Optimal Algorithms for Learning Quantum Phase States

Srinivasan Arunachalam
IBM Quantum, Thomas J Watson Research Center, Yorktown Heights, NY, USA

Sergey Bravyi
IBM Quantum, Thomas J Watson Research Center, Yorktown Heights, NY, USA

Arkopal Dutt
IBM Quantum, Thomas J Watson Research Center, Yorktown Heights, NY, USA
MIT-IBM Watson AI Lab, Cambridge, MA, USA
Department of Physics, Co-Design Center for Quantum Advantage, Massachusetts Institute of Technology, Cambridge, MA, USA

Theodore J. Yoder
IBM Quantum, Thomas J Watson Research Center, Yorktown Heights, NY, USA

Abstract

We analyze the complexity of learning n-qubit quantum phase states. A degree-d phase state is defined as a superposition of all 2^n basis vectors x with amplitudes proportional to $(-1)^{f(x)}$, where f is a degree-d Boolean polynomial over n variables. We show that the sample complexity of learning an unknown degree-d phase state is $\Theta(n^d)$ if we allow separable measurements and $\Theta(n^{d-1})$ if we allow entangled measurements. Our learning algorithm based on separable measurements has runtime $\mathsf{poly}(n)$ (for constant d) and is well-suited for near-term demonstrations as it requires only single-qubit measurements in the Pauli X and Z bases. We show similar bounds on the sample complexity for learning generalized phase states with complex-valued amplitudes. We further consider learning phase states when f has sparsity-s, degree-d in its $\mathbb{F}_2$ representation (with sample complexity $O(2^d sn)$), f has Fourier-degree-t (with sample complexity $O(2^{2t})$), and learning quadratic phase states with ε-global depolarizing noise (with sample complexity $O(n^{1+\varepsilon})$). These learning algorithms give us a procedure to learn the diagonal unitaries of the Clifford hierarchy and IQP circuits.

2012 ACM Subject Classification Theory of computation → Quantum complexity theory

Keywords and phrases Tomography, binary phase states, generalized phase states, IQP circuits

Digital Object Identifier 10.4230/LIPIcs.TQC.2023.3

Related Version *Full Version*: `https://arxiv.org/abs/2208.07851` [6]

Funding SA, SB, and TY were supported in part by the Army Research Office under Grant Number W911NF-20-1-0014.
Arkopal Dutt: AD was supported in part by the MIT-IBM Watson AI Lab, and in part by the U.S. Department of Energy, Office of Science, National Quantum Information Science Research Centers, Co-Design Center for Quantum Advantage under contract DE-SC0012704.

Acknowledgements AD thanks Isaac L Chuang for suggesting applications of the learning algorithms presented here and for useful comments on the manuscript. SA thanks Giacomo Nannicini and Chinmay Nirkhe for useful discussions.

18th Conference on the Theory of Quantum Computation, Communication and Cryptography (TQC 2023).
Editors: Omar Fawzi and Michael Walter; Article No. 3; pp. 3:1–3:24
Leibniz International Proceedings in Informatics
Schloss Dagstuhl – Leibniz-Zentrum für Informatik, Dagstuhl Publishing, Germany

1 Introduction

Quantum state tomography is the problem of learning an unknown quantum state ρ drawn from a specified class of states by performing measurements on multiple copies of ρ. The preeminence of this problem in verification of quantum experiments has motivated an in-depth study of state tomography protocols and their limitations for various classes of quantum states [23, 40, 5, 46]. The main figure of merit characterizing a state tomography protocol is its *sample complexity* defined as the number of copies of ρ consumed by the protocol in order to learn ρ. Of particular interest are classes of n-qubit quantum states that can be learned efficiently, such that the sample complexity grows only polynomially with n. Known examples of efficiently learnable states include Matrix Product States describing weakly entangled quantum spin chains [17], output states of Clifford circuits [36], output states of Clifford circuits with a single layer of T gates [30], and high-temperature Gibbs states of local Hamiltonians [4, 24]. Apart from their potential use in experiments, efficiently learnable quantum states are of great importance for quantum algorithm design. For example, a quantum algorithm for solving the dihedral hidden subgroup problem [7] can be viewed as a tomography protocol for learning so-called hidden subgroup states (although this protocol is efficient in term of its sample complexity, its runtime is believed to be super-polynomial [7]).

A natural question to then ask is: What are other classes of n-qubit quantum states that are ubiquitous in quantum computing, which can be learned efficiently? In this work, we consider the problem of state tomography for *phase states* associated with (generalized) Boolean functions. Phase states are encountered in quantum information theory [26], quantum algorithm design [7], quantum cryptography [29, 11], and quantum-advantage experiments [13, 15].

By definition, an n-qubit, degree-d binary phase state has the form

$$|\psi_f\rangle = 2^{-n/2} \sum_{x\in\{0,1\}^n} (-1)^{f(x)}|x\rangle, \tag{1}$$

where $f : \{0,1\}^n \to \{0,1\}$ is a degree-d polynomial, that is,

$$f(x) = \sum_{J\subseteq[n],\, |J|\le d} \alpha_J \prod_{j\in J} x_j \pmod 2, \tag{2}$$

for some coefficients $\alpha_J \in \{0,1\}$. Phase states associated with homogeneous degree-2 polynomials $f(x)$ coincide with graph states that play a prominent role in quantum information theory [26]. Such states can be alternatively represented as

$$|\psi_f\rangle = \prod_{(i,j)\in E} \mathsf{CZ}_{i,j}|+\rangle^{\otimes n},$$

where n qubits live at vertices of a graph, E is the set of graph edges, $\mathsf{CZ}_{i,j}$ is the controlled-Z gate applied to qubits i, j, and $|+\rangle = (|0\rangle + |1\rangle)/\sqrt{2}$. It is known that the output state of any Clifford circuit is locally equivalent to a graph state for a suitable graph [44]. Our results imply that graph states can be learned efficiently using only single-qubit gates and measurements. The best previously known protocol for learning graph states [36] requires entangled measurements across two copies of $|\psi_f\rangle$. Other examples of circuits producing phase states include measurement-based quantum computing [42] and a subclass of IQP circuits (Instantaneous Quantum Polynomial-time), which correspond to degree-3 phase states [37]. IQP circuits are prevalent in quantum-advantage experiments [13, 15] and are believed to be hard to simulate classically.

We also consider generalized degree-d phase states

$$|\psi_f\rangle = 2^{-n/2} \sum_{x\in\{0,1\}^n} \omega_q^{f(x)}|x\rangle, \qquad \omega_q = e^{2\pi i/q} \tag{3}$$

where $q \geq 2$ is an even integer and $f : \{0,1\}^n \to \mathbb{Z}_q$ is a degree-d polynomial, that is,

$$f(x) = \sum_{J\subseteq[n],\,|J|\leq d} \alpha_J \prod_{j\in J} x_j \pmod{q}. \tag{4}$$

for coefficients $\alpha_J \in \mathbb{Z}_q = \{0,1,\ldots,q-1\}$. It is also known that generalized degree-d phase states with $q = 2^d$ can be prepared from diagonal unitary operators [18] in the d-th level of the Clifford hierarchy [22]. Additionally, 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 [12, Appendix D]. Binary and generalized phase states have also found applications in cryptography [29, 11], and complexity theory [28] (we discuss this in the next section).

In this work, we consider learning phase states through two types of tomography protocols based on *separable* and *entangled* measurements. The former can be realized as a sequence of M independent measurements, each performed on a separate copy of $|\psi_f\rangle$ (furthermore our learning algorithms only require single *qubit* measurements). The latter performs a joint measurement on the state $|\psi_f\rangle^{\otimes M}$. Our goal is to then derive upper and lower bounds on the sample complexity M of learning f, as a function of n and d. In the next section, we state our main results. Interestingly, our protocols based on separable measurements require only single-qubit gates and single-qubit measurements making them well suited for near-term demonstrations.

1.1 Summary of contributions and applications

We first introduce some notation before giving an overview of our contributions. For every n and $d \leq n/2$, let $\mathcal{P}(n,d)$ be the set of all degree-d polynomials of the form Eq. (2). Let $\mathcal{P}_q(n,d)$ be the set of all degree-d $\mathbb{Z}_q$-valued polynomials of the form Eq. (3). By definition, $\mathcal{P}_2(n,d) \equiv \mathcal{P}(n,d)$. To avoid confusion, we shall refer to states defined in Eq. (1) as binary phase states and in Eq. (3) as generalized phase states. Our learning protocol takes as input integers n, d and M copies of a degree-d phase state $|\psi_f\rangle$ with unknown $f \in \mathcal{P}(n,d)$ (or $f \in \mathcal{P}_q(n,d)$). The protocol outputs a classical description of a polynomial $g \in \mathcal{P}(n,d)$ (or $g \in \mathcal{P}_q(n,d)$) such that $f = g$ with high probability.

The main result in this work are optimal algorithms for learning phase states if the algorithm is allowed to make separable or entangled measurements. Prior to our work, we are aware of only two works in this direction (i) algorithms for efficiently learning degree-1 and degree-2 phase states; (ii) Montanaro [35] considered learning multilinear polynomials f, assuming we have *query access* to f, which is a stronger learning model than the sample access model that we assume for our learning algorithm. In this work, we show that if allowed separable measurements, the *sample* complexity of learning binary phase states and generalized phase states is $O(n^d)$. If allowed entangled measurements, we obtain a sample complexity of $O(dn^{d-1})$ for learning binary phase states. We further consider settings where the unknown function f we are trying to learn is known to be sparse, has a small Fourier-degree and the setting when given noisy copies of the quantum phase state. In Table 1, we summarize all our main results (except the first two rows, which include the main prior work in this direction).

Table 1 Upper and lower bounds of sample complexity for exact learning of n-qubit phase states with degree-d. For precise statements of the bounds, we refer the reader to the theorem statements in this work and in the full version of the paper [6].

	Sample complexity	Time complexity	Measurements
Binary phase state $\mathbb{F}_2$-degree-1 [10]	$\Theta(1)$	$O(n^3)$	Separable
Binary phase state $\mathbb{F}_2$-degree-2 [36, 43]	$O(n)$	$O(n^3)$	Entangled
Binary phase state $\mathbb{F}_2$-degree-d	$\Theta(n^d)$ Theorem 7, 10	$O(n^{3d-2})$	Separable
Binary phase state $\mathbb{F}_2$-degree-d	$\Theta(n^{d-1})$ Theorem 9	$O(\exp(n^d \log 2))$	Entangled
Generalized phase states degree-d	$\Theta(n^d)$ Theorem 11	$O(\exp(n^d \log q))$	Separable
Sparse Binary phase state $\mathbb{F}_2$-degree-d, $\mathbb{F}_2$-sparsity s	$O(2^d sn)$ [6, Theorem 6]	$O(2^{3d}s^3n)$	Separable
Binary phase state $\mathbb{F}_2$-degree-2 with global depolarizing noise ε	$n^{1+O(\varepsilon)}$ [6, Theorem 9]	$O(2^{n/\log n})$	Entangled
Binary phase state $\mathbb{F}_2$-degree-2 with local depolarizing noise ε	$\Theta((1-\varepsilon)^n)$ [6, Theorem 11]	$O(2^{n/\log n})$	Entangled
Binary phase state Fourier-degree-d	$O(2^{2d})$ [6, Theorem 7]	$O(\exp(n^2))$	Entangled

Before we give a proof sketch of these results, we first discuss a couple of motivations for considering the task of learning phase states and corresponding applications.

