A survey on the complexity of learning quantum states

Anurag Anshu
Harvard University
anuraganshu@fas.harvard.edu

Srinivasan Arunachalam
IBM Quantum, Almaden Research Center
Srinivasan.Arunachalam@ibm.com

Abstract

We survey various recent results that rigorously study the complexity of learning quantum states. These include progress on quantum tomography, learning physical quantum states, alternate learning models to tomography and learning classical functions encoded as quantum states. We highlight how these results are paving the way for a highly successful theory with a range of exciting open questions. To this end, we distill 25 open questions from these results.

1 Introduction

In the last decade, machine learning has received tremendous attention with the success of deep neural networks (or in more generality deep learning) in practically relevant tasks such as natural language processing, speech recognition and image processing. Some popular applications of deep learning include AlphaGo and AlphaZero (to play the games of Go and chess), chatGPT (to mimic a human conversation) and AlphaFold (for solving instances of protein folding) [JEP+21, VSP+17, SSS+17]. Although these machine learning techniques work very well in practice, they are not well understood from a theoretical standpoint. In a seminal work in 1984, Valiant [Val84] introduced the well-known probability approximately correct (PAC) model of learning, which laid the mathematical foundation to understand machine learning from a computational complexity theory perspective. Since then, several mathematical models for machine learning have been proposed, some of which have theoretically justified the successes of practical learning algorithms. The study of machine learning from this complexity theoretic perspective is often referred to as *computational learning theory*.

In another line of research, a century old quest which includes physicists, mathematicians and - now - computer scientists, is understanding the dividing line between simple and complex quantum states. Some prominent measures of complexity have been formulated in this process - for instance, correlation length and entanglement entropy [ECP10] from the physics point of view; quantum circuit size and description size [Aar16] from the computer science point of view. A recent revolution in quantum information - inspired by practical implementations of quantum devices and incredible success of machine learning - has brought another measure in picture: *learnability*. In the last decade, there have been several works to understand what classes of quantum states are learnable efficiently and why some classes of states are hard to learn. Here we argue that learnability as a complexity-theoretic metric is remarkably powerful and has been revealing fundamentally new properties of physically and computationally relevant quantum states. This is akin to the aforementioned PAC learning framework used to understand machine learning from a complexity theoretic framework.

A general formalism for learning quantum states is as follows. A learning algorithm (which we often refer to as a learner) receives many independent copies of an unknown quantum state - guaranteed to be within a "class" of states (known to the learner). Using quantum measurements, the learner extracts information about the unknown state, and then outputs a sufficiently accurate description of the quantum state. We stress on the three defining notions in this general framework: the class of states, the type of measurement done by the learner and the metric for accuracy. Modifying any one of these parameters can change the quantum learning model in an interesting way and we discuss these models in this survey. The complexity metric associated with these learning models is the quantum *sample complexity*, defined as the number of copies of the unknown state used by the learning algorithm and quantum *time complexity*, defined as the total number of gates used by the algorithm.[1]

1.1 Organization of this survey

Our survey discusses these learning models that come with rigorous guarantees on the sample and time complexity, as detailed below.

1. *Learning arbitrary quantum states.* Here the goal is to learn an arbitrary n-qubit quantum state ρ, given copies of ρ, up to small trace distance. Given the generality of this task, the sample complexity of this task is known to be exponential in n. We discuss this in Section 2.
2. *Learning physical quantum states.* A natural followup question is, can we learn *interesting* subclasses of quantum states efficiently? In this direction we look at stabilizer states, states from the Clifford hierarchy, Gibbs states at different temperature regimes and matrix product states. We discuss this in Section 3.
3. *Learning states in alternate models.* Suppose the goal of the learner was to still learn an unknown quantum state, can we weaken the requirement for the learner and still learn the unknown ρ? To this end, there are models of learning called PAC learning, online learning, shadow tomography and several equivalences between them. We discuss this in Section 4.
4. *Learning classical functions encoded as states.* Suppose the unknown state ρ encodes a classical function, what is the complexity of learning? Here we discuss known results on quantum PAC learning, agnostic learning, statistical query learning and kernel methods which encode classical data into quantum states, and exhibit the strengths and weaknesses of quantum examples for learning classical functions. We discuss this in Section 5

Finally we conclude in Section 6 with some perspective on other works related to sample and time complexity of learning quantum states. Throughout this survey we have put together several open questions that would improve our understanding on the complexity of quantum states from the perspective of learning theory.

2 Tomography

Quantum state tomography (QST) is the following task: given many independent copies of an unknown n-qubit quantum state ρ living in $\mathbb{C}^d$ where $d = 2^n$,[2] output a $\hat{\rho}$ such that $\|\hat{\rho}-\rho\|_{tr} \leq \delta$ (where $\|\cdot\|_{tr}$ is the trace norm). Understanding the sample complexity of QST has been a fundamental

[1] In this survey, we will also discuss classical sample and time complexity and it's definition will be clear when we discuss these complexities.

[2] An n-qubit quantum state is a positive semi-definite operator on $\mathbb{C}^{2^n}$ such that $\mathsf{Tr}[\rho] = 1$.

question in quantum information theory with applications in tasks such as verifying entanglement [KvBE+21], understanding correlations in quantum states [CPF+10a], and is useful for understanding, calibrating and controlling noise in quantum devices. A simple protocol for QST uses $T = O(d^6)$ copies: simply let $P_1, \ldots, P_{d^2}$ be all the d-dimensional Pauli matrices, use $O(d^2/\delta)$ copies of ρ to estimate $\mathsf{Tr}(P_i\rho)$ up to error δ/d^2. Using a technique of linear inversion, this is sufficient to produce $\hat{\rho}$ that satisfies $\|\hat{\rho} - \rho\|_{tr} \leq \delta$. The overall sample complexity is $d^2 \cdot O(d^4/\delta) = O(d^6/\delta)$. The dependence on the error δ is intuitive as more accurate description requires more measurements. Subsequently [FGLE12] used techniques from compressive sensing to improve the complexity to $O(d^4/\delta^2)$ and after that Kueng et al. [KRT17] used more sophisticated techniques to improve the sample complexity to $O(d^3/\delta^2)$ and it was open for a while what was the right sample complexity of tomography. Two breakthrough works by Haah et al. [HHJ+17] and O'Donnell and Wright [OW16] finally obtained optimal bounds for the sample complexity of QST.

Theorem 1. *The sample complexity of QST up to trace distance δ is $O(d^2/\delta^2)$. Promised that the state is rank r, the sample complexity of QST up to infidelity ε is $\tilde{\Theta}(dr/\varepsilon)$.*[3]

We now give a proof overview of a special case of this theorem - when the quantum state is pure (rank $r = 1$). It makes use of the symmetric subspace and achieves a sample complexity of $\tilde{O}(d/\varepsilon)$. This is tight in d, as shown in [HHJ+17].

Special case of Theorem 1. Given an unknown d dimensional pure state $|\psi\rangle^{\otimes k}$, with k yet undetermined, note that the state lives inside the symmetric subspace $\Pi^{d,k}_{sym}$. To determine the state, one can perform the so-called *pretty-good measurement* [EF01], which has (continuous) POVM elements $\{|\phi\rangle\langle\phi|^{\otimes k}\}_{|\phi\rangle \in \mathbb{C}^d}$. Note that this measurement has infinitely many outcomes, which is ill defined, but we can address this by appropriate discretization. As a consequence, the measurement to be performed is

$$X \to \binom{d+k-1}{k} \int_{\phi} d\phi \, |\phi\rangle\langle\phi|^{\otimes k} X |\phi\rangle\langle\phi|^{\otimes k} \otimes |\text{description of } \phi\rangle\langle\text{description of } \phi|,$$

which is a valid POVM whenever X is in the symmetric subspace. The factor $\binom{d+k-1}{d}$ is the dimension of the symmetric subspace and ensures that the measurement is trace-preserving. Given $|\psi\rangle\langle\psi|^{\otimes k}$ as input, observe that a state $|\phi\rangle\langle\phi|$ is output with probability

$$\binom{d+k-1}{k} \langle\phi|^{\otimes k} |\psi\rangle\langle\psi|^{\otimes k} |\phi\rangle^{\otimes k} = \binom{d+k-1}{k} |\langle\phi|\psi\rangle|^{2k}.$$

Thus, the probability that $|\langle\phi|\psi\rangle| \leq 1 - \varepsilon$ is at most

$$\binom{d+n-1}{n} \int_{\phi : |\langle\phi|\psi\rangle| \leq 1-\varepsilon} d\phi |\langle\phi|\psi\rangle|^{2k} \leq \binom{d+k-1}{k} \cdot (1-\varepsilon)^{2k} \leq \left(e \cdot \frac{k+d-1}{d}\right)^d e^{-2k\varepsilon}.$$

Choosing $k = \frac{10d}{\varepsilon} \log \frac{1}{\varepsilon}$, we can guarantee that RHS is small. □

In order to go from the special case to the theorem above, [HHJ+17, OW16] consider a generalization of this argument and proceed by looking at subspaces that are invariant under permutations of registers and local unitary action. We refer the interested reader to [Wri16, OW18] for a detailed exposition of the general proof. Very recently, the work of Flammia and O'Donnell [FO23]

[3] Infidelity between quantum states ρ, σ is defined as $1 - \|\sqrt{\rho}\sqrt{\sigma}\|_{tr}$.

considered the sample complexity of tomography under various distance metrics. A drawback of these tomography algorithms is that the time complexity of the procedure scales exponentially in d (i.e., doubly-exponentially in n). A natural question that was open from their work was, is there a *time-efficient* procedure for tomography? In particular, is it possible to solve QST using only single-copy measurements? There were a few works in this direction recently [Yue22, LN22] and very recently Chen et al. [CHL+22a] answered this question with a surprisingly short proof.

Theorem 2. *The sample complexity of QST using single copy measurements is* $\Theta(d^3/\delta^2)$.

The upper bound comes from the result of Kueng et al [KRT17] and Chen et al. [CHL+22a] proved the lower bound of $\Omega(d^3/\delta^2)$ for QST with single-copy (and, adaptive) measurements. We now sketch their lower bound. A technical challenge they had to overcome was the following: prior works that established sample lower bounds, proved this in the context of *property testing*, where they proved the hardness between distinguishing two hard distributions over states whose statistics (on separable measurements) were far apart. However, for tomography there are not too many techniques that we know to prove lower bounds against separable measurements. In this paper they use the so-called "learning-tree framework" (which was first used in the prior work of Chen et al. [CCHL21] and inspired by classical decision trees which is used to analyze query complexity [BW02]) to prove their lower bounds. Here there is a tree where each node corresponds to a measurement operator applied onto a copy of the unknown state and the leaves are given by classical bit string, corresponding to measurement labels. Based on the classical output in the leaves, the algorithm outputs an hypothesis state σ. The depth of the tree is the sample complexity of the learning algorithm. Chen et al. [CHL+22a] construct a hard distribution of quantum states based on Gaussian ensemble matrices and their main technical contribution is to show the following: if a separable tomography protocol is run on this hard instance, the leaves of the decision tree above (i.e., the output quantum state σ) is anti-concentrated around the unknown target quantum state if the depth of the tree is $o(d^3)$. Proving this anti-concentration is non-trivial, however the proof is fairly short and we refer to their work for more.

We conclude this section by discussing a simpler problem than QST: quantum *spectrum estimation*. Here the goal is to learn the spectrum of an unknown quantum state ρ, given copies of ρ. It was showed [OW16] that $O(d^2/\varepsilon^2)$ copies of ρ suffices to estimate the spectrum of ρ up to ℓ_1 distance ε and they also showed a lower bound of $\Omega(d/\varepsilon^2)$.[4] Spectrum learning has been an important subroutine in several property testing algorithms [OW15, OW16, Wri16, OW17]. One question that remains open is the following:

Question 1. *What is the tight sample complexity of quantum spectrum estimation?*

3 Learning physical quantum states

In the previous section we saw that fully learning arbitrary quantum states could require exponentially many copies of the unknown state. A natural question is, are there *physical* subclasses of quantum states which can be learned using polynomially many copies (and even polynomial time)? In this section, we discuss a few classes of physical states that can be learned using polynomial sample or time complexity.

[4] They also showed that a *class of algorithms* based on Schur sampling require an $\Omega(d^2/\varepsilon^2)$ sample complexity.

3.1 Stabilizer states

A natural candidate class that was considered for efficient tomography were states that are known to be classically simulable. To this end, one of the first classes of states that were known to be learnable in polynomial time are stabilizer states. These are n-qubit states produced by the action of n-qubit Clifford circuits acting on $|0^n\rangle$. Aaronson and Gottesman [AG04, AG08] considered this question and showed the following theorem.

Theorem 3. *The sample complexity of exactly learning n-qubit stabilizer states is $O(n)$ and the time complexity is $O(n^3)$.*

In their paper, [AG04] also showed that with single-copy measurements $O(n^2)$ copies of a stabilizer state $|\psi\rangle$ suffice to learn $|\psi\rangle$. Subsequently, Montanaro [Mon17a] gave a fairly simple procedure to learn stabilizer states using $O(n)$ copies that only uses entangled measurements over 2 copies (prior to his work, Low [Low09] showed how to learn stabilizer states when one is allowed to make queries to the Clifford circuit preparing the unknown stabilizer state). We now discuss Montanaro's protocol: it is well-known [DDM03, Nes08] that every n-qubit stabilizer state can be written as $|\psi\rangle = \frac{1}{\sqrt{|A|}} \sum_{x \in A} i^{\ell(x)} (-1)^{q(x)} |x\rangle$, where $A \subseteq \{0,1\}^n$ is a subspace and ℓ (resp. q) is a linear (resp. quadratic) polynomial over $\mathbb{F}_2$ in the variables $x_1, \ldots, x_n$.

Special case of Theorem 3. We consider the case when $\ell(x) = 1$ for all x. Without loss of generality we can assume that $A = \{0,1\}^n$ as well: a learning algorithm can measure $\tilde{O}(n)$ copies of $|\psi\rangle$ in the computational basis, learn the basis for A and apply an invertible transformation to convert $|\psi\rangle$ to $\sum_{x \in \{0,1\}^k \times 0^{n-k}} (-1)^{q(x)} |x\rangle$ where $\mathrm{rank}(A) = k$ and now apply a learning procedure on states of the form $|\phi\rangle = \frac{1}{\sqrt{2^k}} \sum_{x \in \{0,1\}^k} (-1)^{q(x)} |x\rangle$. With this assumption, the learning algorithm uses the so-called Bell-sampling procedure: given two copies of $|\phi_q\rangle = \frac{1}{\sqrt{2^n}} \sum_x (-1)^{q(x)} |x\rangle$ where $q(x) = x^\top B x$ (where $B \in \mathbb{F}_2^{n \times n}$), perform n CNOTs between the first copy and second copy, and measure the second copy. One obtains a uniformly random $y \in \mathbb{F}_2^n$ and the state

$$\frac{1}{\sqrt{2^n}} \sum_x (-1)^{f(x)+f(x+y)} |x\rangle = \frac{(-1)^{y^\top A y}}{\sqrt{2^n}} \sum_x (-1)^{x^\top (B+B^\top) \cdot y} |x\rangle.$$

The learning algorithm then applies the n-qubit Hadamard transform and measures to obtain bit string $(B + B^\top) \cdot y$. Repeating this process $O(n \log n)$ many times, one can learn n linearly independent constraints about B. Using Gaussian elimination, allows one to learn the off-diagonal elements of B. To learn the diagonal elements of B, a learner applies the operation $|x\rangle \to (-1)^{x_{ij}} |x\rangle$ if $B_{ij} = 1$ for $i \neq j$. Repeating this for all $i \neq j$, the resulting quantum state is $\sum_x (-1)^{\sum_i x_i B_{ii}} |x\rangle$. Again applying the n-qubit Hadamard transform, the learner learns the diagonal elements of B. □

Given that stabilizer states are learnable using $O(n)$ copies, a followup question which hasn't received much attention is the following.

Question 2. *The stabilizer rank of $|\psi\rangle$ is the minimum k for which $|\psi\rangle = \sum_i \alpha_i |\phi_i\rangle$ where $|\phi_i\rangle$ is an n-qubit stabilizer state. Can we learn stabilizer rank-n states in polynomial time?*

Inspired by a result of Raz [Raz19] who proved time-space tradeoffs for parity learning, we also pose the following question.

Question 3. *The standard Bell-sampling approach for learning stabilizer states uses $O(n)$ copies of the stabilizer state and $O(n^2)$ classical space. If we have $o(n^2)$ classical space, what is the sample*

complexity of learning stabilizer states? Similarly, can we prove sample-space tradeoffs when the algorithm is given quantum space?[5]

3.2 Learning circuits with non-Clifford gates

We saw how to learn the output states of Clifford circuits; a natural question is, if the circuit consists of a few *non-Clifford* T gates, can we still learn the output state? It is known that that Clifford+T circuits are universal for quantum computation and they have received much attention in fault-tolerance, circuit compilation and circuit simulation [SBB98, FMMC12, KMM13, Sel15, RS16, BSS16, BG16a, BBC+19]. An arbitrary quantum circuit can be decomposed as a alternating sequence of Clifford stages and T stages (by Clifford stage, we mean a Clifford circuit and by T-stage we mean either a T gate or identity is applied to each qubit). The number of T stages is the *T-depth* of the circuit. The learning task we consider is: Suppose U is an n-qubit quantum circuit belonging to the class of T depth-one circuits, can one learn U? In particular, if we are allowed to apply U to specified prepared states and measure under a class of POVMs, how many measurements are required for learning U? In [LC22], they proved the following theorem.

