Towards efficient and generic entanglement detection

Jue Xu* and Qi Zhao†

(Dated: September 2, 2022)

Detection of entanglement structure is an indispensable step for practical quantum computation and communication. In this work, we compare complexity and performance of several recently-developed methods, including entanglement witness methods, shadow tomography, classical machine learning, and quantum algorithms (circuits). We illustrate the advantages and limitations of machine learning and quantum algorithms.

CONTENTS

I. INTRODUCTION

Entanglement [1] is the key ingredient of quantum computation [], quantum communication [], and quantum cryptography []. For practical purpose, it is essential to benckmark (characterize) multipartite entanglement structures of target states. We review the recently developed methods to entanglement detection: entanglement witness [2], shadow tomography [3], classical machine learning [4], and quantum (variational/circuit) algorithms [5].

* juexu@cs.umd.edu
† email

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, detecting or benchmarking entanglement is essential in actual experiments. Our goal is to find an efficient and generic way to do it. Roughly speaking, our solution is to make use of both machine learning techniques and some recently-developed quantum algorithms. Explicitly, our pipeline starts with a tomographic entanglement witness ansatz which is the linear combination of all possible n-qubit pauli operators. Next, we generate random density matrices with labels for training. Then, we need to estimate the expectation values of each pauli operator, which are features for training a classical SVM. Probably, we can eliminate unimportant features, which means we might not need all pauli operators in our generic witness. Finally, we test and make predictions with brand new samples. We would test our pipeline in actual experiments.

II. PRELIMINARIES

Notations: The hats on the matrices such as $\hat{A}$, $\hat{H}$, ρ (omitted), $\hat{O}$, $\hat{W}$, emphasize that they play the roles of operators (Hermitian matrices). Denote vector (matrix) $\mathbf{x}$, $\mathbf{K}$ by boldface font. A simle (undirected, unweighted) graph $G = (V, E)$ is described by vertices V and edges E.

For specific purpose, we use different basis (representations) for quantum states. One is the computational basis $\{|z\rangle\}$ with $z \in [2^n]$ where n is the number of qubits, while another useful one is the binary representation of computational basis $\{|\mathbf{x}\rangle \equiv |x_1\rangle |x_2\rangle \dots |x_n\rangle\}$ with $x_j \in \{0, 1\}$. For simplicity, we let $N \equiv 2^n$ and $|\mathbf{0}\rangle \equiv |0^n\rangle \equiv |0\rangle^{\otimes n}$ if no ambiguity. shorthand $|\psi_A\rangle |\psi_B\rangle \equiv |\psi_A\rangle \otimes |\psi_B\rangle$ (we omitted the tensor product for readability).

A. Entanglement structures

1. *Bipartite entanglement*

Large scale entanglement is the (main) resource of quantum advantages in quantum computation and communication. Firstly, we consider the simplest entanglement structure: bipartite separable case.

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

Definition 1 (bipartite separable)**.** A pure state is (bi-)separable if it is in a tensor product form $|\psi_b\rangle = |\phi_A\rangle \otimes |\phi_{\bar{A}}\rangle$, where $\mathcal{P}_2 = \left\{A, \bar{A}\right\}$ is a bipartition of the qubits in the system. Note that the state $|\phi_A\rangle$ may be entangled, thus the state $|\psi\rangle$ is not necessarily fully separable. A mixed state is separable iff it can be written as a convex combination of pure biseparable states $\rho = \sum_i p_i |\psi_i\rangle\langle\psi_i|$ where $|\psi_i\rangle$ may be biseparable with respect to different partitions (otherwise genuine multipartite entanglement). 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.

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. A state ρ_{AB} is *separable* if it can be writen as a convex combination $\rho_{AB} = \sum_i \lambda_i \rho_{A,i} \otimes \rho_{B,i}$ with a probability distribution $\lambda_i \geq 0$ and $\sum_i \lambda_i = 1$. Otherwise, ρ_{AB} is entangled.

Note that each separable state $|\psi_b\rangle$ in the summation can have different bipartitions. The separable state set is denoted as S_b. There is another restricted way for the extension to mixed states. 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 $S_b^{\mathcal{P}_2}$. entangled state?...

Rather than qualitatively determining (bi)separability, there are measures to quantify entanglement

Definition 2 (Schmidt measure)**.** Consider the following bipartite pure state, written in Schmidt form:

$$|\psi\rangle = \sum_i^r \sqrt{\lambda_i} \left|\phi_i^A\right\rangle \otimes \left|\phi_i^B\right\rangle \tag{1}$$

where $\left\{\left|\phi_i^A\right\rangle\right\}$ is a basis for $\mathcal{H}_A$ and $\left\{\left|\phi_i^A\right\rangle\right\}$ 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.

Remark 1. a pure (bipartite) state is entangled iff the reduced state $\rho^A = \mathrm{Tr}_A(\rho)$ is mixed. The mixedness of this reduced state allows one to quantify the amount of entanglement in this state.

Classically, the hardness of determining the bipartite separability.

Theorem 1 ([6]). *The weak membership problem for the convex set of separable normalized bipartite density matrices is* NP*-Hard.* ***Input****: unknown state?? formal definition of the problem [7]*

However, we do not know approximately correct complexity? quantum complexity? machine learning (data)? for entanglement () problem? multipartite?

Theorem 2 (PPT criterion [8]). *The positive partial transpose (*PPT*) criterion saying the state is entangled iff the smallest eigenvalue of partial transpose $\rho_{AB}^{\intercal_A}$ is negative. So, a separable state (bipartite separable) must have PPT (the partially transposed (PT) density matrix $\rho_{AB}^{\intercal_A}$ is positive, semidefinite). Note, it is only necessary and sufficient when $d_A d_B \leq 6$.*

Another widely used one is the k-symmetric extension hierarchy [15, 16], 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, full tomography requires an exponential number of measurements.
the minimal eigenvalue (the absolution of this value for entangled states is named as negativity) , can be obtained analytically

$$\lambda_{\min}\left(\rho_{\theta,\phi}^{\intercal_B}\right) = (1-p)/4 - p\cos(\theta/2)\sin(\theta/2) \tag{2}$$

[9]
However, up to now, no general solution for the separability problem is known. 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. c.f. [5], entanglement spectroscopy

2. Multipartite entanglement structures

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 3 (fully separable). An n-qubit pure state $|\psi_f\rangle$ is *fully separable* iff(, and is (fully) n-separable if it is in S_n). An n-qubit pure state $|\psi_f\rangle$ is $\mathcal{P}$-fully separable iff it can be written as $|\psi_f\rangle = \otimes_i^m |\phi_{A_i}\rangle$. An n-qubit mixed state ρ_f is $\mathcal{P}$-fully separable iff it can be decomposed into a convex mixture of $\mathcal{P}$-fully separable pure states.

$$\rho_f = \sum_i p_i \left|\psi_f^i\right\rangle\!\left\langle\psi_f^i\right|, (\forall i)(p_i \geq 0, \sum_i p_i = 1). \tag{3}$$

P-bi-separable... $S_f^{\mathcal{P}} \subset S_b^{\mathcal{P}}$

define fully- and biseparable states with respect to a *specific partition* $\mathcal{P}_m$

Definition 4 (fully entangled). An n-qubit quantum state ρ is a *fully entangled*, if it is outside of the separable state set $S_b^{\mathcal{P}_2}$ for any partition, $\rho \notin S_b^{\mathcal{P}_2}, \forall \mathcal{P}_2 = \left\{A, \bar{A}\right\}$. is fully entangled if it is neither biseparable nor fully separable.

GME is the strongest form of entanglement, that is, all qubits in the system are indeed entangled with each other. The size of the genuinely entangled quantum system becomes a figure of merit for assessing the advancement of quantum devices in the competition among various realizations.

Definition 5 (genuine multipartite entanglement). A state possesses *genuine multipartite entanglement* (GME) if it is outside of S_2. A state possesses $\mathcal{P}$-genuine entanglement if it is outside of $S_b^{\mathcal{P}}$. A state ρ possesses $\mathcal{P}$-genuine entanglement iff $\rho \notin S_b^{\mathcal{P}}$.

Since $S_b^{\mathcal{P}_2} \subset S_b$, GME is a stronger claim than full entanglement. For a state with full entanglement, it is possible to prepare it by mixing bi-separable states with different bipartitions Compared with genuine entanglement, multipartite entanglement structure still lacks a systematic exploration, due to the rich and complex structures of n-partite system. Unfortunately, it remains an open problem of efficient entanglement-structure detection of general multipartite quantum states.

