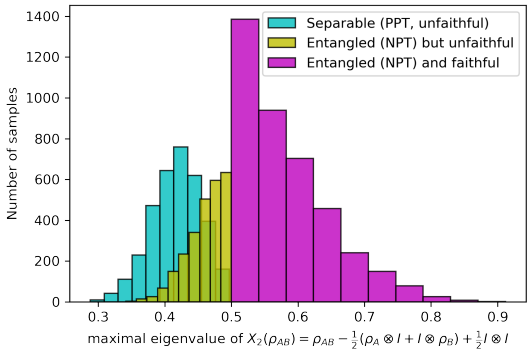

FIG. 1: Faithfulness of random 2-qubit states: ...

For non-stabilizer case, [34] [35] C is hard to compute? non-stabilizer state? SWAP?

$$\rho'_{\text{noise}} = (1 - p_{\text{noise}})\,|\psi_{\text{tar}}\rangle\langle\psi_{\text{tar}}| + p_{\text{noise}}\frac{\mathbb{1}}{2^n} \quad (7)$$

p_{noise} indicates the robustness of the algorithm (witness). the largest noise tolerance p_{limit} just related to the **chromatic number** of the graph (k local measurements)

[26].(**??**) find optimal (robustness) entanglement witness by classical machine learning (quantum circuit?) trade-off between (white noise) tolerance (robustness) and efficiency (number of measurements).

III. END-TO-END ENTANGLEMENT DETECTION PROTOCOL

Different from Entanglement detection, we assume training data

Problem 4 (Learning an entanglement witness)**.** specific entanglement structures

- **Input**: synthetic data consist of density matrices ρ with corresponding labels y
- **Output**: the classifier $\mathbf{w}_{\text{ml}}$ with high training accuracy and minimal features $\mathbf{x} :=$

The idea is to feed the classifier by a large amount of sampled trial states as well as their corresponding class labels.

A. Training a generic witness via SVM

classical machine learning (SVM, NN) with classical shadow [36]??: classify phase, predict ground state, entanglement? The quantum extension of this problem (classficiation/pattern recognition) is to replace the data points $\mathbf{x}_i$ with density matrices of quantum states ρ_i. Specifically, a quantum state classifier outputs a "label" associated with the state, for example, entangled or "unentangled". In actual experiments, we don't know entries of a density matrix. Instead, we need measurements or classical shadow as features of machine learning algorithms.

Lu et. al [37] separability classifier by classical neural network: input: sythetic random density matrices; output: a universal classifier for bipartite separable. (feature: synthetic density matrix with noise flatten as a real vector $\mathbf{x} \in \mathbb{R}^{d_A^2 d_B^2 - 1}$). For the same purpose, Ma and Yung [21] generalize Bell inequlaity and tomographic witness by neural network. an ansatz for entanglement witness

$$W_{\text{ansatz}} := \sum_{\mathbf{p}\in\{I,X,Y,Z\}^n} w_{\mathbf{p}} \bigotimes_i^n \mathbf{p}_i \tag{8}$$

c.f. quantum state tomography a linear Bell-like predictor by generalizing the CHSH operator $W_{\text{ml}} := \mathbf{P} \cdot \mathbf{w}_{\text{ml}}$ in Eq. (2) where the coefficients (or weights) $\mathbf{w}$ are determined by machine learning. tomographic ansatz better performance. training a universal classifier for multi-qubit, high-dimensional system is hard (boundary). it is difficult to generate general GME states or label general states. Tomography is necessary for universal entanglement detection with single-copy observables (non-adaptive schemes) [38]

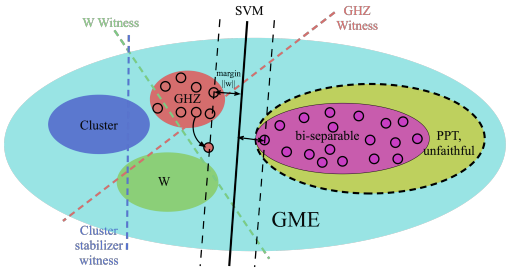

FIG. 2: (a) entanglement witness, PPT criterion, SVM (kernel)?. convex hull... The witness is depicted by the line in state space $\text{Tr}(\rho W) = 0$

classical SVM: [35] This method the ability to obtain witnesses that require only local measurements even when the target state is a **non-stabilizer state** W state (normally need nonlocal measurements). feature: $\mathbf{x}_{\vec{k}}$ expectation of Pauli strings. the training of an SVM is convex; if a solution exists for the given target state and ansatz, the optimal SVM will be found. this SVM formalism allows for the programmatic removal of features, i.e., reducing the number of experimental measurements, in exchange for a lower tolerance to white noise, in a manner similar to [??]. SVM, (universal), 4 qubit [39]. However, these SVM algorithms cannot be directly applied to experiments.