Theorem 4. *Let U be an n-qubit T-depth one quantum circuit comprising of $O(\log n)$ many T gates. There exists a procedure that makes $\mathsf{poly}(n)$ queries to U and outputs a circuit $\tilde{U}$ that is equivalent to U when the input states are restricted to the computational basis.*

We omit the proof of this theorem and refer the reader to [LC22] for more details. Recently, there was a hardness for learning the output distributions of Clifford circuits with a *single* T gate [HIN+22],[6] it is surprising that T-depth 1 circuits are learnable in polynomial time. This theorem naturally motivates the following questions.

Question 4. *What is the complexity of learning circuits with T-depth t (for some $t \geq 2$)?*

Recently, there have been a few works by Grewal et al. [GIKL23a, GIKL23b, GIKL23c] where they showed polynomial-time algorithms for learning states prepared by Clifford circuits with $O(\log n)$ many T gates. They are also able to learn the output states of such circuits in polynomial time.[7]

Question 5. *Can we learn n-qubit states and circuits that consist of $\omega(\log n)$ many T gates? If not, is there a conditional hardness result one could show for learning these states?*

3.3 Learning phase states

In this section, we consider learning classical low-degree Boolean functions encoded as the amplitudes of quantum states, aka *phase states*, which can be viewed as a generalization of stabilizer states. In recent times phase states have found several applications in cryptography, pseudorandomness, measurement-based quantum computing, IQP circuits, learning theory [JLS18, BS19, INN+22, AQY21, RHBM13, TMH19, ABDY22].

A degree-d *binary phase state* is a state of the form $|\psi_f\rangle = 2^{-n/2}\sum_{x\in\{0,1\}^n}(-1)^{f(x)}|x\rangle$ where $f : \{0,1\}^n \to \{0,1\}$ is a degree-d function. Similarly, a degree-d *generalized phase state* is a state of the form $|\psi_f\rangle = 2^{-n/2}\sum_{x\in\{0,1\}^n}\omega_q^{f(x)}|x\rangle$ where $f : \{0,1\}^n \to \mathbb{Z}_q$ is a degree-d polynomial,

[5] Recently, Liu et al. [LRZ23] showed that an algorithm for learning parities needs either $\Omega(n^2)$ classical space, $\Omega(n)$ quantum space or $\Omega(2^n)$ labelled examples.

[6] The hardness result [HIN+22] is in a weaker *statistical query* model which we discuss in Section 5.2, whereas the positive result [LC22] considers the standard tomography model wherein the learner is given copies of the state.

[7] We remark that the states produced by these circuits have stabilizer rank $\leq n$, still leaving open the question we asked in the previous section.

$\omega_q = e^{2\pi i/q}$ and q is a prime. It is known that the output state of a random n-qubit Clifford circuit is a generalized $q = 4$, degree-2 phase state with a constant probability [BG16b], and a generalized degree-d phase states with $q = 2^d$ can be prepared from diagonal unitaries in the d-th level of the Clifford hierarchy [GC99, CGK17]. The learning question is: how many copies of $|\psi_f\rangle$ suffice to learn f exactly? Earlier works [BV97, Mon17a, Röt09] showed $O(n)$ samples suffice for learning degree-1, and $O(n^2)$ suffices for learning degree-2 binary phase states; learning degree-d for $d \geq 3$ has remained open (in fact it was plausible that it was a hard learning task since IQP circuits produce degree-3 phase states [Mon17b, BJS11]). Sample complexity of learning *generalized* phase states had not been studied before. In a recent work [ABDY22] they provided separable and entangled bounds for learning phase states. Below we sketch the upper and lower bounds for the case of separable measurements. We refer the reader to [ABDY22] for the proof of the sample complexity with entangled measurements.

Separable measurements *upper bound.* The proof makes the following simple observation: given $|\psi_f\rangle = 2^{-n/2}\sum_x \omega_q^{f(x)}|x\rangle$, suppose we measure qubits $2, 3, \ldots, n$ in the computational basis and obtain $y \in \{0,1\}^{n-1}$. The post-measurement state is then $|\psi_{f,y}\rangle = (\omega_q^{f(0y)}|0\rangle + \omega_q^{f(1y)}|1\rangle)/\sqrt{2}$. If the base of the exponent was (-1), then applying a Hadamard on $|\psi_{f,y}\rangle$ produces $c = f(0y) - f(1y)$. Their main idea is, it is still possible to obtain a value $b \in \mathbb{Z}_q$ such that $b \neq c$ with certainty. To this end, consider a POVM whose elements are given by $\mathcal{M} = \{|\phi_b\rangle\langle\phi_b|\}_{b\in\mathbb{Z}_q}$, where $|\phi_b\rangle = (|0\rangle - \omega_q^b|1\rangle)/\sqrt{2}$. Applying this POVM $\mathcal{M}$ onto an unknown state $(|0\rangle + \omega_q^c|1\rangle)/\sqrt{2}$ they observe that c is the outcome with probability 0 and furthermore one can show that *every* other outcome $b \neq c$ appears with probability $\Omega(d^{-3})$. Repeating this process $m = O(n^{d-1})$ many times, one obtains $(y^{(k)}, b^{(k)})$ for $k = 1, 2, \ldots, m$ such that $f(1y^{(k)}) - f(0y^{(k)}) \neq b^{(k)}$ for all $k \in [m]$. Let $g(y) = f(1y^{(k)}) - f(0y^{(k)})$ (i.e., $g = \nabla_1 f$). Clearly g is a degree $\leq d-1$ polynomial. A non-trivial analysis in [ABDY22] shows the following: the probability of having more than one polynomial degree-$d-1$ polynomial g satisfying the constraints $g(y^k) \neq b^k$ is exponentially small if we choose $k = \tilde{O}(q^3 n^{d-1})$. Hence k many copies $|\psi_f\rangle$, allows a learning algorithm to learn the derivative $\nabla_1 f$. Repeating this for n directions, we can learn $\nabla_1 f, \ldots, \nabla_n f$, hence f.

Separable measurements *lower bound.* Furthermore, they show that the above protocol is optimal even if allowed single *copy* measurements. The main idea is the following: for a uniformly random degree-d function f, suppose a learning algorithm measures the phase state $|\psi_f\rangle$ in an arbitrary orthonormal basis $\{U|x\rangle\}_x$. One can show that the distribution describing the measurement outcome x is "fairly" uniform. In particular, $\mathbb{E}_f[H(x|f)] \geq n - O(1)$, where $H(x|f)$ is the Shannon entropy of a distribution $P(x|f) = |\langle x|U^*|\psi_f\rangle|^2$. To prove this, they first lower bound the Shannon entropy by Renyi-two entropy and prove a technical statement to bound the latter by deriving an explicit formula for $\mathbb{E}_f[|\psi_f\rangle\langle\psi_f|^{\otimes 2}]$. Thus, for a typical f, measuring one copy of the phase state $|\psi_f\rangle$ provides at most $O(1)$ bits of information about f. Since a random uniform degree-d polynomial f with n variables has entropy $\Omega(n^d)$, one has to measure $\Omega(n^d)$ copies of $|\psi_f\rangle$ in order to learn f.

In [ABDY22] they also constructed a procedure to learn circuits (consisting of diagonal gates in the Clifford hierarchy) which produce phase states, leaving open the following:

Question 6. *What is the complexity of learning circuits consisting of* non-diagonal *gates in the Clifford hierarchy?*[8]

To this end, when given query access to the circuit, Low [Low09] gave a procedure to learn the

[8] When given query access to the circuit, Low [Low09] gave a procedure to learn the Clifford hierarchy.

Clifford hierarchy. However given only copies of $C|0^n\rangle$ where C consists of *non-diagonal* gates in the Clifford hierarchy, the question we ask is open. Liang [Lia22] recently showed a conditional hardness of learning Clifford circuits in the *proper learning* setting. An open question from [ABDY22] which might improve our understanding of phase states is, how many copies suffice to *test* phase states.

Question 7. *What is the complexity of property testing degree-d phase states? In particular, given copies of a state $|\psi\rangle$ promised it is either a degree-d phase state or ε-far from the set of all degree-d phase states, how many copies are necessary and sufficient to distinguish these cases?*

3.4 Gibbs states of local Hamiltonians

In this section we discuss the problem of learning a Hamiltonian given copies of its Gibbs state. The setup of this learning problem is as follows: let H be a local Hamiltonian $H = \sum_{\alpha=1}^{m} \mu_\alpha E_\alpha$ on n qubits, where E_α is some local orthogonal operator basis such as the Pauli matrices, an algorithm receives copies of the Gibbs state $\rho_\beta(H) = \frac{e^{-\beta H}}{\mathsf{Tr}(e^{-\beta H})}$ and the goal is to output a list of numbers $\mu' := \{\mu'_1, \mu'_2, \ldots, \mu'_m\}$ that are close to $\mu := \{\mu_1, \mu_2, \ldots, \mu_m\}$ in either the ℓ_∞ or ℓ_2 distance metric. We make the natural assumption that $\mu_1, \ldots, \mu_m \in (-1, 1)$, which simply says that each local interaction has bounded strength. Learning an unknown Hamiltonian from its Gibbs state has been studied in statistical physics and machine learning [CL68, HS+86, Tan98, AS14] for many decades, known as the "inverse Ising problem".

For machine learning, one is often interested in Ising interaction (that is, each E_α is a Pauli operator of the form $Z \otimes Z$) where the underlying interaction graph[9] is sparse and unknown [Bre15, VMLC16, KM17]. Learning the Ising model also learns the very important underlying graph structure. In the quantum regime, we are far from being able to learn the underlying graph, solely under the sparsity assumption. Thus, we will assume that the underlying graph is known. For most physics applications, the graph can also respect the geometric constraints that arise from living in a low dimensional space. Before discussing algorithms for Hamiltonian learning, we first discuss motivation for considering this learning question.

Hamiltonian learning can be a useful experimental tool in a variety of settings.

- **Understanding the lattice structure:** Suppose we wish to know whether interactions in a given quantum material respect a Kagome lattice structure or a square lattice structure, assuming one of them is the case. This knowledge can significantly affect the physical properties, such as the electronic behaviour as looked at by [JYD+21]. If our Hamiltonian learning algorithm guarantees that $\|\mu' - \mu\|_\infty \leq \frac{1}{3}$, then we can figure out which edge is present or absent, in turn the lattice structure.
- **Estimating the spectral gap of a Hamiltonian:** Another key quantity of interest is the spectral gap of a Hamiltonian, which dictates a myriad of ground state properties. For learning the spectral gap up to constant precision error (say 0.1), we need to know the Hamiltonian really well and the right regime to consider is $\|\mu' - \mu\|_1 \leq 0.1$.
- **Effective Hamiltonians:** Local Hamiltonians are - after all - models of real interactions happening in physics. Effective Hamiltonians regularly arise when we wish to consider interactions between a specific set of particles or quasi-particles. These interactions can be hard to precisely determine theoretically, motivating the use of Hamiltonian learning [SHB+22].

[9] An interaction graph has the qubits in the Hamiltonian as its vertices and each Ising interaction as its edge.

- **Entanglement Hamiltonian:** Li-Haldane conjecture states that the marginals of a 2D gapped ground state are Gibbs state of a local Hamiltonian with temperature that depends on the location of the local term. Learning this Hamiltonian is directly relevant to understanding the entanglement structure of the system [KvBE+21].

We remark that in the applications above, we did not specify the inverse temperature β of the Gibbs state. In some cases, the temperature can be controlled, and then setting β to be a small constant leads to optimal algorithms - see below. In other cases, such as for effective or entanglement Hamiltonians, temperature can be very low at the boundary of the region. Thus, efficient algorithms for Hamiltonian learning at all finite temperatures has interesting consequences in quantum computing.

3.4.1 Sufficient statistics

We now turn to designing algorithms for the learning task above. One natural question is: given an instance of Hamiltonian learning problem, is there any data about the Gibbs state that would suffice to learn the Hamiltonian? In other words, what are the '*sufficient statistics*' for the Hamiltonian? The answer turns out to be very simple: they are the set of expectation values $f_\alpha = \mathsf{Tr}(E_\alpha \cdot \rho_\beta(H))$.

There are two ways to prove that sufficient statistics suffice for learning

- **Information theoretic argument:** Let's consider two Gibbs quantum states $\rho_\beta(H)$ and $\rho_\beta(G)$, where $H = \sum_\alpha \mu_\alpha E_\alpha$ and $G = \sum_\alpha \nu_\alpha E_\alpha$. We will argue that their "distance" is characterized by the expectation values. For this, we evaluate the symmetric relative entropy

$$\begin{aligned} \mathrm{S}\left(\rho_\beta(H)\|\rho_\beta(G)\right) + \mathrm{S}\left(\rho_\beta(G)\|\rho_\beta(H)\right) &= \beta \mathsf{Tr}((H-G)(\rho_\beta(G)-\rho_\beta(H))) \\ &= \beta \sum_\alpha (\mu_\alpha - \nu_\alpha)\mathsf{Tr}(E_\alpha(\rho_\beta(G)-\rho_\beta(H))), \end{aligned}$$

  where the first equality follows by routine calculation. Thus, if $\mathsf{Tr}(E_\alpha \rho_\beta(H)) = \mathsf{Tr}(E_\alpha \rho_\beta(G))$ for all α, the right hand side vanishes. The above argument also says that if $\mathsf{Tr}(E_\alpha \rho_\beta(H)) \approx \mathsf{Tr}(E_\alpha \rho_\beta(G))$ then the relative entropy between the Gibbs quantum states is small. This is good enough to 'learn' the Gibbs state up to small error in total variational distance. More precisely,

$$\begin{aligned} &\mathrm{S}\left(\rho_\beta(H)\|\rho_\beta(G)\right) + \mathrm{S}\left(\rho_\beta(G)\|\rho_\beta(H)\right) \\ &\leq \beta \max_\alpha |\mu_\alpha - \nu_\alpha| \cdot \left(\sum_\alpha |\mathsf{Tr}(E_\alpha \rho_\beta(H)) - \mathsf{Tr}(E_\alpha \rho_\beta(G))| \right) \\ &\leq 2\beta \left(\sum_\alpha |\mathsf{Tr}(E_\alpha \rho_\beta(H)) - \mathsf{Tr}(E_\alpha \rho_\beta(G))| \right). \end{aligned}$$

  However, this estimate is not sufficient to guarantee the closeness of the Hamiltonians.

- **Convexity of log partition function:** A more useful argument - for our problem description - is based on the convexity of the log partition function. The observation here is simply that the function $\log \mathsf{Tr}(e^{-\beta H})$ is a convex function in the parameters $\{\mu_1, \mu_2, \ldots, \mu_m\}$. The vector $(f_1, f_2, \ldots, f_m)$ of trace expectations then forms the gradient of this function. Furthermore, precise knowledge of the gradient can be used to identify the parameters $\mu_1, \ldots, \mu_m$. See Figure 1 (a).

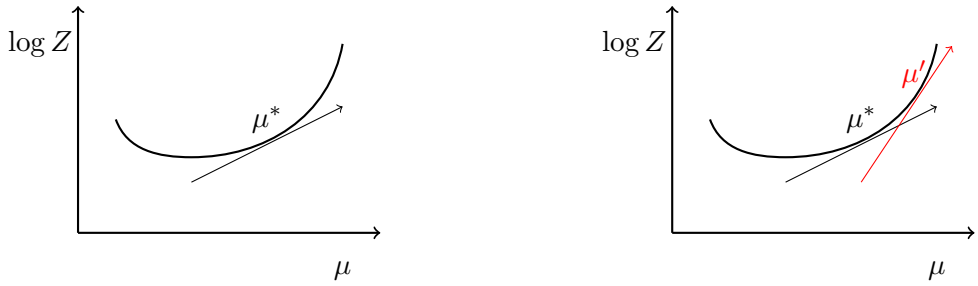


Figure 1: (a) Given the gradient of a convex function - such as the log partition function - there is a unique point that matches the gradient. (b) Strong convexity ensures that good knowledge of the gradient leads to good enough closeness to the desired point.

The quantities f_α can only be known approximately in experiments, due to statistical errors in estimation. Thus, a robust version of sufficient statistics is needed to develop an algorithm for Hamiltonian learning. In [AAKS21b], strong convexity of the log partition function was established. This roughly says that the log partition function "curves well" (see Figure 1 (b)). An algorithm - based on gradient descent - was constructed which uses $O\left(m^3 \cdot 1/\varepsilon^2 \cdot \mathsf{poly}(1/\beta) \cdot \exp(\mathsf{poly}(\beta))\right)$ copies of the Gibbs state to learn the Hamiltonian with guarantee $\|\mu' - \mu\|_2 \leq \varepsilon$. The time complexity of the algorithm depends on computing the gradient of the partition function. An efficient computation at high temperatures, for stoquastic Hamiltonians and 1D Hamiltonians - but requiring large run-time for low temperatures and arbitrary Hamiltonians.