By going through all possible partitions, one can investigate higher level entanglement structures, such as entanglement intactness (non-separability), which quantifies how many pieces in the n-partite state are separated.

Remark 2. P-... can be viewed as generalized versions of regular fully separable, biseparable, and genuinely entangled states, respectively. In fact, when $m = n$, these pairs of definitions are the same. By definitions, one can see that if a state is P_m-fully separable, it must be m-separable. Of course, an m-separable state might not be P_m-fully separable, for example, if the partition is not properly chosen.

entanglement structure measures. To benchmark our technological progress towards the generation of largescale genuine multipartite entanglement, it is thus essential to determine the corresponding entanglement depth.

Definition 6 (Entanglement intactness, depth)**.** the entanglement intactness of a state ρ to be m, iff$\rho \notin S_{m+1}$ and $\rho \in S_m$. When the entanglement intactness is 1, the state possesses genuine multipartite entanglement; and when the intactness is n, the state is fully separable. k-producible.

Example 1 (GHZ)**.** bipartite: Bell states; nontrivial multipartite: tripartite. Greenberger-Horne-Zeilinger (GHZ) state: $|\mathrm{GHZ}\rangle := \frac{1}{\sqrt{2}}(|0\rangle^{\otimes n} + |1\rangle^{\otimes n})$ (eight-photon) produce the five different entangled states (one from each entanglement structure/partition?):

$$|\mathrm{GHZ}_8\rangle\,, |\mathrm{GHZ}_{62}\rangle\,, |\mathrm{GHZ}_{44}\rangle\,, |\mathrm{GHZ}_{422}\rangle\,, |\mathrm{GHZ}_{2222}\rangle\,.$$

The GHZ-state is generally considered as the state with the genuine 3-partite entanglement, while the W-state has the peculiar property of having the maximal expected amount of two-partite entanglement if one party is traced out.

Schmidt rank, PPT criteria, entanglement witness...

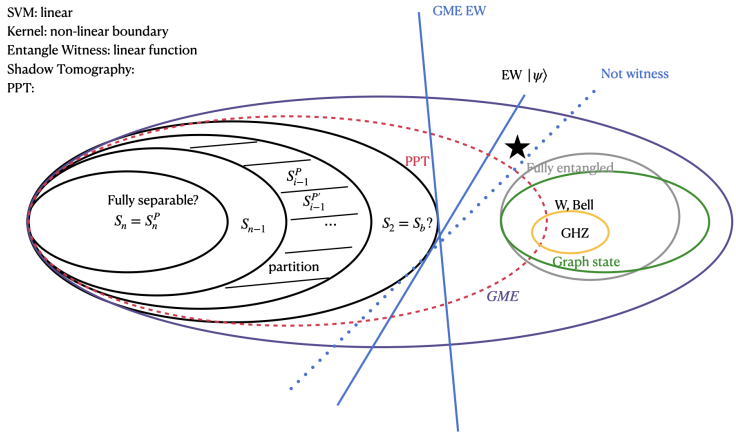


FIG. 1: (a) entanglement witness, PPT criterion [8], SVM (kernel)?. convex hull...

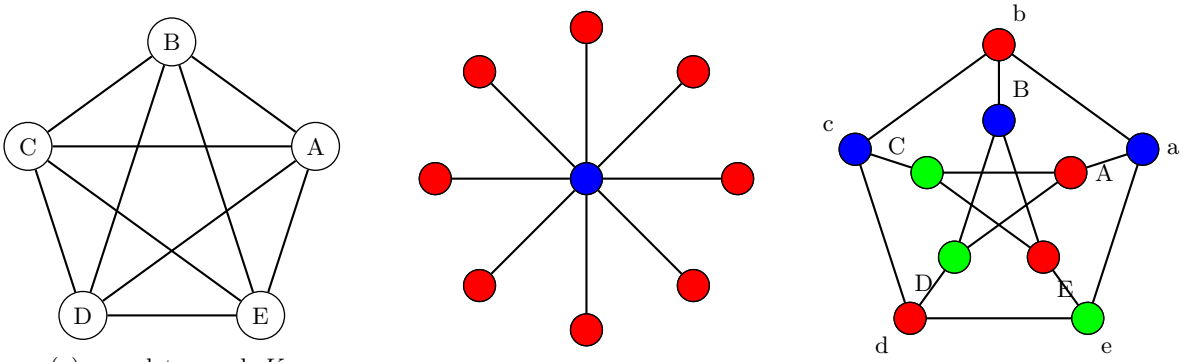
3. *Graph state*

graph state is an important (large?) class of multipartite states in quantum information. Typical graph states include cluster (lattice) states, GHZ states, and the states involved in error correction (toric code?). It worth noting that 2D cluster state is the universal resource for the measurement based quantum computation (MBQC) [10].

Definition 7 (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

$$|G\rangle := \prod_{(i,j)\in E} \mathsf{cZ}_{(i,j)} |+\rangle^{\otimes n}\,, \quad |+\rangle := (|0\rangle + |1\rangle)/\sqrt{2}. \tag{4}$$

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, $|G\rangle\langle G| = \prod_{i=1}^{n} \frac{S_i+\mathbb{1}}{2}$. Stabilizer formalism the graph states themselves represent already a large class of genuine multipartite



(a) complete graph K_5 (b) star graph (2-colorable) - GHZ (c) Petersen graph

FIG. 2: graph states

entangled states that are relatively easy to survey even in the regime of many parties.

Example 2 (graph states)**.** Any connected graph state is fully entangled state. The GHZ state corresponds to the star graph and the complete gtaph (Fig. 2). This is easily seen by applying Hadamard unitaries $\hat{U}_{\mathrm{H}}^{V\setminus a}$ to all but one qubit a in the GHZ-state, which yields the star graph state with a as the central qubit. (line, ring; hypercube, Petersen graph; cluster state in two dimensinos, which corresponds to a rectangular lattice.) The Petersen graph is not LC-equivalent to its isomorphism (exchanging the labels at each ed of the five "spokes"). However, the lists of Schmidt ranks (or, equivalently, the connectivity functions) of these graphs coincide. The class of CSS (error correction) states corresponds to the class of 2-colorable graphs. [11]

Definition 8 (cluster state)**.**

Remark 3. LU, LC equivalence, local operations and classical communication (LOCC)

the entanglement in a graph state is related to the topology of its underlying graph.

Remark 4 (??)**.** The entanglement entropy $S(\rho_A)$ equals the rank of the adjacency matrix of the underlying bipartite graph, which can be efficiently calculated. For graph states, the reduced density matrices can be represented efficiently in terms of their stabilizer elements or their adjacency matrix.

Proposition 1 (Bounds to the Schmidt measure of graph states)**.** *For any graph state $|G\rangle$, the Schmidt measure E_A is bounded from below by the maximal Schmidt rank* $\mathrm{SR}_{\max}$ *and from above by the Pauli persistency PP or the minimal vertex cover, i.e.*

$$\mathrm{SR}_{\max}(G) \leq E_S(|G\rangle) \leq \mathrm{PP}(G) \leq \mathrm{VC}(G). \tag{5}$$

???

B. Entanglement detection

PPT criterion [8]

1. *Bell inequality witness*

A usual approach for detecting entanglement is using Bell inequalities [??]

Definition 9 (Bell inequality)**.** Bell inequalities for graph states $\left|\sum_{\sigma\in S} \langle\sigma\rangle\right| \leq C$?...

Definition 10 (CHSH inequality)**.** CHSH inequality (game) ...;

$$\{\langle a_0 b_0\rangle, \langle a_0 b_0'\rangle, \langle a_0' b_0\rangle, \langle a_0' b_0'\rangle\} \tag{6}$$

features as the input

$$a_0 = \hat{\sigma}_z, a_0' = \hat{\sigma}_x, b_0 = (\hat{\sigma}_x - \hat{\sigma}_z)/\sqrt{2}, b_0 = (\hat{\sigma}_x + \hat{\sigma}_z)/\sqrt{2}, \tag{7}$$

a linear Bell-like predictor by generalizing the CHSH operator

$$\hat{W}_{ml} := w_0 + w_1 a_0 b_0 + w_2 a_0 b_0' + w_3 a_0' b_0 + w_4 a_0' b_0' \quad , \mathbf{w} = \{\pm 2, 1, -1, 1, 1\} \tag{8}$$

where the coefficients (or weights) $\mathbf{w}$ are determined by machine learning. However, even for two-qubit systems there exist entangled states which do not violate any Bell inequality. Bell inequalities are not suited to this aim in general. Multiseparable and biseparable states violate known Bell inequalities less than n-partite GHZ states. However, for $n > 3$ there exist even pure n-partite entangled states with a lower violation than biseparable states. [12]

Example 3 (inequality)**.** [9] First, there exist entangled states not violating the Bell inequalities. To be more specific, the maximally-entangled state, such as $|\psi_0\rangle = (|00\rangle - |11\rangle)/\sqrt{2}$ for a pair of qubits, can maximally violate the CHSH inequality. However, this tool fails under the circumstances of noise, in the form of a quantum channel. After passing through a depolarizing channel, the resulting state,

$$\rho = p\,|\psi_-\rangle\langle\psi_-| + (1-p)\frac{\mathbb{1}}{4} \tag{9}$$

where $0 \le p \le 1$, violates the CHSH inequality only if $p > 1/\sqrt{2}$. However, the state is entangled when $p > 1/3$.

