

Information-Computation Gaps in Quantum Learning via Low-Degree Likelihood

Sitan Chen

Harvard University

Weiyuan Gong

Harvard University

Jonas Haferkamp

Ruhr-University Bochum

Yihui Quek

EPFL

SITAN@SEAS.HARVARD.EDU

WGONG@G.HARVARD.EDU

JONAS.HAFERKAMP@RUB.DE

YIHUI.QUEK@EPFL.CH

Editors: Steve Hanneke and Tor Lattimore

Abstract

In a variety of physically relevant settings for learning from quantum data, there is an established recipe for measuring polynomially many copies of that data such that the resulting measurement readouts contain enough information to reconstruct the underlying system. Yet designing protocols that can *computationally efficiently* extract that information remains largely an art, and there are important cases where we believe this to be impossible, that is, where there is an *information-computation gap*. While there is a large array of tools in the classical literature for giving evidence for average-case hardness of statistical inference problems, the corresponding tools in the quantum literature are far more limited.

One such framework in the classical literature, the *low-degree method*, makes predictions about hardness of inference problems based on the failure of estimators given by low-degree polynomials. In this work, we extend this framework to the quantum setting and show a number of new information-computation gaps for quantum learning. We establish a general connection between state designs and low-degree hardness. We use this to obtain the first information-computation gaps for learning Gibbs states of random, sparse, non-local Hamiltonians. We also use it to prove hardness for learning random shallow quantum circuit states in a challenging model where states can be measured in round-based *adaptively* chosen bases. To our knowledge, the ability to model adaptivity within the low-degree framework was open even in classical settings. In addition, we also obtain a low-degree hardness result for quantum error mitigation against strategies with single-qubit measurements. We define a new quantum generalization of the planted biclique problem and identify the threshold at which this problem becomes computationally hard for protocols that perform local measurements. Interestingly, the complexity landscape for this problem shifts when going from local measurements to more entangled single-copy measurements. We show average-case hardness for the “standard” variant of Learning Stabilizers with Noise (Poremba et al., 2024) and for agnostically learning product states (Bakshi et al., 2024a).

Keywords: Low-degree likelihood ratio, Information-computation gap, quantum learning.

1. Introduction

Given access to quantum data, how efficiently can we learn a classical description of the underlying system that generates that data? Motivated by the exciting prospect of using quantum computers to learn about the physical world, there has been a recent surge of interest in charting out what is possible to learn and where the fundamental limitations lie. The gold standard is a learning protocol with the following three properties:

- (A) **Low statistical complexity:** Only a polynomial amount of quantum data (e.g., copies) is needed
- (B) **Efficient quantum processing:** The measurements that are performed on the data are bounded in complexity, e.g., only a few copies are measured at a time, and each measurement can be implemented by a shallow quantum circuit.
- (C) **Efficient classical post-processing:** A $\text{poly}(n)$ -time classical algorithm takes in the readouts, i.e., the measurement outcomes, and outputs a classical description of the underlying system.¹

The statistical complexity for a wide range of quantum learning problems is by now very well-understood (O’Donnell and Wright, 2015, 2016; Haah et al., 2016; Chen et al., 2023b). There has also been substantial progress on quantifying how the efficiency of quantum processing impacts statistical complexity, for instance in the form of tradeoffs between the number of copies needed and various quantum resources like number of ancilla, locality of the measurements, etc (King et al., 2024; Chen et al., 2024d, 2022a; Aharonov et al., 2022; Huang et al., 2021). *Yet to date, we understand surprisingly little about how these factors interact with axis (C).*

Information-computation gaps. On the positive side, the community has identified a number of learning problems that admit protocols which are efficient along all three axes, like tomography of Gibbs states (Haah et al., 2022; Bakshi et al., 2024b), shallow circuit states (Huang et al., 2024), stabilizer states (Aaronson and Gottesman, 2008; Montanaro, 2017), and matrix product states (Cramer et al., 2010).

In these settings, it is straightforward to achieve axes (A) and (B). Indeed, physically relevant states are typically described by polynomially many degrees of freedom, so with a polynomial amount of quantum data there is in principle enough quantum information to learn the underlying system. And in many cases of interest this quantum information can be processed with inexpensive measurements (Huang et al., 2020) (see Remark 18) into classical data that is information-theoretically sufficient for learning. As such, the core difficulty behind achieving the aforementioned results lies in taming the complexity of classical post-processing, or axis (C).

On the negative side, there are situations where even though this classical data is information-theoretically sufficient, no polynomial-time classical algorithm can harness it to learn the underlying system. This pernicious regime, where learning is “possible-but-hard” (Wein, 2025), is known in the classical literature on statistical inference as an *information-computation gap*. It is crucially important to identify where these gaps lie – it permits us to cease the search for efficient algorithms in these regimes, knowing it will be futile. In the classical literature on statistical inference, this is called an *information-computation gap*.

Pseudorandom quantum objects like pseudorandom states (Ji et al., 2018) and unitaries (Ma and Huang, 2024) cleanly illustrate how such a gap can arise. Given $\text{poly}(n)$ copies of a state which is promised to be either pseudorandom or Haar-random, if one measures each copy in the computational basis, the resulting classical readouts contain enough information to tell apart these two cases, yet no polynomial-time algorithm can do so. Apart from reducing from such cryptographic contexts however, there are few alternative approaches for establishing – or even predicting – the onset of information-computation gaps. In this work we ask:

Is there a general framework for identifying information-computation gaps in quantum learning?

1. King et al. (2024) put forth a similar three-part taxonomy, which they refer to as “triple efficiency,” specifically in the context of shadow tomography.

The low-degree conjecture. We first observe that this state of affairs is in stark contrast to the one in classical learning theory, where numerous frameworks exist for giving evidence for information-computation gaps. Most of these frameworks operate by selecting a family of estimators (e.g., statistical query algorithms, belief propagation, Markov Chain Monte Carlo) that contain the most powerful known algorithms for a large class of problems, and showing that no algorithm from this family can solve the problem at hand. One framework that has proven especially successful at accurately predicting IC gaps is the *low-degree method* (Barak et al., 2019; Hopkins and Steurer, 2017; Hopkins et al., 2017; Hopkins, 2018). Roughly, the family of estimators considered here is the set of all low-degree polynomials in the data (i.e., samples from some probability distribution), and the given learning problem is deemed to be computationally hard if no such low-degree polynomial estimator succeeds (see Section A.2 for a detailed overview). Polynomial estimators are expressive enough to capture a number of algorithmic recipes like spectral methods, message passing, and moment methods and are also closely related to sum-of-squares algorithms for inference. With few exceptions, this framework has obtained the “right” (according to widely believed conjectures) predictions for a slew of problems in statistical inference, notable examples of which include planted clique (Barak et al., 2019), community detection (Hopkins and Steurer, 2017), tensor PCA (Hopkins et al., 2017), sparse PCA (Ding et al., 2024), and combinatorial group testing (Coja-Oghlan et al., 2022).

Low-degree quantum learning algorithms abound. In Section B.5, we observe that a wide swath of known algorithms in quantum learning can also be realized as low-degree polynomial estimators, where the classical data that these estimators take as input is given by readouts from simple measurements of the quantum data. These algorithms include ones for estimating observables, learning shallow circuits (Landau and Liu, 2024; Kim et al., 2024; Huang et al., 2024), learning Hamiltonians via polynomial system solving (Haah et al., 2022), and agnostic tomography of product states (Bakshi et al., 2024a; Chen et al., 2024e). Motivated by this, in this work we extend the low-degree method to the quantum setting. Our contribution is thus to unify these algorithms under a single umbrella, and develop general machinery to prove lower bounds against all of them.

1.1. Our contributions

We begin by informally describing the quantum low-degree formalism that we develop in this work, deferring a formal treatment to Section D. Throughout, we will focus on *hypothesis testing* problems. Here, one is given polynomially many copies of a state ρ and would like to distinguish between the *null hypothesis* that ρ is a simple prescribed state (e.g., the maximally mixed state) and the *alternative hypothesis* that ρ was sampled from some ensemble $\mathcal{E}$ over “interesting” states from some class. Although there exists a gap between the lower bounds for testing and lower bounds for learning, we will focus on learning tasks with efficient certification so that a lower bound for testing imply a lower bound for learning.

By definition, any strategy for measuring the copies of ρ induces a classical distribution over readouts. It is then well-defined to consider low-degree polynomial estimators that take as input the bits in all of the readouts and output a prediction for whether ρ came from the null or the alternative hypothesis (see Figure 1 for a graphical illustration). Given a class $\mathcal{C}$ of measurement strategies, we say that a hypothesis testing problem is *low-degree hard for $\mathcal{C}$* if for *all* strategies in that class, the resulting classical distributions over readouts under the null hypothesis and under the alternative hypothesis cannot be distinguished using any low-degree polynomial estimator.

Table 1: **Summary of our main results and applications.** The first header row separates local measurements from more general single-copy PVM measurements. The second header row refines the latter into ancilla-free non-adaptive, bounded-ancilla non-adaptive, and bounded adaptive settings. Here, bounded adaptive refers to the restricted round-based adaptive models treated in Theorem 5.

Application	Local measurements	Single-copy PVM measurements		
		Ancilla-free non-adaptive	$O(n)$ -ancilla non-adaptive	Bounded adaptive
General testing ρ vs. $I/2^n$	Small k -qubit marginals $\Rightarrow$ degree- k hardness. (Cor. 20)	–	Approx. 2-designs $\Rightarrow$ degree- $n^{O(1)}$ hardness (Thm. 4)	Approx. higher designs $\Rightarrow$ degree- $n^{O(1)}$ hardness (Thm. 5)
Random circuit states $\text{polylog}(n)$ depth	–	–	Degree- $\text{polylog}(n)$ hardness (Cor. 1)	Degree- $\text{polylog}(n)$ hardness (Cor. 1)
Quantum planted biclique	degree- $n^{o(1)}$ hardness at $\lambda \leq \tilde{o}(n^{1/2}d^{1/4})$; IC gap at $d = o(n^{2/3})$ (Thm. 6)	–	–	–
Quantum error mitigation	Low-degree hardness at noise rate $O(1/n^{1-\delta'})$ (Cor. 3)	–	–	–
Learning sparse Hamiltonians from Gibbs states	–	–	Degree- $O(n)$ hardness at $O(n^3)$ sparsity (Cor. 2)	–
Learning Stabilizers with Noise	–	$\text{polylog}(n)$ -degree avg-case hardness (Thm. 8)	–	–
Agnostic learning of product states	–	Degree- $O(n)$ hardness at fidelity $1/\text{poly}(n)$ (Thm. 9)	–	–

This quantification over measurement strategies is the key novelty in passing from classical to quantum and poses a number of technical hurdles that we will overcome in the sequel.