3.4.2 Commuting Hamiltonians

While the above algorithm based on sufficient statistics takes exponential time at low temperatures, classical Hamiltonians can be learned time-efficiently using more refined techniques - as noted earlier [Bre15, VMLC16, KM17]. In fact, here we argue that *commuting* Hamiltonians - that include classical Hamiltonians - can also be efficiently learned at any temperature, as long as the interaction graph is known. The algorithm is fundamentally different from the previous one that was based on estimating the expectation values $f_\alpha = \mathsf{Tr}(E_\alpha \rho_\beta(H))$. Consider $H = \sum_\ell h_\ell$, where the commutator $[h_\ell, h_{\ell'}] = 0$. We note that this notation is different from the one we used earlier, in particular here the h_ℓs need not be an *orthogonal* basis. The algorithm, sketched in [AAKS21a] is based on the following theorem. See also Figure 2

Theorem 5. *[AAKS21a] For any region R on the lattice, define the effective reduced Hamiltonian $H_R = \frac{-1}{\beta} \log \mathrm{Tr}_{R^c}(\rho_\beta)$.[10] Let ∂R be the boundary of R, and $\partial_- R$ be the inner boundary of R (which is the set of qubits in R that interact with a qubit outside R). Then*

$$H_R = \alpha_R I + h_R + \Phi,$$

where Φ is only supported on $\partial_- R$ and $[\Phi, h_R] = 0$. Here, α_R is some real number and $\|\Phi\| \leq 2|\partial R|$.

Using this theorem, the learning algorithm is straightforward: perform good enough tomography of the region around an interaction h_ℓ to reconstruct the marginal to very high accuracy. Then take log of the marginal, followed by computing each h_ℓ up to error ε (i.e., output a h'_ℓ such that $\|h'_\ell - h_\ell\| \leq \varepsilon$). This is good enough to estimate the unknown Hamiltonian H. The resulting

[10] Here the subscript in Tr refers to the registers being traced out. Moreover, R^c is the set of qubits not in R.

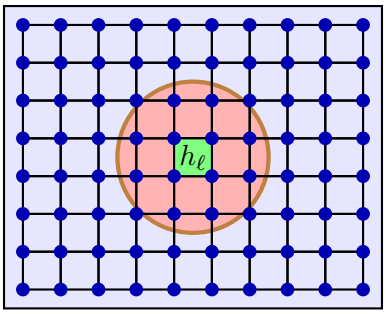


Figure 2: Consider the marginal of the Gibbs state in the brown circle. For commuting Hamiltonians, this marginal is the Gibbs state of a Hamiltonian that is the boundary correction to the Hamiltonian strictly within the region.

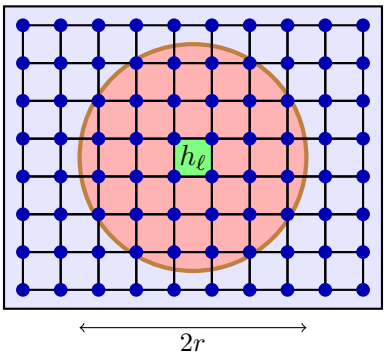


Figure 3: In the high temperature regime, it has been shown by [KKBa20] that the marginal of a Gibbs state is still the Gibbs state of the original Hamiltonian (within the brown circle of radius r) up to boundary correction. However, the boundary term has some support within the circle and has strength $\approx e^{-r}$ near the center. Thus, to learn h_ℓ, we need to make sure that e^{-r} is small enough.

sample complexity [AAKS21a] is $\exp(\mathcal{O}(\beta k^D)) \cdot \mathcal{O}\left(1/\varepsilon^2 \cdot \log(m/\delta)\right)$, where k is the locality of the Hamiltonian, D is the degree of the underlying interaction graph and δ is the probability of failure. Time complexity is $m \cdot \exp(\mathcal{O}(\beta k^D)) \cdot \mathcal{O}\left(1/\varepsilon^2 \cdot \log(m/\delta)\right)$.

3.4.3 High temperature Gibbs states

The idea of using effective reduced Hamiltonian in Theorem 5 can also be applied to non-commuting Hamiltonians, as long as the temperature is high enough (or β smaller than the critical temperature β_c, which is a constant). This follows from a similar result as Theorem 5 shown by [KKBa20] using cluster expansion, with H_R approximated by $h_\ell + \Phi$ as $\|H_R - h_\ell - \Phi\|_\infty \leq \exp(-\Omega(r))$ for a spherical region R of radius r around ℓ (see Figure 3 for an example).

If we use the same approach as above, to estimate each h_ℓ with error ε, we thus need $r = O(\log 1/\varepsilon)$. The sample complexity now incurs an additional factor of $\exp(r^D)$, where D is the lattice dimension or the degree of the graph. Thus, the sample complexity is $O\Big(\exp(\beta(\log \frac{1}{\varepsilon})^D) \cdot 1/\varepsilon^2 \cdot \log(m/\delta)\Big)$ and time complexity is $O\Big(m \cdot \exp((\log \frac{1}{\varepsilon})^D) \cdot 1/\varepsilon^2 \cdot \log(m/\delta)\Big)$. For constant ε, this is very efficient; however for $\varepsilon = \frac{1}{m}$, in which case the ℓ_1 error of learning is small enough, the sample complexity is super-polynomial in m. This is somewhat unsatisfactory, as this approach

seems worse than [AAKS21b] in the regime where each local term has to be learned very accurately. In a subsequent work [HKT22a] provided a unifying and complete answer for $\beta < \beta_c$. Employing the cluster expansion method of [KKBa20, KS20], the authors directly express the sufficient statistics $\mathsf{Tr}(E_\alpha \rho_\beta(H))$ as an infinite series in β with coefficients polynomial in the local Hamiltonian terms. Approximate knowledge of the sufficient statistics is then inverted to estimate the Hamiltonian terms. They achieve tight sample and time complexity of $\mathcal{O}\left(1/\varepsilon^2 \cdot \log(m/\delta)\right)$ and $\mathcal{O}\left(m/\varepsilon^2 \cdot \log(m/\delta)\right)$ respectively.[11]

3.4.4 Discussion

The problem of time efficient Hamiltonian learning - on a fixed geometry and at arbitrary β - remains open. The fact that this is possible in the commuting case is encouraging, as there is no prior reason to expect that the commuting and non-commuting cases would be fundamentally different. Indeed, very good heuristic methods exist for the task [BAL19, QR19]. We end this section with two relevant open questions.

Question 8. *Can we achieve Hamiltonian learning under the assumption that the Gibbs states satisfy an approximate conditional independence?*[12]

Approximate conditional independence is known to hold in 1D [KB19] and conjectured to hold for every dimension.

Question 9. *Pseudorandomness is a bottleneck for learnability. If a family of quantum states are pseudo-random, then polynomially many copies of the state are indistinguishable from Haar random states by any efficient quantum algorithm. Could low temperature Gibbs states be pseudorandom, which would explain the difficulty in finding time efficient algorithm?*

3.5 Matrix product states

Matrix Product States (MPS) are a widely used representation of quantum states on a spin-chain. Mathematically, a state $|\psi\rangle \in (\mathbb{C}^d)^{\otimes n}$ is a *matrix product state* (MPS) if $|\psi\rangle$ can be written as

$$|\psi\rangle = \sum_{i_1,\ldots,i_n \in [d]} \mathsf{Tr}(A^{(1)}_{i_1} \cdot A^{(2)}_{i_2} \cdots A^{(n)}_{i_n}) \, |i_1, \ldots, i_n\rangle,$$

where for all $j \in [n], i \in [d]$, $A^{(j)}_i$ is a $D_j \times D_{j+1}$ matrix. We call the set of matrices $\{A^{(j)}_i\}$ an *MPS representation of* $|\psi\rangle$. We refer to $D = \max_j D_j$ as the *bond dimension* of $|\psi\rangle$, when minimized over all MPS representations. Many physically relevant n-qubit quantum states - such as gapped ground states - can be approximated by MPS with bond dimension polynomial in n. A learning algorithm for an MPS state ρ takes as input, copies of ρ promised to be an MPS of certain bond dimension D and outputs an MPS of bond dimension D' that approximates ρ in fidelity. The goal is learn these states with polynomial sample and time complexity along with keeping D' close to D.

Unlike Gibbs states, local observable statistics do not always determine an MPS. For example, consider the CAT states $\frac{1}{\sqrt{2}}|00\cdots0\rangle \pm \frac{1}{\sqrt{2}}|11\cdots1\rangle$, which are MPS of bond dimension 2. These states can't be distinguished on any set of $n-1$ qubits. Thus, any algorithm for MPS must make global measurements. Indeed, [LCLP10, CPF+10b] gave a polynomial time algorithm to learn

[11] We note that there were gaps in the above results - that originated in [KS20] - were fixed in [WA22]. See [HKT22a] for a discussion.

[12] Given a quantum state ρ on registers A, B, C, $\mathrm{I}(A:C|B)_\rho = S(\rho_{AB}) + S(\rho_{BC}) - S(\rho_B) - S(\rho_{ABC})$ is the quantum conditional mutual information. We say that ρ satisfies approximate conditional independence if $\mathrm{I}(A:C|B)_\rho \approx 0$.

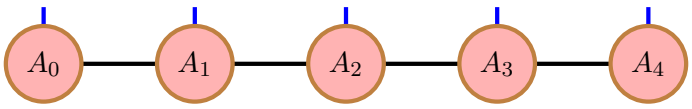


Figure 4: A matrix product state is specified by its bond dimension D and a sequence of $D \times D$ matrices. The virtual bonds (black lines) indicate the amount of entanglement and the physical blue lines represent qudits.

an MPS, using global-but-efficient measurements. Their algorithm learns a sequential circuit that prepares the MPS. The resulting output has bond dimension $D' = \text{poly}(D)$.

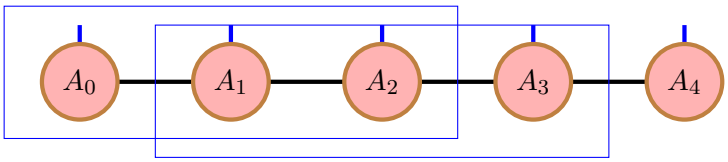


Figure 5: Under coarse graining (blue rectangles) the physical dimension exceeds the bond dimension. For example, if $D = 4$ and each physical blue line is a qutrit, the physical dimension of blue regions is $3^3 = 27$, which is larger than the total bond dimension at the boundary $4^2 = 16$. For typical tensors $A_1, \ldots, A_n$, this makes the map from the virtual bonds to physical qudits invertible. Such an MPS is injective.

However, local measurements are more ideal in the experimental settings. Under the assumption of injectivity (see Figure 5) learning an MPS with just local measurements may be possible. Let us first observe that a 'local' sufficient statistics holds for injective MPS. This is because the marginals on $O(\log D)$ qudits determine the parent Hamiltonian which has the MPS as its unique ground state. Barring the statistical errors - which can be addressed using injectivity - the knowledge of parent Hamiltonian allows one to reconstruct an approximation to the MPS state using the rigorous algorithm in [LVV15, ALVV17]. A drawback of this approach is that the algorithm in [LVV15, ALVV17] outputs an MPS with bond dimension $D' = \text{poly}(n)$, which may be much larger than a constant D. Thus, we see a large blowup in bond dimension of the output MPS. Cramer et al. [CPF+10b] proposes a heuristic efficient algorithm based on the singular value thresholding algorithm in which the output bond dimension does not suffer such blow-up; however there is no guarantee that the output MPS is close to the input MPS. This leads us to the following question:

Question 10. *Can an injective MPS be learned efficiently using local measurements, with the output bond dimension $D' = \text{poly}(D)$?*

A possible direction is to improve [LVV15, ALVV17] under suitable guarantees.

Question 11. *Promised that the ground state is an MPS of bond dimension D, can the algorithm in [LVV15, ALVV17] be improved to produce an output MPS that also has $\text{poly}(D)$ bond dimension?*

Projected Entangled Pair States (PEPS) are higher dimensional generalizations of Matrix Product States. Learnability of PEPS is far from clear, even in terms of sample complexity. Observe that the parent Hamiltonians of PEPS are frustration-free and locally-gapped. A natural question is, can the learning task become simpler assuming injectivity?

Question 12. *Given an injective PEPS, an approximation to the parent Hamiltonian can be learned with polynomial sample and time complexity. Can this be used to write down a description of another PEPS that represents a similar state?*

The key bottleneck above is that there is no two dimensional analogue of [LVV15], despite an area law for locally-gapped frustration-free spin systems [AAG22].

4 Alternate models of learning quantum states

In order to perform full state tomography on n qubits we saw that it is necessary and sufficient to obtain $\Theta(2^{2n})$ many copies of the unknown state. The exponential scaling of the complexity is prohibitive for experimental demonstrations. of course a natural question is, is it necessary to approximate the unknown state up to small trace distance? In particular, do there exist weaker but still practically useful learning goals, with smaller sample complexity? These questions have led some to consider learning only the 'useful' properties of a unknown quantum state. There have been several models of learning quantum states (inspired by computational learning theory) where exponential savings in sample complexity is possible and we discuss these models in this section.

4.1 PAC learning and online learning

PAC learning. In a seminal work, Valiant [Val84] introduced the *Probably Approximately Correct* (PAC) model of learning which lays the foundation for computational learning theory. In this model, there is a concept class consisting of Boolean functions $\mathcal{C} \subseteq \{c : \{0,1\}^n \to \{0,1\}\}$ and an underlying distribution $D : \{0,1\}^n \to [0,1]$. The learning algorithm is provided with labelled examples of the form $(x, c(x))$ where x is sampled from the distribution D.[13] We say a learning algorithm (ε, δ)-learns a concept class $\mathcal{C}$ if it satisfies the following:

> For every $c \in \mathcal{C}$, distribution $D : \{0,1\}^n \to [0,1]$, given labelled examples $(x, c(x))$ where x is sampled from D: with probability at least $1-\delta$, the algorithm outputs $h : \{0,1\}^n \to \{-1,1\}$ such that $\Pr_{x\sim D}[h(x) = c(x)] \geq 1-\varepsilon$.

The sample complexity of a learning algorithm $\mathcal{A}$ is the maximum number of labelled examples, over all the concepts $c \in \mathcal{C}$ and distributions D. The (ε, δ)-sample complexity of a concept class $\mathcal{C}$ is the minimum sample complexity over all (ε, δ)-PAC learners $\mathcal{A}$ for $\mathcal{C}$. Similarly, one can define the sample complexity (resp. time complexity) of (ε, δ)-learning $\mathcal{C}$ as the samples used (resp. time taken) by the (ε, δ)-learning algorithm. There have been many works in classical literature that have looked at *distribution-dependent* PAC models wherein the distribution D is known to the learner and the algorithm needs to perform well under D.

Aaronson [Aar07] considered the natural analog of learning quantum states in the PAC model. In this model of learning, the concept class $\mathcal{C}$ is a collection of functionals described by an unknown quantum states, $\rho \in \mathcal{C}$ acting on the class of measurements operators $\mathcal{E}$ and $D : \mathcal{E} \to [0,1]$ is an unknown distribution over all possible 2-outcome measurements. A quantum learning algorithm obtains several examples of the form $(E_i, \mathsf{Tr}(\rho E_i))$ where E_i is drawn from the distribution D and the goal is to approximate ρ. We say a learning algorithm $(\varepsilon, \delta, \gamma)$-learns $\mathcal{C}$ if it satisfies the following:

> For every $\rho \in \mathcal{C}$, given examples $(E_i, \mathsf{Tr}(\rho E_i))$ where $E_i \sim D$, with probability at least $1-\delta$, output σ that satisfies $\Pr_{E\sim D}[|\mathsf{Tr}(\rho E) - \mathsf{Tr}(\sigma E)| \leq \varepsilon] \geq 1-\gamma$.

In contrast to tomography where the output state σ is close to the unknown ρ in trace distance, i.e., σ should satisfy $\max_E |\mathsf{Tr}(E\rho) - \mathsf{Tr}(E\sigma)| \leq \varepsilon$, in PAC learning the goal is for $\mathsf{Tr}(E\rho)$ to be close to

[13] We assume that the concept class is Boolean here, one could also consider real-valued classes where $c(x)$ is then specified up to certain bits of precision.

$\mathsf{Tr}(E\sigma)$ for *most* E*s*. In a surprising result, Aaronson showed that the class of all n-qubit quantum states can be PAC-learned using just $O(n)$ samples.

Theorem 6. *The sample complexity of* PAC *learning* n*-qubit quantum states is* $O(n \cdot \mathsf{poly}(1/\varepsilon, 1/\delta, 1/\gamma))$.