2. *Entanglement witness*

Another approach for detecting multipartite entanglement is using entanglement witnesses. Different Bell inequalities can be regarded as entanglement witness for different types of entanglement in a multi-party entangled state. These witnesses can be quite useful to detect entanglement in the vicinity of graph states.

Problem 1 (separability)**.** given (input) an **unknown** state, to determine (output) separability.

Problem 2 (Entanglement witness with prior)**.** Entanglement witness

- **Input**: a **known** state $|\psi\rangle$, with noise
- **Output**: separable or not ENTANGLED ??? $S_f^{\mathcal{P}}$? $S_b^{\mathcal{P}}$

difficulty: multi(n)-partite, high-dimensional (qudit) [13], pure/mixed state, with/out prior knowledge, universal?, non-stabilizer [14], certain partition

Problem 3 (Certify entanglement)**.** Multipartite entanglement-structure detection

- **Input**: an (actual) state ρ' from experiment that is close to a **known/target** (general multipartite) state $|\psi\rangle$, certain partition?
- **Output**: the certified lower-order entanglement among several subsystems could be still useful for some quantum information tasks. entanglement structure

Remark 5 (universal entanglement witness)**.** [13] For example, a witness specifically designed for a four-qubit compact cluster state [16] confirms, when its expectation value is negative, the presence of that particular state having a very specific density function, while a positive measured expectation value of that operator only provides information that the tested state is not a compact cluster state. Indeed, the same witness, if applied to a four-qubit linear cluster or GHZ [17] states, would result in a positive measured expectation value, even though these two states are both highly entangled [17, 18]. Hence, a witness is a threshold test that can only detect the presence of a specific state. In contrast to an entanglement monotone (e.g. the entanglement entropy [6]), which determines the amount of entanglement, a witness cannot be used to quantify entanglement.

see Fig. 1 for relations. entanglement detection [15].

Definition 11 (entanglement witness)**.** Given an (unknown? known target state) quantum state (density matrix) ρ, the *entanglement witness* $\hat{W}$ is an obseverable such that

$$\mathbb{E}_\rho[\hat{W}] \equiv \left\langle \hat{W} \right\rangle \equiv \mathrm{Tr}\left(\hat{W}\rho\right) \ge 0, \forall \text{ separable} ; \quad \mathrm{Tr}\left(\hat{W}\rho\right) < 0, \text{ for some entangled} \tag{10}$$

[16], see Fig. 1

3. Fidelity, projector-based witness, and stabilizer state

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

$$\hat{W}_\psi = c\mathbb{1} - |\psi\rangle\langle\psi| \tag{11}$$

where c is the smallest constant such that for every product state $\mathrm{Tr}\left(\rho\hat{W}\right) \geq 0$

Proposition 2 (Section 6.3 of [17]). *A state ρ is separable iff $\forall\hat{W}, \mathrm{Tr}\left[\rho\hat{W}\right] \geq 0$. Corollary, a state ρ is entangled iff $\exists\hat{W}, \mathrm{Tr}\left[\rho\hat{W}\right] < 0$. There is no entanglement witness that detects all entangled states.*

Tomography is necessary for universal entanglement detection with single-copy observables (non-adaptive schemes) [18]

However, it is generally difficult to evaluate the quantity $\mathrm{Tr}(\rho_{\mathrm{pre}}\,|\psi_{\mathrm{tar}}\rangle\langle\psi_{\mathrm{tar}}|)$ by the direct projection, as it is an entangled state. In order to measure the witness in an experiment, it must 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.[??]

Example 4 (entanglement witness for GHZ). three-qubit GHZ state [19]

$$\hat{W}_{\mathrm{GHZ}_3} := \frac{3}{2}\mathbb{1} - \hat{\sigma}_x^{(1)}\hat{\sigma}_x^{(2)}\hat{\sigma}_x^{(3)} - \frac{1}{2}\left[\hat{\sigma}_z^{(1)}\hat{\sigma}_z^{(2)} + \hat{\sigma}_z^{(2)}\hat{\sigma}_z^{(3)} + \hat{\sigma}_z^{(1)}\hat{\sigma}_z^{(3)}\right] \tag{12}$$

This witness requires the measurement of the $\left\{\hat{\sigma}_x^{(1)}, \hat{\sigma}_x^{(2)}, \hat{\sigma}_x^{(3)}\right\}$ and $\left\{\hat{\sigma}_z^{(1)}, \hat{\sigma}_z^{(2)}, \hat{\sigma}_z^{(3)}\right\}$ settings. The projector based witness $\hat{W}_{\mathrm{GHZ}_3} = \mathbb{1}/2 - |\mathrm{GHZ}\rangle\langle\mathrm{GHZ}|$ requires four measurement settings. detect genuine n-qubit entanglement close to GHZ_n

$$\hat{W}_{\mathrm{GHZ}_n} = (n-1)\mathbb{1} - \sum_{k=1}^{n} S_k^{(\mathrm{GHZ}_n)} \tag{13}$$

$\hat{S}_k$ is the Stabilizer ... [14] Detecting Genuine Multipartite Entanglement with Two Local Measurements [19]

It is natural to ask how nonlinear entanglement witness [20] and the kernel method (nonlinear boundary) in machine learning can be applied.

Proposition 3. *Given a state $|\psi\rangle$, the entanglement witness operator $\hat{W}_\psi$ can witness genuine multipartite entanglement near $|\psi\rangle$ with $c = 5/8$ in Eq. (11) that is, $\left\langle\hat{W}_\psi\right\rangle \geq 0$ for any separable state in S_b.*

If the fidelity (quantum kernel?) of the prepared state ρ_{pre} with the target state $|\psi\rangle$, i.e., $\mathrm{Tr}(\rho_{pre}\,|\psi\rangle\langle\psi|)$, exceeds 5/8, ρ_{pre} possesses GME. It is generally difficult to evaluate the quantity $\mathrm{Tr}(\rho_{pre}\,|\psi\rangle\langle\psi|)$ by the direct projection on $|\psi\rangle$, as it is an entangled state.

	$\|\mathrm{GHZ}_3\rangle$	$\|W_3\rangle$	$\|CL_3\rangle$	$\|\psi_2\rangle$	$\|\mathcal{D}_{2,4}\rangle$	$\|\mathrm{GHZ}_n\rangle$	$\|W_3\rangle$	$\|G_n\rangle$
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	$1/2\cdot(1-1/2^n)^{-1}$		
local measurements	4	5	9	15	21	N+1	$2N-1$	depend on graphs

TABLE I: [15]

graph state, stabilizer [2] generalize [21] stabilizer state, neural network state [22]?

Theorem 3. *k local measurements. Here, k is the chromatic number (minimal colorable) of the corresponding graph, typically, a small constant independent of the number of qubits.*

Theorem 4. *[23]*

$$\max_{\rho\in SEP}\sum_{i=1}^{n}\langle A_i\otimes B_i\rangle \le \sqrt{\theta(\bar{G}_A)\theta(\bar{G}_B)} =: \theta_{AB} \tag{14}$$

where $\theta(\bar{G}_A)$ is the Lovasz number of the anticommutativity graph of $\{A_i\}$ and similarly for $\theta(\bar{G}_B)$. This separability criterion naturally leads to entanglement witnesses of the form

$$\hat{W}_{\mathbb{E}} = \theta_{AB}\mathbb{1} - \sum_i A_i\otimes B_i \tag{15}$$

For non-stabilizer case, [21] [24] C is hard to compute? non-stabilizer state? SWAP?
[25, 26]

Definition 12 (unfaithful state)**.**

C. Tomography and trace estimation

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 [27] that extract (recover) all information of a state requires exponential copies (samples/measurements).

