Towards efficient entanglement structure detection

Jue Xu* and Qi Zhao†

(Dated: August 4, 2022)

Verification (detection) of entanglement structure is an indispensable step for practical quantum computation (communication). In this work, we compare complexity and performance of several recently-developed methods, including conventional entanglement witness methods, shadow tomography, classical machine learning, and quantum algorithms (trace estimation). Machine learning algorithms and quantum advantages ...

CONTENTS

I. INTRODUCTION

Entanglement [1] is the key ingredient of quantum computation [], quantum communication [], and quantum cryptography []. It is essential to benckmark (characterize) entanglement structures of target states. multipartite

II. PRELIMINARIES

Notations: The (classical) training data 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 $\mathbf{x} := (x_1, x_2, \ldots, x_d) \in \mathbb{R}^d$ 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, we assume $\Sigma = \{-1, 1\}$ (binary classification).

a graph $G = (V, E)$ with vertices V and edges E; a group $\mathbb{G}$ with a subgroup $\mathbb{H}$. The hats on the matrices such as $\hat{A}$, $\hat{H}$, ρ, $\hat{O}$ (omitted), $\hat{W}$, emphasize that they play the roles of operators. denote vector (matrix) $\mathbf{x}$, $\mathbf{K}$ by boldface font.

* juexu@cs.umd.edu
† email

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 the other useful one is the binary representation of computational basis $\{|\mathbf{x}\rangle \equiv |x_1, x_2, \ldots, x_n\rangle\}$ 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. $|+\rangle := (|0\rangle + |1\rangle)/\sqrt{2}$

A. Entanglement detection

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

Definition 1 (Entangled state)**.** entangled pure state is a quantum state that cannot be written as a product state, i.e.,

$$|\psi\rangle = \sum \tag{1}$$

mixed state is convex combination of entangled pure state ...

Definition 2 (Bipartite state)**.**

1. entanglement measures

Definition 3 (Schmidt coefficient/rank/measure)**.** Schmidt decomposition

Definition 4 (entropy)**.** In quantum mechanics (information), von Neumann *entropy* of a density matrix is $H_N(\rho) := -\operatorname{Tr}(\rho \log \rho)$; In classical information (statistical) theory, Shannon entropy of a probability distribution P is $H_S(P) := -\sum_i P(x_i) \log P(x_i)$. relative entropy (divergence)

2. entanglement structures

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.

Definition 5 (genuine entangled)**.** A state possesses *genuine multipartite entanglement* (GME) if it is outside of S_2, and is (fully) n-separable if it is in S_n. A state possesses P-genuine entanglement if it is outside of S_b^P. A state ρ possesses P-genuine entanglement iff $\rho \notin S_b^P$.

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.

Definition 6 (Multipartite state)**.** denote the partition $\mathcal{P}_m = \{A_i\}$ and omit the index m when it is clear from the context.

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

Definition 7 (fully separable state)**.** An n-qubit pure state $|\psi_f\rangle$ is *fully separable* iff . An n-qubit pure state $|\psi_f\rangle$ is *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 P-fully separable iff it can be decomposed into a conex mixture of P-fully separable pure states

$$\rho_f = \sum_i p_i \left|\psi_f^i\middle\rangle\!\middle\langle\psi_f^i\right|, (\forall i)(p_i \geq 0, \sum_i p_i = 1). \tag{2}$$

P-bi-separable... $S_f^P \subset S_b^P$

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 1. 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. Moreover, for some systems, such as distributed quantum computing, multiple quantum processor, and quantum network, natural partition exists due to the system geometric configuration. Therefore, it is practically interesting to study entanglement structure under partitions.

entanglement structure measures

Definition 8 (Entanglement intactness, depth)**.** the entanglement intactness of a state ρ to be m, iff $\rho \notin S_{m+1}$ and $\rho \in S_m$. k-producible

When the entanglement intactness is 1, the state is genuine entangled; and when the intactness is n, the state is fully separable.