1.1.1. APPLICATIONS AND GENERAL CONDITIONS FOR LOW-DEGREE HARDNESS

In this section we give a general, user-friendly framework under which we establish low-degree hardness. Before describing this framework, to illustrate its strength we first enumerate a series of applications.

Applications. We are able to deduce new computational hardness results for learning well-studied classes of states like random quantum circuit states, Gibbs states, and states arising in the context of quantum error mitigation.

Corollary 1 (Informal, see Theorem 53, Theorem 55, and Theorem 56) *For any non-adaptive, single-copy measurement strategy where each measurement is implemented using at most $O(n)$ ancilla, it is degree- $\text{polylog}(n)$ hard to distinguish between the maximally mixed state and a state prepared by a random geometrically local quantum circuit of $\text{polylog}(n)$ -depth, given $\text{poly}(n)$ copies.*

This extends prior work of [Nietner et al. \(2023\)](#) which showed that the *output distributions* of these circuits, when measured in the computational basis, are indistinguishable from uniformly random. That result, while not in the low-degree framework, can be interpreted as saying that random polylog-depth circuit states cannot be computationally distinguished from maximally mixed by measuring polynomially many copies in the computational basis. In contrast, our result can be interpreted as showing that this distinguishing problem is hard for a much more general family of single-copy measurement strategies, even ones which have some level of adaptivity. In [Section B.4](#) we discuss the connection with a *quantum statistical query* lower bound for this problem that was proved in [Arunachalam et al. \(2023\)](#). Taken together, these results can be viewed as modest steps towards the long-standing conjecture (see for instance Conjecture 1.1 of [Fefferman et al. \(2024\)](#) or Conjecture 1 of [Bostanci et al. \(2024\)](#)) that states prepared by superlog-depth random quantum circuits are pseudorandom.

Next we state our hardness result for learning Gibbs states:

Corollary 2 (Informal, see [Corollary 60](#)) *For any non-adaptive, single-copy measurement strategy where each measurement is implemented using at most $O(n)$ ancilla, it is degree- $\Omega(n)$ hard to distinguish between the maximally mixed state and the Gibbs state of a Hamiltonian given by a random signed sum of $O(n^3)$ random Pauli strings.*

To our knowledge, this is the first evidence that learning *sparse* (but not necessarily local) Hamiltonians from Gibbs state access is computationally hard. In contrast, recent works have shown that such Hamiltonians can be learned given access to their dynamics ([Ma et al., 2024](#); [Hu et al., 2025](#)). Note that [Gu et al. \(2024b\)](#) has proven a no-go theorem for the task of distinguishing between two ensembles of Hamiltonians which are both sparse in the *computational* basis (and thus not sparse in the Pauli basis) given access to their *time evolutions*.

We also show evidence that quantum error mitigation is computationally hard for *geometrically local* noisy quantum circuits in any architecture with constant number of layers of single-qubit depolarizing noise at extremely low noise rate $O(1/n^{1-\delta'})$ for any $\delta' > 0$. In contrast, quantum error mitigation is only known to be statistically hard for geometrically local noisy quantum circuits with $\omega(\log n)$ layers of noise at constant noise rate ([Quek et al., 2024](#)). The low noise rate of $O(1/n^{1-\delta'})$ is far from making the quantum error mitigation task statistically hard: there is a statistically efficient algorithm using multi-copy joint measurement as the trace distance between the two ensembles is large. However, whether there exists a statistically efficient protocol using single-qubit measurements for quantum error mitigation within the same parameter regime is unknown, leaving the existence of a strict information-computation gap an open question.

Corollary 3 (Informal, see [Theorem 48](#)) *For any non-adaptive measurement strategy using $m = \text{poly}(n)$ single-qubit measurements, there is a $D = \omega(\log n)$ such that it is degree- D hard to distinguish between the maximally mixed state and geometrically local noisy quantum circuits on any architecture of depth $\omega(\log n)$ at extremely low local depolarization noise rate $O(1/n^{1-\delta'})$ for any $\delta' > 0$. Furthermore, the low-degree hardness result still holds even if the noise is only applied in constantly many layers instead of every layer in the noisy quantum circuit.*

These applications are nearly immediate applications of the following general conditions under which a hypothesis testing problem is low-degree hard for various classes of measurement strategies.

From weak designs to strong indistinguishability. We show that if the ensemble $\mathcal{E}$ is an approximate state 2-design, that is, if its second-order moments approximately match those of the

Haar measure, then hypothesis testing is low-degree hard. At this juncture we stress that while the general connection between design property and other notions of hardness has previously been explored, as we will see there are subtle differences in how designs relate to low-degree hardness specifically.

Theorem 4 (Informal, see Theorem 25) *Distinguishing whether $\rho = I/2^n$ or whether ρ was sampled from an ensemble $\mathcal{E}$, given $m = \text{poly}(n)$ copies, is degree- k hard for any non-adaptive, single-copy measurement strategy where each measurement is implemented using at most $O(n)$ ancilla, provided $\mathcal{E}$ forms a $2^{-\tilde{\Omega}(k \log n)}$ -approximate state 2-design.*

We also obtain similar results for other strategies like local (i.e. single-qubit) measurements (see Theorem 20) and projective measurements implemented by shallow geometrically local circuits (see Theorem 22).

One interesting feature of Theorem 4 is that the condition on the ensemble $\mathcal{E}$ for degree- k hardness is merely a condition on its *degree-2 moments*. The takeaway is that, starting from the relatively weak condition that *two* copies of ρ drawn from $\mathcal{E}$ are *statistically* indistinguishable from two copies of a Haar-random state, we can upgrade to show that under single-copy measurement strategies, *polynomially* many copies of ρ drawn from $\mathcal{E}$ are *computationally* indistinguishable from polynomially many copies of a Haar-random state. This “degree magnification” effect appears to be an inherently quantum phenomenon.

Adaptive measurements. One distinctive aspect of the quantum setting is that the measurements can also be chosen adaptively based on previous measurement outcomes. Modeling adaptivity in the low-degree framework however turns out to be surprisingly subtle. As we explain in Section E.5 (see Figure 2 for a graphical illustration), the issue is that arbitrary adaptivity can easily hide expensive classical computation that is not captured by the eventual degree of the classical post-processing, and natural attempts at constraining the complexity of the adaptive choices either fail to circumvent this or result in models where proving low-degree hardness is at least as hard as proving circuit lower bounds.

Nevertheless, we are able to identify two nontrivial settings in which we provide general conditions under which low-degree hardness holds even under adaptive measurements:

Theorem 5 (Informal, see Theorem 33 and Theorem 37) *Distinguishing whether $\rho = I/2^n$ or whether ρ was sampled from an ensemble $\mathcal{E}$, given $m = \text{poly}(n)$ copies, is degree- k hard for single-copy projective measurements in bases that are adaptively chosen under either of the following conditions:*

- Full adaptivity within blocks of $m_0 = \text{polylog}(n)$ copies each: provided $\mathcal{E}$ is a $2^{-\text{poly}(k, \log n)}$ -approximate state $O(km_0)$ -design.
- Full adaptivity across $\text{polylog}(n)$ blocks: provided $\mathcal{E}$ is a $2^{-\text{poly}(k, \log n)}$ -approximate state $O(k)$ -design.

The reader may wonder why this is not immediate given that, for k -designs, degree- k polynomials in ρ look Haar-random by definition. However, the key point is that expensive computations possibly hidden in even a single adaptive choice can involve high-degree polynomials. In this context, Theorem 5 surfaces a subtle trade-off between the design-degree (and the level of approximation) and the adaptivity allowed in the framework.

We note that recent work in the low-degree literature has also identified several interesting *classical* settings where there is an option to make adaptive choices, e.g., sublinear-time planted clique (Mardia et al., 2024) and geometric graph inference (Bangachev and Bresler, 2024). These works only prove hardness when the choices are non-adaptive, and capturing adaptivity in the low-degree framework here remains an open question, for which we believe our results shed light both technically and from a modeling perspective.

Our results on adaptivity rely on the development of a stronger notion of *copy-wise* degree which allows for polynomial estimators whose monomials, while each depending on a small number of copies, can have higher degree over the qubits of those copies. This mirrors a notion studied in the classical low-degree literature (Brennan et al., 2021). In Section E.4.2, we also leverage connections, first shown in that work, to an alternative computational model called the statistical query model to show that these two models are equivalent in the realistic setting where there is stochastic noise in the measurement readouts.

1.1.2. BEYOND DESIGNS: FINE-GRAINED THRESHOLDS FOR QUANTUM PLANTED BICLIQUE

One selling point of the classical low-degree method that the results above do not yet capture is its ability to pinpoint *fine-grained* thresholds for the signal-to-noise ratio at which a problem transitions from being computationally tractable to computationally intractable but information-theoretically easy. More precisely, in addition to problems for which computational hardness holds for any polynomial sample complexity, the low-degree method also provides accurate predictions for problems for which computational hardness only manifests with a *specific* polynomial scaling for the sample complexity. While identifying phase transitions is integral to understanding physical systems in other quantum information contexts (Skinner et al., 2019; Ippoliti et al., 2021; Ippoliti and Khemani, 2024), the field currently lacks frameworks for establishing such fine-grained statements when it comes to the onset of hardness for learning.