Problem 4 (full tomography)**.** In contrast to shadow tomography, we refer to *full tomography* here

- **Input**: Given a **unknown** N-dimensional mixed state ρ
- **Output**: a complete description? of ρ (decomposition coefficients) with error? Stokes parameter $S_i \equiv \mathrm{Tr}(\hat{\sigma}_i\rho)$

$$\rho = \frac{1}{2^n}\sum_{i_1,i_2,\ldots,i_n=0}^{3} S_{i_1,i_2,\ldots,i_n}\hat{\sigma}_{i_1}\otimes\hat{\sigma}_{i_2}\otimes\cdots\otimes\hat{\sigma}_{i_n},\ \boldsymbol{\sigma}\in\{\mathbb{1},\hat{\sigma}_x,\hat{\sigma}_y,\hat{\sigma}_z\}^n \tag{16}$$

However, 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 $M=\mathcal{O}(d^{2N})$ measurements.

Theorem 5 (lower bound of full tomography?[28])**.** *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.*

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.

Problem 5 (trace estimation)**.** related problems defined as follows

- **Input**: Given an observable (Hermitian) $\hat{O}$ and (copies of) a mixed state ρ or several states $(\rho',\ldots,\rho_m)$,
- **Output**: with error ϵ measured by trace distance (fidelity...)
  - linear function (mostly): the expectation value $\left\langle\hat{O}\right\rangle = \mathrm{Tr}\left(\hat{O}\rho\right)$; entanglement witness $\mathrm{Tr}\left(\hat{W}\rho\right)$; shadow tomography $\mathrm{Tr}(E_M\rho)=\mathbb{E}[E_M]=\mathbb{P}[E_i \text{ accept } \rho]$??; full tomography; $\mathrm{Tr}\left(\hat{O}\mathcal{E}(|\psi\rangle\langle\psi|)\right)$ where $\mathcal{E}$ is a CPTP (completely positive and trace preserving) map two-point correlation $\langle O_iO_j\rangle$
  - nonlinear function: entropy (non-linear); quadratic $\mathrm{Tr}\left(\hat{O}\rho_i\otimes\rho_j\right)$; fidelity $F(\rho,\rho')$, distance??;
  - multivariate: quantum kernel $\mathrm{Tr}(\rho\rho')$; multivariate $\mathrm{Tr}(\rho_1\cdots\rho_m)$, nonlinear function?? linear;

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.

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

- **Input**: an **unknown** N-dimensional mixed state ρ, M known 2-outcome measurements $E_1,\ldots,E_M$

- **Output**: estimate $\mathbb{P}[E_i \text{ accept } \rho]$ to within additive error ϵ, $\forall i \in [M]$, with $\geq 2/3$ success probability.

Theorem 6 (bounds of shadow tomography [29]). *It is possible to do shadow tomography using* $\tilde{\mathcal{O}}(\frac{\log^4 M \cdot \log N}{\epsilon^4})$ *copies. [no construction algorithm?] sample complexity lower bound* $\Omega(\log(M) \cdot \epsilon^{-2})$,

more details in Section III A 1

Remark 6 (compare shadow tomography and classical shadow [3]). While very efficient in terms of samples, Aaronson's procedure is very demanding in terms of quantum hardware — a concrete implementation of the proposed protocol requires **exponentially long quantum circuits** that act collectively on all the copies of the unknown state stored in a quantum memory.??

III. CLASSICAL-QUANTUM HYBRID, END-TO-END PROTOCOL

In this paper, we focus on the entanglement structure dectection for graph states.

Problem 7 (detect graph state entanglement structure?). problem with/without training data

- **Input**: a graph G encoding in a graph state $|G\rangle$; adjacency matrix A?
- **Output**: entanglement structure: genuine multipartite entanglement??

Question 1. *how to relate graph state entanglement to graph property test ..??. (graph kernel??) [11]. witness; bounds; graph property? vertex cover? Hamiltonian cycle of a graph state?*

A. Classical features of quantum states

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.

1. *Classical shadow*

Inspired by Aaronson's shadow tomography [29], Huang et. al [3] introduce classical shadow. 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 13 (classical shadow). classical shadow (snapshots) ρ_{cs}

$$\rho_{cs} = \mathcal{M}^{-1}\left(U^\dagger \left|\hat{b}\right\rangle\!\left\langle\hat{b}\right| U\right) \tag{17}$$

such that we can predict the linear function with classical shadows

$$o_i = \mathrm{Tr}(O_i \rho_{cs}) \text{ obeys } \mathbb{E}[o] = \mathrm{Tr}(O_i \rho) \tag{18}$$

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.1: Classical Shadow (tomography)

input : an (unknown) density matrix ρ (many copies, black-box access to a circuit preparing a state)
output: classical shadow ρ_{cs}

for $i = 1, 2, \ldots, N$ **do**
  $\rho \mapsto \hat{U}\rho\hat{U}^\dagger$ // apply a random unitary to rotate the state
  $\mapsto |b\rangle \ldots$ // perform a computational-basis measurement
  $\rho_{cs} = \mathcal{M}^{-1}\left(\hat{U}^\dagger |b\rangle\!\langle b| \hat{U}\right)$ // measurement outcome $|b\rangle \in \{0,1\}^n$, $\mathcal{M}$ quantum channel
return $S(\rho, N) = \left\{\rho_{cs_1} = \mathcal{M}^{-1}\left(\hat{U}_1^\dagger |b_1\rangle\!\langle b_1| \hat{U}_1\right), \ldots, \rho_{cs_N}\right\}$ // call this array the classical shadow of ρ

A classical shadow is created by repeatedly performing a simple procedure: Apply a unitary transformation $\rho \mapsto \hat{U}\rho\hat{U}^\dagger$, 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 $\left\{\hat{O}_1, \ldots, \hat{O}_M\right\}$.

Lemma 1. *predict linear function with shadow shadow: the variance*

$$\mathrm{Var}[o] = \mathbb{E}[(o - \mathbb{E}[o])^2] \leq \left\| O - \frac{\mathrm{Tr}(O)}{2^n}\mathbb{1} \right\|^2_{\mathrm{shadow}} \tag{19}$$

Theorem 7. *Fix a measurement primitive $\mathcal{U}$??, a collection $\left\{\hat{O}_1, \ldots, \hat{O}_M\right\}$ of $2^n \times 2^n$ Hermitian matrices and accuracy parameters $\epsilon, \delta \in [0, 1]$. Set*

$$K = 2\log(2M/\delta), \quad N = \frac{34}{\epsilon^2} \max_{1 \leq i \leq M} \left\| \hat{O}_i - \frac{\mathrm{Tr}(O_i)}{2^n}\mathbb{1} \right\|^2_{\mathrm{shadow}} \tag{20}$$

where $\|\cdot\|_{\mathrm{shadow}}$ is shadow norm. Then, a collection of NK independent classical shadows allow for accurately predicting all features via median of means prediction

$$|o_i(N, K), -\mathrm{Tr}(O_i\rho)| \leq \epsilon, \ \forall i \leq i \leq M \tag{21}$$

wieth probability at least $1 - \delta$. $o_i(N, K) = \mathrm{median}\left\{\mathrm{Tr}\left(\hat{O}_i\rho_{(1)}\right), \ldots, \mathrm{Tr}\left(\hat{O}_i\rho_{(K)}\right)\right\}$

Definition 14 (shadow norm)**.** $\|\cdot\|_{\mathrm{shadow}}$ is shadow norm that only depends on the measurement primitive:

$$\|O\|_{\mathrm{shadow}} = \max_{\sigma:state} \left(\mathbb{E}_{U\sim\mathcal{U}} \sum_{b\in\{0,1\}^n} \langle b|\hat{U}\sigma\hat{U}^\dagger|b\rangle \ \langle b|\hat{U}\mathcal{M}^{-1}(O)\hat{U}^\dagger|b\rangle^2 \right)^{1/2} \tag{22}$$

(nonnegative, homogeneous, triangle inequality)

Theorem 8 (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 $\mathrm{Tr}\left(\hat{O}_i\rho\right)$, requires at least (lower bound) $\Omega(\log(M)3^k/\epsilon^2)$ copies of the state ρ. $\Omega(\log(M)\max_i \mathrm{Tr}\left(\hat{O}_i^2\right)/\epsilon^2)$*

Consider a simple family of entanglement witnesses with compatible structure: (ansatz??)

$$O := O(V_A, V_B, V_C) = V_A \otimes V_B \otimes V_C \left|\psi^+_{\mathrm{GHZ}}\right\rangle\!\left\langle\psi^+_{\mathrm{GHZ}}\right| V_A^\dagger \otimes V_B^\dagger \otimes V_C^\dagger \tag{23}$$

the single-qubit unitaries V_A, V_B, V_C parametrize differenet witnesses. for any state ρ_s with only bipartite entanglement, $\mathrm{Tr}\left(\hat{O}\rho_s\right) \leq 0.5$, while for any state ρ_s with at most W-type entanglement, $\mathrm{Tr}\left(\hat{O}\rho_s\right) \leq 0.75$. Therefore verifying that $\mathrm{Tr}\left(\hat{O}\rho\right) \geq 0.5$ certifies that ρ has tripartite entanglement, while $\mathrm{Tr}\left(\hat{O}\rho\right) > 0.75$ certifies that ρ has GHZ-type entanglement.