Example 1 (GHZ)**.** bipartite: Bell states; nontrivial multipartite: tripartite. 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):

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

W state

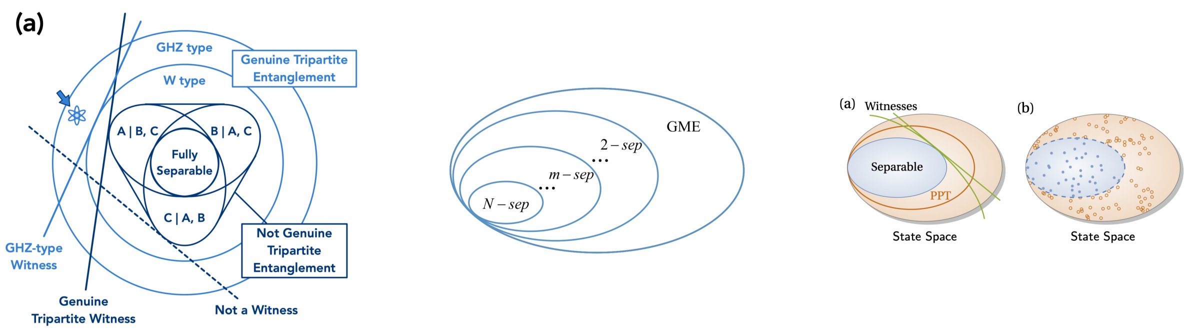

FIG. 1: (a) entanglement witness, PPT criteria, SVM (kernel)?. (c) convex hull...

3. *Entanglement witness*

Theorem 1 ([2])**.** *The weak membership problem for the convex set of separable normalized bipartite density matrices is NP-Hard.* ***Input****: ??*

Question 1. *specific cases? approximately correct? quantum computation? machine learning (data)?*

Theorem 2 (PPT criterion)**.** *the positive partial transpose (PPT) criterion, saying that a separable state must have PPT. Note, it is only necessary and sufficient when $d_A d_B \leq 6$.*

see Fig. 1

Definition 9 (entanglement witness)**.** Given an (unknown) quantum state (density matrix) ρ, the *entanglement witness* $\hat{W}$ is an obseverable such that

$$\mathrm{Tr}\left(\hat{W}\rho\right) \geq 0, \forall \text{ separable} ; \quad \mathrm{Tr}\left(\hat{W}\rho\right) < 0, \text{ for some entangled} \tag{3}$$

It is natural to ask nonlinear entanglement witness [3] kernel ML

Problem 1 (Entanglement witness with prior)**.** with/out prior knowledge

- **Input**: a **known** state $|\psi\rangle$, with noise
- **Output**: separable or not ??? (decision problem?? find problem)

4. Graph state

graph state is an important class of multipartite states in quantum information. Typical graph states include cluster states, GHZ state, and the states involved in error correction. 2D cluster state is the universal resource for the measurement based quantum computation (MBQC) [4].

Definition 10 (graph state)**.** Given a graph $G = (V, E)$, a graph state is constructed as

- vertices: $|+\rangle^{\otimes n}$
- edges: apply controlled-Z to every edge, that is $|G\rangle = \prod_{(i,j)\in E} \mathsf{cZ}_{(i,j)} |+\rangle^{\otimes n}$

An n-partite 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 $\forall i, S_i |G\rangle = |G\rangle$.

Example 2 (graph states)**.** GHZ; complete graph, hypercube, Petersen graph; cluster state

Question 2. *efficient? meaningful encoding of graph data/input?? Hamiltona cycle of a graph state? vertex cover*

Problem 2 (Certify entanglement)**.** Multipartite entanglement-structure detection

- **Input**: Given a state close to a **known** (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 2. The graph state is the unique eigenstate wtih eignevalue of $+1$ for all the n stabilizers. As a result, a graph state can be writteb as a product of stailizer projectors, $|G\rangle\langle G| = \prod_{i=1}^{n} \frac{S_i + \mathbb{1}}{2}$. stabilizer formalism?;

Remark 3. The entanglement entropy $S(\rho_A)$ equals the rank of the adjacency matrix of the underlying bipartite graph, which can be efficiently calculated.