In the second part of this work, we take a step towards filling this gap by identifying the first average-case quantum learning problem, parametrized by a natural notion of signal-to-noise ratio, for which one can establish such fine-grained bounds. The problem we propose, which we call *Quantum Planted Biclique*, can be thought of as a quantum generalization of the well-studied problem of *planted biclique* (Feldman et al., 2017). In classical planted biclique, a learner observes a bipartite $n \times n$ graph that is either truly random, or a complete bipartite graph has been planted in some subgraph of expected size $\lambda \times \lambda$, and the goal is to distinguish between the two cases. For this problem, distinguishing is information-theoretically possible but strongly believed to be computationally intractable when $\log(n) \ll \lambda \ll \sqrt{n}$ (Feldman et al., 2017; Brennan et al., 2021), and λ/n is known as the “signal-to-noise” ratio. This problem plays a significant role as a reduction hub in the classical literature (Berthet and Rigollet, 2013; Brennan et al., 2018; Brennan and Bresler, 2020): many inference problems are shown hard by reductions from it, and we believe our quantum planted biclique problem is a natural candidate for the same role on quantum data.

In our Quantum Planted Biclique problem, the goal is to distinguish, given n copies of an unknown n -qudit state $\sigma \in (\mathbb{C}^{d \times d})^{\otimes n}$, whether σ is the maximally mixed state or was generated as follows:

- Nature samples a Haar-random d -dimensional state ρ and an unknown subsystem $S \subseteq [n]$ of expected size λ at the outset.
- Each copy of σ is constructed by “planting” ρ into the qudits indexed by S in an otherwise maximally mixed state, and then corrupting the state with global depolarizing noise of rate λ/n .

We defer the formal definition of this problem to Section G.

For intuition, consider the case of $d = 2$. If the “plant” ρ were $|1\rangle$, rather than Haar-random, and every copy was measured in the computational basis, then the readouts correspond to a random bipartite $n \times n$ graph where a complete bipartite graph is planted in some subgraph of expected size $\lambda \times \lambda$. In contrast, if σ were maximally mixed, the readouts would correspond to a random bipartite $n \times n$ graph. This is exactly the problem of (*classical*) *planted biclique*. For that problem, distinguishing is information-theoretically possible but strongly believed to be computationally intractable when $\log(n) \ll \lambda \ll \sqrt{n}$ (Feldman et al., 2017; Brennan et al., 2021).

When ρ is instead Haar-random, the resulting classical distinguishing problem that arises after measuring in the computational basis is very similar, but instead of a complete bipartite graph, what is planted is a bipartite graph whose edge probabilities are a constant bounded away from $1/2$. This is the problem of *planted dense subgraph*, which exhibits the same information-computation gap as planted biclique.

Like classical planted biclique, quantum planted biclique has a provable, fine-grained detection threshold that matches the predictions of our low-degree framework. It is straightforward to show that when $\lambda = \omega(n^{1/2}d^{1/4})$, there is a low-degree protocol based on local measurements in the computational basis which succeeds at distinguishing the two cases (see Theorem 65). In this work, we show a matching lower bound under our low-degree framework:

Theorem 6 (Informal, see Theorem 61) *If $\lambda \leq \tilde{o}(n^{1/2}d^{1/4})$, quantum planted biclique is degree- $n^{o(1)}$ hard for non-adaptive local measurements.*

When $\lambda = \omega(\min(n^{1/2}d^{1/4}, d))$, one can verify that there is a computationally *inefficient* protocol using the same measurements which succeeds (see Theorem 65). This means that for any $d = o(n^{2/3})$, there is an information-computation gap for this problem.

Remark 7 *When $d = \omega(n^{2/3})$, we conjecture that quantum planted biclique is not just low-degree hard, but information-theoretically impossible with local measurements. Interestingly, we observe that in this regime, there is a low-degree algorithm based on non-local single-copy measurements which works provided $\lambda = \omega(n^{2/3})$ (see Lemma 64). We leave as an intriguing open question resolving the low-degree complexity of quantum planted biclique under non-local single-copy measurement strategies. For instance, is there such an algorithm that works even below $\lambda \asymp n^{2/3}$? This would yield a fine-grained computational separation between local and non-local single-copy measurement strategies.*

Given the central role played by the planted clique conjecture within the classical literature on complexity of statistical inference, in particular the extensive web of reductions from this problem to other inference problems (Berthet and Rigollet, 2013; Brennan et al., 2018; Brennan and Bresler, 2020), our definition of the quantum version of this problem may also be of independent interest.

1.1.3. QUANTUM HARDNESS FOR HYBRID AND CLASSICAL PROBLEMS.

In the last part of this work, we explore whether the low-degree framework can also be used to establish quantum hardness for problems where a portion or all of the input is *classical*.

Our first result along these lines concerns the *Learning Stabilizers with Noise (LSN)* problem (Poremba et al., 2024), recently proposed as a hybrid classical-quantum analogue of Learning Parity with Noise (LPN). Whereas the latter can be viewed as the task of decoding a random classical linear code in the presence of bit flip noise, LSN is that of decoding a random *quantum*

stabilizer code in the presence of local depolarizing noise. Importantly, in this problem, one is given a *classical* description of a random stabilizer subgroup S , in addition to a noisy quantum encoded state given by passing an unknown state in the computational basis through an encoding circuit for the stabilizer code associated to S . We establish the following average-case hardness result:

Theorem 8 (Informal, see Theorem 66) *Learning Stabilizers with Noise is $\text{polylog}(n)$ -degree hard in the average case for projective measurements.*

Interestingly, unlike the results of Poremba et al. (2024), ours applies to the “standard” average-case formulation of the problem, where the underlying stabilizer code is uniformly random.

Finally, we also consider purely classical problems. In this work, we observe that a natural family of approaches based on reducing to a quantum state learning problem by applying a low-degree polynomial encoding of the classical input into copies of a quantum state, applying non-adaptive single-copy measurements, and running low-degree classical post-processing on the readouts will fail provided the original problem is low-degree hard in the classical sense. By applying this reasoning to tensor PCA (Montanari and Richard, 2014), we obtain as a corollary the following average-case hardness result, which complements a recent result of Bakshi et al. (2024a) that showed that agnostically learning product states is NP-hard in the worst case.

Theorem 9 (Informal, see Theorem 67) *There is a simple ensemble over $4n$ -qubit pure states for which, given $\text{poly}(n)$ copies of a state ρ drawn from this ensemble, for any $D = o(n)$ it is degree- D hard to approximately learn the best product state approximation to ρ , even though such an approximation exists with fidelity at least $1/\text{poly}(n)$.*

2. Overview of techniques

2.1. Low-degree hardness with non-adaptive measurements

Warmup: single-qubit measurements. To start with, we consider a direct extension of the classical low-degree model for quantum data as a warm-up in Appendix D.1. The learner makes non-adaptive single-qubit measurements $\{\otimes_{r=1}^n |\phi_{s_i,r}\rangle\langle\phi_{s_i,r}|\}_{s_i}$ for $i \in [m]$ on $m = \text{poly}(n)$ copies, and computes the output as a function of degree k on the classical output in $\mathbf{s} = (s_1, \dots, s_m) \in \{0, 1\}^{m \cdot n}$ of measurements.

As the classical outputs of single-qubit measurements are Boolean, for any degree k we can write the m -sample low-degree likelihood ratio $\mathbb{E}_\rho(\overline{D}_\rho^{\otimes m})^{\leq k}$ as a Fourier expansion $\mathbb{E}_\rho(\overline{D}_\rho^{\otimes m})^{\leq k}(\mathbf{s}) = \sum_T \hat{D}[T] \chi_T(\mathbf{s})$ for $T \subseteq [mn]$, $|T| \leq k$. Here $\chi_T(\mathbf{s}) = (-1)^{\sum_{(i,r) \in T} s_{i,r}}$ is the usual Fourier basis. We then compute the degree- k advantage $\chi_{\leq k}^2 = \sum_{0 \leq |T| \leq k} \hat{D}[T]^2$. This quantity is the optimal distinguishing advantage achieved by any distinguisher which takes the bits of the measurement outcomes and applies a polynomial of degree at most k .

As there are at most $\sum_{i=1}^k \binom{mn}{i} \leq (mn)^k$ choices of T , we can bound the degree- k advantage by $o(1)$ if we can ensure $\hat{D}[T] \leq 2^{-\tilde{\Omega}(k \log n)}$. We then show that, if every subsystem T of $\rho^{\otimes m}$ of at most k qubits is $2^{-\tilde{\Omega}(k \log n)}$ -close to the maximally mixed state in trace distance, we can ensure $\hat{D}[T] \leq 2^{-\tilde{\Omega}(k \log n)}$ and thus that weak detection is degree- k hard (see Theorem 20 for details).

Application: Quantum error mitigation. As an application of the above warmup, we show in Appendix F.1 that it is degree- $\omega(\log n)$ hard to distinguish between the maximally mixed state and output state of geometrically local noisy quantum circuits of depth $\omega(\log n)$ at extremely low local depolarization noise rate $O(1/n^{1-\delta'})$ for any $\delta' > 0$ as in Theorem 48. To obtain this result, we consider the ensemble of quantum circuits that interleaves a constant number of random circuits of depth $\omega(\log(n))$ that form approximate unitary 2-designs and local depolarization noise of noise rate $O(1/n^{1-\delta'})$. We then show that the local subsystems of output states of random states in this ensemble are indistinguishable from the maximally mixed state. The hardness result then follows by combining the local indistinguishability and the above warmup calculation for single-qubit measurements.

Ancilla-assisted single-copy measurements. In Appendix D.2, we consider the setting in Theorem 4 when the protocol can perform $m = \text{poly}(n)$ non-adaptive single-copy PVMs $\{|\phi_{s_i}\rangle\langle\phi_{s_i}|\}_{s_i}$ for $i \in [m]$, where each measurement is implemented using at most $n' = O(n)$ ancillary qubits. The challenge compared to the previous setting is that here, the measurements can nontrivially entangle different qubits of the system, precluding a direct argument about the reduced density matrices of $\rho^{\otimes m}$.