2. Derandomization

Derandomization can and should be viewed as a refinement of the original classical shadows idea. [30] [31]

3. training data and classical kernel methods

$\sigma_T(\rho(x_l))$ is the classical shadow representation of $\rho(x_l)$, a $2^n \times 2^n$ matrix that reproduces $\rho(x_l)$ in expectation over random Pauli measurements.

$$\{x_l \to \sigma_T(\rho(x_l))\}_{l=1}^N \tag{24}$$

Definition 15 (shadow kernel)**.** given two density matrices (quantum states) ρ and ρ', *shadow kernel* [3] is

$$k_{\text{shadow}}(S_T(\rho), S'_T(\rho')) := \exp\left(\frac{\tau}{T^2}\sum_{t,t'=1}^{T}\exp\left(\frac{\gamma}{n}\sum_{i=1}^{n}\text{Tr}\left(\sigma_i^{(t)}\sigma_i^{(t')}\right)\right)\right) \tag{25}$$

where $S_T(\rho)$ is the classical shadow representation of ρ. The computation time for the inner product is $\mathcal{O}(nT^2)$, linear in the system size n and quadratic in T, the number of copies of each quantum state which are measured to construct the classical shadow.

Proposition 4 ([4])**.** *exist quantum advantage in machine learning (not significant, practical) ... discrete log, factoring...*

Algorithm III.2: Classical learning (SVM) + classical shadow

input : classical shadow? (features) with label (training data); ansatz $\hat{W}_{\text{ml}} = \mathbf{w}\cdot\vec{\sigma}$
output: entanglement structure? decision separble; classify phase

```
// ---------- training phase ----------
for i = 1, 2, ..., m do
    kernel estimation                                   // classical kernel
    SVM                                                 // SVM
    return w                                            // parameters of the separating hyperplane in the feature space
// ========== testing phase ========== //
return w · σ⃗ < 0: separable                             // predict
```

4. *Estimate entanglement witness by quantum machine learning*

though, that while there is no large advantage in query complexity, a substantial quantum advantage in computational complexity is possible.

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 $\text{Tr}(P_x\rho), \forall x \in \{I, X, Y, Z\}^n$.

Algorithm III.3: entanglement witness by quantum ML

input : (copies of) density matrix ρ, an entanglement witness (observable) $\hat{W}$
output: $\text{Tr}(P_x\rho), \forall x \in \{I, X, Y, Z\}^n$

```
for i = 1, 2, ..., m do
    P_x                                                 // estimate entanglement witness by quantum ML
return estimation of Tr(Ŵρ)
```

Theorem 9 ([32])**.** *For M Pauli opertors, there is a (quantum) procedure estimate every expectation value $\text{Tr}(P_x\rho), \forall i = 1, \ldots, M$ within error ϵ under probability at least $1-\delta$ by performing POVM measurements on $\mathcal{O}(\log(M/\delta)\epsilon^{-4})$ copies of the unknown quantum state ρ. ($M = 4^n$ implies linear copy for full tomography???)*

Theorem 10 ([32])**.** *We rigorously show that, for any quantum process $\mathcal{E}$, observables $\hat{O}$, and distribution $\mathcal{D}$, and for any quantum ML model, one can always design a classical ML model achieving a similar average prediction error such that N_C (number of experiments?) is larger than N_Q by at worst a small polynomial factor.*

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.*

	circuits	number of copies (measurements)
classical shadow		
derandomized		
quantum circuit		

TABLE II: trace estimation

B. Quantum trace (kernel) estimation by quantum circuits

1. *Entanglement spectroscopy via quantum trace estimation*

For $\mathrm{Tr}(\rho_A^m)$, an important application of multivariate trace estimation [5] is to entanglement spectroscopy [33] [34] [31] - deducing the full set of eigenvalues of ρ_A. The smallest eigenvalue of diagnoses whether ψ_{AB} is separable or entangled [35]. The well-known identity (related to the replica trick originating in spin glass theory)

$$\mathrm{Tr}\Big(\hat{U}^{\pi}(\rho_1 \otimes \cdots \otimes \rho_m)\Big) = \mathrm{Tr}(\rho_1 \cdots \rho_m) \tag{26}$$

where the RHS is the multivariate trace and $\hat{U}^{\pi}$ is a unitary representation of the cyclic shift permutation. While the left-hand-side is ... We can estimate the quantities $\mathrm{Re}[\mathrm{Tr}(\rho_1 \cdots \rho_m)]$ and $\mathrm{Im}[\mathrm{Tr}(\rho_1 \cdots \rho_m)]$ respectively. to generalize the estimation beyond single-qubit states, we can either increase the width or the depth of the circuit described above.

Remark 7 ([35])**.** Direct entanglement detections, can be **employed as sub-routines in quantum computation**. For example, one may consider performing or not performing a quantum operation on a given quantum system conditioned on some part of quantum data being entangled or not. In fact direct entanglement detections can be viewed as quantum computations solving an inherently quantum decision problem: given as an input n copies of decide whether is entangled. Here the **input data is quantum** and such a decision problem cannot even be even formulated for classical computers.

For the sake of completeness we should also mention here that there are two-particle observables, called entanglement witnesses which can detect quantum entanglement is some special cases (see [20,8]). They have positive mean values on all separable states and negative on some entangled states. Therefore **any individual entanglement witness leaves many entangled states undetected**. When is unknown we need to check infinitely many witnesses, which effectively reduces this approach to the quantum state estimation. However, let us point out any witness defines a positive map which can be used in our test.

Definition 16 (entanglement spectroscopy)**.** the concept of the entanglement spectrum which is the energy spectrum of the "entanglement Hamiltonian" $\hat{H}_E$ defined through $\rho_A = \exp\Big(\hat{H}_E\Big)$. They pointed out that the largest eigenvalues of ρ_A [30] contain more universal signatures than the von Neumann entropy or S_2 alone. (SWAP trick, Quantum phase estimation) [34]:

Theorem 11 ([5])**.** *multivariate trace estimation can be implemented in* ***constant depth****, with only linearly-many controlled two-qubit gates and a linear amount of classical pre-processing. copies* $\mathcal{O}(\epsilon^{-2}\log(\delta^{-1}))$ *...*

Theorem 12. *Let* $\{\rho_1, \ldots, \rho_m\}$ *be a set of p-qubit states, and fix* $\epsilon > 0$ *and* $\delta \in (0,1)$*. There exists a random variable* $\hat{T}_p$ *that can be computed using* $\mathcal{O}(\mathcal{O}(\epsilon^{-2}\log(\delta^{-1})))$ *repetitions of a* ***constant-depth*** *quantum circuit consisting of* $\mathcal{O}(mp)$ *three-qubit gates, and satisfies*

$$\mathbb{P}\Big(\Big|\hat{T}_p - \mathrm{Tr}(\rho_1 \cdots \rho_m)\Big| \leq \epsilon\Big) \geq 1 - \delta. \tag{27}$$

Algorithm III.4: entanglement spectroscopy by ... quantum trace estimation

input : (copies of) density matrix (graph state?) ρ,...
output: spectrum of entanglement Hamiltonian

for $i = 1, 2, \ldots, m$ **do**
    GHZ // prepare GHZ
    parallel // estimate real and imaginary part respectively
    return λ
return smallest eigenvalue of ρ_A