	# observables	weights	input state
fidelity witness	few local	fixed	known
Bell (CHSH) inequality	constant	fixed	unknown
tomographic classifier	$4^n - 1$	trained	unknown
SVM witness	$\ll 4^n - 1$	trained	partial

TABLE I: ansatz

An SVM allows for the construction of a hyperplane $\langle W \rangle = \sum_k w_k \mathbf{x}_k$ that clearly delineates between separable states and the target entangled state (bipartite and **tripartite qubit and qudit**); this hyperplane is a **weighted sum of observables ('features') whose coefficients are optimized during the training of the SVM**. We focus on kernel methods, as they not only provide provable guarantees, but are also very flexible in the functions they can learn. For example, recent advancements in theoretical machine learning show that training neural networks with large hidden layers is equivalent to training an ML model with a particular kernel, known as the neural tangent kernel [40].

Algorithm III.1: train witness via SVM

input : dataset, states with labels $\{(\rho_i, y_i)\}^m$
output: classifier $\mathbf{w}_{\rm ml}$

Evaluate all Pauli observables $\mathbf{x} := \mathrm{Tr}(P_{\mathbf{x}}\rho_i)\ \forall \rho_i$
for $j = 1, 2, \ldots, 4^n$ **do**
  while *accuracy not high enough* **do**
    randomly select i features $\tilde{\mathbf{x}}$ from $\mathbf{x}$
    accuracy, classifier = $\mathrm{SVM}(\{(\tilde{\mathbf{x}}_j, y_j)\}^m)$
  return classifier $\mathbf{w}_{\rm ml}$
Test the classifier on test dataset

B. Sample-efficient expectation estimation methods

To make use of classical machine learning method, we need the classical **features** of quantum states. We cannot directly process quantum data (raw data). In our pipeline, we focus on classical shadow.

The brute force approach is to fully characterize a system by performing quantum state tomography and calculating separability measures from the recovered density matrix. Intuitively, a general tomography [41] that extract (recover) all information of a state requires exponential copies (samples/measurements).

Problem 5 (quantum state tomography)**.** Informally, quantum state tomography refers to the task of estimating complete description (density matrix) of an unknown target D-dimensional quantum mixed state ρ within some error, given the ability to prepare and measure m copies $\rho^{\otimes m}$. Stokes decomposition

$$\rho = \frac{1}{2^n} \sum_{i_1, i_2, \ldots, i_n = 0}^{3} S_{i_1, i_2, \ldots, i_n} \hat{\sigma}_{i_1} \otimes \hat{\sigma}_{i_2} \otimes \cdots \otimes \hat{\sigma}_{i_n} \quad (9)$$

where $\boldsymbol{\sigma} \in \{\mathbb{1}, \hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z\}^n$.

However, full quantum state tomography is experimentally and computationally demanding; for a state consisting of n particles, with each residing in a d-dimensional Hilbert space, we would have to perform $M = \mathcal{O}(d^{2n})$ copies/measurements [42]. $\Omega(D^2)$ is required. Intermediate between independent measurements and unrestricted (also called "collective" or "entangled") measurements are adaptive measurements in which the copies of ρ are measured individually, but the choice of measurement basis can change in response to earlier measurements. Known fundamental lower bounds [66, 73] state that classical shadows of exponential size (at least) $T = \Omega(2^n/\epsilon^2)$ are required to ϵ-approximate ρ in trace distance. [36] ... the required amount of training data scales badly with ϵ. This unfortunate scaling is not a shortcoming of the considered ML algorithm, but a necessary feature.

In quantum mechanics, interesting properties are often linear functions of the underlying density matrix ρ. For example, the fidelity with a pure target state, entanglement witnesses fit this framework. Given an observable (Hermitian) O and (copies of) a mixed state ρ or several states $(\rho', \ldots, \rho_m)$, to estimate linear functions $\langle O \rangle = \mathrm{Tr}(O\rho)$, entanglement witness; nonlinear functions: entropy; multivariate functions: $\mathrm{Tr}(\rho_1 \cdots \rho_m)$, quantum kernel $\mathrm{Tr}(\rho\rho')$, quadratic $\mathrm{Tr}(O\rho_i \otimes \rho_j)$, fidelity $F(\rho, \rho')$.

Nevertheless, we usually only need specific properties of a target state rather than full classical descriptions about the state. This enables the possibility to shadow tomography [43].

Problem 6 (shadow tomography)**.** *shadow tomography*

- **Input**: copies of an unknown N-dimensional state ρ, M known 2-outcome measurements $E_1, \ldots, E_M$
- **Output**: estimate $\mathbb{P}[E_i \text{ accept } \rho] \stackrel{?}{=} \mathrm{Tr}(E_i\rho)$ to within additive error ϵ, $\forall i \in [M]$, with $\geq 2/3$ success probability.

classical shadow [5]: estimate entanglment witness (fixed but unknown target state, e.g., tripartite GHZ) **Classical shadows (Clifford measurements) of logarithmic size allow for checking a large number of potential entanglement witnesses simultaneously**. Directly measuring M different entanglement witnesses requires a number of quantum measurements that scales (at least) linearly in M. In contrast, classical shadows get by with $\log(M)$-many measurements only. classical shadows are based on random Clifford measurements and do not depend on the structure of the concrete witness in question. In contrast, direct estimation crucially depends on the concrete witness in question and may be considerably more difficult to implement.