As each single-copy PVM with n' ancillary qubits can be written as $\{U_i |s\rangle\langle s| U_i^\dagger\}_{s \in \{0,1\}^{n+n'}}$ for a unitary U_i , the protocol can be regarded as performing computational-basis single-qubit measurements on the “rotated” batch of m samples $\{U_i \cdot \rho \otimes |0\rangle\langle 0|^{\otimes n'} \cdot U_i^\dagger\}_{i=1}^m$. According to the previous discussion on protocols with single-qubit measurements, weak detection is degree- k hard if any subsystem (can be chosen across samples) in the m rotated samples up to k qubits is $2^{-\tilde{\Omega}(k \log n)}$ indistinguishable from the maximally mixed state.

To prove Theorem 4, the main step is then to show that an ensemble $\mathcal{E}$ that forms a $2^{-\tilde{\Omega}(k \log n)}$ -approximate state 2-design satisfies such local indistinguishability for any sequence of unitaries $\{U_i\}_{i=1}^m$. We provide the intuition here and leave the detailed proof to Theorem 25.

Our first step is to show that, given m copies of ρ sampled from a state 2-design ensemble and a sequence of n -qubit unitaries $\{V_i\}_{i=1}^m$, any subsystem in the m rotated samples $\{V_i \rho V_i^\dagger\}_{i=1}^m$ up to k qubits is $2^{-\Theta(n)}$ -close to maximally mixed state. We compute trace distance here using the norm inequality and the square 2-norm. As ρ forms a state 2-design, for the purposes of our analysis which implicitly only uses degree-2 moments of $\mathcal{E}$, it behaves as if Haar-random chosen and is thus invariant after rotation by $\{V_i\}_{i=1}^m$. For each single copy, the local subsystem is $2^{-\Theta(n)}$ -close to maximally mixed state (Hayden et al., 2006) and local indistinguishability then follows by a union bound over copies.

This is already enough to establish the special case of Theorem 4 when there are $n' = 0$ ancilla. But projective measurements without ancilla preclude even very simple two-outcome measurements. To address this, we use tools from unitary decomposition (Shende et al., 2005) to decompose each $(n + n')$ -qubit unitary U_i in $\{U_i \cdot \rho \otimes |0\rangle\langle 0|^{\otimes n'} \cdot U_i^\dagger\}$ into $2^{\Theta(n')}$ terms of product between the n qubits and n' ancillary qubits. This decomposition makes it possible for us to ensure local indistinguishability for any sequence of unitaries $\{U_i\}_{i=1}^m$ given $n' \leq O(n)$; we defer the details to Appendix D.2.

Remark 10 Our findings about approximate 2-design property implying local indistinguishability can be interpreted as a version of the statement that states from a 2-design are approximately k -uniform states up to any quantum experiment that takes $\text{poly}(n)$ samples. A k -uniform state of n qudits is one on which any subsystem of k -qudits is maximally-mixed (Goyeneche et al., 2015;

Goyeneche and Życzkowski, 2014; Raissi et al., 2020). While our finding follows from standard calculations on designs, we were not able to find an articulation of this point in the literature and believe it is of conceptual interest.

Application: Gibbs states of sparse Hamiltonians. As an application of Theorem 4, we show in Appendix F.3 that there is a $D = O(n)$ such that it is degree- D hard to distinguish between the maximally mixed state and the Gibbs state of a Hamiltonian given by a random signed sum of $O(n^3)$ random Pauli strings using single-copy PVMs with $O(n)$ ancillary qubits as claimed in Theorem 2. This can be shown by combining the following two observations. Firstly, the eigenvalues of a random Hamiltonian concentrate within the interval $[-3, 3]$ with high probability (Chen et al., 2024a) (see (16) for details). Secondly, the Hamiltonian ensemble matches the Gaussian unitary ensemble up to the third moment. The eigenvectors of the Hamiltonian thus form a state 3-design. Therefore, the Gibbs states of Hamiltonian given by a random signed sum of $O(n^3)$ random Pauli strings can be regarded as a classical combination of states chosen from a state 3-design.

2.2. Low-degree hardness with restricted adaptivity

Here, we consider protocols that can perform $m = \text{poly}(n)$ adaptive single-copy PVMs, in which each measurement can be chosen adaptively based on previous measurement outcomes. The case when the learners have adaptivity is surprisingly subtle. As a tool to analyze the performance, we introduce the quantum copy-wise low-degree model, a quantum analog of the sample-wise degree model, which includes all post-processing as degree- $(\mathcal{D}, k)$ functions that are degree $\mathcal{D}$ in at most k copies (note that all degree- k functions in the standard low-degree model are degree- $(\mathcal{D}, k)$ functions). In Theorem 28 and Theorem 29 (see Appendix E.1), we extend the reasoning from the previous section to show that weak detection is degree- $(\mathcal{D}, k)$ hard for non-adaptive single-copy PVMs provided the ensemble $\mathcal{E}$ forms a $2^{-\tilde{\Omega}(k \log n)}$ -approximate state k -design. To see this, we consider the terms in the Fourier expansion of the degree- $(\mathcal{D}, k)$ advantage $\chi_{\leq(\mathcal{D}, k)}^2$, recalling that each is a square of a Fourier coefficient of likelihood ratio $\mathbb{E}_\rho(\overline{D}_\rho^{\otimes m})$. For each fixed T , the Fourier coefficient $\hat{D}[T]$ is bounded by the $|T|$ -th (the number of copies in T) moment of the degree- $\leq \mathcal{D}$ advantage of a single copy by Hölder’s inequality (see (6) for details). We show via a moment calculation that if $\mathcal{E}$ is an approximate k -design, this ensures the $|T|$ -th moment of the degree- $\mathcal{D}$ advantage of a single copy is bounded by $2^{-\tilde{\Omega}(k \log n)}$ for any $\mathcal{D}$.

Round-based model with adaptivity within blocks. The most important benefit of using the copy-wise low-degree model is that it enables analyses on protocols with adaptivity. In this work, we consider adaptive protocols under two different “round-based models,” which divide the m samples into m_1 blocks, each of m_0 samples.

Let us first consider the round-based model with adaptivity *within* blocks in Appendix E.2, where we have $m_1 = \text{poly}(n)$ blocks, each consisting of $m_0 = \text{polylog}(n)$ copies that can be measured adaptively. While adaptivity is allowed *within* the blocks, no adaptivity is allowed *between* blocks. The first step is to note that if we can ensure a small k -th moment of the degree- $(\mathcal{D}, k)$ advantage for each *individual block*, we can hope to bound the degree- $(\mathcal{D}, k)$ advantage of the overall protocol using the prior non-adaptive analysis, where we now replace individual copies of ρ with *blocks* of copies. So let us focus on the k -th moment of the degree- $(\mathcal{D}, k)$ advantage for each individual block of m_0 copies. As the measurements within a block can be adaptively chosen, we cannot trace out the copies that are not included in T when we are computing the Fourier coefficient

$\hat{D}[T]$. To address this issue, we propose a delicate inductive argument. We divide all subsets $T \subset [m_0]$ into two classes: the subsets that contain the last copy and the subsets that do not. The contribution for the first class of T 's sums up to the degree- $(\mathcal{D}, k)$ advantage of a block containing $m_0 - 1$ copies. To compute the contribution for the second class of T 's, we break all such T 's into a union of the last copy and the remaining copies in (9). By one application of Cauchy-Schwarz, we can bound the contribution by a value that depends on the $2m_0$ -th moment of a single-copy degree- $\mathcal{D}$ advantage and the $2(m_0 - 1)$ -th moment of the likelihood ratio for a single-copy outputting a particular classical measurement outcome (see Eq. (10)). By induction, we are able to compute the degree- $(\mathcal{D}, k)$ advantage for each individual block of m_0 copies using these quantities. Finally, we generalize the above analysis to compute the degree- $(\mathcal{D}, k)$ advantage of the whole protocol using the $2km_0$ -th moment of a single-copy degree- k advantage and the $2k(m_0 - 1)$ -th moment of the likelihood ratio for a single copy outputting a particular classical measurement outcome, where the k overhead in the order of moments comes from the fact that each T can participate in at most k blocks.

To prove the first bullet point of Theorem 5, we employ the fact that a $2^{-\text{poly}(k, \log n)}$ -approximate state $O(km_0)$ -design provides a bound on both the $2km_0$ -th moment of a single-copy degree- $\mathcal{D}$ advantage and the $2k(m_0 - 1)$ -th moment of the likelihood ratio for a single copy outputting a particular classical measurement outcome. The detailed calculation is provided in Theorem 33.

Round-based model with adaptivity among blocks. We also consider the round-based model with adaptivity among blocks in Appendix E.3, where we have $m_1 = \text{polylog}(n)$ adaptive blocks of $m_0 = \text{poly}(n)$ non-adaptive copies within each block. At the start of each block, we can prepare a fixed measurement strategy to apply to the next $\text{poly}(n)$ copies, where the strategy is chosen fully adaptively based on all readouts from the previous blocks.

In this model, we use the combination of the recursive induction and the moment bounds in *reverse order* relative to what we did in the previous model. Given a new block, as the measurements in this block can be adaptively predetermined based on the previous history, we compute the Fourier coefficient for block-wise subsets $T \subset [m_1]$. As a dual to the previous inductive argument, we divide all subsets $T \subset [m_1]$ into two classes: the subsets that contain the last block and the subsets that do not. The contribution for the first class of T 's sums up to the degree- $(\mathcal{D}, k)$ advantage of another round-based protocol with adaptivity among blocks containing $m_1 - 1$ copies. To compute the contribution for the second class of T 's, we break all such T 's into a union of the last block and the remaining blocks in (9). Again by an application of Cauchy-Schwarz, we can bound the contribution by a value that depends on the $2k$ -th moment of a single-copy degree- $\mathcal{D}$ advantage and the $2(m_1 - 1)$ -th moment of the likelihood ratio for a single *block* outputting a particular classical measurement outcome. To prove the second bullet point of Theorem 5, we employ the fact that a $2^{-\text{poly}(k, \log n)}$ -approximate state $O(m_1)$ -design provides a bound on both the $2k$ -th moment of a single-copy degree- $\mathcal{D}$ advantage and the $2(m_1 - 1)$ -th moment of the likelihood ratio for a single block outputting a particular classical measurement outcome. The detailed calculation is provided in Theorem 37.