Quantum complexity. Recently, there has been a few results in quantum cryptography [29, 3, 11] and complexity theory [28] which used the notion of phase states.

Ji et al. [29] introduced the notion of *pseudorandom quantum states* as states of the form $|\phi\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} \omega_N^{F(x)} |x\rangle$ where F is a pseudorandom function.[1] Ji et al. showed that states of the form $|\phi\rangle$ are efficiently preparable and statistically indistinguishable from a Haar random state, which given as input to a polynomial-time quantum algorithm. A subsequent work of Brakerski [11] showed that it suffices to consider $|\phi'\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} (-1)^{F(x)} |x\rangle$ (where F again is a pseudorandom function) and such states are also efficiently preparable and statistically indistinguishable from Haar random states. Subsequently, these states have found applications in proposing many cryptosystems [3]. Although none of these works discuss the degree of the phase function F, our result shows implicitly that when F is low-degree, then $|\phi\rangle$ is exactly learnable and hence distinguishable from Haar random states, implying that they cannot be quantum pseudorandom states. In another recent work, Irani et al. [28] considered the power of quantum witnesses in proof systems. In particular, they showed that in order to construct the witness to a QMA complete problem, say the ground state

[1] We do not discuss the details of pseudorandom functions here, we refer the interested reader to [29].

$|\phi\rangle$ to a local-Hamiltonian problem, it suffices to consider a phase state $\frac{1}{\sqrt{2^n}}\sum_x(-1)^{F(x)}|x\rangle$ which has a good overlap to $|\phi\rangle$. To this end, they show a strong property that, for *every* state $|\tau\rangle$ and a random Clifford operator U (or, more generally, an element of some unitary 2-design), the state $U|\tau\rangle$ has constant overlap with a phase state [28, Lemma A.5]. Our learning result implicitly shows that, assuming $\mathsf{QMA} \neq \mathsf{QCMA}$, then the phase state that has constant overlap with the ground space energy of the local Hamiltonian problem, cannot be of low degree.

Learning quantum circuits. Given access to a quantum circuit U, the goal of this learning task is to learn a circuit representation of U. The sample complexity for learning a general n-qubit quantum circuit is known to be $2^{\Theta(n)}$ [16, 34], which is usually impractical.

If we restrict ourselves to particular classes of quantum circuits, there are some known results for efficient learnability. Low [31] showed that an n-qubit Clifford circuit can be learned using $O(n)$ samples. However, this result was only an existential proof and requires access to the conjugate of the circuit. Constructive algorithms were given in Low [31], and Lai and Cheng [30], both of which showed that Clifford circuits can be learned using $O(n^2)$ samples. Both these algorithms require entangled measurements with the former algorithm using pretty-good measurement [25], and the latter using Bell sampling. In this work, we show that Clifford circuits producing degree-2 binary phase states, can be learned in $O(n^2)$ samples, matching their result but only using separable measurements. Moreover, Low [31] also gave an existential proof of algorithms for learning circuits in the d-th level of the Clifford hierarchy, using $O(n^{d-1})$ samples. In this work, we give constructive algorithms for learning the diagonal elements of the Clifford hierarchy in $O(n^d)$ samples using separable measurements. A direct result of this is that a subset of IQP circuits, which are also believed to be hard to simulate classically [13, 14], are shown to be efficiently learnable. Our learning result thus gives an efficient method for verifying IQP circuits that may be part of quantum-advantage experiments [15, 39].

Learning hypergraph states. We finally observe that degree-3 (and higher-degree) phase states have appeared in works [42, 45] on measurement-based quantum computing (MBQC), wherein they refer to these states as *hypergraph states*. These works show that single-qubit measurements in the Pauli X or Z basis performed on a suitable degree-3 hypergraph state are sufficient for universal MBQC. Our learning algorithm gives a procedure for learning these states in polynomial-time and could potentially be used as a subroutine for verifying MBQC.

1.2 Proof sketch

In this section we briefly sketch the proofs of our main results.

1.2.1 Binary phase states

As we mentioned earlier, Montanaro [36] and Roettler [43] showed how to learn degree-2 phase states using $O(n)$ copies of the state. Crucial to both their learning algorithms was the following so-called Bell-sampling procedure: given two copies of $|\psi_f\rangle = \frac{1}{\sqrt{2^n}}\sum_x(-1)^{f(x)}|x\rangle$ where $f(x) = x^\top Ax$ (where $A \in \mathbb{F}_2^{n\times n}$), perform n CNOTs from the first copy to the second, and measure the second copy. One obtains a uniformly random $y \in \mathbb{F}_2^n$ and the state

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

Using Bernstein-Vazirani [10] one can apply n-qubit Hadamard transform to obtain the bit string $(A + A^\top) \cdot y$. Repeating this process $O(n \log n)$ many times, one can learn n linearly independent constraints about A, and along with Gaussian elimination, allows to learn $A + A^\top$. Diagonal elements of A can be learned with one additional copy of $|\psi_f\rangle$. Applying a controlled-Z gate between all pairs of qubits $i > j$ for which $(A + A^\top)_{ij} = 1$ results in the state $\sum_x (-1)^{\sum_i x_i A_{ii}} |x\rangle$, which can be learned using Bernstein-Vazarani.

Applying this same Bell-sampling procedure to degree-3 phase states does not easily learn the phase function. In this direction, from two copies of the degree-3 phase state $|\psi_f\rangle$ one obtains a uniformly random $y \in \mathbb{F}_2^n$ and the state $|\psi_{g_y}\rangle = \frac{1}{\sqrt{2^n}} \sum_x (-1)^{g_y(x)} |x\rangle$ for a degree-2 polynomial $g_y(x) = f(x) + f(x + y)$. One might now hope to apply the degree-2 learning algorithm from above, but since the single copy of $|\psi_{g_y}\rangle$ was randomly generated, it takes $\Omega(\sqrt{2^n})$ copies of $|\psi_f\rangle$ to obtain enough copies of $|\psi_{g_y}\rangle$. Our main idea is to circumvent this Bell-sampling approach and instead propose two techniques that allow us to learn binary phase states using separable and entangled measurements which we discuss further below.

Separable measurements, upper bound. Our first result is that we are able to learn binary phase states using separable measurements with sample complexity $O(n^d)$. In order to prove our upper bounds of sample complexity for learning with separable measurements, we make a simple observation. Given one copy of $|\psi_f\rangle = \frac{1}{\sqrt{2^n}} \sum_x (-1)^{f(x)} |x\rangle$, measure qubits $2, 3, \ldots, n$ in the computational basis. Suppose the resulting string is $y \in \{0,1\}^{n-1}$. The post-measurement state of qubit 1 is then given by

$$|\psi_{f,y}\rangle = \frac{1}{\sqrt{2}} \left[(-1)^{f(0y)} |0\rangle + (-1)^{f(1y)} |1\rangle \right].$$

By applying a Hadamard transform to $|\psi_{f,y}\rangle$ and measuring, the algorithm obtains $p_1(y) = f(0y) + f(1y) \mod 2$, which can be viewed as the derivative of f in the first direction at point y. Furthermore observe that p_1 is a degree $\leq d - 1$ polynomial over $(n - 1)$ variables. Hence, the learning algorithm repeatedly measures the last $(n - 1)$ qubits and obtains $y^{(1)}, \ldots, y^{(M)}$ for $M = n^{d-1}$ and obtains $(y^{(k)}, p_1(y^{(k)}))$ for all $k = 1, 2, \ldots, M$ using the procedure above, which suffices to learn p_1 completely. Then the algorithm repeats the same procedure by measuring all the qubits except the second qubit in the computational basis and learns the derivative of f in the second direction. This is repeated over all the n qubits. Through this procedure, a learning algorithm learns the partial derivatives of f in the n directions and a simple argument shows that this is sufficient to learn f completely. This gives an overall sample complexity of $O(n^d)$. The procedure above only uses single qubit measurements in the $\{X, Z\}$ basis.

Separable measurements, lower bound. Given the algorithm for learning binary phase states using separable measurements, a natural question is: Is the upper bound on sample complexity we presented above tight? Furthermore, suppose the learning algorithm was allowed to make arbitrary n-qubit measurements on a single copy of $|\psi_f\rangle$, instead of *single qubit* measurements (which are weaker than single *copy* measurements), then could we potentially learn f using fewer than $O(n^d)$ copies?

Here we show that if we allowed *arbitrary* single copy measurements, then a learning algorithm needs $\Omega(n^d)$ many copies of $|\psi_f\rangle$ to learn f. In order to prove this lower bound, our main technical idea is the following. Let f be a degree-d polynomial with n variables sampled uniformly at random. Suppose a learning algorithm measures the phase state $|\psi_f\rangle$ in an arbitrary orthonormal basis $\{U|x\rangle\}_x$. We show that the distribution describing the measurement outcome x is "fairly" uniform. In particular,

$$\mathop{\mathbb{E}}_f[H(x|f)] \geq n - O(1), \tag{5}$$

where $H(x|f)$ is the Shannon entropy of a distribution $P(x|f) = |\langle x|U^*|\psi_f\rangle|^2$. 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 ψ_f in order to learn f. To prove Eq. (5), we first lower bound the Shannon entropy by Renyi-two entropy and bound the latter by deriving an explicit formula for $\mathbb{E}_f[|\psi_f\rangle\langle\psi_f|^{\otimes 2}]$.

Entangled measurements. After settling the sample complexity of learning binary phase states using separable measurements, one final question question remains: Do entangled measurements help in reducing the sample complexity? For the case of quadratic polynomials, we know that Bell measurements (which are entangled measurements) can be used to learn these states in sample complexity $O(n)$. However, as mentioned earlier, it is unclear how to extend the Bell measurement procedure for learning larger degree polynomials.