Similarly, Cheng et al. [CHY16] considered the "dual problem" of learning a quantum measurement in the PAC learning framework. We do not prove these theorems here, we refer the reader to the survey [AdW17, Theorem 4.16]. A natural question left open by Aaronson was, what classes are states are *time-efficiently* PAC learnable? To this end, Rocchetto [Roc17] observed that the class of stabilizer states is PAC learnable in polynomial time. The learning algorithm of Rocchetto assumed that the distribution D was over Pauli observables and he crucially used that $\mathsf{Tr}(P\rho) \in \{-1, 1, 0\}$ when ρ was a stabilizer state. This allowed Rocchetto to learn the stabilizers of the unknown ρ and with some extra work, the entire stabilizer state ρ. A natural question that remains open is the following.

Question 13. *What is the time complexity of* PAC *learning states prepared by Clifford circuits with* t *many* T *gates? What is the* PAC *sample complexity of learning stabilizer-rank* k *states?*

Gollakota and Liang [GL22] considered a natural question of learning stabilizer states in the presence of noise. They looked at a restrictive version of PAC learning, called *statistical query* learning (we discuss this model in further detail in Section 5.2). Here the learning algorithm is allowed to make single-copy "queries" to learn the unknown *noisy* stabilizer state (the noise model they consider is the depolarizing noise). In this model, [GL22] showed that learning stabilizer states with noise is as hard as learning parities with noise (LPN) using classical samples (which is believed to require exponentially many samples [BKW03]).

Online learning. Subsequently, Aaronson et al. [ACH+18], Chen et al. [CHL+22b] looked at the setting of online learning quantum states (inspired by the classical model of online learning functions). The online model can be viewed as a variant of tomography and PAC learning. Consider the setting of tomography, suppose it is infeasible to possess T-fold tensor copies of a quantum state ρ, but instead we can obtain only sequential copies of ρ. The quantum online learning model consists of repeating the following rounds of interaction: the learner obtains a copy of ρ and a description of measurement operator E_i (possibly adversarially) and uses it to predict the value of $\mathsf{Tr}(\rho E_i)$. In the ith round, if the learners prediction was α_i and α_i satisfies $|\mathsf{Tr}(\rho E_i) - \alpha_i| \leq \varepsilon$ then it is correct, otherwise it has made a mistake. The goal of the learner is the following: minimize m so that after making m mistakes (not necessarily consecutively), it makes a correct prediction on *all* future rounds. Aaronson [ACH+18] showed that it suffices to let m be the *sequential fat-shattering dimension* of $\mathcal{C}$, denoted $\mathsf{sfat}(\mathcal{C})$ (a combinatorial parameter introduced in [RST15] to understand classical online learning), which in turn can be upper bounded by $O(n/\varepsilon^2)$ for the class of n-qubit quantum states.

4.2 Shadow tomography

A caveat of the quantum PAC learning model is that the learning algorithm has to only perform well under a distribution and it is a priori unclear if the PAC model is a natural model of learning. Aaronson [Aar18] introduced another learning model called *shadow tomography*. Here, the goal of a learning algorithm algorithm is as follows: let $E_1, \ldots, E_m$ be positive semi-definite operators satisfying $\|E_i\| \leq 1$, how many copies of an n-qubit state ρ are necessary and sufficient in order to estimate $\mathsf{Tr}(\rho E_1), \ldots, \mathsf{Tr}(\rho E_m)$ up to additive error ε. There are two naive protocols for this task: (i) either do quantum state tomography which takes $\exp(n)$ many copies and allows to estimate $\mathsf{Tr}(\rho E_i)$

for all i, or (ii) take $O(m/\varepsilon^2)$ many copies of ρ and estimate up to error ε each of the $\mathsf{Tr}(\rho E_i)$s. Surprisingly, Aaronson showed that one can perform the task of shadow tomography exponentially better in both m and n in *sample complexity*, but still running in time exponential in n.

Theorem 7. *There is a protocol for shadow tomography that succeeds with probability* $\geq 2/3$ *using* $\widetilde{O}((n\log^4 m)/\varepsilon^4)$ *many copies of* ρ.

We now sketch a proof of this theorem. For simplicity, we let ε be a constant, say $1/3$. Aaronson's proof is based on the technique of post selected learning [Aar07] which was introduced in the context of communication complexity. In this communication task, there are two players Alice and Bob: Alice has a d-dimensional quantum state ρ (unknown to Bob) and together they know a set of m many operators $\{E_1,\ldots,E_m\}$. The goal is for Alice to send a classical message to Bob, who should output $\mathsf{Tr}(\rho E_1),\ldots,\mathsf{Tr}(\rho E_m)$ up to error $1/3$. The same two trivial protocols we mentioned earlier would work here, giving a communication upper bound of $O(m+d^2)$. Surprisingly, Aaronson [Aar07] showed that there exists a communication protocol with cost $\mathsf{poly}(\log d,\log m)$ which solves the communication task, whose proof we sketch first. Bob starts by guessing the state Alice possesses. To this end, he lets $\rho_0=\mathbb{I}/d$, the maximally mixed state, and updates his guess in every round. At the tth round, suppose Bob's guess is ρ_t (whose classical description is known to Alice), Alice communicates to Bob a $j\in[m]$ for which $|\mathsf{Tr}(\rho E_j)-\mathsf{Tr}(\rho_t E_j)|$ is the largest and sends him $b=\mathsf{Tr}(E_j\rho)$. With this, Bob updates $\rho_t\to\rho_{t+1}$ as follows: let $q=O(\log\log d)$ and F_t be a two-outcome measurement on $\rho_t^{\otimes q}$ that applies the POVM $\{E_j,\mathbb{I}-E_j\}$ to each of the q copies of ρ_t and accepts if and only if the number of 1-outcomes was at least $(b-1/3)q$. Suppose σ_{t+1} is the state obtained by post-selecting on F_t accepting $\rho_t^{\otimes q}$, then ρ_{t+1} is the state obtained by tracing out the last $q-1$ registers of σ_{t+1}. Aaronson showed that after $T=O(\log d)$ rounds, Bob will have ρ' which satisfies $|\mathsf{Tr}(E_i\rho)-\mathsf{Tr}(E_i\rho')|\leq 1/3$ for $i\in[m]$. Returning to shadow tomography, observe that there is no Alice, and Bob is replaced by a quantum learner. So, at the tth stage, without any assistance, the learner needs to figure out $j\in[m]$ for which $|\mathsf{Tr}(E_j\rho_t)-\mathsf{Tr}(E_j\rho)|$ is large. To this end, Aaronson used a variant of the Quantum OR lemma [HLM17], which uses $O(\log m)$ copies of ρ and outputs "yes" if there exists a $j\in[m]$ for which $|\mathsf{Tr}(E_j\rho)-\mathsf{Tr}(E_j\rho)|\geq 2/3$ and outputs "no" if $|\mathsf{Tr}(E_j\rho)-\mathsf{Tr}(E_j\rho)|\leq 1/3$ for every $j\in[m]$. However, in order to use the ideas from the communication protocol, in the "yes" instance of the OR lemma, Bob needs to know j (not just the existence of j) in order to update ρ_t to ρ_{t+1}. Aaronson shows how to do this by using a simple binary search over $\{E_1,\ldots,E_m\}$ to find such a j. Putting these ideas together, Aaronson shows the sample complexity upper bound for the shadow tomography.

4.3 Max-entropy principle and Matrix Multiplicative Weight Update

Recall that to solve shadow tomography, the goal is to find a quantum state σ that satisfies $\mathrm{Tr}(\sigma E_i)\approx\mathrm{Tr}(\rho E_i)$ for all i. Further, one would like to minimize the number of copies of ρ, suggesting that σ should be no more informative than matching the above expectations. This is an ideal ground to invoke the *max entropy principle*, which states that the quantum state σ maximizing $S(\sigma)$ (maximum uncertainty) subject to the constraints $\mathrm{Tr}(\sigma E_i)=\mathrm{Tr}(\rho E_i)$ is the Gibbs quantum state $\frac{e^{-\sum_i\alpha_i E_i}}{\mathrm{Tr}(e^{-\sum_i\alpha_i E_i})}$. Here, α_is are determined by the expectations $\mathrm{Tr}(\rho E_i)$. From here, an algorithm for shadow tomography can start with a trivial guess for σ - the maximally mixed state - which is then updated as new knowledge from ρ arrives. Since the maximally mixed state is the Gibbs state of the trivial Hamiltonian '0', the updates can be done directly to the Hamiltonian. To see how this update can be determined, consider a technical theorem from [FBaK21].

Theorem 8. *Consider a Hamiltonian* G *and an operator* E *with* $\|E\|_\infty\leq 1$. *For* $\eta\in\mathbb{R}$, *consider*

the Gibbs states $\sigma = \frac{e^{-\beta H}}{\mathrm{Tr}(e^{-\beta H})}$ *and* $\sigma' = \frac{e^{-\beta(H+\eta E)}}{\mathrm{Tr}(e^{-\beta(H+\eta E)})}$. *It holds that for any quantum state* ρ,

$$\mathrm{S}\left(\rho\|\sigma'\right) - \mathrm{S}\left(\rho\|\sigma\right) \leq \beta \cdot \eta\Big(\beta\eta e^{|\beta\eta|} + \mathrm{Tr}(E(\rho-\sigma))\Big).$$

In particular, setting $\eta = -\frac{\mathrm{Tr}(E(\rho-\sigma))}{4\beta}$, *we find that*

$$\mathrm{S}\left(\rho\|\sigma'\right) - \mathrm{S}\left(\rho\|\sigma\right) \leq -\mathrm{Tr}(P(\rho-\sigma))^2/8.$$

Proof. To prove this result, a direct calculation reveals that

$$\mathrm{S}\left(\rho\|\sigma'\right) - \mathrm{S}\left(\rho\|\sigma\right) = \beta\eta\,\mathrm{Tr}(\rho E) + \log\frac{\mathrm{Tr}(e^{-\beta(H+\eta E)})}{\mathrm{Tr}(e^{-\beta H})}.$$

Using the Golden-Thompson inequality, we find that

$$\mathrm{S}\left(\rho\|\sigma'\right) - \mathrm{S}\left(\rho\|\sigma\right) \leq \beta\eta\,\mathrm{Tr}(\rho E) + \log\frac{\mathrm{Tr}(e^{-\beta H}e^{-\beta\eta E})}{\mathrm{Tr}(e^{-\beta H})} = \beta\eta\,\mathrm{Tr}(\rho E) + \log\mathrm{Tr}(\sigma e^{-\beta\eta E}).$$

Since $\|E\|_\infty \leq 1$, we can estimate $\mathrm{Tr}(\sigma e^{-\beta\eta E}) \leq 1 - \beta\eta\,\mathrm{Tr}(\sigma E) + \beta^2\eta^2 e^{|\beta\eta|}$, which implies

$$\mathrm{S}\left(\rho\|\sigma'\right) - \mathrm{S}\left(\rho\|\sigma\right) \leq \beta\eta\,\mathrm{Tr}(\rho P) + \log(1 - \beta\eta\,\mathrm{Tr}(\sigma E) + \beta^2\eta^2 e^{|\beta\eta|}) \leq \beta\eta\,\mathrm{Tr}((\rho-\sigma)P) + \beta^2\eta^2 e^{|\beta\eta|}.$$

This proves the theorem statement. □

Thus, the alternate algorithm for shadow tomography proceeds by identifying an E_i that still does not satisfy $\mathrm{Tr}(E_i\rho) = \mathrm{Tr}(E_i\sigma)$ and then updating the weight of such an E_i in σ. In order to find such an E_i with $\mathrm{poly}(\log m)$ sample complexity, one can use the 'quantum OR lemma' as described earlier. We also highlight that this procedure can be used to learn the Hamiltonian. Assuming that ρ itself is a Gibbs state, we update the weights of the basis operators E_α until the expectation values are close. In such a case, the strong convexity from [AAKS21b] ensures that the Hamiltonian is learned up to desired error.

4.4 Subsequent works building on shadow tomography

There have been several subsequent works that have built upon Aaronson's shadow tomography protocol which we discuss in this section.

4.4.1 Classical shadows

A subsequent work of Huang, Kueng and Preskill [HKP20] presented an alternate protocol for a restricted version of shadow tomography that is more time efficient than Aaronson's original shadow tomography protocol. In particular, they proved the following.

Theorem 9. *Let* $B > 0$ *be an integer and* $\varepsilon, \delta \in [0,1]$. *Given* $O(B/\varepsilon^2 \log(1/\delta))$ *copies of* ρ, *there exists a procedure that satisfies the following: for every observable* M *that satisfies* $\mathsf{Tr}(M^2) \leq B$, *with probability* $\geq 1-\delta$, *the quantity* $\mathsf{Tr}(\rho M)$ *can be computed to error* ε.

To compare this procedure and shadow tomography, suppose the algorithm needs to estimate m many expectation values, then by letting $\delta \sim 1/m$ with success probability $\geq 2/3$, the overall sample complexity scales as $O((\log m)\cdot B/\varepsilon^2)$. Additionally, observe that the procedure above is *independent* of the observables M, unlike Aaronson's protocol [Aar18] which used the observables

$E_1, \ldots, E_m$ in a crucial way to learn $\mathsf{Tr}(\rho E_i)$. We now give a proof sketch of the theorem: they first give a polynomial-time procedure for generating *classical shadows* of the unknown quantum state ρ using $T = O(B/\varepsilon^2 \log(1/\delta))$ copies of ρ. These classical shadows are generated by running the following procedure: given copies of ρ, the algorithm samples a uniformly random Clifford C, computes $C\rho C^\dagger$ and measures the state in the computational basis to get an n-bit string b. So the classical shadows is the set $\{(C_i, b_i\}_{i\in[T]}$. Using these classical shadows, [HKP20] use a simple median of means estimation procedure to estimate $\mathsf{Tr}(\rho M)$ for an arbitrary observable M satisfying $\mathsf{Tr}(M^2) \leq B$. Thus the sample complexity is $O(B/\varepsilon^2 \log(1/\delta))$.

4.4.2 Improved shadow tomography and agnostic learning

Bădescu and O'Donnell [BO21] improved the complexity of shadow tomography to $\tilde{O}((n\cdot\log^2 m)/\varepsilon^2)$, which simultaneously obtains the best known dependence on each of the parameters n, m, ε. We do not sketch their protocol, but remark on one interesting corollary of shadow tomography is a procedure which they call *quantum hypothesis selection*. Although not phrased in this language, quantum hypothesis selection can be viewed as *agnostic* learning quantum states. The setup for quantum agnostic learning states is the following: $\mathcal{C}$ is a collection of known quantum states $\{\rho_1, \ldots, \rho_m\}$, a learning algorithm is provided with copies of an unknown state σ and needs to find $\rho_k \in \mathcal{C}$ which is closest to σ in the following sense: output $\rho_k \in \mathcal{C}$ such that

$$\|\rho_k - \sigma\|_1 \leq \alpha \cdot \min_{\rho\in\mathcal{C}} \|\rho - \sigma\|_1 + \varepsilon, \tag{1}$$

for some α. We briefly sketch the reduction from quantum agnostic learning to shadow tomography: consider the two states ρ_i, ρ_j in the concept class $\mathcal{C}$, by Holevo-Helstrom's theorem there exists an optimal measurement measurement $\{A_{ij}, \mathbb{I} - A_{ij}\}$ such that $\mathsf{Tr}(A_{ij} \cdot (\rho_i - \rho_j)) = \|\rho_i - \rho_j\|_{tr}$. Now perform shadow tomography using $\tilde{O}((n \cdot \log^2 m)/\varepsilon^2)$ copies of σ along with the operators $\{A_{ij}\}_{i,j\in[m]}$ to obtain α_{ij}s satisfying $|\alpha_{ij} - \mathsf{Tr}(A_{ij}\sigma)| \leq \varepsilon/2$. At this point, [BO21] simply goes over all $\rho \in \mathcal{C}$ to find a ρ_k that minimizes the quantity $\max_{i,j} |\mathsf{Tr}(\rho_k A_{ij}) - \alpha_{ij}|$ (this is inspired by classical hypothesis selection [Yat85]). Let $\eta = \min_{\rho\in\mathcal{C}} \|\rho - \sigma\|_1$ and $i^* = \operatorname{argmin}_{\rho\in\mathcal{C}} \|\rho - \sigma\|_1$. Observe that

$$\begin{aligned}
\|\rho_k - \sigma\|_{tr} &\leq \eta + \|\rho_k - \rho_{i^*}\|_{tr} \\
&= \eta + |\mathsf{Tr}(A_{i^*k}\rho_k) - \mathsf{Tr}(A_{i^*k}\rho_{i^*})| \\
&\leq \eta + |\mathsf{Tr}(A_{i^*k}\rho_k) - \alpha_{i^*k}| + |\mathsf{Tr}(A_{i^*k}\rho_{i^*}) - \alpha_{i^*k}| \leq 3\eta + \varepsilon.
\end{aligned}$$

Hence the resulting ρ_k satisfies Eq. (1) with $\alpha = 3$. As far as we are aware, [BO21, CL21, FQR22, Car21] are the only few works to look at agnostic learning of quantum states. These works gives rise to the following two interesting questions.