Theorem 3 ([43])**.** *It is possible to do shadow tomography using $\tilde{\mathcal{O}}(\log^4 M \cdot \log N \cdot \epsilon^{-4})$ copies. sample complexity lower bound $\Omega(\log(M) \cdot \epsilon^{-2})$, (additive error $\epsilon \ll 1/D$).*

1. *Classical shadow and derandomized version*

Inspired by Aaronson's shadow tomography, Huang et. al [5] introduce classical shadow method that is more friendly to experiments. A classical shadow is a succinct classical description of a quantum state, which can be extracted by performing reasonably simple single-copy measurements on a reasonably small number of copies of the state. The classical shadow attempts to approximate this expectation value by an empirical average over T independent samples, much like Monte Carlo sampling approximates an integral.

Definition 7 (classical shadow)**.** Given a quantum state ρ, its classical shadow (snapshots) ρ_{cs} is

$$\rho_{cs} := \mathcal{M}^{-1}\left(U^\dagger \left|\hat{b}\right\rangle\!\left\langle\hat{b}\right| U\right) \quad (10)$$

where $|b\rangle$ is ... $U\rho U^\dagger$ such that we can predict the linear function with classical shadows

$$o_i = \mathrm{Tr}(O_i \rho_{cs}) \text{ obeys } \mathbb{E}[o] = \mathrm{Tr}(O_i \rho) \tag{11}$$

The classical shadow size required to accurately approximate all reduced r-body density matrices scales exponentially in subsystem size r, but is independent of the total number of qubits n.

Algorithm III.2: Estimate features by classical shadow

input : samples of ρ and $W_{\mathrm{ml}} := \mathbf{P} \cdot \mathbf{w}_{\mathrm{ml}}$
output: estimation of $\mathrm{Tr}(W_{\mathrm{ml}}\rho)$

for $i = 1, 2, \ldots, N$ **do**
  $\rho \mapsto U\rho U^\dagger$ // apply a random unitary
  $\mapsto |b\rangle ...$ // measurement
  $\rho_{cs} = \mathcal{M}^{-1}\left(U^\dagger |b\rangle\langle b| U\right)$ // measurement outcome
  $|b\rangle \in \{0,1\}^n$, $\mathcal{M}$ quantum channel
$S(\rho, N) = \{\rho_{cs_1}, \ldots, \rho_{cs_N}\}$ // call this array the classical shadow of ρ
return estimate $\mathrm{Tr}(W\rho)$ // estimate features for SVM

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. The number of times this procedure is repeated is called the size of the classical shadow. The transformation U is randomly selected from an ensemble of unitaries, and different ensembles lead to different versions of the procedure that have characteristic strengths and weaknesses. Classical shadows with size of order $\log(M)$ suffice to predict M target functions $\{O_1, \ldots, O_M\}$.

Theorem 4 (Pauli/Clifford measurements)**.** *Any procedure based ona fixed set of single-copy local measurements that can predict, with additive error ϵ, M arbitrary k-local linear function $\mathrm{Tr}(O_i\rho)$, requires at least (lower bound) $\Omega(\log(M)3^k/\epsilon^2)$ copies of the state ρ.* $\Omega(\log(M)\max_i \mathrm{Tr}(O_i^2)/\epsilon^2)$

Derandomization [44] can and should be viewed as a refinement of the original classical shadows idea, but not necessarily guarantees better performance for global observables. [17]

2. *Estimate expectation by machine learning*

The quantum ML algorithm accesses the quantum channel $\mathcal{E}_\rho$ multiple times to obtain multiple copies of the underlying quantum state ρ. Each access to $\mathcal{E}_\rho$ allows us to obtain one copy of ρ. Then, the quantum ML algorithm performs a sequence of measurements on the copies of ρ to accurately predict $\mathrm{Tr}(P_{\mathbf{x}}\rho), \forall \mathbf{x} \in \{I, X, Y, Z\}^n$.

Theorem 5 ([45])**.** *In contrast, for achieving accurate prediction on all inputs, we prove that* ***exponential quantum advantage is possible****. For example, to predict expectations of all* ***Pauli observables*** *(entanglement witness??) in an n-qubit system ρ, classical ML models require $2^{\Omega(n)}$ copies of ρ , but we present a quantum ML model using only $\mathcal{O}(n)$ copies. $\mathcal{O}(\log(M/\delta)\epsilon^{-4})$ copies of the unknown quantum state ρ. ($M = 4^n$ implies linear copy for full tomography???)*

[5] [36] [45] [46] generative neural network [47]

	circuit/sample complexity
shadow tomography	exp circuit? Theorem 3
classical shadow	Theorem 4
quantum/classical ML	Q. advantage Theorem 5

TABLE II: complexity (measures) of different trace estimation methods

IV. NUMERICAL SIMULATION

A. Data preparation and state generation

multi-partite entangled state: generate synthetic (engineered) data from (random graph?). separable state from randomly ...