Here, we give a learning algorithm based on the so-called pretty-good measurements (PGM) that learns $|\psi_f\rangle$ for a degree-d polynomial f using $O(n^{d-1})$ copies of $|\psi_f\rangle$. In order to prove this bound, we follow the following three step approach: (a) we first observe that in order to learn degree-d binary phase states, the *optimal* measurement is the pretty good measurement since the ensemble $\mathcal{S} = \{|\psi_f\rangle\}_f$ is geometrically uniform. By geometrically uniform, we mean that $\mathcal{S}$ can be written as $\mathcal{S} = \{U_f|\phi\rangle\}_f$ where $\{U_f\}_f$ is an Abelian group. (b) We next observe a property about the geometrically uniform state identification problem (which is new as far as we are aware): suppose $\mathcal{S}$ is a geometrically uniform ensemble, then the success probability of the PGM in correctly identifying f, given copies of $|\psi_f\rangle$, is *independent* of f, i.e., every element of the ensemble has the same probability of being identified correctly when measured using the PGM. (c) Finally, we need one powerful tool regarding the the weight distribution of Boolean polynomials: it was shown in [1] that for any degree-d polynomial f, the following relation on $\mathsf{wt}(f)$ or the fraction of strings in $\{0,1\}^n$ for which f is one holds:

$$|\{f \in \mathcal{P}(n,d) : \mathsf{wt}(f) \leq (1-\varepsilon)2^{-\ell}\}| \leq (1/\varepsilon)^{C\ell^4 \cdot \binom{n-\ell}{\leq d-\ell}},$$

for every $\varepsilon \in (0, 1/2)$ and $\ell \in \{1, \ldots, d-1\}$. Using this statement, we can comment on the average inner product of $|\langle\psi_f|\psi_g\rangle|$ over all ensemble members with $f \neq g \in \mathcal{P}(n,d)$. Combining this with a well-known result of PGMs, we are able to show that, given $M = O(n^{d-1})$ copies of $|\psi_f\rangle$ for $f \in \mathcal{S}$, the PGM identifies f with probability ≥ 0.99. Combining observations (a) and (b), the PGM also has the same probability of acceptance given an arbitrary $f \in \mathcal{S}$. Hence, we get an overall upper bound of $O(n^{d-1})$ for sample complexity of learning binary phase states using entangled measurements.

The lower bound for entangled measurement setting is straightforward: each quantum sample $\frac{1}{\sqrt{2^n}}\sum_{x\in\{0,1\}^n}(-1)^{f(x)}|x\rangle$ provides n bits of information and the goal is to learn f which contains $O(n^d)$ bits of information, hence by Holevo's bound, we need at least n^{d-1} quantum samples in order to learn f with high probability.

1.2.2 Generalized phase states

As far as we are aware, ours is the first work that considers the learnability of generalized phase states (using either entangled or separable measurements). The sample complexity upper bounds follow the same high-level idea as that in the binary phase state setting. However, we need a few more technical tools for this setting which we discuss below.

Separable bounds. At a high-level, the learning procedure for generalized phase states is similar to the procedure for learning binary phase states with the exception of a couple of subtleties that we need to handle here. Suppose we perform the same procedure as in binary phase states by measuring the last $(n-1)$ qubits in the computational basis. We then obtain a uniformly random $y \in \mathbb{F}_2^{n-1}$, and the post-measurement state for a generalized phase state is given by

$$|\psi_{f,y}\rangle = \frac{1}{\sqrt{2}}(\omega_q^{f(0y)}|0\rangle + \omega_q^{f(1y)}|1\rangle).$$

This state is proportional to $(|0\rangle + \omega_q^c|1\rangle)/\sqrt{2}$, where $c = f(1y) - f(0y) \pmod q$. In the binary case, $q = 2$, the states associated with $c = 0$ and $c = 1$ are orthogonal, so that the value of c can be learned with certainty by measuring $|\psi_{f,y}\rangle$ in the Pauli X basis. However, in the generalized case, $q > 2$, the states $(|0\rangle + \omega_q^c|1\rangle)/\sqrt{2}$ with $c \in \mathbb{Z}_q$ are not pairwise orthogonal. It is then unclear how to learn c given a single copy of $|\psi_{f,y}\rangle$. However, we observe that 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 $|\phi_b\rangle = \frac{1}{\sqrt{2}}(|0\rangle - \omega_q^b|1\rangle)$. Applying this POVM $\mathcal{M}$ onto an unknown state $(|0\rangle + \omega_q^c|1\rangle)/\sqrt{2}$ we observe that c is the outcome with probability 0 and furthermore *every* other outcome $b \neq c$ appears with non-negligible probability $\Omega(q^{-3})$.

Hence with one copy of $\frac{1}{\sqrt{2^n}}\sum_{x\in\{0,1\}^n} \omega_q^{f(x)}|x\rangle$, we obtain uniformly random $y \in \{0,1\}^{n-1}$ and $b \in \mathbb{Z}_q$ such that $f(1y) - f(0y) \neq b$. We now repeat this process $m = O(n^{d-1})$ many times and obtain $(y^{(k)}, b^{(k)})$ for $k = 1, 2, \ldots, M$ such that $f(1y^{(k)}) - f(0y^{(k)}) \neq b^{(k)}$ for all $k \in [M]$. We next show a variant of the Schwartz-Zippel lemma in the following sense: that for every $f \in \mathcal{P}_q(n,d)$ and $c \in \mathbb{Z}_q$, then either f is a constant function or the fraction of $x \in \mathbb{F}_2^n$ for which $f(x) \neq c$ is at least 2^{-d}. Using this, we show that after obtaining $O(2^d n^{d-1})$ samples, we can find a polynomial $g \in \mathcal{P}_q(n-1, d-1)$ for which $f(1y) - f(0y) = g(y)$. We now repeat this protocol for n different directions (by measuring each of the n qubits in every iteration) and we learn all the n directional derivatives of f, which suffices to learn f completely.

Entangled bounds. We do not give a result on learning generalized phase states with entangled measurements. We expect the proof of the sample complexity upper bound for learning generalized phase states using entangled measurements should proceed similarly to our earlier analysis of learning binary phase states using entangled measurements. However, we need a new technical tool that generalizes the earlier work on the weight distribution [2] of Boolean functions $f : \mathbb{F}_2^n \to \mathbb{F}_2$ to those of form $f : \mathbb{F}_2^n \to \mathbb{Z}_q$ with $q = 2^d$.

1.2.3 Learning with further constraints

Learning sparse and low-Fourier degree states. A natural constraint to put on top of having low $\mathbb{F}_2$-degree in the polynomial is the sparsity, i.e., number of monomials in the $\mathbb{F}_2$ decomposition of f. Sparse low-degree phase states appear naturally when learning circuits with few gates. In particular, suppose we are learning a quantum circuit U with s gates from $\{\mathsf{Z}, \mathsf{CZ}, \ldots, \mathsf{C}^{d-1}\mathsf{Z}\}$ (where $\mathsf{C}^m\mathsf{Z}$ is the controlled-Z gate with m controls), then the output of $U|+\rangle^{\otimes n}$ is a phase state with sparsity-s and degree-d.

One naive approach to learn sparse $\mathbb{F}_2$ polynomials is to directly apply our earlier learning algorithm for binary phase states but this ignores the $\mathbb{F}_2$-sparsity information, and doesn't improve the sample complexity. Instead, here we use ideas from compressed sensing [20] to propose a linear program that allows us to improve the sample complexity to $O(2^d sn)$. Finally we make an observation that, if the function has *Fourier*-degree d, then one can learn f, given only $O(2^d \log n)$ many copies of $|\psi_f\rangle$, basically using the fact that there are only 2^{2^d} many such functions, each having at least a 2^{-d} distance between them.

Learning with depolarizing noise. One motivation for learning stabilizer states was potential experimental demonstrations of the learning algorithm [41]. Here, we consider a theoretical framework in order to understand the sample complexity of learning degree-2 phase states under global and local depolarizing noise. In this direction, we present two results. Under global depolarizing noise, i.e., when we are given $\rho_f = (1-\varepsilon)|\psi_f\rangle\langle\psi_f| + \varepsilon \cdot \mathbb{I}$, then it suffices to take $O(n^{1+\varepsilon})$ many copies ρ_f in order to learn f. The crucial observation is that one can use Bell sampling to reduce learning ρ_f to learning parities with noise, which we can accomplish using $O(n^{1+\varepsilon})$ samples and in time $2^{n/(\log\log n)}$ [32]. Additionally, however, a simple argument reveals that under local depolarizing noise, the sample complexity of learning stabilizer states is exponential in n.

1.3 Organization

In Section 2, we introduce phase states, discuss separable and entangled measurements. In Section 3, we prove our upper and lower bounds for learning binary phase states with separable and entangled measurements. We omit our results on learning sparse and low-Fourier-degree phase states, and binary phase states under depolarizing noise from this version of the paper (see [6]). In Section 4, we prove our upper bound for learning generalized phase states using separable and entangled measurements. Our algorithms for learning quantum phase states can be used to learn the corresponding circuits that produce them. We explicitly discuss the connection between phase states, and the diagonal unitaries in the d-th level of the Clifford hierarchy and IQP circuits in [6].

2 Preliminaries

2.1 Notation

Let $[n] = \{1, \ldots, n\}$. Let e_i be an n-dimensional vector with 1 in the ith coordinate and 0s elsewhere. We denote the finite field with the elements $\{0,1\}$ as $\mathbb{F}_2$ and the ring of integers modulo q as $\mathbb{Z}_q = \{0, 1, \ldots, q-1\}$ with q usually being a power of 2 in this work. For a Boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$, the bias of f is defined as

$$\text{bias}(f) = \mathop{\mathbb{E}}_{x}[(-1)^{f(x)}],$$

where the expectation is over a uniformly random $x \in \{0,1\}^n$. For $g : \mathbb{F}_2^n \to \mathbb{Z}_{2^d}$, the bias of g in the coordinate $j \in \mathbb{F}_{2^d}^\star$ is defined as $\text{bias}_j(g) = \mathbb{E}_x[(\omega_{2^d})^{j \cdot g(x)}]$. For a function $f : \mathbb{F}_2^n \to \mathbb{F}_2$, $y \in \mathbb{F}_2^{n-1}$ and $k \in [n]$, we denote $(D_k f)(y) = f(y^{k=1}) + f(y^{k=0})$, where $y^{i=1}, y^{i=0} \in \mathbb{F}_2^n$ is defined as: the ith bit of $y^{i=1}$ equals 1 and $y^{i=0}$ equals 0 and otherwise equals y.

2.2 Boolean Functions

A Boolean function $f : \mathbb{F}_2^n \to \mathbb{F}_2$ can be uniquely represented by a polynomial over $\mathbb{F}_2$ as follows (which we call its $\mathbb{F}_2$ representation):

$$f(x) = \sum_{J \subseteq [n]} \alpha_J \prod_{i \in J} x_i \pmod 2, \tag{6}$$

where $\alpha_J \in \{0, 1\}$. Similar to Eq. (6), we can write Boolean functions $f : \mathbb{F}_2^n \to \mathbb{Z}_q$ as

$$f(x) = \sum_{J \subseteq [n]} \alpha_J \prod_{i \in J} x_i \pmod q \tag{7}$$

for some integer coefficients $\alpha_J \in \{0, 1, \ldots, q - 1\}$. Throughout this paper, unless explicitly mentioned, we will be concerned with writing Boolean functions as a decomposition over $\mathbb{F}_2$ or $\mathbb{Z}_q$ with $q = 2^d$. The $\mathbb{F}_2$ degree of f is defined as

$$\deg(f) = \max\{|J|: \alpha_J \neq 0\}.$$

Similarly for polynomials over $\mathbb{Z}_{2^d}$, we can define the degree as the size of the largest monomial whose coefficient α_J is non-negative.