Question 14. *What is the sample complexity of quantum agnostic learning if we require $\alpha = 1$?*[14]

Question 15. *What classes of states can be agnostic learned time-efficiently? Can we learn stabilizer states efficiently in the quantum agnostic model?*

4.4.3 Shadow tomography with separable measurements

Chen et al. [CCHL21] considered the problem of shadow tomography if one was only allowed *separable* measurements. In this setting, they showed that $\tilde{\Omega}(\min\{m, d\})$ many copies are necessary

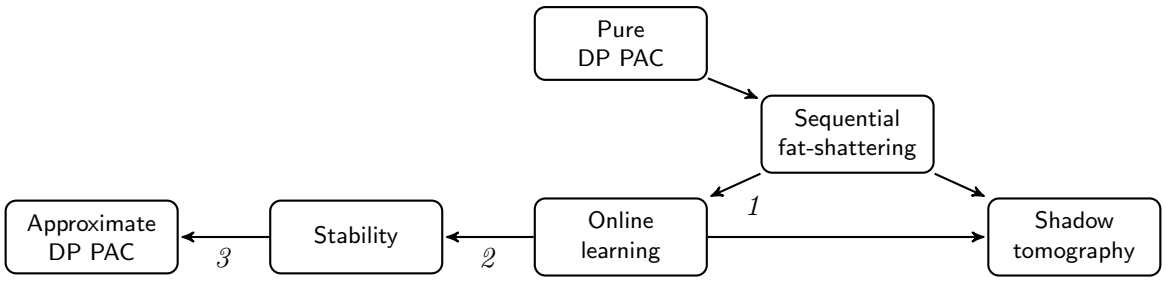
[14] In classical computational learning theory, reducing the value of α to 1 has been resolved for certain Boolean function classes in the seminal works [LMN93, GKK08].

for shadow tomography, matching the upper bound of Huang et al. [HKP20] of $\tilde{O}(\min\{m, d\})$. They in fact show that, in order to estimate the expectation values of all 4^n many n-qubit Pauli observables, one needs $\Omega(2^n)$ copies of ρ (given access to only separable measurements). The proof of this lower bound follows the following three-step approach (i) They first consider the learning tree framework that we discussed below Theorem 2, wherein there is a tree with each node corresponding to a measurement applied to the unknown state ρ and the leaves of the tree correspond to the m many expectation values. (ii) Using this learning tree technique, the main technical lemma they show is that, in order to prove the hardness of estimating $\mathsf{Tr}(\rho Q_i)$ for arbitrary Q_i, using separable measurements, it suffices to upper bound $\delta(Q_1, \ldots, Q_{2^n}) = \frac{1}{m} \sup_{|\psi\rangle} \sum_i \langle\psi|Q_i|\psi\rangle^2$. (iii) Finally they show that for the Paulis P_i, we have that $\delta(P_1, \ldots, P_{2^n})$ is exactly $1/(2^n + 1)$, which immediately gives them their sample complexity lower bound of $\Omega(2^n)$.

Additionally they also consider settings wherein the learning algorithm is adaptive (i.e., the learner can perform measurements based out of previous measurement outcomes) and non-adaptive (i.e., the learning algorithm needs to decide at the beginning a sequence of measurements to carry out). Similarly, a followup work of Gong and Aaronson [GA22] showed how to perform shadow tomography when given m many k outcome measurements using $\mathsf{poly}(k, \log m, n, 1/\varepsilon)$ copies of ρ.

4.5 Equivalence between quantum learning models

So far, we saw many many seemingly (unrelated) models of computation aimed at learning an unknown quantum state such as, shadow tomography, PAC learning, communication complexity, online learning. Aaronson and Rothblum [AR19] also considered differential privacy in learning quantum states and used this notion to prove new bounds on online learning and shadow tomography.[15] A natural question is, is there a connection between these models? In [AQS21] they showed "equivalences" between all these models of computation. A high-level overview of the results in their work is summarized in the figure below. For technical reasons, we do not discuss pure and approximate DP in detail, we simple remark that *pure* DP is a stronger requirement than *approximate* DP and refer the reader to [AQS21] for more details. In particular, these equivalences imply that algorithms in one framework gives rise to quantum learning algorithms in other frameworks. We remark that only a few of these arrows are efficient in both sample and time complexity, otherwise these implications are primarily information-theoretic.



Although *a priori*, it seems that PAC learning, online learning and DP learning have little to do with one another, classically there have been a sequence of works establishing tight connections between these three fields [KLN+11]. The main center piece in establishing these connections is the notion of stability which was introduced in a recent breakthrough work of Bun et al. [BLM20].

[15] Classically differential privacy was formalized in the seminal works by Dwork [Dwo06, DR+14]: we say a learning algorithm is differentially private if it behaves approximately the same when given two training datasets which differ in only limited number of entries.

In [AQS21] they "quantize" these connections. Below we give a sketch of their proofs and refer to their work for a detailed overview.

It is well-known classically that if there is a DP PAC learning algorithm for a class $\mathcal{C}$ then the *representation dimension* of the class is small. Representation dimension then upper-bounds classical communication complexity and $\mathsf{sfat}(\mathcal{C})$. In [AQS21] they show that this connection carries over in a simple way to the quantum setting.

(1): Let $\mathcal{C}$ be a concept class of states with finite $\mathsf{sfat}(\mathcal{C})$. In order to describe an online-learner for $\mathcal{C}$ making at most $\mathsf{sfat}(\mathcal{C})$ mistakes, in [AQS21] they construct a *robust* standard optimal algorithm (denoted RSOA) whose accuracy guarantees are robust to adversarial imprecision in the training feedback. The RSOA algorithm is inspired by the classical standard optimal algorithm for Boolean functions (however in the quantum setting it needs to work for real functions as well as with adversarial noise). Aaronson et al. [ACH+18] showed an upper bound of $\mathsf{sfat}(\mathcal{C}) \leq n$ on the number of mistakes in this setting asking if there is an *explicit* algorithm that achieves this bound (their RSOA made explicit this algorithm).

(2): Here, they show that a concept class $\mathcal{C}$ with $\mathsf{sfat}(\mathcal{C}) = d$ can be learned by a stable algorithm. To prove this, they follow the technique of [BLM20] which feeds a standard optimal algorithm (which they replace with RSOA) with a specially-tailored input sample. The tailoring algorithm deliberately injects "mistake" examples into the sample, each of will force a prediction mistake in RSOA. Since the RSOA *completely* identifies the target concept after making at most d prediction mistakes, the injection step allows the stable algorithm to output the correct hypothesis. In [AQS21], their quantum-focused adaptation of this technique handles the twin challenges of accurately engineering the mistake examples for real-valued functions, and having ε-uncertainty in the adversary's feedback (both of which are not present in the Boolean setting).

(3): Now that one has a stable algorithm established above, one need to make it differentially private. In the Boolean setting, given a stable algorithm $\mathcal{A}$, there is a well-known "Stable Histograms" algorithm may be used a 'wrapper' around $\mathcal{A}$, to privately identify $\mathcal{A}$'s high-probability output functions. This involves running $\mathcal{A}$ many times and outputting its most frequent output, while adding Laplacian noise to make it DP. However, they encounter an additional complication in the quantum setting: outputting the "most frequent" quantum state doesn't make sense, since two quantum states could be arbitrarily close and qualify as valid outputs. Addressing this, [AQS21] modify Stable Histograms and show that it can be used to make the quantum stable learning algorithm DP.

5 Learning classical functions through quantum encoding

5.1 Learning Boolean functions

The *quantum* PAC model for learning a concept class of Boolean functions $\mathcal{C} \subseteq \{c : \{0,1\}^n \to \{0,1\}\}$ was introduced by Bshouty and Jackson [BJ95]. In this model, instead of access to labelled examples $(x, c(x))$ where x is sampled from D, the quantum learning algorithm is provided with copies of the *quantum example* state $|\psi_c\rangle = \sum_{x\in\{0,1\}^n} \sqrt{D(x)}|x, c(x)\rangle$. Quantum examples are a natural generalization of classical labelled examples (by measuring a single quantum example, we obtain a classical labelled example). A quantum PAC learner is given copies of the quantum example state, performs a POVM (where each outcome of the POVM is associated with an hypothesis) and outputs the resulting hypothesis. The sample complexity of the learner here is measured as the number of *copies* of $|\psi_c\rangle$ and the (ε, δ)-quantum sample complexity of learning $\mathcal{C}$ is defined

similarly to the classical PAC learning. There have been a few works that have looked at quantum PAC learning function classes [BJ95, AS07, ACL+21, AdW18a] and showed some strengths and weakness of quantum examples in the PAC model of learning: under the distribution independent setting, we know that quantum examples are not useful for learning [AdW18b], for uniform and product distributions we know quantum examples are useful [AS07, ACL+21, BJ95, KRS18],[16] for the uniform distribution we know they are not useful for learning circuit families [AGG+22] and for certain applications such as learning parities with noise, quantum examples are known to be useful [GKZ19]. For further details, we refer the reader to [AdW17].

Question 16. *Almost all quantum learning speedups are in the uniform distribution setting, is there a quantum learning speedup in the distribution-independent model in terms of sample or time complexity?*

In [AdW18b] they also considered two other models of learning (motivated by classical computational learning theory): (*i*) *random classification noise learning*: here, the learner is given copies of $\sum_x \sqrt{D(x)}|x\rangle \otimes (\sqrt{1-\eta}|c(x)\rangle + \sqrt{\eta}|\bar{c}(x)\rangle$ and the goal of the learning algorithm is the same as the PAC learner, (*ii*) *agnostic learning*: here $D : \{0,1\}^{n+1} \to [0,1]$ is an unknown distribution, the learner is given copies of $\sum_{(x,b)\in\{0,1\}^{n+1}} \sqrt{D(x,b)}|x,b\rangle$ and needs to find the concept $c \in \mathcal{C}$ that best approximates D, i.e., output c that satisfies $\mathrm{err}_D(c) \leq \min_{c'\in\mathcal{C}}\{\mathrm{err}_D(c')\} + \varepsilon$, where $\mathrm{err}_D(c') = \Pr_{(x,b)\sim D}[c'(x) \neq b]$. In both these distribution-independent learning models, [AdW18b] showed that quantum sample complexity of learning is equal to classical sample complexity of learning up to constant factors. A natural question is, what *can* be learned in polynomial time in these models? As far as we are aware, only parities are known to be learnable in the classification noise model [GKZ19, Car20] when $D = \{0,1\}^n$ and agnostic learning interesting concept classes has not received any attention in literature.

Question 17. *Can we learn DNF formulas in the quantum agnostic model in polynomial time?*[17]

5.2 Statistical query model

The quantum statistical query model was introduced in [AGY20], inspired by the classical statistical query model introduced by Kearns [Kea98]. Classically, it is well-known that *many* algorithms used in practice can be implemented using a statistical query oracle, for example, expectation maximization, simulated annealing, gradient descent, support vector machine, Markov chain Monte carlo methods, principal component analysis, convex optimization (see [Rey20, FGR+17] for these applications). We first discuss the classical SQ model for learning an unknown concept c from the concept class $\mathcal{C} \subseteq \{c : \{0,1\}^n \to \{-1,1\}\}$ under an unknown distribution $D : \{0,1\}^n \to [0,1]$. The SQ learner has access to a *statistical query oracle* which takes as input two quantities: *tolerance* $\tau \geq 0$, a function $\phi : \{0,1\}^n \times \{-1,1\} \to \{-1,1\}$ and returns $\alpha \in \mathbb{R}$ satisfying $\left|\alpha - \mathbb{E}_{x\sim D}[\phi(x,c(x))]\right| \leq \tau$. The SQ learning algorithm adaptively chooses a sequence $\{(\phi_i,\tau_i)\}$, and based on the responses of the statistical oracle $\{\alpha_i\}_i$, it outputs an hypothesis $h : \{0,1\}^n \to \{-1,1\}$ that approximates c, similar to the setting of PAC learning.

The QSQ model is similar to the quantum PAC model, except that the learning algorithm isn't allowed entangled measurements on several copies of the quantum example state. More formally, let $\mathcal{C} \subseteq \{c : \{0,1\}^n \to \{-1,1\}\}$ be a concept class, $D : \{0,1\}^n \to [0,1]$ be a distribution and

[16] We remark that almost all known quantum speedups are based on a version of quantum Fourier sampling.

[17] A positive answer to this question would imply a polynomial-time quantum algorithm for PAC learning depth-3 circuits in the uniform distribution model [Fel09].

let $|\psi_c\rangle = \sum_x \sqrt{D(x)}|x, c(x)\rangle$. In the QSQ model, a learning algorithm specifies an operator M satisfying $\|M\| \leq 1$, tolerance $\tau \in [0,1]$ and obtains a number $\beta \in [\langle\psi_c|M|\psi_c\rangle - \tau, \langle\psi_c|M|\psi_c\rangle + \tau]$. An intuitive way to think about the QSQ model is, a learning algorithm can specify a two-outcome measurement $\{M, \mathbb{I} - M\}$ and obtains a τ-approximation of the probability of this measurement accepting $|\psi_c\rangle$. Ideally, one would want a QSQ algorithm for which $\tau = 1/\mathsf{poly}(n)$ and M can be implemented using $\mathsf{poly}(n)$ gates. A QSQ algorithm is amenable to near-term implementation since unlike the quantum PAC framework, it works only by making single copy measurements on the quantum example state $|\psi_c\rangle$. Surprisingly, in [AGY20], they show that positive results for quantum PAC learning that we discussed in the previous section (such as learning parities, DNF formulas, $(\log n)$ juntas) can actually be implemented in the QSQ framework.

Theorem 10. *The concept classes consisting of parities, juntas, DNF formulas, sparse functions, can be learned under the uniform distribution in the* QSQ *model.*

The crucial (and simple) observation in order to see this theorem is that computing the Fourier mass of a subset can be done in the QSQ model. Given that learning algorithms for parities, juntas, DNF formulas, sparse functions are via Fourier sampling [AdW17], this observation implies the theorem. To see the observation, let $M = \sum_{S \in T} |S\rangle\langle S|$ and consider the observable

$$M' = \mathsf{H}^{\otimes(n+1)} \cdot \left(\mathbb{I}^{\otimes n} \otimes |1\rangle\langle 1|\right) \cdot M \cdot \left(\mathbb{I}^{\otimes n} \otimes |1\rangle\langle 1|\right) \cdot \mathsf{H}^{\otimes(n+1)}.$$

Operationally, M' corresponds to first applying the Fourier transform on $|\psi_f\rangle$, post-selecting on the last qubit being 1 and finally applying M to the first n qubits. In order to see the action of M' on $|\psi_f\rangle$, first observe that $\mathsf{H}^{\otimes(n+1)}|\psi_f\rangle$ yields $\frac{1}{\sqrt{2^n}} \sum_x |x, f(x)\rangle \to \frac{1}{2^n} \sum_{x,y} \sum_{b\in\{0,1\}} (-1)^{x\cdot y + b\cdot f(x)} |y, b\rangle$. Conditioned on the $(n+1)$-th qubit being 1, we have that the resulting quantum state is $|\psi'_f\rangle = \sum_Q \widehat{f}(Q)|Q\rangle$. The expectation value of M with respect to the resulting state is given by $\langle\psi'_f|M|\psi'_f\rangle = \sum_{S\in T} \widehat{f}(S)^2$. Therefore, one quantum statistical query with measurement M', tolerance τ produces a τ-approximation of $\sum_{S\in T} \widehat{f}(S)^2$. In [AGY20], they use this observation to prove Theorem 10. We pose the following question, which would serve as a tool to understand the fundamental question "is entanglement needed for quantum learning Boolean functions?".[18]

Question 18. *Is there a concept class separating* QSQ *and quantum* PAC *learning with separable measurements?*

More recently, there have been few works that considered the "diagonal-QSQ" framework: here, the QSQ learner can only specify a *diagonal* measurement operator M, i.e., the QSQ learner specifies a $\phi(x) \in [-1, 1]$ and makes a QSQ query with $M = \sum_x \phi(x)|x\rangle\langle x|$ for the unknown state $|\phi\rangle$. Recently [HIN+21, HIN+22, NIS+23] looked at learning unknown circuits U given diagonal-QSQ access to $|\psi_U\rangle = U|0^n\rangle$ (these learning algorithms allow to learn the output distributions $\{\langle x|U|0^n\rangle^2\}_x$ of unknown quantum circuits U in the computational basis). In particular, [HIN+22] showed that distributions induced by Clifford circuits can be learned in the QSQ framework, however, if we add a single T gate, then classical SQ learning the output distribution is as hard as learning parities with noise. Subsequent works [HIN+21, NIS+23] looked at larger circuit families showing stronger lower bounds. One interesting question left open by their work is the following

Question 19. *What is* QSQ *complexity of learning output distributions of constant-depth circuits in the diagonal-*QSQ *framework?*

[18] For learning the general class of quantum states, the recent work of Chen et al. [CCHL21] showed entanglement is needed for learning quantum states.

In another direction Du et al. [DHL+21] showed that the QSQ model can be effectively simulated by noisy quantum neural networks (QNN). Since we saw above that the QSQ model can learn certain concept classes in polynomial time, their result suggests that QNNs implemented on a noisy device could potentially retain the quantum speed-up.