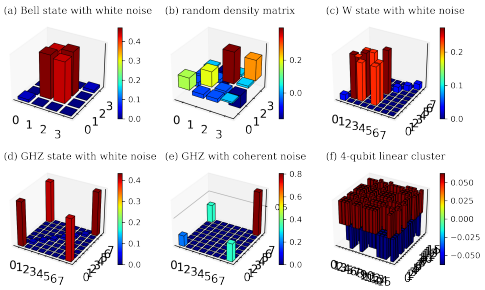

FIG. 3: Data praparation: (a) Bell state with white noise, (b) random 2-qubit density matrix, (c) 3-qubit W state with white noise (d) 3-qubit GHZ state with white noise (e) GHZ state with coherent noise (f) 4-qubit linear cluster.

QuTiP library [48]; quantum circuit [49]
entangled states generation: Bell, GHZ, W state, graph (cluster) state
separable states generation

- 2-qubit: bipartite $\rho_A \otimes \rho_B$ where ρ_A is a random density matrix sampled (Haar measure)
- white noise Eq. (7), coherent noise Eq. (5)
- Training data for $\rho_b^{(1,2,3)}$ generated by sampling over the Hilbert-Schmidt-distributed space of single-qubit and bi-qubit density matrices. As for the entangled

state, we again use the Werner state to generate the training data for that class of states.

$$\rho_{A|B|C},\ \rho_{A|BC},\ \rho_{AB|C},\ \rho_{B|AC} \tag{12}$$

B. Classification accuracy and comparison

We consider a set of different regularization parameters,...

The goal of RFE is to eliminate non-essential features by recursively considering smaller and smaller subsets of the original features using a greedy algorithm. Initially, RFE takes the SVM we trained and ranks the coefficients by their magnitudes, with the lowest one pruned away; then the model is trained again with the remaining features.

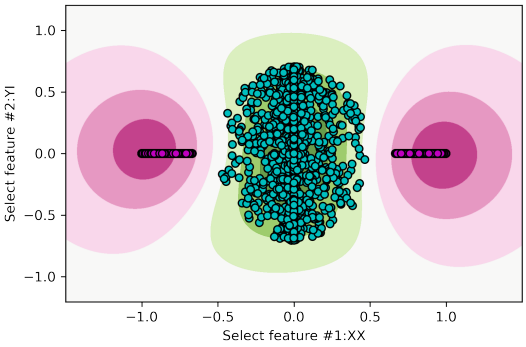

FIG. 4: two-dimensional embedding (non-linear kernel): feature space

FIG. 5: accuracies (variance) VS different data sizes

FIG. 6: number of features VS number of qubits (n). feature elimination

performance of different methods: for any state ρ_s with only bipartite entanglement, $\mathrm{Tr}(O\rho_s) \leq 0.5$, while for any state ρ_s with at most W-type entanglement, $\mathrm{Tr}(O\rho_s) \leq 0.75$. Therefore verifying that $\mathrm{Tr}(O\rho) \geq 0.5$ certifies that ρ has tripartite entanglement, while $\mathrm{Tr}(O\rho) > 0.75$ certifies that ρ has GHZ-type entanglement. [50]

V. EXPERIMENTS

Related experiments: photonic implementation with a few qubits (generation, verification) [51]; fully entangled graph state (ring of 16 qubits) IBM by measuring negativity [52]; optical lattice [53] (homogeneous, restricted

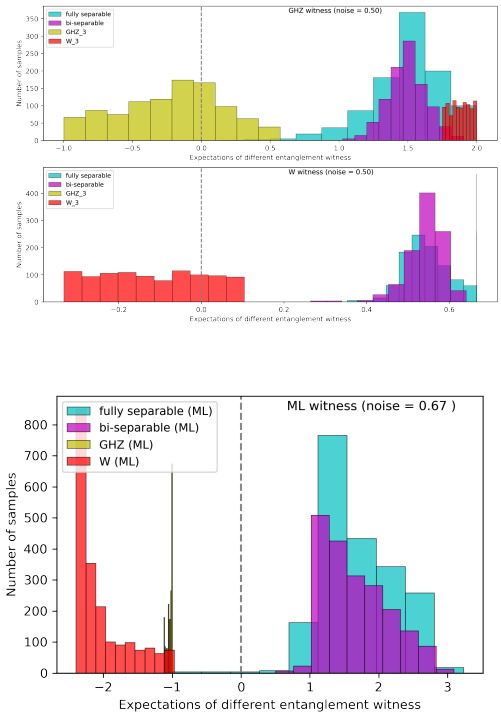

FIG. 7: (a) compare different methods: Bell inequality, witness, ML ansatz; different white noise limit, unfaithful state; (b) ML witness for unfaithful (large white noise), non-stabilizer (W) state

measurement, different noise channels; detect GME, full entanglement); general classical shadow experiments [54]; evaluate p3-PPT by classical shadow [17].