Proposition 1 ([5])**.** *Given a graph state $|G\rangle$ and a partition $\mathcal{P} = \{A_i\}$, the fidelity between $|G\rangle$ and any fully separable state is upper bounded by*

$$\mathrm{Tr}(|G\rangle\langle G| \rho_f) \leq \min_{\{A,\bar{A}\}} 2^{-S(\rho_A)} \tag{4}$$

where $S(\rho_A)$ is the von Neumann entropy of the reduced density matrix $\rho_A = \mathrm{Tr}_{\bar{A}}(|G\rangle\langle G|)$.

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

Proposition 2 (Entanglement of graph state)**.** *[6]. witness; bounds; graph property? vertex cover?*

generalize [7] stabilizer state, neural network state?

B. Shadow tomography

Intuitively, a general tomography [8] that extract all information about a state requires exponential copies (samples/measurements).

Problem 3 (Full tomography)**.** we refer to *full tomography* in this paper, contrast to shadow tomography

- **Given (Input): unknown** D-dimensional mixed state ρ
- **Goal (Output):** a complete description of ρ (decomposition coefficients) with error?

Theorem 4 (lower bound of tomography?[9])**.** *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.*

Definition 11 (fidelity)**.** Given a pair of states (target and real),

$$F(\rho, \rho') := \mathrm{Tr}\sqrt{\sqrt{\rho}\rho'\sqrt{\rho}} \tag{5}$$

Definition 12 (distance)**.** trace distance

$$d_{tr}(\rho, \rho') := \frac{1}{2}\|\rho - \rho'\|_1 \tag{6}$$

Definition 13 (norm)**.** Schatten p-norm $\|x\|_p := (\sum_i |x_i|^p)^{1/p}$. Euclidean norm l_2 norm; Spectral (operator) norm ; Trace norm $\|A\|_{\mathrm{Tr}} \equiv \|A\|_1 := \mathrm{Tr}\left(\sqrt{A^\dagger A}\right)$, $p = 1$; Frobenius norm $\|A\|_F := \sqrt{\mathrm{Tr}(A^\dagger A)}$, $p = 2$; Hilbert-Schmidt norm $\|A\|_{HS} := \sqrt{\sum_{i,j} A_{ij}^2}$

Problem 4 (Fidelity estimate)**.** defined as follows

- **Input**: Given two density matrices ρ and ρ',
- **Output**: fidelity with error ϵ

Problem 5 (Trace/expectaton estimate)**.** defined as follows

- **Input**: Given an observable $\hat{O}$ and a mixed state ρ in density matrix,
- **Output**: the expectation value $\mathrm{Tr}\left(\hat{O}\rho\right) =$ with error ϵ (trace distance)

Nevertheless, we usually only need specific property of a target state rather than all information about the state. This enables the possibility to Inspired by Aaronson's shadow tomography [10], Huang et. al [11]

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

- **Given (Input): unknown** D-dimensional mixed state ρ, known 2-outcome measurements $E_1, \ldots, E_M$
- **Goal (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 5 ([10])**.** *It is possible to do shadow tomography using $\tilde{\mathcal{O}}(\frac{\log^4 M \cdot \log D}{\epsilon^4})$ copies. [no construction algorithm?] sample complexity lower bound $\Omega(\log M \cdot \epsilon^{-2})$,*

random Pauli measurements

Definition 14 (classical shadow)**.** classical shadow

$$\rho_{cs} = \mathcal{M}^{-1}\left(U^\dagger \left|\hat{b}\middle\rangle\middle\langle\hat{b}\right| U\right) \tag{7}$$

predict linear function with classical shadows

$$o_i = \mathrm{Tr}(O_i \rho_{cs}) \text{ obeys } \mathbb{E}[o] = \mathrm{Tr}(O_i \rho) \tag{8}$$