5.3 Kernel Methods

So far we discussed a family of quantum algorithms that implicitly assumed the learning algorithm could learn a classical function by given access to quantum examples that encode classical information. Furthermore, these quantum examples use a number of qubits that is only logarithmic in the size of the unknown function. In this framework there have been several quantum machine learning algorithms that are able to achieve polynomial or even exponential speed-ups over classical approaches [HHL09, WBL12, LMR13, LMR14, RML14, LGZ16, CD16, KP17, BKL+19, RSML18, ZFF19]. However, it is not known whether data can be efficiently provided this way in practically relevant settings. This raises the question of whether the advantage comes from the quantum algorithm, or from the way data is provided [Aar15]. Indeed, recent works have shown that if classical algorithms have an analogous sampling access to data, then some of the proposed exponential speed-ups do no longer exist [Tan19, Tan21, GLT18, CLW18, DBH19, CGL+20].

A natural question is, if we demand classical input and classical output, but let the intermediate operation be a quantum operations, can one hope for a quantum speedup? To this end, a powerful technique called the *quantum kernel method* was introduced [HCT+19, SK19]. These papers proposed obtaining a quantum speedup via the use of a *quantum-enhanced feature space*, where each data point is mapped non-linearly to a quantum state and then classified by a linear classifier in the high-dimensional Hilbert space. The advantage of the quantum learner stems from its ability to recognize classically intractable complex patterns using the quantum feature map, which maps each classical data point non-linearly through a parameterized family of unitary circuits to a quantum state, $x \mapsto |\phi(x)\rangle = U(x)|0^n\rangle$, in both training and testing. The learning algorithm proceeds by finding the optimal separating hyperplane for the training data in the high-dimensional feature space. To do so efficiently, they use the standard kernel method in *support vector machines* (SVMs), a well-known family of supervised classification algorithms [Vap13]. More specifically, their algorithm only uses the quantum computer to estimate a kernel function and then implement a conventional SVM on a classical computer. In general, kernel functions are constructed from the inner products of the feature vectors for each pair of data points, which can be estimated as the transition amplitude of a quantum circuit as $|\langle\phi(x_j)|\phi(x_i)\rangle|^2 = |\langle 0^n|U^\dagger(x_j)U(x_i)|0^n\rangle|^2$ (see Figure 6 for the quantum circuit implementation of this). One can therefore estimate each kernel entry up to a small additive error using the quantum computer – a procedure that is referred to as *quantum kernel estimation* (QKE). Then, the kernel matrix is given to a classical optimizer that efficiently finds the linear classifier that optimally separates the training data in feature space by running a convex quadratic program.

Despite the popularity of these quantum kernel methods, it was unclear if, quantum algorithms using kernel methods could provide a provable advantage over classical machine learning algorithms. There have been several proposal for interesting feature maps [GGC+21, HCT+19] based on group covariant maps, but their utility and implementability is unclear. In [LAT21], they constructed a classification task based on the discrete logarithm problem and showed that given classical access to data, quantum kernel methods can provably solve a classically intractable learning problem, even in the presence of finite sampling noise. While the particular problem they consider does require a fault-tolerant quantum computer, the learning algorithm is general in nature which suggests

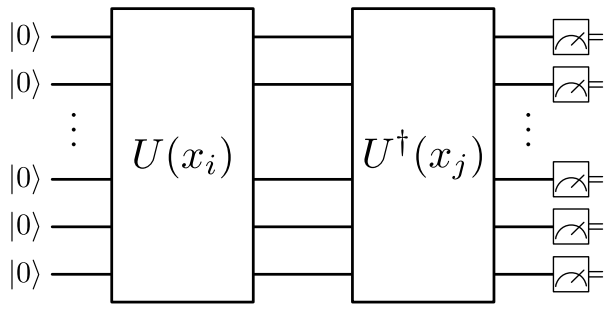


Figure 6: Quantum kernel estimation. A quantum feature map $x \mapsto \Phi(x) = |\phi(x)\rangle\langle\phi(x)|$ is represented by a circuit where $|\phi(x)\rangle = U(x)|0^n\rangle$. Each kernel entry $K(x_i, x_j)$ is obtained using a quantum computer by running the circuit $U^\dagger(x_j)U(x_i)$ on input $|0^n\rangle$, and then estimating $\left|\langle 0^n|U^\dagger(x_j)U(x_i)|0^n\rangle\right|^2$ by counting the frequency of the 0^n output.

potential to find near-term implementable problems and is suitable for error-mitigation techniques. Their result can be viewed as one of the first formal evidence of quantum advantage using *quantum kernel methods*, a widely-studied family of quantum learning algorithms that can be applied to a wide range of problems.

Question 20. *Can quantum kernel methods give an unconditional polynomial advantage over classical learning for a natural problem?*

6 Perspective on other works

The theory of quantum information and computation lies at the intersection of computer science and physics. Quantum learning theory has evolved in the same spirit, addressing questions that are native to both bodies of knowledge. The field is inspired, on one hand, from the notions of PAC learning and statistical query learning from theoretical computer science and on the other hand, from the experimental goal of learning physics of a system from natural quantum states. This survey adopts the view that the most exciting questions in the field lie precisely at this intersection.

The success of learning theory lies in its adaptation of the 'number of samples' as a natural complexity measure - which is well motivated from the point of view of practical machine learning. It is remarkable that we can obtain sample efficient algorithms - in many cases even time efficient - for a wide class of learning problems. These successful results are accompanied by new insights into the structure of corresponding families of quantum states, such as phase states, Gibbs quantum states and Matrix product states. Here, we list several notable works related to learning quantum states that haven't been covered in this survey. These include results on learning quantum noise [FGLE12] in quantum experiments, tomography of quantum channels [HCP22, Car22, CL21, Car21, FQR22, CD20, HKOT23], learning properties of ground states of the Hamiltonians [LHT+23, ORFW23, HKT+22b, RF21], provable bounds for learning parametrized quantum circuits, quantum dynamics, simulation [CGFM+21, CHC+22, GHC+22, CHE+22], learning matrix product states [CPF+10a, GSG+23, KR21], investigation into quantum Born machines [CMDK20, ZGYN22, GYN22], learning Hamiltonians in a heuristic manner [WGFC14a, WBL12, VMN+19], power of quantum neural networks [ASZ+21, BBF+20], learning unitaries defined by time evolution - e^{-iHt}, where several recent works [WGFC14b, HKT22a, HTFS23, DPW+21] have given efficient algorithms. We refer

the interested reader to the references for more details.

Sample and time complexity beyond learnability. Finally, we highlight that - beyond learnability - the notion of sample complexity is well motivated even in quantum information problems that are not canonical learning tasks. We discuss a few directions and questions here.

1. *Sample complexity as a measure in quantum communication.* In the standard quantum communication complexity setting, Alice and Bob compute a classical function with classical inputs, using quantum resources. One can also define a model where inputs are quantum and functions of quantum inputs are to be computed. An example of this is: Alice's input is a quantum state $|\psi\rangle$, Bob's input is a quantum state $|\phi\rangle$, and they wish to estimate $\langle\psi|M|\phi\rangle$ for a given M. A single copy of each input is insufficient and unbounded number of inputs render the problem classical. An interesting intermediate regime is to allow several independent copies of inputs and minimize the sample complexity. The work [ALL22] first considered this for $M = \mathbb{I}$ and showed exponential separation in sample complexity between classically communicating Alice-Bob and quantumly communicating Alice-Bob.

Question 21. *What is the sample complexity of evaluating $\langle\psi|M|\phi\rangle$ when Alice and Bob are only allowed classical communication, and how does it relate to the sample complexity when quantum communication is allowed?*

2. *Sample complexity as a measure in the Local Hamiltonian problem:* A canonical Quantum Merlin-Arthur complete problem is the Local Hamiltonian Problem, with the goal of determining if the ground energy of a n-qubit local Hamiltonian is small or large. The proof - that certifies that the ground energy is small - is a quantum state, and there is some evidence that the proofs cannot be polynomial sized classical strings. In fact, the famous result of Marriot and Watrous shows that one copy of the witness suffices [MW05]. Now, let's restrict the proof to be a simple quantum state, such as a state that can be prepared by a low-depth circuit or a stabilizer state. We can find local Hamiltonians whose ground states have very small overlap with one such state [AN22]. Thus, many copies of the simple witness would be needed to eventually reach a complex witness of the ground state (via phase estimation algorithm). But it is not clear if such a simple witness could be useful in other ways to estimate the ground energy.

Question 22. *Can we provide a sample complexity lower bound for interesting class of simple witness states, when the goal is to use them to estimate the ground energy of a Hamiltonian? Is this problem easier if the Hamiltonian itself is a sparse Hamiltonian with oracle access?*

3. *Time-efficient learning coset states.* One way to view the Hidden subgroup problem (HSP) is in terms of sample complexity of learning the coset state. In the HSP, there is a group G. Let $\mathcal{H}(G)$ be the set of all subgroups $H \leq G$ of G. We say a function $f_H : G \to S$ hides a subgroup H if $f(x_1) = f(x_2)$ for all $x_1, x_2 \in H$ and is distinct for different cosets. Given quantum query access to f, the goal is to learn H. The so-called *standard approach* (which has been the focus of almost all known HSP algorithms) is the following: prepare $\frac{1}{\sqrt{|G|}}\sum_{x\in G}|x\rangle$, query f to produce $\frac{1}{\sqrt{|G|}}\sum_{x\in G}|x, f(x)\rangle$ and discard the second register to obtain the state $\rho_H = \frac{|H|}{|G|}\sum_{g\in K}|gH\rangle\langle gH|$ where $|gH\rangle = \frac{1}{\sqrt{|H|}}\sum_{h\in H}|gh\rangle$ and K is a complete set of left coset representatives of the subgroup $H \leq G$. The state ρ_H is called the *coset state* and the question is: what is the sample complexity and time complexity of learning H given copies of ρ_H? A well-known result [EHK04] shows that the sample complexity of learning H is $O(\log^2|G|)$. However, time-efficient learning H for arbitrary groups has remained a long-standing open question. We know time efficient implements for special

groups [BCvD06, HRTS00, Kit96, RB98, FIM+14]. Given that learning H given copies of ρ_H for arbitrary groups has been open for decades, this motivates the following questions.

Question 23. *For arbitrary groups, can we time-efficiently learn coset states in the alternate models of learning that we discussed in Section 4? What other groups can we time-efficiently learn H given copies of ρ_H?*

Additionally, we remark that all known HSP algorithms following the standard approach where they measure the second register, which leads to the following question.

Question 24. *Does there exist a proposal for non-Abelian* HSP *that doesn't measure the second register in $|\psi_f\rangle = \frac{1}{\sqrt{|G|}}\sum_{x\in G}|x, f(x)\rangle$ and takes advantage of the function register to learn the unknown subgroup H? Similarly, can we extend the lower bounds in [HRS10] to the setting where the learning algorithm has access to copies of $|\psi_f\rangle$?*

4. *Sample complexity of generalizing LMR.* Lloyd, Mohseni, and Rebentrost [LMR14] understood the following question (in the context of Hamiltonian simulation): How many copies of an unknown quantum state ρ are required to simulate a unitary $U = e^{-i\rho t}$ which encodes ρ for some $t \in \mathbb{R}$? The LMR protocol [LMR14] showed that the sample complexity of implementing U up to diamond norm δ is $O(t/\delta^2)$, and has found several applications in quantum computing. Subsequently the sample complexity obtained by the LMR protocol was shown to be optimal [KLL+17]. A natural followup question is the following.

Question 25. *What is the sample complexity of approximately implementing $e^{-if(\rho)t}$ for other functions f acting on density matrices?*

Acknowledgements. We thank Matthias Caro and the anonymous reviews of Nature Reviews Physics for several comments improving the presentation of this work and Abhinav Deshpande for useful comments. We thank Iulia Georgescu for commissioning this survey for the Nature Reviews Physics. AA acknowledges support through the NSF CAREER Award No. 2238836 and NSF award QCIS-FF: Quantum Computing & Information Science Faculty Fellow at Harvard University (NSF 2013303).

References

[AAG22] Anurag Anshu, Itai Arad, and David Gosset. An area law for 2d frustration-free spin systems. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, page 12–18, New York, NY, USA, 2022. Association for Computing Machinery. 14

[AAKS21a] Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, and Mehdi Soleimanifar. Efficient learning of commuting hamiltonians on lattices, 2021. https://anuraganshu.seas.harvard.edu/links. 10, 11

[AAKS21b] Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, and Mehdi Soleimanifar. Sample-efficient learning of interacting quantum systems. *Nature Physics*, 17(8):931–935, 2021. 10, 12, 17

[Aar07] Scott Aaronson. The learnability of quantum states. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 463(2088):3089–3114, Sep 2007. 14, 16

[Aar15] Scott Aaronson. Read the fine print. *Nature Physics*, 11(4):291–293, Apr 2015. 23

[Aar16] Scott Aaronson. The complexity of quantum states and transformations: from quantum money to black holes. *arXiv:1607.05256*, 2016. 1

[Aar18] Scott Aaronson. Shadow tomography of quantum states. In *Proceedings of the 50th Annual ACM SIGACT Symposium on Theory of Computing, STOC*, pages 325–338. ACM, 2018. 15, 17

[ABDY22] Srinivasan Arunachalam, Sergey Bravyi, Arkopal Dutt, and Theodore J. Yoder. Optimal algorithms for learning quantum phase states, 2022. arXiv:2208.07851v1. 6, 7, 8

[ACH+18] Scott Aaronson, Xinyi Chen, Elad Hazan, Satyen Kale, and Ashwin Nayak. Online learning of quantum states. In *Advances in Neural Information Processing Systems*, pages 8962–8972, 2018. 15, 20

[ACL+21] Srinivasan Arunachalam, Sourav Chakraborty, Troy Lee, Manaswi Paraashar, and Ronald De Wolf. Two new results about quantum exact learning. *Quantum*, 5:587, 2021. 21

[AdW17] Srinivasan Arunachalam and Ronald de Wolf. Guest column: A survey of quantum learning theory. *SIGACT News*, 48(2):41–67, 2017. 15, 21, 22

[AdW18a] Srinivasan Arunachalam and Ronald de Wolf. Optimal quantum sample complexity of learning algorithms. *Journal of Machine Learning Research*, 19(71):1–36, 2018. 21

[AdW18b] Srinivasan Arunachalam and Ronald de Wolf. Optimal quantum sample complexity of learning algorithms. *Journal of Machine Learning Research*, 19:71:1–71:36, 2018. 21

[AG04] Scott Aaronson and Daniel Gottesman. Improved simulation of stabilizer circuits. *Physical Review A*, 70(5):052328, 2004. 5

[AG08] S. Aaronson and D. Gottesman. Identifying stabilizer states, 2008. http://pirsa.org/08080052. 5

[AGG+22] Srinivasan Arunachalam, Alex B Grilo, Tom Gur, Igor C Oliveira, and Aarthi Sundaram. Quantum learning algorithms imply circuit lower bounds. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 562–573. IEEE, 2022. 21

[AGY20] Srinivasan Arunachalam, Alex B Grilo, and Henry Yuen. Quantum statistical query learning. *arXiv:2002.08240*, 2020. 21, 22

[ALL22] Anurag Anshu, Zeph Landau, and Yunchao Liu. Distributed quantum inner product estimation. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, STOC 2022, page 44–51. Association for Computing Machinery, 2022. 25

[ALVV17] Itai Arad, Zeph Landau, Umesh Vazirani, and Thomas Vidick. Rigorous rg algorithms and area laws for low energy eigenstates in 1d. *Communications in Mathematical Physics*, 356(1):65–105, Nov 2017. 13

[AN22] Anurag Anshu and Chinmay Nirkhe. Circuit Lower Bounds for Low-Energy States of Quantum Code Hamiltonians. In Mark Braverman, editor, *13th Innovations in Theoretical Computer Science Conference (ITCS 2022)*, volume 215 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 6:1–6:22, Dagstuhl, Germany, 2022. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. 25

[AQS21] Srinivasan Arunachalam, Yihui Quek, and John A. Smolin. Private learning implies quantum stability. In *Advances in Neural Information Processing Systems 34: Annual Conference on Neural Information Processing Systems 2021, NeurIPS 2021, December 6-14, 2021, virtual*, pages 20503–20515, 2021. 19, 20

[AQY21] Prabhanjan Ananth, Luowen Qian, and Henry Yuen. Cryptography from pseudorandom quantum states. *arXiv:2112.10020*, 2021. 6

[AR19] Scott Aaronson and Guy N Rothblum. Gentle measurement of quantum states and differential privacy. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 322–333, 2019. 19

[AS07] Alp Atıcı and Rocco A Servedio. Quantum algorithms for learning and testing juntas. *Quantum Information Processing*, 6(5):323–348, 2007. 21

[AS14] Joseph Albert and Robert H. Swendsen. The inverse Ising problem. *Physics Procedia*, 57:99–103, 2014. Proceedings of the 27th Workshop on Computer Simulation Studies in Condensed Matter Physics (CSP2014). 8

[ASZ+21] Amira Abbas, David Sutter, Christa Zoufal, Aurélien Lucchi, Alessio Figalli, and Stefan Woerner. The power of quantum neural networks. *Nature Computational Science*, 1(6):403–409, 2021. 24