VI. CONCLUSION AND DISCUSSION

Possible future research directions: (1) rigorous proof for dataset size and number of features (required for high training accuracy) scaling with the system size; (2) better kernel options such as graph kernel, quantum kernel, shadow kernel and neural tangent kernel; (3) quantum machine learning for estimating all classical features (tomography) efficiently; (4) if we have all classical features, is it possible to train a universal classifier or with weaker promise; (5) can we estimate concurrence by quantum circuit

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

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

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

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

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

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

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

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

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

[13] F. G. S. L. Brandao, M. Christandl, and J. Yard, A quasipolynomial-time algorithm for the quantum separability problem (2011), arXiv:1011.2751 [quant-ph].

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

[15] 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.

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

[17] 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].

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

[19] S. Sciara, C. Reimer, M. Kues, P. Roztocki, A. Cino, D. J. Moss, L. Caspani, W. J. Munro, and R. Morandotti, Phys. Rev. Lett. **122**, 120501 (2019).

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

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

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

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

[24] 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.

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

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

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

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

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

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

[31] 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].

[32] 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).

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

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

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

[36] H.-Y. Huang, R. Kueng, G. Torlai, V. V. Albert, and J. Preskill, Provably efficient machine learning for quantum many-body problems (2021), arXiv:2106.12627 [quant-ph].

[37] 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].

[38] 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].

[39] 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].

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

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

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

[43] 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.

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

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

[46] 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].

[47] 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].
[48] J. R. Johansson, P. D. Nation, and F. Nori, Computer Physics Communications **184**, 1234 (2013), arXiv:1110.0573.
[49] B. Li, S. Ahmed, S. Saraogi, N. Lambert, F. Nori, A. Pitchford, and N. Shammah, Quantum **6**, 630 (2022), arXiv:2105.09902 [quant-ph].
[50] A. Acin, D. Bruss, M. Lewenstein, and A. Sanpera, Phys. Rev. Lett. **87**, 040401 (2001), arXiv:quant-ph/0103025.
[51] 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).
[52] Y. Wang, Y. Li, Z.-q. Yin, and B. Zeng, npj Quantum Inf **4**, 46 (2018), arXiv:1801.03782.
[53] Y. Zhou, B. Xiao, M.-D. Li, Q. Zhao, Z.-S. Yuan, X. Ma, and J.-W. Pan, npj Quantum Inf **8**, 1 (2022).
[54] 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].
[55] M. Hein, W. Dür, J. Eisert, R. Raussendorf, M. V. den Nest, and H.-J. Briegel, Entanglement in Graph States and its Applications (2006), arXiv:quant-ph/0602096.
[56] L. Bai, L. Rossi, A. Torsello, and E. R. Hancock, Pattern Recognition **48**, 344 (2015).
[57] X. Gao and L.-M. Duan, Nat Commun **8**, 662 (2017), arXiv:1701.05039 [cond-mat, physics:quant-ph].

Appendix A: Definitions

Definition 8 (density matrix)**.** A quantum (mixed) state ρ can be represented by a density matrix which is a Hermitian, PSD operator (matrix) of trace one. If the rank of ρ is 1, then the state is a pure state $\rho \equiv |\psi\rangle\langle\psi|$.

Definition 9 (POVM)**.** A positive-operator valued measurement (POVM) M consists of a set of positive operators that sum to the identity operator $\mathbb{1}$. When a measurement $M = \{E_1, \ldots, E_k\}$ is applied to a quantum state ρ, the outcome is $i \in [k]$ with probability $p_i = \mathrm{tr}(\rho E_i)$. observables ... $\mathbb{E}[x] \equiv \langle O_x \rangle := \mathrm{tr}(O_x \rho)$

Definition 10 (PSD)**.** A matrix (operator) is positive, semidefinite (PSD) if all its eigenvalues are non-negative.

Definition 11 (partial trace)**.** reduced density matrix $\rho_A = \mathrm{Tr}_B(\rho_{AB})$

Definition 12 (partial transpose)**.** [8] The partial transpose (PT) operation - acting on subsystem A - is defined as

$$|k_A, k_B\rangle\langle l_A, l_B|^{\intercal_A} := |l_A, k_B\rangle\langle k_A, l_B| \tag{A1}$$

where $\{|k_A, k_B\rangle\}$ is a product basis of the joint system $\mathcal{H}_{AB}$.

Definition 13 (Schmidt measure)**.** Consider the following bipartite pure state, written in Schmidt form:

$$|\psi\rangle = \sum_i^r \sqrt{\lambda_i}\, |\phi_i^A\rangle \otimes |\phi_i^B\rangle \tag{A2}$$

where $\{|\phi_i^A\rangle\}$ is a basis for $\mathcal{H}_A$ and $\{|\phi_i^A\rangle\}$ for $\mathcal{H}_B$. The strictly positive values $\sqrt{\lambda_i}$ in the Schmidt decomposition are its *Schmidt coefficients*. The number of Schmidt coefficients, counted with multiplicity, is called its *Schmidt rank*, or Schmidt number. (Schmidt rank ?? $\mathrm{SR}^A(\psi) = \mathrm{rank}(\rho_\psi^A)$) Schmidt measure is minimum of $\log_2 r$ where r is number of terms in an expansion of the state in product basis.