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. 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 II.1: Shadow tomography

input : density matrix ρ, ..
output: classical shadow

```
for i = 1, 2, ..., m do
    random Pauli measurements        // a comment
    return "?"
return?
```

Lemma 1. *the variance*

$$\mathrm{Var}[o] = \mathbb{E}[(o - \mathbb{E}[o])^2] \leq \left\| O - \frac{\mathrm{Tr}(O)}{2^n}\mathbb{1} \right\|_{shadow}^2 \tag{9}$$

sample complexity

$$N_{tot} = \mathcal{O}\left(\frac{\log(M)}{\epsilon^2}\max_{1\leq i\leq M}\left\|O_i - \frac{\mathrm{Tr}(O_i)}{2^n}\mathbb{1}\right\|^2_{\text{shadow}}\right) \tag{10}$$

Theorem 6 (Pauli/Clifford measurements). *additive error ϵ, M arbitrary k-local linear function $\mathrm{Tr}\left(\hat{O}_i\rho\right)$, $\Omega(\log(M)3^k/\epsilon^2)$ copies of the state ρ.*

III. CLASSICAL, DATA-POWERED, AND QUANTUM ALGORITHMS

We consider the problem

Problem 7 (?data). problem without training data

- **Input**: a graph G encoding in a graph state $|G\rangle$
- **Output**: entanglement structure

with training data

- **features**: classical shadow?
- label:

A. Quantum-classical (ML) hybrid method

1. *Classical machine learning*

separability classifier by neural network [12]. rigorous quantum advantage of quantum kernel method in SVM [13]. classical machine learning with classical shadow [14].

Definition 15 (SVM). find a hyperplane (a linear function) such that maximize the margin

nonlinear boundary. map to a higher dimensional (feature) space, in which data is linearly separable.

Remark 4. input encoding:

- qubit (basic) encoding?: given a vector $\mathbf{z}\in\mathbb{Z}^N$, $|\mathbf{z}\rangle = \bigotimes_i^N |\mathbf{x}_i\rangle$ where $\mathbf{x}_i$ is the binary representation of z_i, need $\mathcal{O}(N\cdot\log(\max(z_i)))$ (not space efficient)
- amplitude encoding: given a normalized vector $\mathbf{x}\in\mathbb{R}^N$, the quantum state $|\mathbf{x}\rangle = \sum_z^N x_z |z\rangle$. need $\log N \equiv n$ qubits for a data point;
- graph state encoding: graph state, discrete, efficient? space (time)
- quantum data $|\psi\rangle$ or ρ: quantum many-body state from real-world experiments: quantum state from quantum circuits $\hat{U}$
- (quantum) oracle: $\hat{O}|i,a\rangle = |i, a+x_i\rangle$; quantum oracle: unitary and controlled unitary

Definition 16 (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

$$k(\mathbf{x},\mathbf{x}') := \langle\phi(\mathbf{x}),\phi(\mathbf{x}')\rangle \tag{11}$$

with the feature map $\phi(\mathbf{x}):\mathbb{R}^d\to\mathbb{R}^n$ ($n>d$) from a low dimensional space to a higher dimensional space. The corresponding kernel matrix $\mathbf{K}$ should be a positive, semidefinite (PSD) matrix.

- *Gaussian kernel*;

$$k(\mathbf{x}, \mathbf{x}') := \exp\Big(-\|\mathbf{x} - \mathbf{x}'\|^2/(2\beta)\Big) \tag{12}$$

note that infinite dimensional feature map

- *graph kernel*: given a pair of graphs (G, G')

$$k(G, G') = \tag{13}$$

quantum graph kernel $k(G, G') = |\langle G|G'\rangle|^2$??