Learning random quantum circuits. As an application, we show in Appendix F.2 that weak detection is degree- $\text{polylog}(n)$ hard for non-adaptive and round-based adaptive protocols to distinguish between the maximally mixed state and a state prepared by a random polylogarithmic-depth geometrically local quantum circuit. This is because random polylogarithmic-depth geometrically

local quantum circuits form a polylog-design up to inverse superpolynomially small approximation factor according to our calculation in Appendix F.2.1 and Schuster et al. (2024).

2.3. Quantum planted biclique

Here we sketch the key technical ingredients in our proof of low-degree hardness for quantum planted biclique. In this problem, one is either given n copies of the maximally mixed state on n d -dimensional qudits, or n copies of a state which is given by planting λ/n copies of the same Haar-random pure state ρ into a random subset S of the qudits and corrupting the state with global depolarizing noise. As mentioned in the introduction, when $d = 2$ and ρ is simply $|1\rangle$, measuring these states in the computational basis will result in an instance of the classical planted biclique problem.

For general d and general measurements, the starting point is the fact that if one measures a d -dimensional Haar-random state in the computational basis, the likelihood of seeing one of the first $d/2$ measurement outcomes is roughly $1/2 + \Theta(1/\sqrt{d})$. The algorithmic upper bound for local measurements (see Theorem 65) simply uses this to reduce quantum planted biclique to an instance of classical *planted dense subgraph* where the edges in the planted subgraph have elevated probability $1/2 + \Theta(1/\sqrt{d})$ whereas all other edges have probability $1/2$. A standard second moment calculation shows that this would give rise to the $n^{1/2}d^{1/4}$ scaling in Theorem 6, and indeed, an existing calculation for this classical problem (see Claim 8.7 in Brennan et al. (2021)) would predict the same threshold on the lower bound side *assuming the protocol simply measures every qudit in the computational basis and checks whether the readout is within $\{1, \dots, d/2\}$ or $\{d/2+1, \dots, d\}$.*

The key challenge lies in going from this rough intuition to a lower bound against all local measurements. *A priori* the above strategy appears to be wasting significant information by throwing all measurement outcomes into one of two buckets, and furthermore it might seem advantageous to measure the different qudits in different bases. The content of our lower bound is that neither of these modifications would help, and that the naive protocol above is actually optimal. The proof is ultimately a delicate moment calculation, but here we highlight some ingredients unique to the quantum setting that may be of independent interest.

One point is that in the classical setting of planted dense subgraph, it is trivial to compute the Fourier coefficients of the likelihood ratio as an explicit function of the edges that appear in the overall graph. The Fourier coefficients of the *expected* likelihood ratio, over the randomness of the planted structure, can be read off easily from this. In the quantum setting, this is far from the case as the likelihood ratio is ultimately an unwieldy multilinear polynomial in the measurement vectors. Instead of getting a clean expression for the likelihood ratio for a given ρ, S , we have to carry around this multilinear polynomial for most of the proof. When we expect over the planted structure at the end by taking a Haar integral, this results in a weighted linear combination of permutation operators applied to the measurement vectors. While similar expressions have appeared in the related context of *purity testing* in Chen et al. (2022a), there the challenge was to derive *lower bounds* on these quantities, whereas we need to derive *upper bounds*. Finally, another hurdle that we encounter is the need to work with the harmonic basis over $\mathbb{Z}_d$ instead of the Boolean Fourier basis, as local measurement of any qudit in the null hypothesis results in the uniform distribution over $\{1, \dots, d\}$. Bounding the low-degree coefficients under this harmonic basis requires careful bookkeeping to avoid d^k factors. We defer the technical details of all of this to Section G.1.

Acknowledgments

We are grateful to Tim Hsieh for generously sharing his expertise on low-degree hardness during this project. We also thank Alex Wein for his thoughtful and thorough feedback on an early version of this manuscript. We also thank Scott Aaronson, Anurag Anshu, John Bostanci, Adam Bouland, Hsin-Yuan Huang, Jerry Li, Yunchao Liu, Zidu Liu, Qi Ye, and Henry Yuen for illuminating discussions on learning and cryptography, learning shallow circuits, matrix product states, and agnostic learning of product states. SC acknowledges support from NSF Award 2430375.

References

- Scott Aaronson and Daniel Gottesman. Identifying stabilizer states. *Perimeter Institute Recorded Seminar Archive*, page <https://pirsa.org/08080052>, 2008.
- Scott Aaronson, Adam Bouland, Bill Fefferman, Soumik Ghosh, Umesh Vazirani, Chenyi Zhang, and Zixin Zhou. Quantum pseudoentanglement. In *15th Innovations in Theoretical Computer Science Conference (ITCS 2024)*. Schloss-Dagstuhl-Leibniz Zentrum für Informatik, 2024.
- Emmanuel Abbe, Pritish Kamath, Eran Malach, Colin Sandon, and Nathan Srebro. On the power of differentiable learning versus pac and sq learning. *Advances in Neural Information Processing Systems*, 34:24340–24351, 2021.
- Dorit Aharonov, Jordan Cotler, and Xiao-Liang Qi. Quantum algorithmic measurement. *Nature communications*, 13(887):1–9, 2022.
- Greg W Anderson, Alice Guionnet, and Ofer Zeitouni. *An introduction to random matrices*. Number 118. Cambridge University Press, 2010.
- Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, and Mehdi Soleimanifar. Sample-efficient learning of quantum many-body systems. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 685–691. IEEE, 2020.
- Anurag Anshu, Srinivasan Arunachalam, Tomotaka Kuwahara, and Mehdi Soleimanifar. Efficient learning of commuting Hamiltonians on lattices. *Unpublished notes available at Anurag Anshu’s website,(link to note)*, 2021. URL https://anuraganshu.seas.harvard.edu/sites/g/files/omnuum6156/files/anshu/files/learning_commuting_hamiltonian.pdf.
- Srinivasan Arunachalam, Alex B Grilo, and Henry Yuen. Quantum statistical query learning. *arXiv preprint arXiv:2002.08240*, 2020.
- Srinivasan Arunachalam, Vojtech Havlicek, and Louis Schatzki. On the role of entanglement and statistics in learning. *Advances in Neural Information Processing Systems*, 36:55064–55076, 2023.
- Ainesh Bakshi, John Bostanci, William Kretschmer, Zeph Landau, Jerry Li, Allen Liu, Ryan O’Donnell, and Ewin Tang. Learning the closest product state. *arXiv:2411.04283*, 2024a.

- Ainesh Bakshi, Allen Liu, Ankur Moitra, and Ewin Tang. Learning quantum Hamiltonians at any temperature in polynomial time. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1470–1477, 2024b.
- Afonso S Bandeira, March T Boedihardjo, and Ramon van Handel. Matrix concentration inequalities and free probability. *Inventiones mathematicae*, 234(1):419–487, 2023.
- Kiril Bangachev and Guy Bresler. On the fourier coefficients of high-dimensional random geometric graphs. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 549–560, 2024.
- Boaz Barak, Samuel Hopkins, Jonathan Kelner, Pravesh K Kothari, Ankur Moitra, and Aaron Potechin. A nearly tight sum-of-squares lower bound for the planted clique problem. *SIAM Journal on Computing*, 48(2):687–735, 2019.
- Quentin Berthet and Philippe Rigollet. Complexity theoretic lower bounds for sparse principal component detection. In *Conference on learning theory*, pages 1046–1066. PMLR, 2013.
- Avrim Blum, Merrick Furst, Michael Kearns, and Richard J Lipton. Cryptographic primitives based on hard learning problems. In *Proceedings of the 13th Annual International Cryptology Conference on Advances in Cryptology*, pages 278–291, 1994.
- John Bostanci, Jonas Haferkamp, Dominik Hangleiter, and Alexander Poremba. Efficient quantum pseudorandomness from hamiltonian phase states. *arXiv preprint arXiv:2410.08073*, 2024.
- Tatiana Brailovskaya and Ramon van Handel. Universality and sharp matrix concentration inequalities. *Geometric and Functional Analysis*, 34(6):1734–1838, 2024.
- Zvika Brakerski, Vadim Lyubashevsky, Vinod Vaikuntanathan, and Daniel Wichs. Worst-case hardness for lpn and cryptographic hashing via code smoothing. In *Annual international conference on the theory and applications of cryptographic techniques*, pages 619–635. Springer, 2019.
- Fernando GSL Brandao, Aram W Harrow, and Michał Horodecki. Local random quantum circuits are approximate polynomial-designs. *Communications in Mathematical Physics*, 346:397–434, 2016.
- Matthew Brennan and Guy Bresler. Reducibility and statistical-computational gaps from secret leakage. In *Conference on Learning Theory*, pages 648–847. PMLR, 2020.
- Matthew Brennan, Guy Bresler, and Wasim Huleihel. Reducibility and computational lower bounds for problems with planted sparse structure. In *Conference On Learning Theory*, pages 48–166. PMLR, 2018.
- Matthew S Brennan, Guy Bresler, Sam Hopkins, Jerry Li, and Tselil Schramm. Statistical query algorithms and low degree tests are almost equivalent. In Mikhail Belkin and Samory Kpotufe, editors, *Proceedings of Thirty Fourth Conference on Learning Theory*, volume 134 of *Proceedings of Machine Learning Research*, pages 774–774. PMLR, 15–19 Aug 2021. URL <https://proceedings.mlr.press/v134/brennan21a.html>.