We will call $g : \mathbb{F}_2^n \to \mathbb{F}_2$ with $g = \prod_{i \in J} x_i$ as monic monomials over n variables of at most degree-d, characterized by set $J \subseteq [n]$, $|J| \leq d$. We will denote the set of these monic monomials by $\mathcal{M}(n, d)$. Note that $|\mathcal{M}(n, d)| = \sum_{j=0}^{d} \binom{n}{j} = O(n^d)$. We will denote the set of polynomials over n variables of $\mathbb{F}_2$-degree d as $\mathcal{P}(n, d)$. Note that these polynomials are just linear combinations of monomials in $\mathcal{M}(n, d)$. We will denote the set of polynomials over n variables of $\mathbb{F}_2$-degree d with sparsity s as $\mathcal{P}(n, d, s)$. Similarly, we will denote $\mathcal{P}_q(n, d)$ as the set of all degree-d Boolean polynomials $f : \mathbb{F}_2^n \to \mathbb{Z}_q$ with n variables. In particular, one can specify any polynomial $f \in \mathcal{P}_q(n, d)$ by $O(dn^d)$ bits and $|\mathcal{P}_q(n, d)| \leq 2^{O(dn^d)}$.

Consider a fixed d, and any $x \in \mathbb{F}_2^n$. Let the d-evaluation of x, denoted by $\text{eval}_d(x)$, be a column vector in $\mathbb{F}_2^{|\mathcal{M}(n,d)|}$ with its elements being the evaluations of x under different monomials $g \in \mathcal{M}(n, d)$. This can be expressed as follows:

$$\text{eval}_d(x) = \left(\prod_{i \in J \subseteq [n], |J| \leq d} x_i \right)^{\top} \tag{8}$$

For a set of points $\mathbf{x} = (x^{(1)}, x^{(2)}, \ldots, x^{(m)}) \in (\mathbb{F}_2^n)^m$, we will call the matrix in $\mathbb{F}_2^{|\mathcal{M}(n,d)| \times m}$ with its kth column corresponding to d-evaluations of $x^{(k)}$, as the d-evaluation matrix of $\mathbf{x}$, and denote it by $Q_{\mathbf{x}}$.

2.3 Useful Lemmas

Let $e_i \in \mathbb{F}_2^n$ denote the vector of all zeros except for a 1 in the i^{th} coordinate.

▶ **Fact 1.** *Let $d \in [n]$, $s \leq |\mathcal{M}(n, d)| = \sum_{k=1}^{d} \binom{n}{k}$, and $f \in \mathcal{P}(n, d, s)$. There exists $g_i \in \mathcal{P}(n, d - 1, s)$ such that $g_i(x) = f(x + e_i) + f(x) \pmod 2$ for all $x \in \{0, 1\}^n$.*

The proof of this fact is straightforward. Without loss of generality, consider $i = 1$. For every $f(x) = \sum_S \alpha_S \prod_{i \in S} x_i$, we can express it as

$$f(x) = x_1 p_1(x_2, \ldots, x_n) + p_2(x_2, \ldots, x_n),$$

where p_1 has degree $\leq d-1$ and p_2 has degree $\leq d$. Observe that $f(x+e_1)-f(x)$ is either $p_1(x_2,\ldots,x_n)$ or $-p_1(x_2,\ldots,x_n)$ which has degree $d-1$ and corresponds to the polynomial g_1 in the fact statements. This applies for every coordinate i.

Note that the polynomial g_i above is also often called the *directional* derivative of f in direction w and is denoted as $D_i f$.

▶ **Fact 2.** *Let $N, s \geq 1$ such that $\gamma = s/N \leq 1/2$. Then we have*

$$\sum_{\ell=1}^{s}\binom{N}{\ell} \leq 2^{H_b(\gamma)N} \leq 2^{2\gamma\log(1/\gamma)}.$$

where we used above that $H_b(\gamma) = \gamma\log\frac{1}{\gamma} + (1-\gamma)\log\frac{1}{1-\gamma} \leq 2\gamma\log\frac{1}{\gamma}$ (for $\gamma \leq 1/2$).

▶ **Lemma 1** (The Schwartz-Zippel Lemma). *Let $p(y_1,\ldots,y_n)$ be a nonzero polynomial on n variables with degree d. Let S be a finite subset of $\mathbb{R}$, with at least d elements in it. If we assign $y_1,\ldots,y_n$ values from S independently and uniformly at random, then*

$$\Pr[p(y_1,\ldots,y_n)=0] \leq \frac{d}{|S|}. \tag{9}$$

▶ **Lemma 2** ([38]). *Let $p(x_1,\ldots,x_n)$ be a non-zero multilinear polynomial of degree d. Then*

$$\Pr_{x\in\{0,1\}^n}[p(x)=0] \leq 1-2^{-d},$$

where the probability is over a uniformly random distribution on $\{0,1\}^n$.

We will also need the following structural theorem about Reed-Muller codes which comments on the weight distribution of Boolean functions $f : \mathbb{F}_2^n \to \mathbb{F}_2$.

▶ **Theorem 3** ([2, Theorem 3]). *Let $n \geq 1$ and $d \leq n/2$. Define $|f|= \sum_{x\in\{0,1\}^n}[f(x)=1]$ and $\mathsf{wt}(f) = |f|/2^n$. Then, for every $\varepsilon \in (0,1/2)$ and $\ell \in \{1,\ldots,d-1\}$, we have that*

$$|\{f \in P(n,d) : \mathsf{wt}(f) \leq (1-\varepsilon)2^{-\ell}\}|\leq (1/\varepsilon)^{C\ell^4\cdot\binom{n-\ell}{\leq d-\ell}}.$$

Fix $w = (1-\varepsilon)2^{n-\ell}$ and we get

$$|\{f \in P(n,d) : |f|\leq w\}|\leq (1-w/2^{n-\ell})^{-C\ell^4\cdot\binom{n-\ell}{\leq d-\ell}}.$$

▶ **Lemma 4** (Fano's inequality). *Let A and B be classical random variables taking values in $\mathcal{X}$ (with $|\mathcal{X}|= r$) and let $q = \Pr[\mathsf{A} \neq \mathsf{B}]$. Then,*

$$H(\mathsf{A}|\mathsf{B}) \leq H_b(q) + q\log(r-1),$$

where $H(\mathsf{A}|\mathsf{B})$ is the conditional entropy and $H_b(q)$ is the standard binary entropy.

2.4 Measurements

Throughout this paper we will be concerned with learning algorithms that use either separable or entangled measurements. Given $|\psi_f\rangle^{\otimes k}$, a learning algorithm for f is said to use separable measurements if it only measure each copy of $|\psi_f\rangle$ separately in order to learn f. Similarly, a learning algorithm for f is said to use entangled measurements if it makes an entangled measurement on the k-fold tensor product $|\psi_f\rangle^{\otimes k}$. In this direction, we will often use two techniques which we discuss in more detail below: sampling random partial derivatives in order to learn from separable measurements and Pretty Good Measurements in order to learn from entangled measurements.

2.4.1 Separable Measurements

Below we discuss a subroutine that we will use often to learn properties about $f : \mathbb{F}_2^n \to \mathbb{F}_2$: given a single copy of $|\psi_f\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} (-1)^{f(x)}|x\rangle$, the subroutine produces a uniformly random $y \in \mathbb{F}_2^{n-1}$ and $f(1y) + f(0y) \pmod 2$. To this end, suppose we measure qubits $2, 3, \ldots, n$ of $|\psi_f\rangle$ in the usual Z basis. We denote the resulting string as $y \in \{0,1\}^{n-1}$. The post-measurement state of qubit 1 is then given by

$$|\psi_{f,y}\rangle = \frac{1}{\sqrt{2}} \Big[(-1)^{f(0y)}|0\rangle + (-1)^{f(1y)}|1\rangle\Big]. \tag{10}$$

We note that $|\psi_{f,y}\rangle$ is then an X-basis state ($|+\rangle$ or $|-\rangle$) depending on the values of $f(1y)$ and $f(0y)$. If $f(1y) = f(0y)$, then $|\psi_{f,y}\rangle = |+\rangle$ and if $f(1y) = f(0y) + 1 \pmod 2$, then $|\psi_{f,y}\rangle = |-\rangle$. Measuring qubit 1 in the X-basis and qubits $2, 3, \ldots, n$ in the Z-basis thus produces examples of the form (y, b) where $y \in \{0,1\}^{n-1}$ is uniformly random and $b = f(0y) + f(1y) \pmod 2$. Considering Fact 1 with the basis of e_1, we note that theses examples are of the form $(y, D_1 f(y))$, where $D_1 f(y) = f(1y) + f(0y) \pmod 2$ is the partial derivative of f along direction e_1. Changing the measurement basis chosen above to $ZZ \cdots X_k \cdots Z$ such that we measure all the qubits in the Z basis except for the kth qubit which is measured in the X basis, will allow us to obtain random samples of the form $(y, D_k f(y))$. Accordingly, we introduce a new subroutine.

▶ **Definition 1** (Random Partial Derivative Sampling (RPDS) along e_k). *For every $k \in [n]$, measuring every qubit of $|\psi_f\rangle$ in the Z basis, except the kth qubit which is measured in the X basis, we obtain a uniformly random $y \in \mathbb{F}_2^{n-1}$ and $(D_k f)(y)$.*

2.4.2 Entangled Measurements

In general one could also consider a joint measurement applied to multiple copies of $|\psi_f\rangle$, which we refer to as entangled measurements. In this work, we consider two types of entangled measurements, Bell sampling and the pretty-good measurement. We omit a detailed discussion on Bell sampling as we do not include the corresponding results for learning binary phase states under depolarizing noise (see the full version [6] for more).

Pretty Good Measurements. Consider an ensemble of states, $\mathcal{E} = \{(p_i, |\psi_i\rangle)\}_{i\in[m]}$, where $p = \{p_1, \ldots, p_m\}$ is a probability distribution. In the quantum state identification problem, a learning algorithm is given an unknown quantum state $|\psi_i\rangle \in \mathcal{E}$ sampled according to the distribution p and the learning algorithm needs to identity i with probability $\geq 2/3$. In this direction, we are interested in maximizing the average probability of success to identify i. For a POVM specified by positive semidefinite matrices $\mathcal{M} = \{M_i\}_{i\in[m]}$, the probability of obtaining outcome j equals $\langle\psi_i|M_j|\psi_i\rangle$ and the average success probability is given by

$$P_{\mathcal{M}}(\mathcal{E}) = \sum_{i=1}^{m} p_i \langle\psi_i|M_i|\psi_i\rangle.$$

Let $P^{opt}(\mathcal{E}) = \max_{\mathcal{M}} P_{\mathcal{M}}(\mathcal{E})$ denote the optimal average success probability of $\mathcal{E}$, where the maximization is over the set of valid m-outcome POVMs. For every ensemble $\mathcal{E}$, the so-called *Pretty Good Measurement* (PGM) is a specific POVM (depending on the ensemble $\mathcal{E}$) that does *reasonably* well against $\mathcal{E}$. In particular, it is well-known that

$$P^{opt}(\mathcal{E})^2 \leq P^{PGM}(\mathcal{E}) \leq P^{opt}(\mathcal{E}).$$

We now define the POVM elements of the pretty-good measurement. Let $|\psi_i'\rangle = \sqrt{p_i}|\psi_i\rangle$, and $\mathcal{E}' = \{|\psi_i'\rangle : i \in [m]\}$ be the set of states in $\mathcal{E}$, renormalized to reflect their probabilities. Define $\rho = \sum_{i\in[m]}|\psi_i'\rangle\langle\psi_i'|$. The PGM is defined as the set of measurement operators $\{|\nu_i\rangle\langle\nu_i|\}_{i\in[m]}$ where $|\nu_i\rangle = \rho^{-1/2}|\psi_i'\rangle$ (the inverse square root of ρ is taken over its non-zero eigenvalues). We will use the properties of these POVM elements later on and will also need the following theorems about PGMs.