Definition 14 (entropy)**.** In quantum mechanics (information), the von Neumann *entropy* of a density matrix is $H_N(\rho) := -\mathrm{Tr}(\rho \log \rho) = -\sum_i \lambda_i \log(\lambda_i)$; In classical information (statistical) theory, the Shannon entropy of a probability distribution P is $H_S(P) := -\sum_i P(x_i) \log P(x_i)$.

Definition 15 (entanglement entropy)**.** The bipartite *von Neumann entanglement entropy* S is defined as the von Neumann entropy of either of its reduced density matrix ρ_A. For a pure state $\rho_{AB} = |\Psi\rangle\langle\Psi|_{AB}$, it is given by

$$E(\Psi_{AB}) = S(\rho_A) = -\mathrm{Tr}(\rho_A \log \rho_A) = -\mathrm{Tr}(\rho_B \log \rho_B) = S(\rho_B) \tag{A3}$$

where $\rho_A = \mathrm{Tr}_B(\rho_{AB})$ and $\rho_B = \mathrm{Tr}_A(\rho_{AB})$ are the reduced density matrices for each partition. With Schmidt decomposition (Eq. (A2)), the entropy of entanglement is simply $-\sum_i p_i^2 \log(p_i)$. the nth Renyi entropy, $S_n = \frac{1}{n-1}\log(R_n)$ where $R_n = \mathrm{Tr}(\rho_A^n)$

Example 1. The Schmidt measure for any multi-partite GHZ states is 1, because there are just two terms. Schmidt measure for 1D, 2D, 3D-cluster state is $\lfloor \frac{N}{2} \rfloor$. Schmidt measure of tree is the size of its minimal vertex cover[??]. other entanglement measures...

Definition 16 (fidelity)**.** Given a pair of states (target ρ and prepared ρ'), Uhlmann fidelity $F(\rho,\rho') := \mathrm{Tr}\big(\sqrt{\sqrt{\rho}\rho'\sqrt{\rho}}\big) \equiv \big\|\sqrt{\rho}\sqrt{\rho'}\big\|_1$, where $\sqrt{\rho}$ dentoes the positive semidefinite square root of the operator ρ. (infidelity $1-F(\rho,\rho')$) For any mixed state ρ and pure state $|\psi\rangle$, $F(\rho,|\psi\rangle\langle\psi|) = \sqrt{\langle\psi|\rho|\psi\rangle} \equiv \sqrt{\mathrm{Tr}(\rho\,|\psi\rangle\langle\psi|)}$ which can be obtained by the Swap-test[?]. linear fidelity or overlap $F(\rho,\rho') := \mathrm{tr}(\rho\rho')$.

Notation 2 (norm)**.** Schatten p-norm $\|x\|_p := (\sum_i |x_i|^p)^{1/p}$. Euclidean norm l_2 norm; Spectral (operator) norm $\|\mathbf{x}\|_\infty$; Trace norm $\|A\|_{\mathrm{Tr}} \equiv \|A\|_1 := \mathrm{Tr}(|A|) \equiv \mathrm{Tr}\Big(\sqrt{A^\dagger A}\Big)$, $|A| := \sqrt{A^\dagger A}$, $p=1$; Frobenius norm $\|A\|_F := \sqrt{\mathrm{Tr}(A^\dagger A)}$, $p=2$; Hilbert-Schmidt norm $\|A\|_{HS} := \sqrt{\sum_{i,j} A_{ij}^2} =? \sum_{i\in I}\|Ae_i\|_H^2$; Hilbert-Schmidt inner product $\langle A,B\rangle_{\mathrm{HS}} := \mathrm{Tr}\big(A^\dagger B\big)$, Frobenius inner product $\langle A,B\rangle_{\mathrm{F}} := \mathrm{Tr}\big(A^\dagger B\big)$? (in finite-dimensionala Euclidean space, the HS norm is identical to the Frobenius norm) Although the Hilbert-Schmidt distance is arguably not too meaningful, operationally, one can use Cauchy-Schwarz to relate it to the very natural trace distance. shadow norm ...

Definition 17 (distance)**.** For mixed states, trace distance $d_{\mathrm{tr}}(\rho,\rho') := \frac{1}{2}\|\rho-\rho'\|_1$. For pure states, $d_{\mathrm{tr}}(|\psi\rangle,|\psi'\rangle) := \frac{1}{2}\||\psi\rangle\langle\psi| - |\psi'\rangle\langle\psi'|\|_1 = \sqrt{1-|\langle\psi|\psi'\rangle|^2}$. fidelity and trace distance are related by the inequalities

$$1-F \le D_{\mathrm{tr}}(\rho,\rho') \le \sqrt{1-F^2} \tag{A4}$$

variation distance of two distribution $d_{var}(p,p') := \frac{1}{2}\sum_i |p_i - p_i'| = \frac{1}{2}\|p-p'\|_1$. l_2 distance ... Hellinger distance ... HS distance $D_{\mathrm{HS}}(\rho,\rho') := \|\rho-\rho'\|_{\mathrm{HS}} = \sqrt{\mathrm{Tr}((\rho-\rho')^2)}$

Definition 18 (stabilizer)**.** An observable S_k is a stabilizing operator of an n-qubit state $|\psi\rangle$ if the state $|\psi\rangle$ is an eigenstate of S_k with eigenvalue 1, A stabilizer set $S=\{S_1,\ldots,S_n\}$ consisting of n mutually commuting and independent stabilizer operators is called the set of stabilizer "generators".