- quantum kernel (transition amplitude / quantum propagator);

$$k_Q(\rho, \rho') := |\langle \phi(x)|\phi(x')\rangle|^2 = \Big|\langle 0|\hat{U}^\dagger_{\phi(x)}\hat{U}_{\phi(x')}|0\rangle\Big|^2 = \mathrm{Tr}(\rho\rho') \tag{14}$$

with quantum feature map $\phi(x) : \mathcal{X} \to |\phi(x)\rangle\langle\phi(x)|$

- *shadow kernel*: given two density matrices ρ and ρ'

$$k_{\mathrm{shadow}}(\rho, \rho') := \tag{15}$$

- neural tagent kernel [15]:

measures

Definition 17 (divergence)**.** KL divergence (relative entropy): measure the distance (similarity) between two probability distributions:

$$D_{\mathrm{KL}}(P||Q) := \sum P(x)\log(P(x)/Q(x)) \tag{16}$$

symmetric version: Jensen-Shannon divergence (machine learning)

$$D_{\mathrm{JS}}(P||P') := (D_{\mathrm{KL}}(P||M) + D_{\mathrm{KL}}(P'||M))/2 \equiv H_S(M) - \frac{1}{2}(H_S(P) - H_S(P')) \tag{17}$$

where $M = (P + P')/2$ and Shannon entropy H_S. Analoguously, quantum Jensen-Shannon divergence

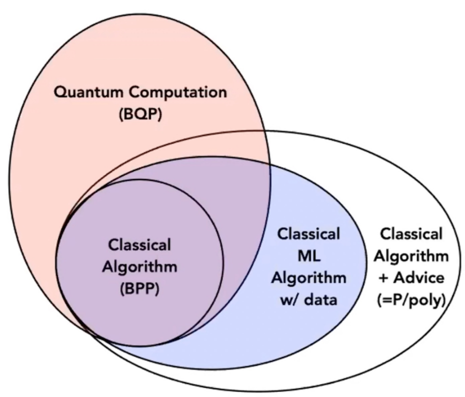

FIG. 2: computational model powered by training data

power of data -

Theorem 7 (informal [16])**.** *data learning*

- *machine learning (strictly) more powerful than BPP*
- *exist quantum advantage in machine learning (not significant, practical)*

Algorithm III.1: Classical learning (SVM)

input : labeled features (data)
output: entanglement structure

```
for i = 1, 2, ..., m do
    kernel estimation                // a comment
    return "?"
return ?
```

2. *Quantum trace estimation and kernel estimation*

Definition 18 (Quantum trace estimation)**.** The task of estimating quantities like

$$\mathrm{Tr}(\rho_1 \cdots \rho_m) \qquad \text{(multivariate traces)}$$

given access to copies of the quantum states ρ_1 through ρ_m.

Theorem 8. *[17] multivariate trace estimation can be implemented in constant quantum depth, with only linearly-many controlled two-qubit gates and a linear amount of classical pre-processing*

B. Variational quantum circuits

1. *Variational quantum kernel estimation*

an ansatz

$$\hat{W}_a := \sum_i a_{..} \bigotimes_i^n \hat{\sigma}^{(i)}, \quad \hat{\sigma} \in \{\hat{\sigma}_x, \hat{\sigma}_y, \hat{\sigma}_z, I\} \tag{18}$$

Algorithm III.2: Entanglement witness by ...

input : density matrix ρ
output: determine entangled structure??

```
for i = 1, 2, ..., m do
    W_i                              // this is a comment
    return "separable?"
return entangled ?
```

2. *Variational trace estimate*

find optimal entanglement witness (qunatum circuit?) [18] [19] [20]

C. Theoretic upper bounds and lower bounds

[16] [11] [10] [21] [13]

Definition 19 (graph property)**.** monotone

Problem 8 (graph property test)**.**

quantum advantages:

- no input encoding problem [22] in most quantum machine learning algorithm.
- contrived problem? for exponential speedup
- convex body query? complexity

obstacles: (i)

	gate/depth/computation	measurements/samples	query?	necessary?sufficient
shadow tomography		Theorem 5	N/A	
indirect? direct (no prior), promise				
entanglement witness (Section II A 3)			constant	
classical ML + SC (Section III A 1)				
quantum (variational) circuits	c-depth?			