- Guy Bresler and Tianze Jiang. Detection-recovery and detection-refutation gaps via reductions from planted clique. In *The Thirty Sixth Annual Conference on Learning Theory*, pages 5850–5889. PMLR, 2023.
- Costin Buadescu and Ryan O’Donnell. Improved quantum data analysis. In *Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing*, pages 1398–1411, 2021.
- Sebastien Bubeck, Sitan Chen, and Jerry Li. Entanglement is necessary for optimal quantum property testing. In *2020 IEEE 61st Annual Symposium on Foundations of Computer Science (FOCS)*, pages 692–703. IEEE, 2020.
- Rares-Darius Buhai, Jun-Ting Hsieh, Aayush Jain, and Pravesh K. Kothari. The quasi-polynomial low-degree conjecture is false, 2025. URL <https://arxiv.org/abs/2505.17360>.
- Siu-On Chan, Ilias Diakonikolas, Paul Valiant, and Gregory Valiant. Optimal algorithms for testing closeness of discrete distributions. In *Proceedings of the twenty-fifth annual ACM-SIAM symposium on Discrete algorithms*, pages 1193–1203. SIAM, 2014.
- Chi-Fang Chen, Alexander M Dalzell, Mario Berta, Fernando GSL Brandão, and Joel A Tropp. Sparse random Hamiltonians are quantumly easy. *Physical Review X*, 14(1):011014, 2024a.
- Chi-Fang Chen, Jeongwan Haah, Jonas Haferkamp, Yunchao Liu, Tony Metger, and Xinyu Tan. Incompressibility and spectral gaps of random circuits. *arXiv:2406.07478*, 2024b.
- Senrui Chen, Changhun Oh, Sisi Zhou, Hsin-Yuan Huang, and Liang Jiang. Tight bounds on pauli channel learning without entanglement. *Physical Review Letters*, 132(18):180805, 2024c.
- Sitan Chen and Weiyan Gong. Efficient Pauli channel estimation with logarithmic quantum memory. *arXiv:2309.14326*, 2023.
- Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. Exponential separations between learning with and without quantum memory. In *2021 IEEE 62nd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 574–585. IEEE, 2022a. URL <https://ieeexplore.ieee.org/abstract/document/9719827/>.
- Sitan Chen, Aravind Gollakota, Adam Klivans, and Raghu Meka. Hardness of noise-free learning for two-hidden-layer neural networks. *Advances in Neural Information Processing Systems*, 35: 10709–10724, 2022b.
- Sitan Chen, Jerry Li, Brice Huang, and Allen Liu. Tight bounds for quantum state certification with incoherent measurements. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1205–1213. IEEE, 2022c.
- Sitan Chen, Jerry Li, and Yuanzhi Li. Learning (very) simple generative models is hard. *Advances in Neural Information Processing Systems*, 35:35143–35155, 2022d.
- Sitan Chen, Jordan Cotler, Hsin-Yuan Huang, and Jerry Li. The complexity of NISQ. *Nature Communications*, 14(1):6001, 2023a.

- Sitan Chen, Brice Huang, Jerry Li, Allen Liu, and Mark Sellke. When does adaptivity help for quantum state learning? In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*. IEEE, 2023b.
- Sitan Chen, Weiyuan Gong, and Qi Ye. Optimal tradeoffs for estimating pauli observables. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1086–1105, 2024d. doi: 10.1109/FOCS61266.2024.00072.
- Sitan Chen, Weiyuan Gong, Qi Ye, and Zhihan Zhang. Stabilizer bootstrapping: A recipe for efficient agnostic tomography and magic estimation. *arXiv:2408.06967*, 2024e.
- Sitan Chen, Jerry Li, and Allen Liu. Optimal high-precision shadow estimation. *arXiv:2407.13874*, 2024f.
- Sitan Chen, Jerry Li, and Allen Liu. An optimal tradeoff between entanglement and copy complexity for state tomography. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1331–1342, 2024g.
- Amin Coja-Oghlan, Oliver Gebhard, Max Hahn-Klimroth, Alexander S Wein, and Ilias Zadik. Statistical and computational phase transitions in group testing. In *Conference on Learning Theory*, pages 4764–4781. PMLR, 2022.
- 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, 2010.
- Abhishek Dhawan, Cheng Mao, and Alexander S Wein. Detection of dense subhypergraphs by low-degree polynomials. *Random Structures & Algorithms*, 66(1):e21279, 2025.
- Ilias Diakonikolas and Daniel Kane. Near-optimal statistical query hardness of learning halfspaces with massart noise. In *Conference on Learning Theory*, pages 4258–4282. PMLR, 2022.
- Ilias Diakonikolas, Daniel M Kane, and Alistair Stewart. Statistical query lower bounds for robust estimation of high-dimensional gaussians and gaussian mixtures. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 73–84. IEEE, 2017.
- Ilias Diakonikolas, Daniel M Kane, Vasilis Kontonis, and Nikos Zarifis. Algorithms and sq lower bounds for pac learning one-hidden-layer relu networks. In *Conference on Learning Theory*, pages 1514–1539. PMLR, 2020.
- Yunzi Ding, Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Subexponential-time algorithms for sparse PCA. *Foundations of Computational Mathematics*, 24(3):865–914, 2024.
- Bill Fefferman, Soumik Ghosh, Makrand Sinha, and Henry Yuen. The hardness of learning quantum circuits and its cryptographic applications. 2024. URL https://www.henryyuen.net/papers/random_circuits_crypto.pdf.
- Vitaly Feldman, Will Perkins, and Santosh Vempala. On the complexity of random satisfiability problems with planted solutions. In *Proceedings of the Forty-seventh Annual ACM Symposium on Theory of Computing*, pages 77–86, 2015.

- 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.
- Surbhi Goel, Aravind Gollakota, Zhihan Jin, Sushrut Karmalkar, and Adam Klivans. Superpolynomial lower bounds for learning one-layer neural networks using gradient descent. In *International Conference on Machine Learning*, pages 3587–3596. PMLR, 2020a.
- Surbhi Goel, Aravind Gollakota, and Adam Klivans. Statistical-query lower bounds via functional gradients. *Advances in Neural Information Processing Systems*, 33:2147–2158, 2020b.
- Weiyuan Gong, Jonas Haferkamp, Qi Ye, and Zhihan Zhang. On the sample complexity of purity and inner product estimation. *arXiv:2410.12712*, 2024.
- Dardo Goyeneche, Daniel Alsina, José I. Latorre, Arnau Riera, and Karol Życzkowski. Absolutely maximally entangled states, combinatorial designs, and multiunitary matrices. *Phys. Rev. A*, 92: 032316, Sep 2015. doi: 10.1103/PhysRevA.92.032316. URL <https://link.aps.org/doi/10.1103/PhysRevA.92.032316>.
- Dardo Goyeneche and Karol Życzkowski. Genuinely multipartite entangled states and orthogonal arrays. *Physical Review A*, 90(2), August 2014. ISSN 1094-1622. doi: 10.1103/physreva.90.022316. URL <http://dx.doi.org/10.1103/PhysRevA.90.022316>.
- Sabee Grewal, Vishnu Iyer, William Kretschmer, and Daniel Liang. Agnostic tomography of stabilizer product states. *arXiv:2404.03813*, 2024a.
- Sabee Grewal, Vishnu Iyer, William Kretschmer, and Daniel Liang. Improved stabilizer estimation via Bell difference sampling. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1352–1363, 2024b.
- Andi Gu, Lorenzo Leone, Soumik Ghosh, Jens Eisert, Susanne F. Yelin, and Yihui Quek. Pseudomagic quantum states. *Physical Review Letters*, 132(21), May 2024a. ISSN 1079-7114. doi: 10.1103/physrevlett.132.210602. URL <http://dx.doi.org/10.1103/PhysRevLett.132.210602>.
- Andi Gu, Yihui Quek, Susanne Yelin, Jens Eisert, and Lorenzo Leone. Simulating quantum chaos without chaos, 2024b. URL <https://arxiv.org/abs/2410.18196>.
- Jeongwan Haah, Aram W Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. In *Proceedings of the Forty-eighth Annual ACM Symposium on Theory of Computing*, pages 913–925, 2016.
- Jeongwan Haah, Robin Kothari, and Ewin Tang. Optimal learning of quantum Hamiltonians from high-temperature Gibbs states. In *2022 IEEE 63rd Annual Symposium on Foundations of Computer Science (FOCS)*, pages 135–146. IEEE, 2022.
- J. Haferkamp. Random quantum circuits are approximate unitary t -designs in depth $O(nt^{5+o(1)})$. *Quantum*, 6:795, 2022.

- Bruce Hajek, Yihong Wu, and Jiaming Xu. Computational lower bounds for community detection on random graphs. In *Conference on Learning Theory*, pages 899–928. PMLR, 2015.
- Aram W Harrow. The church of the symmetric subspace. *arXiv:1308.6595*, 2013.
- Matthew B Hastings. Classical and quantum algorithms for tensor principal component analysis. *Quantum*, 4:237, 2020.
- Patrick Hayden, Debbie W Leung, and Andreas Winter. Aspects of generic entanglement. *Communications in mathematical physics*, 265:95–117, 2006.
- 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, 2021. URL <https://arxiv.org/abs/2110.05517>.
- Marcel Hinsche, Marios Ioannou, Alexander Nietner, Jonas Haferkamp, Yihui Quek, Dominik Hangleiter, J-P Seifert, Jens Eisert, and Ryan Sweke. One t gate makes distribution learning hard. *Physical Review Letters*, 130(24):240602, 2023.
- Justin Holmgren and Alexander S Wein. Counterexamples to the low-degree conjecture. *arXiv:2004.08454*, 2020.
- Samuel Hopkins. *Statistical inference and the sum of squares method*. PhD thesis, Cornell University, 2018.
- Samuel B Hopkins and David Steurer. Efficient bayesian estimation from few samples: community detection and related problems. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 379–390. IEEE, 2017.
- Samuel B Hopkins, Jonathan Shi, and David Steurer. Tensor principal component analysis via sum-of-square proofs. In *Conference on Learning Theory*, pages 956–1006. PMLR, 2015.
- Samuel B Hopkins, Pravesh K Kothari, Aaron Potechin, Prasad Raghavendra, Tselil Schramm, and David Steurer. The power of sum-of-squares for detecting hidden structures. In *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 720–731. IEEE, 2017.
- Hong-Ye Hu, Muzhou Ma, Weiyuan Gong, Qi Ye, Yu Tong, Steven T Flammia, and Susanne F Yelin. Ansatz-free hamiltonian learning with heisenberg-limited scaling. *arXiv preprint arXiv:2502.11900*, 2025.
- Hsin-Yuan Huang, Richard Kueng, and John Preskill. Predicting many properties of a quantum system from very few measurements. *Nature Physics*, 16(10):1050–1057, October 2020. ISSN 1745-2473, 1745-2481. doi: 10.1038/s41567-020-0932-7.
- Hsin-Yuan Huang, Richard Kueng, and John Preskill. Information-theoretic bounds on quantum advantage in machine learning. *Physical Review Letters*, 126(19):190505, 2021. URL <https://journals.aps.org/prl/abstract/10.1103/PhysRevLett.126.190505>.
- Hsin-Yuan Huang, Yunchao Liu, Michael Broughton, Isaac Kim, Anurag Anshu, Zeph Landau, and Jarrod R McClean. Learning shallow quantum circuits. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1343–1351, 2024.