Remark 8 ([5]). We remark that, an alternative way to estimate $\mathrm{Tr}\left[\rho^k\right]$ for each $k \in [m]$ is by using the method of classical shadows to obtain 'classical snapshots' of ρ that can be linearly combined to obtain a classical random variable whose expectation is $\mathrm{Tr}\left[\rho^k\right]$ (see Supplementary Material Section 6 of [3]). However, it is **unclear to us if this method would offer savings in the quantum resources required, as the total number of times the quantum circuit needs to be run in the data acquisition phase should scale with the variance of the corresponding estimator**. We do not know of a concise expression for this variance for arbitrary m. Indeed, calculating it for just a single value of m ($m = 2$) required four pages of calculations in [3].

	observables	weights	input
entanglement witness			known
Bell (CHSH) inequality		unknown	
entangle **spectrum??		unknown	

TABLE III: ansatz

C. Training entanglement witness with SVM

1. *Related works*

separability classifier by classical neural network [36]: input: sythetic random density matrices; output: a classical classifier for bipartite separable (independent of state??). The idea is to feed the classifier by a large amount of sampled trial states (feature: synthetic density matrix with noise flatten as a real vector $\mathbf{x} \in \mathbb{R}^{d_A^2 d_B^2 - 1}$) as well as their corresponding class labels (separable or entangled by PPT criterion [8], CHA), and then train the classifier to predict the class labels of new states that it has not encountered before. Previous methods **only detect a limited part of the state space**, e.g. different entangled states often require different entanglement witness. In contrast, this classifier can handle a variety?? of input states once properly trained Fig. 1.

Bell inequlaity and NN [9]. Overall, for scaling up this method for detecting higher-dimensional quantum entanglement, the major challenge is related to a lack of reliable method for labeling the entanglement. We have constructed such a universal state classifier for a pair of qubits; we find that the performance depends heavily on the testing sets; the major source of error comes from the data near the boundary between entangled and separable states. Tomographic predictors make use of all information of a given quantum state and is used to benchmark the performance of Belllike predictors, which employs a subset of non-orthogonal measurements setting. [36] reported that they independently combined machine learning and semidefinite programming to train their predictors as quantum state classifiers. Using all information without any prior knowledge, the error of their predictor is always around 10% on general 2-qubit system. However, our tomographic predictor performs below 2% on the same ensembles with 3000 hidden neurons.

SVM, (universal), 4 qubit [37]

classical shadow [3]: estimate entanglment witness (fixed but unknown target state, e.g., tripartite GHZ) Entanglement verification. Fidelities with pure target states can also serve as (bipartite) entanglement witnesses. For every (bipartite) entangled state ρ, there exists a constant α and an observable $\hat{O} = |\psi\rangle\langle\psi|$ such that $\mathrm{Tr}\left(\hat{O}\rho\right) > \alpha \geq \mathrm{Tr}\left(\hat{O}\rho_s\right)$, for all (bipartite) separable states ρ_s. Establishing $\mathrm{Tr}\left(\hat{O}\rho\right) > \alpha$ verifies the existence of entanglement in the state ρ. Any $\hat{O} = |\psi\rangle\langle\psi|$ that satisfies the above condition is known as an entanglement witness for the state ρ. **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.

classical SVM: An SVM allows for the construction of a hyperplane $\left\langle \hat{W} \right\rangle = \sum_k w_k \mathbf{x}_k$ 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**. 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 2 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 [??].

classical machine learning (SVM, NN) with classical shadow [38]??: classify phase, predict ground state, entanglement?

2. *Our witness ansatz and optimization*

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.

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 [39].

an ansatz for entanglement witness [24] (graph state entanglement)

$$\hat{W}_{\text{ansatz}} := \sum_{\mathbf{p}\in\{I,X,Y,Z\}^n} w_{\mathbf{p}} \bigotimes_i^n \hat{\sigma}^{(\mathbf{p}_i)} \tag{28}$$

c.f. full tomography (Stokes parameters) Eq. (16)

Algorithm III.5: ansatz + classical shadow (quantum ML) + classical SVM

input : (copies of) ρ, an entanglement witness (observable); ansatz $\hat{W}_{\text{ml}} = \mathbf{w}\cdot\vec{\sigma}$
output: classifier $\mathbf{w}$

for $i = 1, 2, \ldots, m$ **do**
  ρ_{cs} // classical shadow
  wx // optimize SVM
return parameters $\mathbf{w}$ (SVM hyperplane)

Theorem 13. *On quantum computers, evaluating the trace distances is probably hard since even judging whether ρ and ρ' have large or small trace distance is known to be QSZK-complete [40], where QSZK (quantum statistical zero-knowledge) is a complexity class that includes BQP (bounded-error quantum polynomial time). [41]*

D. Theoretic upper bounds and lower bounds

[3] [32] [4] [29]

	gate/depth/computation	measurements/samples	query?	input/unknown?
shadow tomography	exp circuit?	Theorem 6	N/A	unknown
entanglement witness	N/A	Proposition 3 (constant?)	convex?[42]	known
classical shadow	N/A	Theorems 7 and 8	N/A	unknown?
C. ML + C. entanglement witness ansatz	??	Q. advantage Theorem 10	N/A	unknown
QML. entanglement witness ansatz	??	Theorem 9	N/A	unknown
Q. entanglement spectroscopy	Theorem 11 (c-depth?)		property test [43]	unknown
SVM + quantum kernel estimation				??

TABLE IV: complexity (different measures) of different methods

IV. NUMERICAL SIMULATION

A. Data preparation

multi-partite entangled state: generate synthetic (engineered) data from (random graph?). separable state from randomly ... QuTiP library [44]; quantum circuit [45]

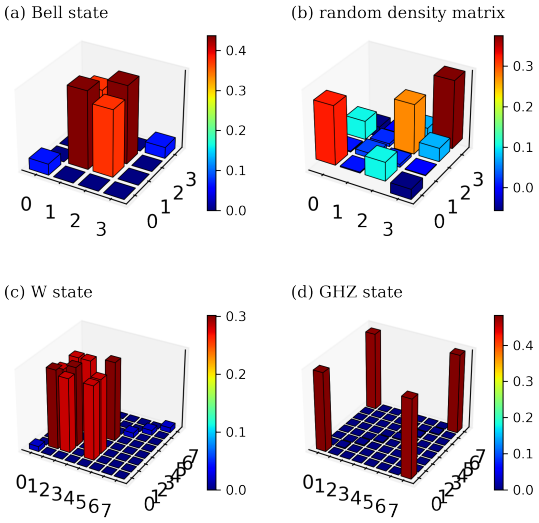


FIG. 3: (a) three-qubit W state, (b) GHZ state with white noise

data generation

- 2-qubit: bipartite
- 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.

$$\rho_{A|B|C},\ \rho_{A|BC},\ \rho_{AB|C},\ \rho_{B|AC} \tag{29}$$

$$|rand\rangle_A \otimes |rand\rangle_B\,,\ |rand\rangle_A \otimes |entangle\rangle_{BC}\,, \tag{30}$$

$$\cos(\theta)\,|00\rangle + \sin(\theta)e^{\mathrm{i}\phi}\,|11\rangle\,,\ \cos(\theta)\,|01\rangle + \sin(\theta)e^{\mathrm{i}\phi}\,|10\rangle \tag{31}$$

B. Classification accuracy and comparison

1. *Hyperparameters and settings*

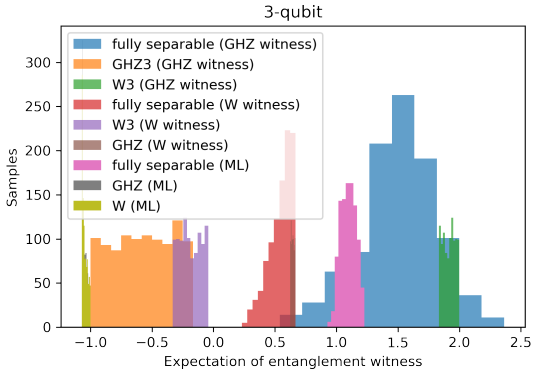
We consider a set of different regularization parameters,...
The goal of RFE is to eliminate non-essential features by recursively considering smaller and smaller subsets of the original features using a greedy algorithm. Initially, RFE takes the SVM we trained and ranks the coefficients by their magnitudes,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

FIG. 4: accuracies (variance) VS different data sizes

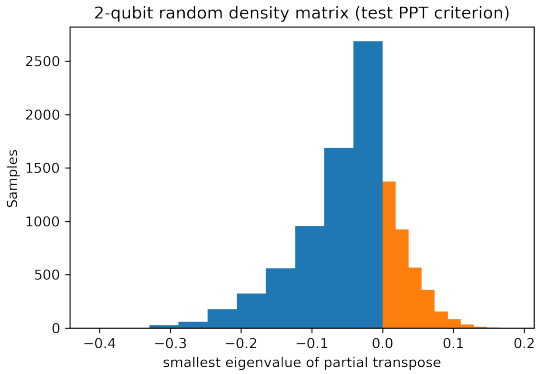
FIG. 5: number of features VS number of qubits (n). feature elimination

2. *Results*

performance of different methods:



(a) compare different methods: Bell inequality, witness, ML ansatz



(b) PPT criterion (2-qubit random density matrix)

C. Robustness to noise

tradeoff between (white noise) tolerance (robustness) and efficiency (number of measurements).

$$\rho'_{\text{noise}} = (1 - p_{\text{noise}}) |G\rangle\langle G| + p_{\text{noise}} \frac{\mathbb{1}}{2^n} \tag{32}$$

p_{noise} indicates the robustness of the algorithm (witness).