TABLE I: complexity measures of different methods

IV. NUMERICAL SIMULATION

A. Classification accuracy

1. Data preparation

multi-partite entangled state: generate synthetic data from

2. Results

performance of different methods:

B. Robustness to noise

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

$$\rho'_{\text{noise}} = (1 - p_{\text{noise}}) |G\rangle\langle G| + p_{\text{noise}} \frac{\mathbb{1}}{2^n} \tag{19}$$

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

Remark 5. the largest noise tolerance p_{limit} just related to the **chromatic number** of the graph.[??] graph property

V. CONCLUSION AND DISCUSSION

todo

- experiment (generation, verification) [23]
- error correction? not benchmark

Acknowledgements

[1] R. Horodecki, P. Horodecki, M. Horodecki, and K. Horodecki, Rev. Mod. Phys. **81**, 865 (2009), arXiv:quant-ph/0702225.
[2] L. Gurvits, Classical deterministic complexity of Edmonds' problem and Quantum Entanglement (2003), arXiv:quant-ph/0303055.
[3] O. Gühne and N. Lütkenhaus, Phys. Rev. Lett. **96**, 170502 (2006).
[4] H. J. Briegel, D. E. Browne, W. Dür, R. Raussendorf, and M. V. den Nest, Nature Phys **5**, 19 (2009), arXiv:0910.1116.
[5] Y. Zhou, Q. Zhao, X. Yuan, and X. Ma, npj Quantum Inf **5**, 83 (2019).
[6] 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.
[7] Y. Zhang, Y. Tang, Y. Zhou, and X. Ma, Phys. Rev. A **103**, 052426 (2021), arXiv:2012.07606 [quant-ph].

[8] J. Altepeter, E. Jeffrey, and P. Kwiat, in *Advances In Atomic, Molecular, and Optical Physics*, Vol. 52 (Elsevier, 2005) pp. 105–159.
[9] J. Haah, A. W. Harrow, Z. Ji, X. Wu, and N. Yu, IEEE Trans. Inform. Theory , 1 (2017).
[10] S. Aaronson, in *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2018 (Association for Computing Machinery, New York, NY, USA, 2018) pp. 325–338, arXiv:1711.01053.
[11] H.-Y. Huang, R. Kueng, and J. Preskill, Nat. Phys. **16**, 1050 (2020), arXiv:2002.08953 [quant-ph].
[12] 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].
[13] Y. Liu, S. Arunachalam, and K. Temme, Nat. Phys. **17**, 1013 (2021), arXiv:2010.02174 [quant-ph].
[14] 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].
[15] A. Jacot, F. Gabriel, and C. Hongler, Neural Tangent Kernel: Convergence and Generalization in Neural Networks (2020), arXiv:1806.07572 [cs, math, stat].
[16] 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].
[17] Y. Quek, M. M. Wilde, and E. Kaur, Multivariate trace estimation in constant quantum depth (2022), arXiv:2206.15405 [hep-th, physics:quant-ph].
[18] 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].
[19] 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.
[20] M. Schuld and N. Killoran, Phys. Rev. Lett. **122**, 040504 (2019), arXiv:1803.07128 [quant-ph].
[21] H.-Y. Huang, R. Kueng, and J. Preskill, Phys. Rev. Lett. **126**, 190505 (2021), arXiv:2101.02464 [quant-ph].
[22] E. Tang, Phys. Rev. Lett. **127**, 060503 (2021), arXiv:1811.00414 [quant-ph].
[23] 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).

Appendix A: Machine learning background

Appendix B: Hardness assumptions