[BAL19] Eyal Bairey, Itai Arad, and Netanel H Lindner. Learning a local hamiltonian from local measurements. *Physical Review Letters*, 122(2):020504, 2019. 12

[BBC+19] Sergey Bravyi, Dan Browne, Padraic Calpin, Earl Campbell, David Gosset, and Mark Howard. Simulation of quantum circuits by low-rank stabilizer decompositions. *Quantum*, 3:181, 2019. 6

[BBF+20] Kerstin Beer, Dmytro Bondarenko, Terry Farrelly, Tobias J Osborne, Robert Salzmann, Daniel Scheiermann, and Ramona Wolf. Training deep quantum neural networks. *Nature communications*, 11(1):808, 2020. 24

[BCvD06] Dave Bacon, Andrew M. Childs, and Wim van Dam. Optimal measurements for the dihedral hidden subgroup problem. *Chic. J. Theor. Comput. Sci.*, 2006, 2006. 26

[BG16a] Sergey Bravyi and David Gosset. Improved classical simulation of quantum circuits dominated by Clifford gates. *Physical Review Letters*, 116:250501, Jun 2016. 6

[BG16b] Sergey Bravyi and David Gosset. Improved classical simulation of quantum circuits dominated by Clifford gates. *Physical Review Letters*, 116(25):250501, 2016. 7

[BJ95] Nader H. Bshouty and Jeffrey C. Jackson. Learning DNF over the uniform distribution using a quantum example oracle. In Wolfgang Maass, editor, *Proceedings of the Eigth Annual Conference on Computational Learning Theory, COLT*, pages 118–127. ACM, 1995. 20, 21

[BJS11] Michael J Bremner, Richard Jozsa, and Dan J Shepherd. Classical simulation of commuting quantum computations implies collapse of the polynomial hierarchy. *Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences*, 467(2126):459–472, 2011. 7

[BKL+19] Fernando G. S. L. Brandão, Amir Kalev, Tongyang Li, Cedric Yen-Yu Lin, Krysta M. Svore, and Xiaodi Wu. Quantum SDP Solvers: Large Speed-Ups, Optimality, and Applications to Quantum Learning. In *46th International Colloquium on Automata, Languages, and Programming (ICALP 2019)*, volume 132 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 27:1–27:14, 2019. 23

[BKW03] Avrim Blum, Adam Kalai, and Hal Wasserman. Noise-tolerant learning, the parity problem, and the statistical query model. *Journal of the ACM (JACM)*, 50(4):506–519, 2003. 15

[BLM20] Mark Bun, Roi Livni, and Shay Moran. An equivalence between private classification and online prediction. In *61st IEEE Annual Symposium on Foundations of Computer Science, FOCS*, pages 389–402. IEEE, 2020. 19, 20

[BO21] Costin Badescu and Ryan O'Donnell. Improved quantum data analysis. In *STOC '21: 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1398–1411. ACM, 2021. 18

[Bre15] Guy Bresler. Efficiently learning Ising models on arbitrary graphs. In *STOC'15—Proceedings of the 2015 ACM Symposium on Theory of Computing*, pages 771–782, 2015. 8, 10

[BS19] Zvika Brakerski and Omri Shmueli. (Pseudo) Random Quantum States with Binary Phase. In *Theory of Cryptography Conference*, pages 229–250. Springer, 2019. 6

[BSS16] Sergey Bravyi, Graeme Smith, and John A. Smolin. Trading classical and quantum computational resources. *Phys. Rev. X*, 6:021043, Jun 2016. 6

[BV97] Ethan Bernstein and Umesh Vazirani. Quantum complexity theory. *SIAM Journal on Computing*, 26(5):1411–1473, 1997. 7

[BW02] Harry Buhrman and Ronald de Wolf. Complexity measures and decision tree complexity: a survey. *Theoretical Computer Science*, 288(1):21–43, 2002. 4

[Car20] Matthias C Caro. Quantum learning boolean linear functions wrt product distributions. *Quantum Information Processing*, 19(6):172, 2020. 21

[Car21] Matthias C Caro. Binary classification with classical instances and quantum labels. *Quantum Machine Intelligence*, 3(1):18, 2021. 18, 24

[Car22] Matthias C Caro. Learning quantum processes and Hamiltonians via the Pauli transfer matrix. *arXiv:2212.04471*, 2022. 24

[CCHL21] Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. Exponential separations between learning with and without quantum memory. In *62nd IEEE Annual Symposium on Foundations of Computer Science, FOCS*, pages 574–585. IEEE, 2021. 4, 18, 22

[CD16] Iris Cong and Luming Duan. Quantum discriminant analysis for dimensionality reduction and classification. *New Journal of Physics*, 18(7):073011, jul 2016. 23

[CD20] Matthias C Caro and Ishaun Datta. Pseudo-dimension of quantum circuits. *Quantum Machine Intelligence*, 2(2):14, 2020. 24

[CGFM+21] Matthias C Caro, Elies Gil-Fuster, Johannes Jakob Meyer, Jens Eisert, and Ryan Sweke. Encoding-dependent generalization bounds for parametrized quantum circuits. *Quantum*, 5:582, 2021. 24

[CGK17] Shawn X Cui, Daniel Gottesman, and Anirudh Krishna. Diagonal gates in the Clifford hierarchy. *Physical Review A*, 95(1):012329, 2017. 7

[CGL+20] Nai-Hui Chia, András Gilyén, Tongyang Li, Han-Hsuan Lin, Ewin Tang, and Chunhao Wang. Sampling-based sublinear low-rank matrix arithmetic framework for dequantizing quantum machine learning. In *Proceedings of the 52nd Annual ACM SIGACT Symposium on Theory of Computing*, STOC, page 387–400, 2020. 23

[CHC+22] Matthias C Caro, Hsin-Yuan Huang, Marco Cerezo, Kunal Sharma, Andrew Sornborger, Lukasz Cincio, and Patrick J Coles. Generalization in quantum machine learning from few training data. *Nature communications*, 13(1):4919, 2022. 24

[CHE+22] Matthias C Caro, Hsin-Yuan Huang, Nicholas Ezzell, Joe Gibbs, Andrew T Sornborger, Lukasz Cincio, Patrick J Coles, and Zoë Holmes. Out-of-distribution generalization for learning quantum dynamics. *arXiv:2204.10268*, 2022. 24

[CHL+22a] Sitan Chen, Brice Huang, Jerry Li, Allen Liu, and Mark Sellke. Tight bounds for state tomography with incoherent measurements. *arXiv:2206.05265*, 2022. 4

[CHL+22b] Xinyi Chen, Elad Hazan, Tongyang Li, Zhou Lu, Xinzhao Wang, and Rui Yang. Adaptive online learning of quantum states. *arXiv:2206.00220*, 2022. 15

[CHY16] Hao-Chung Cheng, Min-Hsiu Hsieh, and Ping-Cheng Yeh. The learnability of unknown quantum measurements. *Quantum Inf. Comput.*, 16(7&8):615–656, 2016. 15

[CL68] C. Chow and C. Liu. Approximating discrete probability distributions with dependence trees. *IEEE Transactions on Information Theory*, 14(3):462–467, 1968. 8

[CL21] Kai-Min Chung and Han-Hsuan Lin. Sample efficient algorithms for learning quantum channels in PAC model and the approximate state discrimination problem. In Min-Hsiu Hsieh, editor, *16th Conference on the Theory of Quantum Computation, Communication and Cryptography, TQC 2021, July 5-8, 2021, Virtual Conference*, volume 197 of *LIPIcs*, pages 3:1–3:22. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2021. 18, 24

[CLW18] Nai-Hui Chia, Han-Hsuan Lin, and Chunhao Wang. Quantum-inspired sublinear classical algorithms for solving low-rank linear systems, 2018. 23

[CMDK20] Brian Coyle, Daniel Mills, Vincent Danos, and Elham Kashefi. The born supremacy: quantum advantage and training of an Ising born machine. *NPJ Quantum Information*, 6(1):60, Jul 2020. 24

[CPF+10a] Marcus Cramer, Martin B Plenio, Steven T Flammia, Rolando Somma, David Gross, Stephen D Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai Liu. Efficient quantum state tomography. *Nature Communications*, 1(1):1–7, 2010. 3, 24

[CPF+10b] Marcus Cramer, Martin B. Plenio, Steven T. Flammia, Rolando Somma, David Gross, Stephen D. Bartlett, Olivier Landon-Cardinal, David Poulin, and Yi-Kai Liu. Efficient quantum state tomography. *Nature Communications*, 1(1):149, Dec 2010. 12, 13

[DBH19] Chen Ding, Tian-Yi Bao, and He-Liang Huang. Quantum-inspired support vector machine, 2019. 23

[DDM03] Jeroen Dehaene and Bart De Moor. Clifford group, stabilizer states, and linear and quadratic operations over gf (2). *Physical Review A*, 68(4):042318, 2003. 5

[DHL+21] Yuxuan Du, Min-Hsiu Hsieh, Tongliang Liu, Shan You, and Dacheng Tao. Learnability of quantum neural networks. *PRX Quantum*, 2(4):040337, 2021. 23

[DPW+21] Arkopal Dutt, Edwin Pednault, Chai Wah Wu, Sarah Sheldon, John Smolin, Lev Bishop, and Isaac L Chuang. Active learning of quantum system hamiltonians yields query advantage. *arXiv:2112.14553*, 2021. 24

[DR+14] Cynthia Dwork, Aaron Roth, et al. The algorithmic foundations of differential privacy. *Foundations and Trends in Theoretical Computer Science*, 9(3–4):211–407, 2014. 19

[Dwo06] Cynthia Dwork. Differential privacy. In *Automata, Languages and Programming: 33rd International Colloquium, ICALP 2006, Venice, Italy, July 10-14, 2006, Proceedings, Part II 33*, pages 1–12. Springer, 2006. 19

[ECP10] J. Eisert, M. Cramer, and M. B. Plenio. Colloquium: Area laws for the entanglement entropy. *Rev. Mod. Phys.*, 82:277–306, Feb 2010. 1

[EF01] Yonina C Eldar and G David Forney. On quantum detection and the square-root measurement. *IEEE Transactions on Information Theory*, 47(3):858–872, 2001. 3

[EHK04] Mark Ettinger, Peter Høyer, and Emanuel Knill. The quantum query complexity of the hidden subgroup problem is polynomial. *Inf. Process. Lett.*, 91(1):43–48, 2004. 25

[FBaK21] Daniel Stilck França, Fernando G.S L. Brandão, and Richard Kueng. Fast and Robust Quantum State Tomography from Few Basis Measurements. In Min-Hsiu Hsieh, editor, *16th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2021)*, volume 197 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 7:1–7:13. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2021. 16

[Fel09] Vitaly Feldman. Distribution-specific agnostic boosting. *arXiv:0909.2927*, 2009. 21

[FGLE12] Steven T Flammia, David Gross, Yi-Kai Liu, and Jens Eisert. Quantum tomography via compressed sensing: error bounds, sample complexity and efficient estimators. *New Journal of Physics*, 14(9):095022, 2012. 3, 24

[FGR+17] Vitaly Feldman, Elena Grigorescu, Lev Reyzin, Santosh S Vempala, and Ying Xiao. Statistical algorithms and a lower bound for detecting planted cliques. *Journal of the ACM (JACM)*, 64(2):1–37, 2017. 21

[FIM+14] Katalin Friedl, Gábor Ivanyos, Frédéric Magniez, Miklos Santha, and Pranab Sen. Hidden translation and translating coset in quantum computing. *SIAM Journal on Computing*, 43(1):1–24, Jan 2014. 26

[FMMC12] Austin G. Fowler, Matteo Mariantoni, John M. Martinis, and Andrew N. Cleland. Surface codes: Towards practical large-scale quantum computation. *Physical Review A*, 86(3), sep 2012. 6

[FO23] Steve Flammia and Ryan O'Donnell. Quantum chi-squared tomography and mutual information testing. *arXiv:2305.18519v1*, 2023. 3

[FQR22] Marco Fanizza, Yihui Quek, and Matteo Rosati. Learning quantum processes without input control. *arXiv:2211.05005*, 2022. 18, 24

[GA22] Weiyuan Gong and Scott Aaronson. Learning distributions over quantum measurement outcomes. *arXiv:2209.03007*, 2022. 19

[GC99] Daniel Gottesman and Isaac L Chuang. Demonstrating the viability of universal quantum computation using teleportation and single-qubit operations. *Nature*, 402(6760):390–393, 1999. 7

[GGC+21] Jennifer R. Glick, Tanvi P. Gujarati, Antonio D. Corcoles, Youngseok Kim, Abhinav Kandala, Jay M. Gambetta, and Kristan Temme. Covariant quantum kernels for data with group structure, 2021. 23

[GHC+22] Joe Gibbs, Zoë Holmes, Matthias C Caro, Nicholas Ezzell, Hsin-Yuan Huang, Lukasz Cincio, Andrew T Sornborger, and Patrick J Coles. Dynamical simulation via quantum machine learning with provable generalization. *arXiv:2204.10269*, 2022. 24

[GIKL23a] Sabee Grewal, Vishnu Iyer, William Kretschmer, and Daniel Liang. Efficient learning of quantum states prepared with few non-clifford gates. 2023. 6

[GIKL23b] Sabee Grewal, Vishnu Iyer, William Kretschmer, and Daniel Liang. Improved stabilizer estimation via bell difference sampling. *arXiv:2304.13915*, 2023. 6

[GIKL23c] Sabee Grewal, Vishnu Iyer, William Kretschmer, and Daniel Liang. Low-stabilizer-complexity quantum states are not pseudorandom. In Yael Tauman Kalai, editor, *14th Innovations in Theoretical Computer Science Conference, ITCS 2023, January 10-13, 2023, MIT, Cambridge, Massachusetts, USA*, volume 251 of *LIPIcs*, pages 64:1–64:20. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2023. 6

[GKK08] Parikshit Gopalan, Adam Tauman Kalai, and Adam R. Klivans. Agnostically learning decision trees. In *Proceedings of the 40th Annual ACM Symposium on Theory of Computing*, pages 527–536. ACM, 2008. 18

[GKZ19] Alex B. Grilo, Iordanis Kerenidis, and Timo Zijlstra. Learning-with-errors problem is easy with quantum samples. *Phys. Rev. A*, 99:032314, Mar 2019. 21

[GL22] Aravind Gollakota and Daniel Liang. On the hardness of PAC-learning stabilizer states with noise. *Quantum*, 6:640, 2022. 15

[GLT18] András Gilyén, Seth Lloyd, and Ewin Tang. Quantum-inspired low-rank stochastic regression with logarithmic dependence on the dimension, 2018. 23

[GSG+23] Valentin Gebhart, Raffaele Santagati, Antonio Andrea Gentile, Erik M. Gauger, David Craig, Natalia Ares, Leonardo Banchi, Florian Marquardt, Luca Pezzè, and Cristian Bonato. Learning quantum systems. *Nature Reviews Physics*, Feb 2023. 24

[GYN22] Abigail McClain Gomez, Susanne F. Yelin, and Khadijeh Najafi. Reconstructing quantum states using basis-enhanced born machines, 2022. 24

[HCP22] Hsin-Yuan Huang, Sitan Chen, and John Preskill. Learning to predict arbitrary quantum processes. *arXiv:2210.14894*, 2022. 24

[HCT+19] Vojtěch Havlíček, Antonio D. Córcoles, Kristan Temme, Aram W. Harrow, Abhinav Kandala, Jerry M. Chow, and Jay M. Gambetta. Supervised learning with quantum-enhanced feature spaces. *Nature*, 567(7747):209–212, Mar 2019. 23

[HHJ+17] Jeongwan Haah, Aram W Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, 63(9):5628–5641, 2017. 3

[HHL09] Aram W. Harrow, Avinatan Hassidim, and Seth Lloyd. Quantum algorithm for linear systems of equations. *Phys. Rev. Lett.*, 103:150502, Oct 2009. 23

[HIN+21] Marcel Hinsche, Marios Ioannou, Alexander Nietner, Jonas Haferkamp, Yihui Quek, Dominik Hangleiter, Jean-Pierre Seifert, Jens Eisert, and Ryan Sweke. Learnability of the output distributions of local quantum circuits. *arXiv:2110.05517*, 2021. 22

[HIN+22] Marcel Hinsche, Marios Ioannou, Alexander Nietner, Jonas Haferkamp, Yihui Quek, Dominik Hangleiter, Jean-Pierre Seifert, Jens Eisert, and Ryan Sweke. A single T-gate makes distribution learning hard. *arXiv:2207.03140*, 2022. 6, 22

[HKOT23] Jeongwan Haah, Robin Kothari, Ryan O'Donnell, and Ewin Tang. Query-optimal estimation of unitary channels in diamond distance. *arXiv:2302.14066*, 2023. 24

[HKP20] Hsin-Yuan Huang, Richard Kueng, and John Preskil. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16:1050–1057, 2020. 17, 18, 19

[HKT22a] Jeongwan Haah, Robin Kothari, and Ewin Tang. Optimal learning of quantum hamiltonians from high-temperature gibbs states. In *63rd IEEE Annual Symposium on Foundations of Computer Science, FOCS 2022, Denver, CO, USA, October 31 - November 3, 2022*, pages 135–146. IEEE, 2022. 12, 24