▶ **Theorem 5** ([25]). *Let $\mathcal{S} = \{\rho_1, \dots, \rho_m\}$. Suppose $\rho \in \mathcal{S}$ is an unknown quantum state picked from $\mathcal{S}$. Let $\max_{i\neq j}\|\sqrt{\rho_i}\sqrt{\rho_j}\|_1 \leq F$. Then, given*

$$M = O((\log(m/\delta))/\log(1/F))$$

copies of ρ, the Pretty good measurement identifies ρ with probability at least $1-\delta$.

The above theorem in fact implies the following stronger statement immediately (also stated in [8]) that we use here.

▶ **Lemma 6.** *Let $\mathcal{S} = \{\rho_1, \dots, \rho_m\}$. Suppose $\rho \in \mathcal{S}$ is an unknown quantum state picked uniformly from $\mathcal{S}$. Suppose there exists k such that*

$$\frac{1}{m}\sum_{i\neq j}\left\|\sqrt{\rho_i^{\otimes k}}\sqrt{\rho_j^{\otimes k}}\right\|_1 \leq \delta,$$

then given k copies of ρ, the Pretty Good Measurement identifies ρ with probability at least $1-\delta$.

3 Learning Binary Phase States

In this section, we consider the problem of learning binary phase states as given by Eq. (1), assuming that f is a Boolean polynomial of $\mathbb{F}_2$-degree d.

3.1 Learning algorithm using separable measurements

We now describe our learning algorithm for learning binary phase states $|\psi_f\rangle$ when f has $\mathbb{F}_2$-degree d, using separable measurements. We carry out our algorithm in n rounds, which we index by t. In the t-th round, we perform RPDS along e_t (Def. 1) in order to obtain samples of the form $(y, D_t f(y))$ where $y \in \{0,1\}^{n-1}$. For an $m \geq 1$ to be fixed later, we use RPDS on m copies of $|\psi_f\rangle$ to obtain $\{\big(y^{(k)}, D_t f(y^{(k)})\big)\}_{k\in[m]}$ where $y^{(k)} \in \{0,1\}^{n-1}$ is uniformly random. We now describe how to learn $D_t f$ using these m samples.

Using Fact 1, we know that $D_t f \in \mathcal{P}(n-1, d-1)$. Thus, there are at most $N = |\mathcal{M}(n-1, d-1)| = \sum_{k=1}^{d-1}\binom{n}{k} = n^{O(d)}$ monomials in the $\mathbb{F}_2$ representation of $D_t f$. Let $A_t \in \mathbb{F}_2^{m\times N}$ be the transpose of the $(d-1)$-evaluation matrix (defined in Eq. (8)), such that the kth row of A_t corresponds to the evaluations of $y^{(k)}$ under all monomials in $\mathcal{M}(n-1, d-1)$, i.e., $(y_S^{(k)})_{|S|\leq d-1}$, where $y_S^{(k)} = \prod_{j\in S} y_j^{(k)}$, and let $\beta_t = (\alpha_S)_{|S|\leq d-1}$ be the vector of unknown coefficients. Obtaining $\{(y^{(k)}, D_t f(y^{(k)}))\}_{k\in[m]}$, allows one to solve $A_t\beta_t = D_t f(\mathbf{y})$ for β_t (where $\mathbf{y} = (y^{(1)}, \dots, y^{(m)})$ and $(D_t f(\mathbf{y}))_k = D_t f(y^{(k)})$) and learn the $\mathbb{F}_2$-representation of $D_t f$ completely. Over n rounds, one then learns $D_1 f, D_2 f, \dots, D_n f$. The $\mathbb{F}_2$-representations of these partial derivatives can then be used to learn f completely, as show in Fact 3. This procedure is shown in Algorithm 1.

▶ **Fact 3.** *Let $f : \mathbb{F}_2^n \to \mathbb{F}_2$ be such that $f \in \mathcal{P}(n, d)$. Learning $D_1 f, \dots, D_n f$ suffices to learn f.*

Proof. Let the $\mathbb{F}_2$-representation of the unknown f be

$$f(x) = \sum_{J\subseteq[n],|J|\leq d} \alpha_J \prod_{i\in J} x_i. \tag{11}$$

The $\mathbb{F}_2$-representation of $D_t f$ for any $t \in \{1, 2, \ldots, n\}$ is then given by

$$D_t f(x) = \sum_{\substack{J\subseteq[n]:\\ t\in J,|J|\leq d}} \alpha_J \prod_{i\in J\setminus t} x_i, \tag{12}$$

where we notice that $D_t f$ only contains those monomials that correspond to sets J containing the component x_t. Let the $\mathbb{F}_2$-representation of $D_t f$ with the coefficient vector β_t be given by

$$D_t f(x) = \sum_{S\subset[n],|S|\leq d-1} (\beta_t)_S \prod_{i\in S} x_i. \tag{13}$$

Suppose an algorithm learns $D_1 f, \ldots, D_n f$. In order to learn f, we must retrieve the coefficients α_J from the learned coefficients $\{\beta_t\}_{t\in\{1,2,\ldots,n\}}$. We accomplish this by noting that $(\beta_t)_S = \alpha_{S\cup t}$ or in other words, $\alpha_J = \{\beta_t\}_{J\setminus t}$, $t \in J$. However, there may be multiple values of t that will allow us retrieve α_J. For example, suppose f contains the monomial term $x_1x_2x_3$ (i.e., $J = \{1, 2, 3\}$) then $\alpha_{\{1,2,3\}}$ could be retrieved from $(\beta_1)_{\{2,3\}}$, $(\beta_2)_{\{1,3\}}$, or $(\beta_3)_{\{1,2\}}$. When $D_t f$ (or β_t) for all t is learned with zero error, all these values coincide and it doesn't matter which learned coefficient is used. When there may be error in learning $D_t f$ (or β_t), we can carry out a majority vote: $\alpha_J = \text{Majority}(\{(\beta_t)_{J\setminus t} | t \in J\})$ for all $J \subseteq [n], |J| \leq d$. The majority vote is guaranteed to succeed as long as there is no error in at least half of the contributing β_t (which is the case in our learning algorithm). ◀

Algorithm 1 Learning binary phase states through separable measurements.

```
Input: Given $M = O((2n)^d)$ copies of $|\psi_f\rangle$ where $f \in \mathcal{P}(n, d)$
for qubit $t = 1, \ldots, n$ do
    Set $m = M/n$
    Perform RPDS along $e_t$ to obtain $\{(y^{(k)}, D_t f(y^{(k)})\}_{k\in[m]}$ by measuring $m$ copies of $|\psi_f\rangle$.
    Solve the linear system of equations $A_t \cdot \beta_t = D_t f(\mathbf{y})$ to learn $D_t f$ explicitly.
end for
Use Fact 3 to learn $f$ using $D_1 f, \ldots, D_n f$ (let $\tilde{f}$ be the output).
Output: Output $\tilde{f}$
```

We now prove the correctness of this algorithm.

▶ **Theorem 7.** *Let $n \geq 2, d \leq n/2$. Algorithm 1 uses $M = O(2^d n^d)$ copies of an unknown $|\psi_f\rangle$ for $f \in \mathcal{P}(n, d)$ and with high probability identifies f using single qubit X, Z measurements.*

Proof. Algorithm 1 learns f by learning $D_1 f, \ldots, D_n f$ and thereby learns f completely. Here we prove that each $D_t f$ can be learned with $m = O(2^d n^{d-1})$ copies of $|\psi_f\rangle$ and an exponentially small probability of error. This results in an overall sample complexity of $O(2^d n^d)$ for learning f and hence $|\psi_f\rangle$. Let us consider round t in Algorithm 1. We generate m constraints $\{(y^k, (D_t f)(y^{(k)})\}_{k\in[m]}$ where $y^{(k)} \in \mathbb{F}_2^{n-1}$ by carrying out RPDS along e_t on m copies of $|\psi_f\rangle$.

We learn the $\mathbb{F}_2$-representation of $D_t f$ by setting up a linear system of equations using these m samples: $A_t\beta_t = D_t f(\mathbf{y})$, where A_t is the transposed $(d-1)$-evaluation matrix in round t, evaluated over $\mathbf{y} = (y^{(1)}, y^{(2)}, \dots y^{(m)})$, and $\beta_t \in \mathbb{F}_2^{|\mathcal{M}(n-1,d-1)|}$ is the collective vector of coefficients corresponding to the monomials in $\mathcal{M}(n-1, d-1)$. By construction, this system has at least one solution. If there is exactly one solution, then we are done. Otherwise, the corresponding system has a non-zero solution, that is, there exists a non-zero degree-$(d-1)$ polynomial $g : \mathbb{F}_2^{n-1} \to \mathbb{F}_2$ such that $g(y^{(j)}) = 0$ for all $j = 1, 2, \dots, m$.

Below we prove that the probability of this bad event can be bounded through the Schwartz-Zippel lemma. Applying Lemma 2 and by noting that $y^j \in \mathbb{F}_2^{(n-1)}$ are independent and uniformly distributed, we have that

$$\Pr[g(y^{(1)}) = g(y^{(2)}) = \cdots = g(y^{(m)}) = 0] \leq (1-2^{-d})^m \leq e^{-m2^{-d}} \tag{14}$$

Let $\mathcal{P}_{\mathrm{nnz}}(n, d)$ be the set of all degree-d polynomials $g : \mathbb{F}_2^n \to \mathbb{F}_2$ which are not identically zero. Define event

$$\mathsf{BAD}(y^1, \dots, y^m) = [\exists g \in \mathcal{P}_{\mathrm{nnz}}(n-1, d-1) \,:\, g(y^1) = \ldots = g(y^m) = 0 \pmod 2]. \tag{15}$$

We note that $|\mathcal{P}_{\mathrm{nnz}}(n-1, d-1)| \leq 2^N$ where $N = O(n^{d-1})$. By union bound and Eq. (14), we have

$$\Pr[\mathsf{BAD}(y^{(1)}, \dots, y^{(m)})] \leq |\mathcal{P}_{\mathrm{nnz}}(n-1, d-1)| \cdot (1-2^{-d})^m \leq 2^{n^{d-1} - m2^{-d}(\ln 2)}. \tag{16}$$

Thus choosing $m = O((2n)^{d-1})$ is enough to learn all coefficients $\{\alpha_J\}_{t\in J}$ (through β_t) in the $\mathbb{F}_2$ representation of f with an exponentially small probability of error. We need to repeat this over all the n qubits in order to learn $D_1 f, \dots, D_n f$ and then use Fact 3 to learn f completely. This gives an overall sample complexity of $O((2n)^d)$ for learning binary phase states. Observe that the only measurements that we needed in this algorithm were single qubit $\{X, Z\}$ measurements. ◀

▶ **Corollary 8.** *An n-qubit state $|\psi_f\rangle$ with the unknown Boolean function f of given Fourier-sparsity s can be learned with Algorithm 1 that consumes M copies of $|\psi_f\rangle$ with probability $1 - 2^{-\Omega(n)}$ provided that $M \geq O(sn^{\log s})$.*

The proof of this corollary simply follows from the following: for a Boolean function, the Fourier sparsity s of f is related to the $\mathbb{F}_2$-degree d of f [9] as $d \leq \log s$. Along with Theorem 7 we obtain the corollary.