Many highly entangled n-qubit states can be uniquely defined by n stabilizing operators which are locally measurable, i.e., they are products of Pauli matrices. A stabilizer S_i is an n-fold tensor product of n operators chosen from the one qubit Pauli operators $\{\mathbb{1},X,Y,Z\}$. An n-partite(qubit) graph state can also be uniquely determined by n independent stabilizers, $S_i := X_i\bigotimes_{j\in n} Z_j$, which commute with each other and $\forall i, S_i|G\rangle = |G\rangle$.?? The graph state is the unique eigenstate wtih eignevalue of $+1$ for all the n stabilizers. As a result, a graph state can be written as a product of stailizer projectors, $|G\rangle\langle G| = \prod_{i=1}^n \frac{S_i+\mathbb{1}}{2}$.

Example 2 (GHZ)**.** For GHZ state: $|\mathrm{GHZ}\rangle := \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} + |1\rangle^{\otimes n})$, the projector based witness

$$W_{\mathrm{GHZ}_3} = \frac{1}{2}\mathbb{1} - |\mathrm{GHZ}\rangle\langle\mathrm{GHZ}| \tag{A5}$$

requires four measurement settings. For three-qubit GHZ state [25], the local measurement witness

$$W_{\mathrm{GHZ}_3} := \frac{3}{2}\mathbb{1} - X^{(1)}X^{(2)}X^{(3)} - \frac{1}{2}\Big(Z^{(1)}Z^{(2)} + Z^{(2)}Z^{(3)} + Z^{(1)}Z^{(3)}\Big) \tag{A6}$$

This witness requires the measurement of the $\Big\{\hat\sigma_x^{(1)},\hat\sigma_x^{(2)},\hat\sigma_x^{(3)}\Big\}$ and $\Big\{\hat\sigma_z^{(1)},\hat\sigma_z^{(2)},\hat\sigma_z^{(3)}\Big\}$ settings. For n-qubit case, detect genuine n-qubit entanglement close to GHZ_n

$$W_{\mathrm{GHZ}_n} = (n-1)\mathbb{1} - \sum_{k=1}^{n} S_k(\mathrm{GHZ}_n) \tag{A7}$$

where $\hat S_k$ is the stabilizer ... [6]

graph state is an important large class of multipartite states in quantum information, because (connected) graph states represent a large class of genuine multipartite entangled states that have concise representations. Typical graph states include cluster states, GHZ states, and the states involved in error correction (toric code). It worth noting that 2D cluster (rectangular lattice graph) state is the universal resource for the measurement based quantum computation (MBQC) [2].

Definition 19 (cluster state)**.** 1D four qubits

$$\big|\psi_4^{1D}\big\rangle = \frac{1}{2}(|{+}0{+}\rangle + |{+}0{-}\rangle + |{-}1{0}+\rangle - |{-}1{-}\rangle) \tag{A8}$$

The entanglement in a graph state is related to the topology of its underlying graph [55].

Remark 1. LU, LC equivalence, local operations and classical communication (LOCC),

Definition 20 (graph state)**.** Given a simple graph (undirected, unweighted, no loop and multiple edge) $G = (V, E)$, a graph state is constructed as from the initial state $|+\rangle^{\otimes n}$ corresponding to n vertices. Then, apply controlled-Z gate to every edge, that is $|G\rangle := \prod_{(i,j)\in E} \mathsf{cZ}_{(i,j)} |+\rangle^{\otimes n}$ with $|+\rangle := (|0\rangle + |1\rangle)/\sqrt{2}$.

	$\|\mathrm{GHZ}_3\rangle$	$\|W_3\rangle$	$\|CL_3\rangle$	$\|\psi_2\rangle$	$\|\mathcal{D}_{2,4}\rangle$	$\|\mathrm{GHZ}_n\rangle$	$\|W_n\rangle$	$\|G_n\rangle$
maximal overlap α	1/2	2/3	1/2	3/4	2/3	1/2	$(n-1)/n$	1/2
maximal p_{noise}	4/7	8/21	8/15	4/15	16/45	$1/2 \cdot (1-1/2^n)^{-1}$	$1/n \cdot (1-1/2^n)^{-1}$	$1/2 \cdot (1-1/2^n)^{-1}$
# local measurements	4	5	9	15	21	$n+1$	$2n-1$	depend on graphs

TABLE III: Results on local decompositions of different entanglement witnesses for different states. [22]

Appendix B: Machine learning background

Notations: The (classical) training data (for supervised learning) is a set of m data points $\left\{(\mathbf{x}^{(i)}, y^{(i)})\right\}_{i=1}^{m}$ where each data point is a pair $(\mathbf{x}, y)$. Normally, the input (e.g., an image) $\mathbf{x} := (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ is a vector where d is the number of *features* and its *label* $y \in \Sigma$ is a scalar with some discrete set Σ of alphabet/categories. For simplicity and the purpose of this paper, we assume $\Sigma = \{-1, 1\}$ (binary classification).