[HKT+22b] Hsin-Yuan Huang, Richard Kueng, Giacomo Torlai, Victor V Albert, and John Preskill. Provably efficient machine learning for quantum many-body problems. *Science*, 377(6613):eabk3333, 2022. 24

[HLM17] Aram W. Harrow, Cedric Yen-Yu Lin, and Ashley Montanaro. Sequential measurements, disturbance and property testing. In *Proceedings of the Twenty-Eighth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA*, pages 1598–1611. SIAM, 2017. 16

[HRS10] Sean Hallgren, Martin Rötteler, and Pranab Sen. Limitations of quantum coset states for graph isomorphism. *Journal of the ACM*, 57(6):34:1–34:33, 2010. 26

[HRTS00] Sean Hallgren, Alexander Russell, and Amnon Ta-Shma. Normal subgroup reconstruction and quantum computation using group representations. In *Proceedings of the thirty-second annual ACM symposium on Theory of computing*, pages 627–635, 2000. 26

[HS+86] Geoffrey E Hinton, Terrence J Sejnowski, et al. Learning and relearning in Boltzmann machines. *Parallel distributed processing: Explorations in the microstructure of cognition*, 1(282-317):2, 1986. 8

[HTFS23] Hsin-Yuan Huang, Yu Tong, Di Fang, and Yuan Su. Learning many-body hamiltonians with heisenberg-limited scaling. *Physical Review Letters*, 130(20):200403, 2023. 24

[INN+22] Sandy Irani, Anand Natarajan, Chinmay Nirkhe, Sujit Rao, and Henry Yuen. Quantum search-to-decision reductions and the state synthesis problem. In *37th Computational Complexity Conference, CCC*, volume 234 of *LIPIcs*, pages 5:1–5:19. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2022. 6

[JEP+21] John Jumper, Richard Evans, Alexander Pritzel, Tim Green, Michael Figurnov, Olaf Ronneberger, Kathryn Tunyasuvunakool, Russ Bates, Augustin Žídek, Anna Potapenko, et al. Highly accurate protein structure prediction with alphafold. *Nature*, 596(7873):583–589, 2021. 1

[JLS18] Zhengfeng Ji, Yi-Kai Liu, and Fang Song. Pseudorandom quantum states. In *Annual International Cryptology Conference*, pages 126–152. Springer, 2018. 6

[JYD+21] Yu-Xiao Jiang, Jia-Xin Yin, M. Michael Denner, Nana Shumiya, Brenden R. Ortiz, Gang Xu, Zurab Guguchia, Junyi He, Md Shafayat Hossain, Xiaoxiong Liu, Jacob Ruff, Linus Kautzsch, Songtian S. Zhang, Guoqing Chang, Ilya Belopolski, Qi Zhang, Tyler A. Cochran, Daniel Multer, Maksim Litskevich, Zi-Jia Cheng, Xian P. Yang, Ziqiang Wang, Ronny Thomale, Titus Neupert, Stephen D. Wilson, and M. Zahid Hasan. Unconventional chiral charge order in kagome superconductor kv3sb5. *Nature Materials*, 20(10):1353–1357, Oct 2021. 8

[KB19] Kohtaro Kato and Fernando G. S. L. Brandão. Quantum approximate markov chains are thermal. *Communications in Mathematical Physics*, 370(1):117–149, Aug 2019. 12

[Kea98] M. J. Kearns. Efficient noise-tolerant learning from statistical queries. *Journal of the ACM*, 45(6):983–1006, 1998. Earlier version in STOC'03. 21

[Kit96] Alexei Y. Kitaev. Quantum measurements and the abelian stabilizer problem. *Electron. Colloquium Comput. Complex.*, 1996. 26

[KKBa20] Tomotaka Kuwahara, Kohtaro Kato, and Fernando G. S. L. Brandão. Clustering of conditional mutual information for quantum Gibbs states above a threshold temperature. *Phys. Rev. Lett.*, 124:220601, Jun 2020. 11, 12

[KLL+17] Shelby Kimmel, Cedric Yen-Yu Lin, Guang Hao Low, Maris Ozols, and Theodore J Yoder. Hamiltonian simulation with optimal sample complexity. *NPJ Quantum Information*, 3(1):1–7, 2017. 26

[KLN+11] Shiva Prasad Kasiviswanathan, Homin K. Lee, Kobbi Nissim, Sofya Raskhodnikova, and Adam D. Smith. What can we learn privately? *SIAM Journal on Computing*, 40(3):793–826, 2011. 19

[KM17] Adam R. Klivans and Raghu Meka. Learning graphical models using multiplicative weights. In Chris Umans, editor, *58th IEEE Annual Symposium on Foundations of Computer Science, FOCS*, pages 343–354. IEEE Computer Society, 2017. 8, 10

[KMM13] Vadym Kliuchnikov, Dmitri Maslov, and Michele Mosca. Fast and efficient exact synthesis of single qubit unitaries generated by Clifford and t gates. *Quantum Information and Computation*, 13(7–8):607–630, 2013. 6

[KP17] Iordanis Kerenidis and Anupam Prakash. Quantum recommendation systems. In *8th Innovations in Theoretical Computer Science Conference, ITCS*, volume 67 of *LIPIcs*, pages 49:1–49:21. Schloss Dagstuhl - Leibniz-Zentrum für Informatik, 2017. 23

[KR21] Behnoush Khavari and Guillaume Rabusseau. Lower and upper bounds on the pseudo-dimension of tensor network models. *Advances in Neural Information Processing Systems*, 34:10931–10943, 2021. 24

[KRS18] Varun Kanade, Andrea Rocchetto, and Simone Severini. Learning DNFs under product distributions via μ-biased quantum Fourier sampling. *arXiv:1802.05690*, 2018. 21

[KRT17] Richard Kueng, Holger Rauhut, and Ulrich Terstiege. Low rank matrix recovery from rank one measurements. *Applied and Computational Harmonic Analysis*, 42(1):88–116, 2017. 3, 4

[KS20] Tomotaka Kuwahara and Keiji Saito. Gaussian concentration bound and ensemble equivalence in generic quantum many-body systems including long-range interactions. *Annals of Physics*, 421:168278, 2020. 12

[KvBE+21] Christian Kokail, Rick van Bijnen, Andreas Elben, Benoît Vermersch, and Peter Zoller. Entanglement hamiltonian tomography in quantum simulation. *Nature Physics*, 17(8):936–942, Aug 2021. 3, 9

[LAT21] Yunchao Liu, Srinivasan Arunachalam, and Kristan Temme. A rigorous and robust quantum speed-up in supervised machine learning. *Nature Physics*, 17(9):1013–1017, 2021. 23

[LC22] Ching-Yi Lai and Hao-Chung Cheng. Learning quantum circuits of some T gates. *IEEE Transactions on Information Theory*, 68(6):3951–3964, 2022. 6

[LCLP10] Olivier Landon-Cardinal, Yi-Kai Liu, and David Poulin. Efficient direct tomography for matrix product states, 2010. 12

[LGZ16] Seth Lloyd, Silvano Garnerone, and Paolo Zanardi. Quantum algorithms for topological and geometric analysis of data. *Nature communications*, 7(1):1–7, 2016. 23

[LHT+23] Laura Lewis, Hsin-Yuan Huang, Viet T Tran, Sebastian Lehner, Richard Kueng, and John Preskill. Improved machine learning algorithm for predicting ground state properties. *arXiv:2301.13169*, 2023. 24

[Lia22] Daniel Liang. Clifford Circuits can be Properly PAC Learned if and only if $\mathsf{RP} = \mathsf{NP}$. *arXiv:2204.06638*, 2022. 8

[LMN93] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, fourier transform, and learnability. *Journal of the ACM*, 40(3):607–620, 1993. 18

[LMR13] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. Quantum algorithms for supervised and unsupervised machine learning. *arXiv:1307.0411*, 2013. 23

[LMR14] Seth Lloyd, Masoud Mohseni, and Patrick Rebentrost. Quantum principal component analysis. *Nature Physics*, 10(9):631–633, Sep 2014. 23, 26

[LN22] Angus Lowe and Ashwin Nayak. Lower bounds for learning quantum states with single-copy measurements. *arXiv:2207.14438*, 2022. 4

[Low09] Richard A Low. Learning and testing algorithms for the Clifford group. *Physical Review A*, 80(5):052314, 2009. 5, 7

[LRZ23] Qipeng Liu, Ran Raz, and Wei Zhan. Memory-sample lower bounds for learning with classical-quantum hybrid memory. In Barna Saha and Rocco A. Servedio, editors, *Proceedings of the 55th Annual ACM Symposium on Theory of Computing, STOC 2023, Orlando, FL, USA, June 20-23, 2023*, pages 1097–1110. ACM, 2023. 6

[LVV15] Zeph Landau, Umesh Vazirani, and Thomas Vidick. A polynomial time algorithm for the ground state of one-dimensional gapped local hamiltonians. *Nature Physics*, 11(7):566–569, Jul 2015. 13, 14

[Mon17a] Ashley Montanaro. Learning stabilizer states by Bell sampling. *arXiv:1707.04012*, 2017. 5, 7

[Mon17b] Ashley Montanaro. Quantum circuits and low-degree polynomials over $\mathbb{F}_2$. *Journal of Physics A: Mathematical and Theoretical*, 50(8):084002, Jan 2017. 7

[MW05] Chris Marriott and John Watrous. Quantum Arthur—Merlin games. *Comput. Complex.*, 14(2):122–152, jun 2005. 25

[Nes08] M Nest. Classical simulation of quantum computation, the Gottesman-Knill theorem, and slightly beyond. *arXiv:0811.0898*, 2008. 5

[NIS+23] Alexander Nietner, Marios Ioannou, Ryan Sweke, Richard Keung, Jens Eisert, Marcel Hinsche, and Jonas Haferkamp. On the average-case complexity of learning output distributions of quantum circuits, 2023. arXiv:2305.05765. 22

[ORFW23] Emilio Onorati, Cambyse Rouzé, Daniel Stilck França, and James D Watson. Efficient learning of ground & thermal states within phases of matter. *arXiv:2301.12946*, 2023. 24

[OW15] Ryan O'Donnell and John Wright. Quantum spectrum testing. In *Proceedings of the forty-seventh annual ACM symposium on Theory of computing*, pages 529–538, 2015. 4

[OW16] Ryan O'Donnell and John Wright. Efficient quantum tomography. In *Proceedings of the forty-eighth annual ACM symposium on Theory of Computing*, pages 899–912, 2016. 3, 4

[OW17] Ryan O'Donnell and John Wright. Efficient quantum tomography ii. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 962–974, 2017. 4

[OW18] Ryan O'Donnell and John Wright. A primer on the statistics of longest increasing subsequences and quantum states. *SIGACT News*, 2018. 3

[QR19] Xiao-Liang Qi and Daniel Ranard. Determining a local Hamiltonian from a single eigenstate. *Quantum*, 3:159, 2019. 12

[Raz19] Ran Raz. Fast learning requires good memory: A time-space lower bound for parity learning. *Journal of the ACM*, 66(1):3:1–3:18, 2019. 5

[RB98] Martin Roetteler and Thomas Beth. Polynomial-time solution to the hidden subgroup problem for a class of non-abelian groups. *quant-ph/9812070*, 1998. 26

[Rey20] Lev Reyzin. Statistical queries and statistical algorithms: Foundations and applications. *arXiv:2004.00557*, 2020. 21

[RF21] Cambyse Rouzé and Daniel Stilck França. Learning quantum many-body systems from a few copies. *arXiv:2107.03333*, 2021. 24

[RHBM13] Matteo Rossi, Marcus Huber, Dagmar Bruß, and Chiara Macchiavello. Quantum hypergraph states. *New Journal of Physics*, 15(11):113022, 2013. 6

[RML14] Patrick Rebentrost, Masoud Mohseni, and Seth Lloyd. Quantum support vector machine for big data classification. *Phys. Rev. Lett.*, 113:130503, Sep 2014. 23

[Roc17] Andrea Rocchetto. Stabiliser states are efficiently PAC-learnable. *arXiv:1705.00345*, 2017. 15

[Röt09] Martin Rötteler. Quantum algorithms to solve the hidden shift problem for quadratics and for functions of large Gowers norm. In *International Symposium on Mathematical Foundations of Computer Science*, pages 663–674. Springer, 2009. 7

[RS16] Neil J. Ross and Peter Selinger. Optimal ancilla-free Clifford+T approximation of z-rotations. *Quantum Information and Computation*, 16(11–12):901–953, 2016. 6

[RSML18] Patrick Rebentrost, Adrian Steffens, Iman Marvian, and Seth Lloyd. Quantum singular-value decomposition of nonsparse low-rank matrices. *Phys. Rev. A*, 97:012327, Jan 2018. 23

[RST15] Alexander Rakhlin, Karthik Sridharan, and Ambuj Tewari. Online learning via sequential complexities. *Journal of Machine Learning Research*, 16(1):155–186, 2015. 15

[SBB98] A. Yu. Kitaev S. B. Bravyi. *Quantum codes on a lattice with boundary*, 1998. 6

[Sel15] Peter Selinger. Efficient Clifford+T approximation of single-qubit operators. *Quantum Information and Computation*, 15(1–2):159–180, 2015. 6

[SHB+22] Henning Schlomer, Timon Hilker, Immanuel Bloch, Ulrich Schollwock, Fabian Grusdt, and Annabelle Bohrdt. Quantifying hole-motion-induced frustration in doped antiferromagnets by hamiltonian reconstruction, 2022. 8

[SK19] Maria Schuld and Nathan Killoran. Quantum machine learning in feature Hilbert spaces. *Phys. Rev. Lett.*, 122:040504, Feb 2019. 23

[SSS+17] David Silver, Julian Schrittwieser, Karen Simonyan, Ioannis Antonoglou, Aja Huang, Arthur Guez, Thomas Hubert, Lucas Baker, Matthew Lai, Adrian Bolton, et al. Mastering the game of go without human knowledge. *nature*, 550(7676):354–359, 2017. 1

[Tan98] Toshiyuki Tanaka. Mean-field theory of Boltzmann machine learning. *Phys. Rev. E*, 58:2302–2310, 1998. 8

[Tan19] Ewin Tang. A quantum-inspired classical algorithm for recommendation systems. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, STOC, page 217–228, 2019. 23

[Tan21] Ewin Tang. Quantum principal component analysis only achieves an exponential speedup because of its state preparation assumptions. *Physical Review Letters*, 127(6):060503, 2021. 23

[TMH19] Yuki Takeuchi, Tomoyuki Morimae, and Masahito Hayashi. Quantum computational universality of hypergraph states with Pauli-X and Z basis measurements. *Scientific Reports*, 9(1):1–14, 2019. 6

[Val84] Leslie G. Valiant. A theory of the learnable. *Communications of the ACM*, 27(11):1134–1142, 1984. 1, 14

[Vap13] Vladimir Vapnik. *The nature of statistical learning theory*. Springer science & business media, 2013. 23

[VMLC16] Marc Vuffray, Sidhant Misra, Andrey Lokhov, and Michael Chertkov. Interaction screening: Efficient and sample-optimal learning of Ising models. In *Advances in Neural Information Processing Systems*, pages 2595–2603, 2016. 8, 10

[VMN+19] Guillaume Verdon, Jacob Marks, Sasha Nanda, Stefan Leichenauer, and Jack Hidary. Quantum hamiltonian-based models and the variational quantum thermalizer algorithm. *arXiv:1910.02071*, 2019. 24

[VSP+17] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 1

[WA22] Dominik S. Wild and Alvaro M. Alhambra. Classical simulation of short-time quantum dynamics, 2022. 12

[WBL12] Nathan Wiebe, Daniel Braun, and Seth Lloyd. Quantum algorithm for data fitting. *Phys. Rev. Lett.*, 109:050505, Aug 2012. 23, 24

[WGFC14a] Nathan Wiebe, Christopher Granade, Christopher Ferrie, and David G Cory. Hamiltonian learning and certification using quantum resources. *Physical review letters*, 112(19):190501, 2014. 24

[WGFC14b] Nathan Wiebe, Christopher Granade, Christopher Ferrie, and David G Cory. Hamiltonian learning and certification using quantum resources. *Physical Review Letters*, 112(19):190501, 2014. 24

[Wri16] John Wright. *How to learn a quantum state*. PhD thesis, Carnegie Mellon University, 2016. 3, 4

[Yat85] Yannis G Yatracos. Rates of convergence of minimum distance estimators and kolmogorov's entropy. *The Annals of Statistics*, 13(2):768–774, 1985. 18

[Yue22] Henry Yuen. An improved sample complexity lower bound for quantum state tomography. *arXiv:2206.11185*, 2022. 4

[ZFF19] Zhikuan Zhao, Jack K. Fitzsimons, and Joseph F. Fitzsimons. Quantum-assisted gaussian process regression. *Phys. Rev. A*, 99:052331, May 2019. 23

[ZGYN22] Weishun Zhong, Xun Gao, Susanne F. Yelin, and Khadijeh Najafi. Many-body localized hidden born machine, 2022. 24