3.2 Learning using entangled measurements

We now consider the problem of learning binary phase states using entangled measurements. We have the following result.

▶ **Theorem 9.** *Let $n \geq 2, d \leq n/2$. There exists an algorithm that uses $M = O((2n)^{d-1})$ copies of an unknown $|\psi_f\rangle$ for $f \in \mathcal{P}(n, d)$ and identifies f using entangled measurements with probability $\geq 2/3$. There is also a lower bound of $\Omega(n^{d-1})$ for learning these states.*

Proof. In order to prove this theorem, we follow the following steps. We first observe that the optimal measurement for our state distinguishing problem is the pretty good measurement (PGM). Second we observe that the success probability of the PGM is the same for *every* concept in the ensemble. We bound the success probability of the PGM using Corollary 6 we get our upper bound.

For $f \in \mathcal{P}(n,d)$, let U_f be the unitary defined as $U_f = \mathsf{diag}(\{(-1)^{f(x)}\}_x)$, that satisfies $U_f|+\rangle^n = |\psi_f\rangle$. Observe that the set $\{U_f\}_{f \in \mathcal{P}(n,d)}$ is an Abelian group. The ensemble we are interested in is $\mathcal{S} = \{U_f|+\rangle^n\}_{f \in \mathcal{P}(n,d)}$ and such an ensemble is called *geometrically uniform* if the $\{U_f\}$ is an Abelian group. A well-known result of Eldar and Forney [21] showed that the optimal measurement for state distinguishing a geometrically uniform (in particular $\mathcal{S}$) is the pretty-good measurement. We now show that the success probability of the PGM is the same for every state in the ensemble. In this direction, for $M \geq 1$, let $\sigma_f = |\psi_f\rangle\langle\psi_f|^{\otimes M}$. The POVM elements of the pretty good measurement $\{E_f : f \in \mathcal{P}(n,d)\}$ is given by the POVM elements $E_f = S^{-1/2}\sigma_f S^{-1/2}$ where $S = \sum_{f \in \mathcal{P}(n,d)} \sigma_f$. The probability that the PGM identifies the unknown σ_f is given by

$$\Pr(f) = \mathrm{Tr}(\sigma_f E_f) = \langle \psi_f^{\otimes M}|S^{-1/2}|\psi_f^{\otimes M}\rangle^2.$$

Our claim is that $\Pr(f)$ is the same for every $f \in \mathcal{P}(n,d)$. Using the Abelian property of the unitaries $\{U_f\}_f$, observe that $U_f|\psi_g\rangle = |\psi_{f\oplus g}\rangle$ for every $f, g \in \mathcal{P}(n,d)$. Thus, we have that $(U_f^{\otimes M})^\dagger S U_f^{\otimes M} = S$, which implies that$(U_f^{\otimes M})^\dagger S^{-1/2} U_f^{\otimes M} = S^{-1/2}$. Hence it follows that

$$\Pr(f) = (\langle +|^{\otimes M}(U_f^{\otimes M})^\dagger S^{-1/2} U_f^{\otimes M}|+\rangle^{\otimes M})^2 = (\langle +|^{\otimes M} S^{-1/2}|+\rangle^{\otimes M})^2 = \Pr(0),$$

for every $f \in \mathcal{P}(n,d)$. Finally, observe that $\langle\psi_f|\psi_g\rangle = \mathbb{E}_x\left[(-1)^{f(x)+g(x)}\right] = 1 - 2\Pr_x[f(x) \neq g(x)]$. Let $\mathcal{P}^*(n,d)$ be the set of non-constant polynomials in $\mathcal{P}(n,d)$. We now have the following

$$\frac{1}{2^{\binom{n}{\leq d}}} \sum_{\substack{f \neq g: \\ f,g \in P(n,d)}} \|\sqrt{\rho_f^{\otimes k}}\sqrt{\rho_g^{\otimes k}}\|_1 = \sum_{g \in P^*(n,d)} (1 - 2\Pr_x[g(x)=1])^{2k} = \sum_{g \in P^*(n,d)} (1 - 2\mathsf{wt}(g))^{2k}$$

which we can further upper bound as follows

$$\begin{aligned}
&\sum_{\ell=1}^{d-1} \sum_{g \in P^*(n,d)} (1 - 2|g|/2^n)^{2k} \cdot \Big[|g| \in [2^{n-\ell-1}, 2^{n-\ell}-1]\Big] \\
&= \sum_{g \in P^*(n,d)} (1 - 2|g|/2^n)^{2k} \cdot \Big[|g| \in [2^{n-2}, 2^{n-1}-1]\Big] \\
&\qquad + \sum_{\ell=2}^{d-1} \sum_{g \in P^*(n,d)} (1 - 2|g|/2^n)^{2k} \cdot \Big[|g| \in [2^{n-\ell-1}, 2^{n-\ell}-1]\Big] \\
&\leq 2^{n-1} 2^{-2k + C\binom{n-1}{\leq d-1}} + \sum_{\ell=2}^{d-1} (1 - \frac{1}{2^\ell})^{2k} \sum_{g \in P^*(n,d)} \Big[|g| \leq 2^{n-\ell}\Big],
\end{aligned}$$

where the first equality used that the PGM has the same success probability for every $f, g \in \mathcal{P}(n,d)$, third equality used that $|g| \geq 2^{n-d}$ for any non-zero polynomial $g \in P(n,d)$ [33] and last inequality used Theorem 3. For $k = O(n^{d-1})$ (by picking a sufficiently large constant in $O(\cdot)$), the first term is at most $\leq 1/100$. To bound the second term, using Theorem 3 we have

$$\sum_{\ell=2}^{d-1} (1 - \frac{1}{2^\ell})^{2k} \sum_{g \in P^*(n,d)} \Big[|g| \leq 2^{n-\ell}\Big] \leq \sum_{\ell=2}^{d-1} 2^{n-\ell} \exp(-2k/2^\ell + (n-\ell)\ell^4 \binom{n-\ell}{\leq d-\ell}).$$

Each term is $\exp(-n^{d-1})$ for $k = O(n^{d-1})$, so the overall sum is $\leq 1/100$. Corollary 6 implies our desired upper bound.

In order to see the lower bound, observe that each state $|\psi_f\rangle$ contains n bits of information and the goal of the learning algorithm is to learn an unknown f, i.e., obtain $O(n^d)$ bits of information. Hence by Holevo's theorem [27], one requires $\Omega(n^{d-1})$ copies of the unknown state for state identification.[2] ◀

3.3 Lower bounds

In the last section we saw that $\Theta(n^{d-1})$ many copies of $|\psi_f\rangle$ with degree-d are necessary and sufficient to learn f if we allowed only entangled measurements. Earlier we saw that $O(n^d)$ many copies of $|\psi_f\rangle$ sufficed to learn f using separable measurements. A natural question is: Can we learn f using fewer copies if we are restricted to using only separable measurements? In the theorem below, we provide a lower bound that complements our upper bound, thereby showing $\Theta(n^d)$ copies are necessary and sufficient to learn f using separable measurements.

▶ **Theorem 10.** *Let $2 \leq d \leq n/2$. Suppose there exists an algorithm that with probability $\geq 1/10$, learns an n-variate polynomial $f \in \mathcal{P}(n,d)$, given M copies of the phase state $|\psi_f\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} (-1)^{f(x)} |x\rangle$, measuring each copy in an arbitrary orthonormal basis, and performing an arbitrary classical processing. Then $M = \Omega(\log|\mathcal{P}(n,d)|) = \Omega(n^d)$.*

Proof. The proof is given in the full version of this paper [6]. ◀

4 Learning generalized phase states

In this section, we consider the problem of learning generalized phase states $|\psi_f\rangle$ as given by Eq. (3), assuming that f is a degree-d $\mathbb{Z}_q$-valued polynomial, $f \in \mathcal{P}_q(n,d)$. Note that since our goal is to learn $|\psi_f\rangle$ up to an overall phase, we shall identify polynomials which differ only by a constant shift.

▶ **Definition 2.** *Polynomials $f, g \in \mathcal{P}_q(n,d)$ are equivalent if $f(x) - g(x)$ is a constant.*

To simplify notation, here and below we omit modulo operations keeping in mind that degree-d polynomials take values in the ring $\mathbb{Z}_q$. Thus all equal or not-equal constraints that involve a polynomial's value are modulo q.

4.1 Learning using separable measurements

Let $q \geq 2$ and $d \geq 1$ be integers. For technical reasons, we shall assume that q is even. Let $\omega_q = e^{2\pi i/q}$. Our main result is as follows.

▶ **Theorem 11.** *Let $d \leq n/2$. There exists an algorithm that uses $M = O(2^d q^3 n^d \log q) = O(n^d)$ copies of a generalized phase state $|\psi_f\rangle = \frac{1}{\sqrt{2^n}} \sum_{x\in\{0,1\}^n} \omega_q^{f(x)} |x\rangle$ with an unknown polynomial $f \in \mathcal{P}_q(n,d)$ and outputs a polynomial $g \in \mathcal{P}_q(n,d)$ such that g is equivalent to f with the probability at least $1 - 2^{-\Omega(n)}$. The quantum part of the algorithm requires only single-qubit unitary gates and measurements in the standard basis.*

Moreover, suppose there exists an algorithm that with probability $\geq 1/10$, learns an n-variate polynomial $f \in \mathcal{P}_q(n,d)$, given k copies of $|\psi_f\rangle$, measuring each copy in an arbitrary orthonormal basis, and performing an arbitrary classical processing. Then $M = \Omega(n^d)$.

[2] We refer the reader to Montanaro [35, Proposition 1] for a detailed exposition of this lower bound proof.

Before stating our learning algorithm and sample complexity, we need the following lemmas.

▶ **Lemma 12.** *Choose any $f \in \mathcal{P}_q(n,d)$ and $c \in \mathbb{Z}_q$. Then either $f(x)$ is a constant function or the fraction of inputs $x \in \{0,1\}^n$ such that $f(x) \neq c$ is at least $1/2^d$.*

Proof. We shall use the following simple fact which is proved in [6].

▶ **Proposition 1.** *Consider a function $f : \{0,1\}^n \to \mathbb{Z}_q$ specified as a polynomial*

$$f(x) = \sum_{J \subseteq [n]} \alpha_J \prod_{j \in J} x_j \pmod{q}. \tag{17}$$

Here $\alpha_J \in \mathbb{Z}_q$ are coefficients. The function f is constant if and only if $\alpha_J = 0 \pmod{q}$ for all non-empty subsets $J \subseteq [n]$.