Remark 9 ([2])**.** the largest noise tolerance p_{limit} just related to the **chromatic number** (graph property) of the graph.

Question 2. *how far white noise? other noise (depolarization)? e.g., flip error, phase error?, local, random unitary transformation? find optimal (robustness) entanglement witness by classical machine learning (qunatum circuit?)*

FIG. 7: robustness: accuracy VS p noise

D. Experiments

future: experimental (photonic) implementation with a few qubits (generation, verification) [46]. fully entangled graph state (ring of 16 qubits) IBM by measuring negativity [47] optical lattice [48] (homogeneous measurement; detect GME, full entanglement).

V. CONCLUSION AND DISCUSSION

Acknowledgements

[1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009), arXiv:quant-ph/0702225.
[2] Y. Zhou, Q. Zhao, X. Yuan, and X. Ma, npj Quantum Inf **5**, 83 (2019).
[3] H.-Y. Huang, R. Kueng, and J. Preskill, Nat. Phys. **16**, 1050 (2020), arXiv:2002.08953 [quant-ph].
[4] 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].
[5] Y. Quek, M. M. Wilde, and E. Kaur, Multivariate trace estimation in constant quantum depth (2022), arXiv:2206.15405 [hep-th, physics:quant-ph].
[6] L. Gurvits, Classical deterministic complexity of Edmonds' problem and Quantum Entanglement (2003), arXiv:quant-ph/0303055.
[7] L. M. Ioannou, Quantum Inf. Comput. **7**, 335 (2007), arXiv:quant-ph/0603199.
[8] M. Horodecki, P. Horodecki, and R. Horodecki, Physics Letters A **223**, 1 (1996), arXiv:quant-ph/9605038.
[9] Y.-C. Ma and M.-H. Yung, npj Quantum Inf **4**, 34 (2018), arXiv:1705.00813 [quant-ph].
[10] H. J. Briegel, D. E. Browne, W. Dür, R. Raussendorf, and M. V. den Nest, Nature Phys **5**, 19 (2009), arXiv:0910.1116.
[11] 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.
[12] 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.
[13] 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).
[14] G. Tóth and O. Gühne, Phys. Rev. A **72**, 022340 (2005).
[15] O. Gühne and G. Toth, Physics Reports **474**, 1 (2009), arXiv:0811.2803 [cond-mat, physics:physics, physics:quant-ph].
[16] B. M. Terhal, Physics Letters A **271**, 319 (2000), arXiv:quant-ph/9911057.
[17] T. Heinosaari and M. Ziman, *The Mathematical Language of Quantum Theory: From Uncertainty to Entanglement*, 1st ed. (Cambridge University Press, 2011).
[18] 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].
[19] G. Toth and O. Guehne, Phys. Rev. Lett. **94**, 060501 (2005), arXiv:quant-ph/0405165.
[20] O. Gühne and N. Lütkenhaus, Phys. Rev. Lett. **96**, 170502 (2006).
[21] Y. Zhang, Y. Tang, Y. Zhou, and X. Ma, Phys. Rev. A **103**, 052426 (2021), arXiv:2012.07606 [quant-ph].
[22] X. Gao and L.-M. Duan, Nat Commun **8**, 662 (2017), arXiv:1701.05039 [cond-mat, physics:quant-ph].
[23] C. de Gois, K. Hansenne, and O. Gühne, Uncertainty relations from graph theory (2022), arXiv:2207.02197 [quant-ph].
[24] E. Y. Zhu, L. T. H. Wu, O. Levi, and L. Qian, Machine Learning-Derived Entanglement Witnesses (2021), arXiv:2107.02301 [quant-ph].
[25] M. Weilenmann, B. Dive, D. Trillo, E. A. Aguilar, and M. Navascués, Phys. Rev. Lett. **124**, 200502 (2020), arXiv:1912.10056 [quant-ph].
[26] Y. Zhan and H.-K. Lo, Detecting Entanglement in Unfaithful States (2021), arXiv:2010.06054 [quant-ph].
[27] J. Altepeter, E. Jeffrey, and P. Kwiat, in *Advances In Atomic, Molecular, and Optical Physics*, Vol. 52 (Elsevier, 2005) pp. 105–159.
[28] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, IEEE Trans. Inform. Theory , 1 (2017).
[29] 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.
[30] H.-Y. Huang, R. Kueng, and J. Preskill, Phys. Rev. Lett. **127**, 030503 (2021), arXiv:2103.07510 [quant-ph].
[31] 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].
[32] H.-Y. Huang, R. Kueng, and J. Preskill, Phys. Rev. Lett. **126**, 190505 (2021), arXiv:2101.02464 [quant-ph].
[33] 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.
[34] S. Johri, D. S. Steiger, and M. Troyer, Phys. Rev. B **96**, 195136 (2017), arXiv:1707.07658.
[35] P. Horodecki and A. Ekert, Phys. Rev. Lett. **89**, 127902 (2002), arXiv:quant-ph/0111064.
[36] 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].
[37] 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].
[38] 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].

[39] A. Jacot, F. Gabriel, and C. Hongler, Neural Tangent Kernel: Convergence and Generalization in Neural Networks (2020), arXiv:1806.07572 [cs, math, stat].
[40] J. Watrous, Quantum Computational Complexity (2008), arXiv:0804.3401 [quant-ph].
[41] R. Chen, Z. Song, X. Zhao, and X. Wang, Quantum Sci. Technol. **7**, 015019 (2022), arXiv:2012.05768 [math-ph, physics:quant-ph].
[42] S. Chakrabarti, A. M. Childs, T. Li, and X. Wu, Quantum **4**, 221 (2020), arXiv:1809.01731 [quant-ph].
[43] A. Montanaro and R. de Wolf, A Survey of Quantum Property Testing (2018), arXiv:1310.2035 [quant-ph].
[44] J. R. Johansson, P. D. Nation, and F. Nori, Computer Physics Communications **184**, 1234 (2013), arXiv:1110.0573.
[45] B. Li, S. Ahmed, S. Saraogi, N. Lambert, F. Nori, A. Pitchford, and N. Shammah, Quantum **6**, 630 (2022), arXiv:2105.09902 [quant-ph].
[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. Wang, Y. Li, Z.-q. Yin, and B. Zeng, npj Quantum Inf **4**, 46 (2018), arXiv:1801.03782.
[48] Y. Zhou, B. Xiao, M.-D. Li, Q. Zhao, Z.-S. Yuan, X. Ma, and J.-W. Pan, npj Quantum Inf **8**, 1 (2022).
[49] C. Bădescu, R. O'Donnell, and J. Wright, Quantum state certification (2017), arXiv:1708.06002 [quant-ph].
[50] S. Ben-David, A. M. Childs, A. Gilyén, W. Kretschmer, S. Podder, and D. Wang, 2020 IEEE 61st Annu. Symp. Found. Comput. Sci. FOCS , 649 (2020), arXiv:2006.12760.
[51] N. M. Kriege, F. D. Johansson, and C. Morris, Appl Netw Sci **5**, 6 (2020), arXiv:1903.11835 [cs, stat].
[52] L. Bai, L. Rossi, A. Torsello, and E. R. Hancock, Pattern Recognition **48**, 344 (2015).
[53] V. Havlicek, A. D. Córcoles, K. Temme, A. W. Harrow, A. Kandala, J. M. Chow, and J. M. Gambetta, Nature **567**, 209 (2019), arXiv:1804.11326.
[54] M. Schuld and N. Killoran, Phys. Rev. Lett. **122**, 040504 (2019), arXiv:1803.07128 [quant-ph].
[55] Y. Liu, S. Arunachalam, and K. Temme, Nat. Phys. **17**, 1013 (2021), arXiv:2010.02174 [quant-ph].
[56] J. R. Glick, T. P. Gujarati, A. D. Corcoles, Y. Kim, A. Kandala, J. M. Gambetta, and K. Temme, Covariant quantum kernels for data with group structure (2021), arXiv:2105.03406 [quant-ph].