1. Support vector machine

SVM is a typical supervised learning algorithm for classification. Taking the example of classifying cat/dog images, supervised learning means we are given a dataset in which every image is labeled either a cat or a dog such that we can find a function classifying new images with high accuracy. More precisely, the training dataset is a set of pairs of features X and their labels y. In the image classification case, features are obtained by transforming all pixels of an image into a vector. In SVM, we want to find a linear function, that is a hyperplane which separates cat data from dog data. So, the prediction label is given by the sign of the inner product (projection) of the hyperplane and the feature vector. We can observe that the problem setting of image classification by SVM is quite analogous to entanglement detection, where input data are quantum states now and the labels are either entangled or separable.

Definition 21 (SVM)**.** Given a set of (binary) labeled data, support vector machine (SVM) is designed to find a hyperplane (a linear function) such that maximize the margin between two partitions...

$$\max_{\mathbf{w}} \|\mathbf{w}\|^2 \text{ s.t. } \forall i, y^{(i)} \cdot (\mathbf{w} \cdot \mathbf{x} + b) \geq 1. \tag{B1}$$

Lagrange multipliers α

$$L = \frac{1}{2}\|\mathbf{w}\|^2 - \sum_i^m \alpha^{(i)} \left(\mathbf{w} \cdot \mathbf{x}^{(i)} + b\right) + \sum_i^m \alpha^{(i)} \tag{B2}$$

a. kernel method

However, note that SVM is only a linear classifier. while most real-world data, such as cat/dog images and entangled/separable quantum states are not linearly separable. For example, with this two dimension dataset, we are unable to find a hyperplane to separate red points from the purple points very well. Fortunately, there is a very useful tool called kernel method or kernel trick to remedy this drawback. The main idea is mapping the features to a higher dimensional space such that they can be linearly separated in the high dimensional feature space. Just like this example, two dimensional data are mapped to the three dimensional space. Now, we can easily find the separating plane. With SVM and kernel methods, we expect to find a generic and flexible way for entanglement detection. kernel

Definition 22 (kernel)**.** In general, the kernel function $k : \mathcal{X} \times \mathcal{X} \to \mathbb{R}$ measures the similarity between two input data points by an inner product

$$k(\mathbf{x}, \mathbf{x}') := \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle \tag{B3}$$

If the input $\mathbf{x} \in \mathbb{R}^d$ (conventional machine learning task, e.g., image classification), the feature map $\phi(\mathbf{x}) : \mathbb{R}^d \to \mathbb{R}^n$ $(d < n)$ from a low dimensional space to a higher dimensional space. The corresponding kernel (Gram) matrix $\mathbf{K}$ is PSD.

Example 3 (kernels)**.** Some common kernels: the polynomial kernel $k_{\text{poly}}(\mathbf{x}, \mathbf{x}') := (1 + \mathbf{x} \cdot \mathbf{x}')^q$ with feature map $\phi(\mathbf{x})$ ... The Gaussian kernel $k_{\text{gaus}}(\mathbf{x}, \mathbf{x}') := \exp\left(-\gamma \|\mathbf{x} - \mathbf{x}'\|_2^2\right)$ with an infinite dimensional feature map $\phi(\mathbf{x})$. An important feature of kernel method is that kernels can be computed efficiently without evaluating feature map (might be infinite dimension) explicitly.

Definition 23 (graph kernel)**.** given a pair of graphs (G, G'), *graph kernel* is $k(G, G') =$. quantum graph kernel $k(G, G') = |\langle G|G' \rangle|^2$?? [56]

b. *Quantum kernel*

Definition 24 (quantum kernel)**.** quantum kernel with quantum feature map $\phi(\mathbf{x}) : \mathcal{X} \to |\phi(\mathbf{x})\rangle\langle\phi(\mathbf{x})|$

$$k_Q(\rho, \rho') := |\langle \phi(\mathbf{x}) | \phi(\mathbf{x}') \rangle|^2 = \left| \langle 0 | U_{\phi(\mathbf{x})}^\dagger U_{\phi(\mathbf{x}')} | 0 \rangle \right|^2 =? \operatorname{Tr}(\rho\rho') \equiv \langle \rho, \rho' \rangle_{\text{HS}} \tag{B4}$$

where $U_{\phi(\mathbf{x})}$ is a quantum circuit or physics process that encoding an input $\mathbf{x}$. In quantum physics, quantum kernel is also known as transition amplitude (quantum propagator);

c. *neural network and kernel*

Definition 25 (neural tagent kernel)**.** neural tangent kernel [40]: proved to be equivalent to deep neural network [57] in the limit ...

$$k_{\text{NT}}\left(S_T(\rho_l), \tilde{S}_T(\rho_{l'})\right) = \left\langle \phi^{(\text{NT})}(S_T(\rho_l)), \phi^{(\text{NT})}(\tilde{S}_T(\rho_l)) \right\rangle \tag{B5}$$