- Matteo Ippoliti and Vedika Khemani. Learnability transitions in monitored quantum dynamics via eavesdropper’s classical shadows. *PRX Quantum*, 5(2), April 2024. ISSN 2691-3399. doi: 10.1103/prxquantum.5.020304. URL <http://dx.doi.org/10.1103/PRXQuantum.5.020304>.
- Matteo Ippoliti, Michael J. Gullans, Sarang Gopalakrishnan, David A. Huse, and Vedika Khemani. Entanglement phase transitions in measurement-only dynamics. *Phys. Rev. X*, 11:011030, Feb 2021. doi: 10.1103/PhysRevX.11.011030. URL <https://link.aps.org/doi/10.1103/PhysRevX.11.011030>.
- Zhengfeng Ji, Yi-Kai Liu, and Fang Song. Pseudorandom quantum states. In *Advances in Cryptology—CRYPTO 2018: 38th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 19–23, 2018, Proceedings, Part III* 38, pages 126–152. Springer, 2018.
- Michael Kearns. Efficient noise-tolerant learning from statistical queries. *Journal of the ACM (JACM)*, 45(6):983–1006, 1998.
- Hyun-Soo Kim, Isaac H Kim, and Daniel Ranard. Learning state preparation circuits for quantum phases of matter. *arXiv:2410.23544*, 2024.
- Robbie King, David Gosset, Robin Kothari, and Ryan Babbush. Triply efficient shadow tomography. *arXiv:2404.19211*, 2024.
- Dmitriy Kunisky, Alexander S Wein, and Afonso S Bandeira. Notes on computational hardness of hypothesis testing: Predictions using the low-degree likelihood ratio. In *ISAAC Congress (International Society for Analysis, its Applications and Computation)*, pages 1–50. Springer, 2019.
- Zeph Landau and Yunchao Liu. Learning quantum states prepared by shallow circuits in polynomial time. *arXiv:2410.23618*, 2024.
- Thibault Lesieur, Léo Miolane, Marc Lelarge, Florent Krzakala, and Lenka Zdeborová. Statistical and computational phase transitions in spiked tensor estimation. In *2017 IEEE International Symposium on Information Theory (ISIT)*, pages 511–515. IEEE, 2017.
- Siqi Liu, Sidhanth Mohanty, Tselil Schramm, and Elizabeth Yang. Testing thresholds for high-dimensional sparse random geometric graphs. In *Proceedings of the 54th Annual ACM SIGACT Symposium on Theory of Computing*, pages 672–677, 2022.
- Fermi Ma and Hsin-Yuan Huang. How to construct random unitaries. *arXiv preprint arXiv:2410.10116*, 2024.
- Muzhou Ma, Steven T Flammia, John Preskill, and Yu Tong. Learning k -body hamiltonians via compressed sensing. *arXiv preprint arXiv:2410.18928*, 2024.
- Zongming Ma and Yihong Wu. Computational barriers in minimax submatrix detection. 2015.
- Jay Mardia, Kabir Aladin Verchand, and Alexander S Wein. Low-degree phase transitions for detecting a planted clique in sublinear time. In *The Thirty Seventh Annual Conference on Learning Theory*, pages 3798–3822. PMLR, 2024.

- Andrea Montanari and Emile Richard. A statistical model for tensor pca. *Advances in neural information processing systems*, 27, 2014.
- Ashley Montanaro. Learning stabilizer states by Bell sampling. *arXiv:1707.04012*, 2017.
- MA Naimark. About second-kind self-adjoint extensions of symmetrical operator. *Izv. Akad. Nauk USSR, Ser. Mat.*, 4:53–104, 1940.
- Shyam Narayanan. Improved algorithms for learning quantum Hamiltonians, via flat polynomials. *arXiv:2407.04540*, 2024.
- Alexander Nietner, Marios Ioannou, Ryan Sweke, Richard Kueng, Jens Eisert, Marcel Hinsche, and Jonas Haferkamp. On the average-case complexity of learning output distributions of quantum circuits. *arXiv:2305.05765*, 2023.
- 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.
- 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.
- Ketan N. Patel, Igor L. Markov, and John P. Hayes. Optimal synthesis of linear reversible circuits. *Quantum Info. Comput.*, 8(3):282–294, March 2008. ISSN 1533-7146.
- Alexander Poremba, Yihui Quek, and Peter Shor. The learning stabilizers with noise problem. *arXiv:2410.18953*, 2024.
- Aaron Potechin and Goutham Rajendran. Sub-exponential time sum-of-squares lower bounds for principal components analysis. *Advances in Neural Information Processing Systems*, 35:35724–35740, 2022.
- Yihui Quek, Daniel Stilck França, Sumeet Khatri, Johannes Jakob Meyer, and Jens Eisert. Exponentially tighter bounds on limitations of quantum error mitigation. *Nature Physics*, 20(10):1648–1658, 2024.
- Zahra Raissi, Adam Teixedó, Christian Gogolin, and Antonio Acín. Constructions of $\mathbb{Z}_2$ -uniform and absolutely maximally entangled states beyond maximum distance codes. *Physical Review Research*, 2(3), September 2020. ISSN 2643-1564. doi: 10.1103/physrevresearch.2.033411. URL <http://dx.doi.org/10.1103/PhysRevResearch.2.033411>.
- Cynthia Rush, Fiona Skerman, Alexander S Wein, and Dana Yang. Is it easier to count communities than find them? In *14th Innovations in Theoretical Computer Science Conference (ITCS 2023)*, pages 94–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2023.
- Alexander Schmidhuber, Ryan O’Donnell, Robin Kothari, and Ryan Babbush. Quartic quantum speedups for planted inference. In *Proceedings of the 2025 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 905–913. SIAM, 2025.
- Thomas Schuster, Jonas Haferkamp, and Hsin-Yuan Huang. Random unitaries in extremely low depth. *arXiv:2407.07754*, 2024.

- Vivek V Shende, Stephen S Bullock, and Igor L Markov. Synthesis of quantum logic circuits. In *Proceedings of the 2005 Asia and South Pacific Design Automation Conference*, pages 272–275, 2005.
- B. Skinner, J. Ruhman, and A. Nahum. Measurement-induced phase transitions in the dynamics of entanglement. *Phys. Rev. X*, 9:031009, Jul 2019. doi: 10.1103/PhysRevX.9.031009. URL <https://link.aps.org/doi/10.1103/PhysRevX.9.031009>.
- Ryota Tomioka and Taiji Suzuki. Spectral norm of random tensors. *arXiv:1407.1870*, 2014.
- Joel A Tropp. Second-order matrix concentration inequalities. *Applied and Computational Harmonic Analysis*, 44(3):700–736, 2018.
- Joel A Tropp et al. An introduction to matrix concentration inequalities. *Foundations and Trends® in Machine Learning*, 8(1-2):1–230, 2015.
- Francisca Vasconcelos and Hsin-Yuan Huang. Learning shallow quantum circuits with many-qubit gates. *arXiv:2410.16693*, 2024.
- Alexander S. Wein. Computational complexity of statistics: New insights from low-degree polynomials, June 2025. URL <https://arxiv.org/abs/2506.10748>.
- Alexander S Wein, Ahmed El Alaoui, and Cristopher Moore. The kikuchi hierarchy and tensor pca. In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1446–1468. IEEE, 2019.
- Eugene P Wigner. On the distribution of the roots of certain symmetric matrices. *Annals of Mathematics*, 67(2):325–327, 1958.
- Ales Wodecki and Jakub Marecek. Learning quantum Hamiltonians at any temperature in polynomial time with Chebyshev and bit complexity. *arXiv:2402.05552*, 2024.
- Yu Yu and Jiang Zhang. Smoothing out binary linear codes and worst-case sub-exponential hardness for lpn. In *Advances in Cryptology—CRYPTO 2021: 41st Annual International Cryptology Conference, CRYPTO 2021, Virtual Event, August 16–20, 2021, Proceedings, Part III 41*, pages 473–501. Springer, 2021.
- Haimeng Zhao, Laura Lewis, Ishaan Kannan, Yihui Quek, Hsin-Yuan Huang, and Matthias C Caro. Learning quantum states and unitaries of bounded gate complexity. *PRX Quantum*, 5(4):040306, 2024.
- Huangjun Zhu. Multiqubit Clifford groups are unitary 3-designs. *Physical Review A*, 96(6):062336, 2017.

Roadmap

In Section A, we review basic notions from quantum information and from the classical low-degree method. In Section B, we discuss relevant prior work and situate our results within the broader literature on the complexity of quantum and classical learning. In Section C, we propose potential future directions of this work.

In Section D, we present our first set of results connecting state 2-designs to low-degree hardness for non-adaptive single-copy measurement protocols. In Section E, we prove our results on adaptive measurements. In Section F, we apply the theory from the preceding sections to derive new information-computation gaps for quantum error mitigation, learning Gibbs states, and random quantum circuits. In Section G, we present the quantum planted biclique problem and prove upper and lower bounds for this problem. In Section H, we prove our results on low-degree hardness for problems with hybrid classical-quantum or fully classical input, namely for Learning Stabilizers with Noise and for agnostically learning product states.

Appendix A. Preliminaries

Here, we collect the basic concepts and technical tools used in quantum information and classical low-degree likelihood ratio throughout this paper. The readers can skip these background results according to their knowledge in these fields.