We shall prove Lemma 12 by induction in n. The base case of induction is $n = d$. Clearly, a non-constant function $f : \{0,1\}^d \to \mathbb{Z}_q$ takes a value different from c at least one time, that is, the fraction of inputs $x \in \{0,1\}^d$ such that $f(x) \neq c$ is at least $1/2^d$.

Suppose $n > d$ and $f \in \mathcal{P}_q(n,d)$ is not a constant function. Let d' be the maximum degree of non-zero monomials in f. Clearly $1 \leq d' \leq d$. Suppose f contains a monomial $\alpha_S \prod_{j \in S} x_j$ where $\alpha_S \in \mathbb{Z}_q \setminus \{0\}$ and $|S|= d'$. Since $|S|< n$, one can choose a variable x_i with $i \in [n] \setminus S$. Let $g_a : \{0,1\}^{n-1} \to \mathbb{Z}_q$ be a function obtained from f by setting the variable x_i to a constant value $a \in \{0,1\}$. Clearly, $g_a \in \mathcal{P}_q(n-1,d)$. The coefficients of the monomial $\prod_{j \in S} x_j$ in g_0 and g_1 are α_S and $\alpha_S + \alpha_{S \cup \{i\}} \pmod{q}$ respectively. However, $\alpha_{S \cup \{i\}} = 0 \pmod{q}$ since otherwise f would contain a monomial $x_i \prod_{j \in S} x_j$ of degree larger than d'. We conclude that both g_0 and g_1 contain a non-zero monomial $\alpha_S \prod_{j \in S} x_j$. By Proposition 1, g_0 and g_1 are not constant functions. Since g_0 and g_1 are degree-d polynomials in $n-1$ variables, the induction hypothesis gives

$$\Pr_y[g_a(y) \neq c] \geq \frac{1}{2^d}. \tag{18}$$

Here $y \in \{0,1\}^{n-1}$ is picked uniformly at random. Thus

$$\Pr_x[f(x) \neq c] = \frac{1}{2}\left[\Pr_y[g_0(y) \neq c] + \Pr_y[g_1(y) \neq c]\right] \geq \frac{1}{2^d}. \tag{19}$$

Here $x \in \{0,1\}^n$ is picked uniformly at random. This proves the induction step. ◀

With this lemma, we are now ready to prove Theorem 11. In the section below we first describe our learning algorithm and in the next section we prove the theorem by proving the sample complexity upper bound.

4.1.1 Learning Algorithm in Theorem 11

We are now ready to state our learning algorithm. As in Section 3.1 for learning binary phase states with separable measurements, we learn generalized phase states through examples containing information about the derivatives of $f(x)$. The crucial difference between the binary phase state learning algorithm and the generalized setting is, in the binary case, we obtained a measurement outcome b_y that corresponded to $b_y = f(0y) - f(1y)$, however in the generalized scenario, we obtain a measurement outcome b'_y that satisfies $f(0y) - f(1y) \neq b'_y$. Nevertheless, we are able to still learn f using such measurement outcomes which we describe in the rest of the section.

We now describe the learning algorithm. We carry out the algorithm in n rounds, which we index by t. For simplicity, we describe the procedure for the first round. Suppose we measure qubits $2, 3, \ldots, n$ of the state $|\psi_f\rangle$ in the Z-basis. Let $y \in \{0,1\}^{n-1}$ be the measured bit string. Note that the probability distribution of y is uniform. The post-measurement state of qubit 1 is

$$|\psi_{f,y}\rangle = \frac{1}{\sqrt{2}}(\omega_q^{f(0y)}|0\rangle + \omega_q^{f(1y)}|1\rangle) \tag{20}$$

For each $b \in \mathbb{Z}_q$ define a single-qubit state

$$|\phi_b\rangle = \frac{1}{\sqrt{2}}(|0\rangle - \omega_q^b|1\rangle) \tag{21}$$

Using the identity $\sum_{b\in\mathbb{Z}_q} \omega_q^b = 0$ one gets

$$I = \frac{2}{q}\sum_{b\in\mathbb{Z}_q} |\phi_b\rangle\langle\phi_b| \tag{22}$$

One can view Eq. (22) as a single-qubit POVM with q elements $(2/q)|\phi_b\rangle\langle\phi_b|$. Let $\mathcal{M}$ be the single-qubit measurement described by this POVM. Applying $\mathcal{M}$ to the state $|\psi_{f,y}\rangle$ returns an outcome $b \in \mathbb{Z}_q$ with the probability

$$\Pr(b|y) := \frac{2}{q}|\langle\phi_b|\psi_{f,y}\rangle|^2 = \frac{1}{2q}\left|1 - \omega_q^{f(1y)-f(0y)-b}\right|^2. \tag{23}$$

Clearly, $\Pr(b|y)$ is a normalized probability distribution, $\sum_{b\in\mathbb{Z}_q}\Pr(b|y) = 1$. Furthermore,

$$f(1y) - f(0y) = b \quad \text{implies } \Pr(b|y) = 0, \tag{24}$$

$$f(1y) - f(0y) \neq b \quad \text{implies } \Pr(b|y) \geq \frac{2}{q}\sin^2(\pi/q) = \Omega(1/q^3). \tag{25}$$

To conclude, the combined n-qubit measurement consumes one copy of the state $|\psi_f\rangle$ and returns a pair $(y,b) \in \{0,1\}^{n-1} \times \mathbb{Z}_q$ such that

$$f(1y) - f(0y) \neq b \tag{26}$$

with certainty and all outcomes b satisfying Eq. (26) appear with a non-negligible probability. Define a function $g : \{0,1\}^{n-1} \to \mathbb{Z}_q$ such that

$$g(y) = f(1y) - f(0y). \tag{27}$$

We claim that g is a degree-$(d-1)$ polynomial, that is, $g \in \mathcal{P}_q(n-1, d-1)$. Indeed, it is clear that $g(y)$ is a degree-d polynomial. Moreover, all degree-d monomials in $f(x)$ that do not contain the variable x_1 appear in $f(1y)$ and $f(0y)$ with the same coefficient. Such monomials do not contribute to $g(y)$. A degree-d monomial in $f(x)$ that contains the variable x_1 contributes a degree-$(d-1)$ monomial to $g(y)$. Thus $g \in \mathcal{P}_q(n-1, d-1)$, as claimed.

From Eq. (26) one infers a constraint $g(y) \neq b$whenever the combined n-qubit measurement of $|\psi_f\rangle$ returns an outcome (y,b). Suppose we repeat the above process m times obtaining constraints

$$g(y^{(k)}) \neq b^{(k)}, \qquad k = 1, 2, \ldots, m. \tag{28}$$

This consumes m copies of $|\psi_f\rangle$. We claim that the probability of having more than one polynomial $g \in \mathcal{P}_q(n-1, d-1)$ satisfying the constraints Eq. (28) is exponentially small if we choose

$$m = O(q^3 \log(q) 2^d n^{d-1}). \tag{29}$$

4.1.2 Sample Complexity bound in Theorem 11

Define a probability distribution $\pi(\vec{y},\vec{b})$ where

$$\vec{z}=(y^{(1)},\ldots,y^{(m)})\in\{0,1\}^{(n-1)m}\quad\text{and}\qquad \vec{b}=(b^{(1)},\ldots,b^{(m)})\in(\mathbb{Z}_q)^{\times m} \tag{30}$$

such that $y^{(j)}$ are picked uniformly at random and $b^{(k)}$ are sampled from the distribution $\Pr(b^{(k)}|y^{(k)})$ defined in Eq. (23). For each polynomial $h\in\mathcal{P}_q(n-1,d-1)$ define an event

$$\mathsf{BAD}(h)=\{(\vec{y},\vec{b})\,:\,h(y^{(k)})\neq b^{(k)}\quad\text{for all}\quad k\in[m]\}. \tag{31}$$

We claim that

$$\Pr[\mathsf{BAD}(h)]:=\sum_{(\vec{y},\vec{b})\in\mathsf{BAD}(h)}\pi(\vec{y},\vec{b})\leq\left[1-\Omega(2^{-d}q^{-3})\right]^m \tag{32}$$

for any $h\neq g$. Indeed, consider some fixed $k\in[m]$. The event $b^{(k)}\neq h(y^{(k)})$ occurs automatically if $h(y^{(k)})=g(y^{(k)})$. Otherwise, if $h(y^{(k)})\neq g(y^{(k)})$, the event $b^{(k)}\neq h(y^{(k)})$ occurs with the probability at most $1-\Omega(1/q^3)$ since $b^{(k)}=h(y^{(k)})$ with the probability at least $\Omega(1/q^3)$ due to Eq. (25). It follows that

$$\Pr_{y^{(k)},b^{(k)}}[h(y^{(k)})\neq b^{(k)}]\leq\Pr_{y^{(k)}}[h(y^{(k)})=g(y^{(k)})]+\Pr_{y^{(k)}}[h(y^{(k)})\neq g(y^{(k)})]\left(1-\Omega(1/q^3)\right) \tag{33}$$

$$=1-\Pr_{y^{(k)}}[h(y^{(k)})\neq g(y^{(k)})]\cdot\Omega(1/q^3). \tag{34}$$

If h and g are equivalent then $h(y)=g(y)+c$ for some constant $c\in\mathbb{Z}_q$. Note that $c\neq 0$ since we assumed $h\neq g$. In this case

$$\Pr_{y^{(k)}}[h(y^{(k)})\neq g(y^{(k)})]=1. \tag{35}$$

If h and g are non-equivalent, apply Lemma 12 to a non-constant degree-$(d-1)$ polynomial $h-g$. It gives

$$\Pr_{y^{(k)}}[h(y^{(k)})\neq g(y^{(k)})]\geq\frac{1}{2^{d-1}}. \tag{36}$$

In both cases we get

$$\Pr_{y^{(k)},b^{(k)}}[h(y^{(k)})\neq b^{(k)}]\leq 1-\Omega(2^{-d}q^{-3}), \tag{37}$$

which proves Eq. (32) since the pairs $(y^{(k)},b^{(k)})$ are i.i.d. random variables.

As noted earlier in the preliminaries, observe that $|\mathcal{P}_q(n-1,d-1)|\leq q^{O(n^{d-1})}=2^{O(\log(q)n^{d-1})}$. By the union bound, one can choose $m=O(2^dq^3\log(q)n^{d-1})$ such that

$$\Pr\left[\bigcup_{h\in\mathcal{P}_q(n-1,d-1)\setminus g}\mathsf{BAD}(h)\right]\leq 2^{O(\log(q)n^{d-1})}\left[1-\Omega(2^{-d}q^{-3})\right]^m\leq 2^{-\Omega(n)}. \tag{38}$$

In other words, the probability that g is the unique element of $\mathcal{P}_q(n-1,d-1)$ satisfying all the constraints Eq. (28) is at least $1-2^{-\Omega(n)}$. One can identify such polynomial g by checking the constraints Eq. (28) for every $g\in\mathcal{P}_q(n-1,d-1)$. If the constraints are satisfied for more than one polynomial, declare a failure.