Appendix A: Definitions

Definition 17 (density matrix)**.** pure state $|\psi\rangle$; A quantum state ρ is defined to be a positive operator $\rho \in \mathrm{End}(V)$ with $\mathrm{Tr}(\rho) = 1$. density matrix ρ (trace one, Hermitian, PSD)...

Definition 18 (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 ... $\mathbb{E}[x] \equiv \left\langle \hat{O}_x \right\rangle := \mathrm{tr}\left(\hat{O}_x \rho\right)$

Definition 19 (positive, semidefinite)**.** denoted $X \preceq Y$ provided $Y - X$ is positive

Definition 20 (partial trace)**.** reduced density matrix $\rho_A = \mathrm{Tr}_B(\rho_{AB})$

Definition 21 (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 AB.

Definition 22 (maximally entangled)**.** a state vector is *maximually entangled* iff the reduced state at one qubit is maximally mixed, i.e., $\mathrm{Tr}_A(|\psi\rangle\langle\psi|) = \frac{1}{2}$.

Definition 23 (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 $H_S(P) := -\sum_i P(x_i) \log P(x_i)$. relative entropy (divergence)

Definition 24 (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

$$E(\Psi_{AB}) = S(\rho_A) = -\mathrm{Tr}(\rho_A \log \rho_A) = -\mathrm{Tr}(\rho_B \log \rho_B) = S(\rho_B) \tag{A2}$$

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. (1)), 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 5. 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 $\left\lfloor \frac{N}{2} \right\rfloor$. Schmidt measure of tree is the size of its minimal vertex cover[??].

Definition 25 (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 $F(\rho,\rho') := \mathrm{tr}(\rho\rho')$.

different distance measures [49]

Definition 26 (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.

Definition 27 (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

$$1 - F \le D_{\mathrm{tr}}(\rho,\rho') \le \sqrt{1-F^2} \tag{A3}$$

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 $D_{\mathrm{HS}}(\rho,\rho') := \|\rho-\rho'\|_{\mathrm{HS}} = \sqrt{\mathrm{Tr}((\rho-\rho')^2)}$

denote a group by $\mathbb{G}$ and a subgroup $\mathbb{H}$.

Definition 28 (Pauli group)**.**

Definition 29 (Clifford group)**.**

Definition 30 (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 $\{\mathbb{1}, X, Y, Z\}$.

Appendix B: Machine learning background

Notations: The (classical) training data (for supervised learning) is a set of m data points $\big\{(\mathbf{x}^{(i)}, y^{(i)})\big\}_{i=1}^m$ 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 31 (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...

$$\max_{\mathbf{w}} .... \tag{B1}$$

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 32 (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

$$k(\mathbf{x}, \mathbf{x}') := \langle \phi(\mathbf{x}), \phi(\mathbf{x}') \rangle \tag{B2}$$

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 $\mathbf{K}$ should be a positive, semidefinite (PSD) matrix, i.e. all eigenvalues are non-negative

Example 6 (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.

similarity measures? advantages? why? (isomorphism?)

Definition 33 (divergence)**.** KL divergence (relative entropy): measure the distance (similarity) between two probability distributions:

$$D_{\text{KL}}(P||Q) := \sum P(x) \log(P(x)/Q(x)) \tag{B3}$$

symmetric version: Jensen-Shannon divergence (machine learning)

$$D_{\text{JS}}(P||P') := \frac{1}{2}(D_{\text{KL}}(P||M) + D_{\text{KL}}(P'||M)) \equiv H_S(M) - \frac{1}{2}(H_S(P) - H_S(P')) \tag{B4}$$

where $M = (P + P')/2$ and Shannon entropy H_S. Analogously, quantum Jensen-Shannon divergence D_{QJS} of two density matrices can be defined...

$$D_{\text{QJS}}(\rho||\rho') := H_V(\rho_M) - \frac{1}{2}(H_V(\rho) - H_V(\rho')) \tag{B5}$$

as a quantum graph kernel (ρ induced by quantum random walk)

Definition 34 (geometric difference)**.**

$$g(K^1||k^2) = \sqrt{\left\| \sqrt{K^2}(K^1)^{-1}\sqrt{K^2} \right\|_\infty} \tag{B6}$$

where $\|\cdot\|_\infty$ is the spectral norm.

b. Graph kernel

Definition 35 (graph property)**.** monotone ...

Example 7 (colorable)**.** k-colorable is a graph property, i.e., allow for a coloring of the vertices with k colors such that no two adjacent vertices have the same color. A graph is bipartite iff 2-colorable. other graph properties: isomorphism; vertex cover; Hamiltonian cycle ...

Problem 8 (graph property test)**. promise**: the input graph either has a property, or is ϵ-far from having the property, meaning that we must change at least an ϵ fraction of the edges to make the property hold.

Theorem 14 (bounds for graph property test).

Question 3. *[43] Is there any graph property which admits an exponential quantum speed-up? [50] depends on input model (query adjacency matrix/list)*

Graphs is another kind of data which is fundamentally different from a real value vector because of vertex-edge relation and graph isomorphism. So, graph kernel [51] need additional attention.

Definition 36 (graph kernel). given a pair of graphs (G, G'), *graph kernel* is $k(G, G') =$. quantum graph kernel $k(G, G') = |\langle G|G'\rangle|^2$?? [52]

c. *Quantum kernel*

related works:

- quantum kernel method: estimate kernels by quantum algorithms (circuits) [53] [54]: for classical problem (data)
- rigorous and robust quantum advantage of quantum kernel method in SVM [55]. group structured data [56]
- power of data in quantum machine learning [4]: input??? projected quantum kernel

Definition 37 (quantum kernel). quantum kernel with quantum feature map $\phi(\mathbf{x}) : \mathcal{X} \to |\phi(\mathbf{x})\rangle\langle\phi(\mathbf{x})|$

$$k_Q(\rho, \rho') := |\langle\phi(\mathbf{x})|\phi(\mathbf{x}')\rangle|^2 = \left|\langle 0|\hat{U}^\dagger_{\phi(\mathbf{x})}\hat{U}_{\phi(\mathbf{x}')}|0\rangle\right|^2 =? \mathrm{Tr}(\rho\rho') \equiv \langle\rho, \rho'\rangle_{\mathrm{HS}} \tag{B7}$$

where $\hat{U}_{\phi(\mathbf{x})}$ 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);

Proposition 5 ([4]). *If a classical algorithm without training data can compute (label) $y = f(x) = \langle x|\hat{U}^\dagger_{\mathrm{QNN}}\hat{O}U_{\mathrm{QNN}}|x\rangle$ (with amplitude encoding) efficiently (poly time in ...) for any $\hat{U}_{\mathrm{QNN}}$ and $\hat{O}$, then* BPP = BQP *(which is believed unlikely).*

Proposition 6 ([4]). *Training an arbitrarily deep quantum neural network $\hat{U}_{\mathrm{QNN}}$ with a trainable observable $\hat{O}$ is equivalent to training a quantum kernel method with kernel $k_Q(\mathbf{x}, \mathbf{x}') = \mathrm{Tr}(\rho(\mathbf{x})\rho'(\mathbf{x}'))$*

Definition 38 (projected quantum kernel). ....

2. Neural network

a. *neural network and kernel*

Definition 39 (neural tagent kernel). neural tangent kernel [39]: proved to be equivalent to deep neural network [22] in the limit ...

$$k_{\mathrm{NT}}\Big(S_T(\rho_l), \tilde{S}_T(\rho_{l'})\Big) = \Big\langle\phi^{(\mathrm{NT})}(S_T(\rho_l)), \phi^{(\mathrm{NT})}(\tilde{S}_T(\rho_l))\Big\rangle \tag{B8}$$

b. *quantum neural network*

Appendix C: Hardness assumptions

Definition 40 (NP). NP, NP-hard, NP-complete

Definition 41 (#P). #P

Definition 42 (QMA). QMA

Definition 43 (BPP). BPP

Definition 44 (BQP). BQP