At this point we have learned a polynomial $g \in \mathcal{P}_q(n-1, d-1)$ such that $f(1y) - f(0y) = g(y)$ for all $y \in \{0,1\}^{n-1}$. For simplicity, we ignore the exponentially small failure probability. Applying the same protocol n times to copies of the quantum state $|\psi_f\rangle$ by a cyclic shift of qubits, one can learn polynomials $g_0, g_1, \ldots, g_{n-1} \in \mathcal{P}_q(n-1, d-1)$ such that

$$f(C^i(1y)) - f(C^i(0y)) = g_i(y) \quad \text{for all} \quad i \in [n] \quad \text{and} \quad y \in \{0,1\}^{n-1}, \tag{39}$$

where C is the cyclic shift of n bits. This consumes $M = O(nm) = O(2^d q^3 \log(q) n^d)$ copies of the state $|\psi_f\rangle$. We can assume wlog that $f(0^n) = 0$ since our goal is to learn $f(x)$ modulo a constant shift. Suppose we have already learned values of $f(x)$ for all bit strings x with the Hamming weight $|x| \leq w$ (initially $w = 0$). Any bit string x with $|x| = w+1$ can be represented as $x = C^i(1y)$ for some $y \in \{0,1\}^{n-1}$ such that $|y| = w$. Now Eq. (39) determines $f(x)$ since $|C^i(0y)| = |y| = w$ so that $f(C^i(0y))$ is already known and the polynomial $g_i(y)$ has been learned. Proceeding inductively one can learn $f(x)$ for all x.

It remains to note that the POVM Eq. (22) is a probabilistic mixture of projective single-qubit measurements whenever q is even. Indeed, in this case the states $|\phi_b\rangle$ and $|\phi_{b+q/2}\rangle = Z|\phi_b\rangle$ form an orthonormal basis of a qubit, see Eq. (21). Thus the POVM defined in Eq. (22) can be implemented by picking a random uniform $b \in \mathbb{Z}_q$ and measuring a qubit in the basis $\{|\phi_b\rangle, Z|\phi_b\rangle\}$. Thus the learning protocol only requires single-qubit unitary gates and measurements in the standard basis.

The lower bound in the proof of Theorem 11 follows in a straightforward manner from the lower bound for binary phase states. Indeed, suppose

$$f'(x) = \sum_{J \in [n]} \alpha_J \prod_{j \in J} x_j \pmod 2$$

is an $\mathbb{F}_2$-valued degree-d polynomial, $f' \in \mathcal{P}(n,d)$. Suppose $q = 2r$ for some integer r. Define a polynomial $f(x) = rf'(x) \pmod q$. Clearly $f \in \mathcal{P}_q(n,d)$ and $\omega_q^{f(x)} = (-1)^{f'(x)}$ for all x, that is the binary phase state corresponding to f' coincides with the generalized phase state corresponding to f. Using Theorem 10, we obtain a lower bound of $M = \log|\mathcal{P}(n,d)| = \Omega(n^d)$ for learning ψ_f. This concludes the proof of Theorem 11.

4.2 Learning stabilizer states

We now describe how the algorithm stated in Theorem 11 could be used to learn any n-qubit stabilizer state (produced by a Clifford circuit applied to $|0^n\rangle$ state) using separable measurements. Note that we can learn a subclass of stabilizer states called graph states (which are simply binary phase states with $d = 2$) using Algorithm 1 with the sample complexity of $O(n^2)$ (as shown in Theorem 7).

From a result in [19], we know that a stabilizer state can be represented as follows

$$|\psi\rangle = \frac{1}{\sqrt{|A|}} \sum_{x \in A} i^{\ell(x)} (-1)^{q(x)} |x\rangle, \tag{40}$$

where A is an affine subspace of $\mathbb{F}_2^n$, $\ell : \mathbb{F}_2^n \to \mathbb{F}_2$ is a linear function and $q : \mathbb{F}_2^n \to \mathbb{F}_2$ is quadratic function. Clearly, an alternate form is a generalized phase state with degree-2

$$|\psi_f\rangle = \frac{1}{\sqrt{|A|}} \sum_{x \in A} i^{f(x)} |x\rangle \tag{41}$$

where the summation is over A instead of the entire $\mathbb{F}_2^n$, and the function $f : \mathbb{F}_2^n \to \mathbb{Z}_4$ has its coefficients corresponding to the quadratic monomials take values in $\{0, 2\}$. We can now learn this using separable measurements as stated in the following statement as opposed to entangled measurements as required by Bell sampling [36].

▶ **Corollary 13.** *There exists an algorithm that uses $M = O(n^2)$ copies of a stabilizer state $|\psi_f\rangle = \frac{1}{\sqrt{|A|}} \sum_{x \in A} i^{f(x)}|x\rangle$ with an unknown polynomial $f \in \mathcal{P}_4(n,2)$ and outputs a polynomial $g \in \mathcal{P}_4(n,2)$ such that g is equivalent to f with the probability at least $1 - 2^{-\Omega(n)}$. The quantum part of the algorithm requires only single-qubit unitary gates and measurements in the standard basis.*

Proof. The subspace A of an unknown stabilizer state can be denoted as $a + S_A$ where $a \in \mathbb{F}_2^n$ is a translation vector and S_A is a linear subspace of $\mathbb{F}_2^n$. To learn a and a basis of the subspace S_A, it is enough to measure $O(n \log n)$ copies of $|\psi_f\rangle$ in the computational basis. This in turn defines a subset of the n directions $\{e_i\}$ along which we need to search for non-zero monomials in the partial derivatives of f. We can now use the learning algorithm in Theorem 11 to learn the unknown stabilizer state using $O(n^2)$ copies with the desired probability. ◀

References

1 Emmanuel Abbe, Amir Shpilka, and Avi Wigderson. Reed-Muller Codes for Random Erasures and Errors. In *Proceedings of the Forty-Seventh Annual ACM Symposium on Theory of Computing*, STOC '15, pages 297–306, New York, NY, USA, 2015. Association for Computing Machinery. doi:10.1145/2746539.2746575.

2 Emmanuel Abbe, Amir Shpilka, and Min Ye. Reed–Muller Codes: Theory and Algorithms. *IEEE Transactions on Information Theory*, 67(6):3251–3277, 2020. doi:10.1109/TIT.2020.3004749.

3 Prabhanjan Ananth, Luowen Qian, and Henry Yuen. Cryptography from pseudorandom quantum states. *arXiv preprint arXiv:2112.10020*, 2021.

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

5 Apeldoorn van Joran, Arjan Cornelissen, Andras Gilyén, and Giacomo Nannicini. Quantum tomography using state-preparation unitaries, 2022. arXiv:2207.08800.

6 Srinivasan Arunachalam, Sergey Bravyi, Arkopal Dutt, and Theodore J Yoder. Optimal algorithms for learning quantum phase states. *arXiv preprint arXiv:2208.07851*, 2022.

7 Dave Bacon, Andrew M Childs, and Wim van Dam. Optimal measurements for the dihedral hidden subgroup problem. *arXiv preprint quant-ph/0501044*, 2005.

8 Howard Barnum and Emanuel Knill. Reversing quantum dynamics with near-optimal quantum and classical fidelity. *Journal of Mathematical Physics*, 43(5):2097–2106, 2002.

9 A. Bernasconi and B. Codenotti. Spectral analysis of Boolean functions as a graph eigenvalue problem. *IEEE Transactions on Computers*, 48(3):345–351, 1999. doi:10.1109/12.755000.

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

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

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

13 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.

14 Michael J Bremner, Ashley Montanaro, and Dan J Shepherd. Average-case complexity versus approximate simulation of commuting quantum computations. *Physical Review Letters*, 117(8):080501, 2016.

15 Michael J Bremner, Ashley Montanaro, and Dan J Shepherd. Achieving quantum supremacy with sparse and noisy commuting quantum computations. *Quantum*, 1:8, 2017.

16 Isaac L Chuang and Michael A Nielsen. Prescription for experimental determination of the dynamics of a quantum black box. *Journal of Modern Optics*, 44(11-12):2455–2467, 1997. doi:10.1080/09500349708231894.

17 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.

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

19 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.

20 Stark C. Draper and Sheida Malekpour. Compressed sensing over finite fields. In *2009 IEEE International Symposium on Information Theory*, pages 669–673. IEEE, 2009. doi:10.1109/ISIT.2009.5205666.

21 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.

22 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.

23 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.

24 Jeongwan Haah, Robin Kothari, and Ewin Tang. Optimal learning of quantum Hamiltonians from high-temperature Gibbs states. *arXiv preprint arXiv:2108.04842*, 2021.

25 Aram W Harrow and Andreas Winter. How many copies are needed for state discrimination? *IEEE Transactions on Information Theory*, 58(1):1–2, 2012.

26 Marc Hein, Jens Eisert, and Hans J Briegel. Multiparty entanglement in graph states. *Physical Review A*, 69(6):062311, 2004.

27 Alexander Semenovich Holevo. Bounds for the quantity of information transmitted by a quantum communication channel. *Problemy Peredachi Informatsii*, 9(3):3–11, 1973.

28 Sandy Irani, Anand Natarajan, Chinmay Nirkhe, Sujit Rao, and Henry Yuen. Quantum search-to-decision reductions and the state synthesis problem. *arXiv preprint arXiv:2111.02999*, 2021.

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

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

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

32 Vadim Lyubashevsky. The parity problem in the presence of noise, decoding random linear codes, and the subset sum problem. In *Approximation, randomization and combinatorial optimization. Algorithms and techniques*, pages 378–389. Springer, 2005.

33 Florence Jessie MacWilliams and Neil James Alexander Sloane. *The theory of error correcting codes*, volume 16. Elsevier, 1977.

34 Masoud Mohseni, Ali T Rezakhani, and Daniel A Lidar. Quantum-process tomography: Resource analysis of different strategies. *Physical Review A*, 77(3):032322, March 2008. doi:10.1103/PhysRevA.77.032322.

35 Ashley Montanaro. The quantum query complexity of learning multilinear polynomials. *Information Processing Letters*, 112(11):438–442, 2012.

36 Ashley Montanaro. Learning stabilizer states by Bell sampling. *arXiv preprint arXiv:1707.04012*, 2017.

37 Ashley Montanaro. Quantum circuits and low-degree polynomials over $\mathbb{F}_2$. *Journal of Physics A: Mathematical and Theoretical*, 50(8):084002, January 2017. doi:10.1088/1751-8121/aa565f.

38 Noam Nisan and Mario Szegedy. On the degree of Boolean functions as real polynomials. *Computational Complexity*, 4(4):301–313, 1994.

39 Leonardo Novo, Juani Bermejo-Vega, and Raúl García-Patrón. Quantum advantage from energy measurements of many-body quantum systems. *Quantum*, 5:465, 2021.

40 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.

41 Andrea Rocchetto, Scott Aaronson, Simone Severini, Gonzalo Carvacho, Davide Poderini, Iris Agresti, Marco Bentivegna, and Fabio Sciarrino. Experimental learning of quantum states. *Science Advances*, 5(3):eaau1946, 2019.

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

43 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.

44 Dirk Schlingemann. Stabilizer codes can be realized as graph codes. *Quantum Info. Comput.*, 2(4):307–323, June 2002.

45 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.

46 Henry Yuen. An improved sample complexity lower bound for quantum state tomography. *arXiv preprint arXiv:2206.11185*, 2022.