Notation. We use $\|A\|$ and $\|A\|_1$ to represent the operator norm and trace norm of a matrix A , and use $\|v\|_p$ to represent the L_p norm of a vector v . We use I_n to denote the $2^n \times 2^n$ identity matrix, and omit the subscript n when it is clear from context. We use $\tilde{O}$, $\tilde{\Omega}$, and $\tilde{\Theta}$ to hide poly-logarithmic dependences in big-O notation. Given two distributions p, q on a discrete domain X , we denote by $d_{\text{TV}}(p, q) := \frac{1}{2} \sum_{x \in X} |p(x) - q(x)|$ the total variation distance between them. We will abbreviate the set $\{1, 2, \dots, c\}$ as $[c]$.

Given a distribution D and an integer m , we denote by $D^{\otimes m}$ the joint distribution over m independent samples from D . Given $f, g \in L^2(D)$, we use $\langle \cdot, \cdot \rangle_D$ and $\|\cdot\|_D$ to denote the inner product and induced norm for $L^2(D)$ and drop the subscript D when it is clear from context. We will occasionally conflate probability distributions with their laws and density functions.

A.1. Quantum information basics

Quantum states. The *density matrix* of a general n -qubit mixed state is given by a psd matrix $\rho \in \mathbb{C}^{2^n \times 2^n}$ with $\text{tr}(\rho) = 1$. We also use $d = 2^n$ to denote its dimension. ρ is a *pure state* if it can be represented as a rank-1 matrix $|\psi\rangle\langle\psi|$ for some unit vector $|\psi\rangle \in \mathbb{C}^{2^n}$, in which case $\text{tr}(\rho^2) = 1$ and we sometimes refer to the state as $|\psi\rangle$. Given an n -qubit quantum state ρ and a subset $S \subseteq [n]$, the quantum state on the subsystem supported by qubits in S (also known as the *reduced density matrix*) is denoted by the *partial trace* $\text{tr}_{[n] \setminus S}(\rho)$. Given two states $\rho, \sigma \in \mathbb{C}^{2^n \times 2^n}$, their *trace distance* is $d_{\text{tr}}(\rho, \sigma) = \frac{1}{2} \|\rho - \sigma\|_1$.

Quantum measurements. Quantum measurements can in general be represented as *positive operator-valued measures* (POVMs). An n -qubit POVM is defined as a set of psd operators $\{F_s\}_s$ with $\sum_s F_s = I$, where each F_s is called a *POVM element* corresponding to the *measurement outcome* s . The probability of obtaining outcome s when we measure a given state ρ using $\{F_s\}_s$ is given by $\text{tr}(F_s \rho)$.

In this work, we often consider the important special case of *projector-valued measures* (PVMs). These are POVMs whose elements are projection matrices. The set of all PVMs captures all measurements that can be performed on n -qubit states without introducing additional ancillary qubits. By Naimark's dilation theorem (Naimark, 1940), an arbitrary 2^m -outcome POVM can be realized by a PVM on an enlarged quantum system by introducing m ancillary qubits.

Given a PVM $\{\Pi_s\}_s$, we can also associate the projectors $\{\Pi_s\}_s$ with an Hermitian observable $M = \sum_s \Pi_s$, so that Π_s is the projector to the eigenspace of M with eigenvalue m . We diagonalize $M = U^\dagger D U$ by a unitary matrix U (not unique). Without loss of generality, we can assume any PVM can be performed by applying a unitary U followed by the PVM given by projectors to the *computational basis* (i.e. standard basis) vectors $\{|x\rangle : x \in \{0, 1\}^n\}$.

We say that PVM $\{\Pi_s\}_s$ can be prepared by a quantum circuit of depth L if there exists such a U that can be prepared by a geometrically-local quantum circuit containing only single- and two-qubit gates on neighboring qubits of depth L .

Haar random unitaries and states. The *Haar measure* μ_{Haar} on the unitary group $U(d)$ is the unique probability measure that is invariant under left- and right-multiplication by any $U \in U(d)$. The *Haar measure on d -dimensional states* is the unique rotation invariant measure on states given by $|\psi\rangle = U|0\rangle$ with $U \sim \mu_{\text{Haar}}$ and $|0\rangle \in \mathbb{S}_{\mathbb{C}}^{d-1}$ the all-zeros state. We will often abuse notation and write $\psi \sim \mu_{\text{Haar}}$. Moments of the Haar measure can be computed with the following folklore formula (see e.g. Harrow (2013)):

$$\mathbb{E}_{\psi \sim \mu_{\text{Haar}}} [(|\psi\rangle\langle\psi|)^{\otimes k}] = \frac{S_k}{\binom{d+k-1}{k}} = \frac{1}{d \cdots (d+k-1)} \sum_{\pi \in S_k} \pi^d,$$

where S_k denotes the projector onto the symmetric subspace $\text{Sym}^k(\mathbb{C}^d)$, S_k is the symmetric group on k elements, and π^d acts on $(\mathbb{C}^d)^{\otimes k}$ via

$$\pi^d |i_1, \dots, i_k\rangle = |i_{\pi^{-1}(1)}, \dots, i_{\pi^{-1}(k)}\rangle.$$

Based on this formula, we can also deduce the following formula for any $|\varphi\rangle$:

$$\mathbb{E}_{\psi \sim \mu_{\text{Haar}}} [|\langle\psi|\varphi\rangle|^{\otimes 2k}] = \mathbb{E}_{\psi \sim \mu_{\text{Haar}}} \left[\text{tr} \left[(|\psi\rangle\langle\psi|)^{\otimes k} (|\varphi\rangle\langle\varphi|)^{\otimes k} \right] \right] = \binom{d-k-1}{k}^{-1}.$$

We will also need the following fact:

Fact 11 (Lemma 25 of Nietner et al. (2023)) *Let $|i_1\rangle, \dots, |i_t\rangle$ with $i_1, \dots, i_t \in [d]$ be mutually orthogonal state vectors and $\lambda = (\lambda_1, \dots, \lambda_t)$ a partition of k for $k \leq d$. We have*

$$\mathbb{E}_{\psi \sim \mu_{\text{Haar}}} \left[\prod_{l=1}^t |\langle\psi|i_l\rangle|^{2\lambda_l} \right] = \frac{\prod_{l=1}^t \lambda_l!}{d \cdots (d+k-1)}.$$

Unitary and state designs. A *unitary (resp. state) k -design* is a probability distribution ν over $U \in U(d)$ (resp. over states $|\psi\rangle \in \mathbb{C}^d$) such that the k -th moments are the same as k -th moments of the Haar measure μ_{Haar} .

Definition 12 (Unitary (resp. state) design) A probability distributions ν over unitary $U \in U(d)$ (pure states in $\mathbb{C}^d$) is called a unitary (resp. state) k -design if the following inequalities are satisfied respectively:

$$\begin{aligned}\mathbb{E}_{U \sim \mu} f(U) &= \mathbb{E}_{U \sim \mu_{\text{Haar}}} f(U), \\ \mathbb{E}_{\psi \sim \mu} (|\psi\rangle\langle\psi|)^{\otimes k} &= \frac{\mathcal{S}_k}{\binom{d+k-1}{k}},\end{aligned}$$

for any polynomial f of degree at most k .

By definition, k -designs are related to low-degree hardness: If all expectation values of degree k -polynomials look Haar, then the ensemble is k -degree hard to distinguish from Haar. We will also need the more subtle concept of *approximate* designs:

Definition 13 (Approximate unitary (state) design) A probability distribution ν over unitaries $U \in U(d)$ (resp. pure states in $\mathbb{C}^d$) is called an ε -approximate unitary (resp. state) k -design if the following inequalities are satisfied respectively:

$$\begin{aligned}\max_{A \in \mathbb{C}^{d^2 \times d^2} : \|A\|_1 \leq 1} \left\| \{(\Phi^{(k)}(\nu) - \Phi^{(k)}(\mu_{\text{Haar}})) \otimes \mathcal{I}_n\}(A) \right\| &\leq \frac{\varepsilon}{d^k}, \\ (1 - \varepsilon) \mathbb{E}_{\psi \sim \mu} (|\psi\rangle\langle\psi|)^{\otimes k} &\leq \frac{\mathcal{S}_k}{\binom{d+k-1}{k}} \leq (1 + \varepsilon) \mathbb{E}_{\psi \sim \mu} (|\psi\rangle\langle\psi|)^{\otimes k},\end{aligned}$$

where $\mathcal{I}_n$ is the identity transformation on a $d = 2^n$ -dimensional ancillary system and

$$\Phi^{(k)}(\mu)(B) = \int U^{\otimes k} B U^{\dagger \otimes k} d\mu(U)$$

for any $B \in \mathbb{C}^{d \times d}$.

A.2. Low-degree likelihood ratio

In this work we focus on *hypothesis testing* problems. In preparation for discussing the particulars of the quantum setting we will consider in Section D, here we review existing frameworks for establishing statistical and computational lower bounds for such problems in the *classical setting*.

In a *many-versus-one distinguishing problem* over $\mathbb{R}^n$, the goal is to distinguish between the following two cases, given samples $x_1, \dots, x_m \in \mathbb{R}^n$:

- H_0 (null case): the samples were drawn i.i.d. from a *null distribution* $D_\emptyset$ over $\mathbb{R}^n$.
- H_1 (alternative case): the samples were drawn i.i.d. from some distribution D_u coming from a family $\mathcal{S} = \{D_u\}_{u \sim \mu}$, with u sampled at the outset from prior distribution μ .

We will assume $m = \text{poly}(n)$ throughout this paper as we only consider sample-efficient algorithms. Denote all m samples as $\mathbf{x} = (x_1, \dots, x_m) \in \mathbb{R}^{m \times n}$.

Definition 14 A distinguisher $f : (\mathbb{R}^n)^m \rightarrow \{0, 1\}$ is said to achieve η -detection if

$$|Pr_{D_\emptyset^{\otimes m}}[f(\mathbf{x}) = 1] - Pr_{u \sim \mu} Pr_{D_\mu^{\otimes m}}[f(\mathbf{x}) = 1]| > \eta.$$

If $\eta = \Omega(1)$ (resp. $\eta = 1 - o_n(1)$), we refer to this as weak (resp. strong) detection.

Given a hypothesis testing problem, the hardness of achieving weak detection for a class of distinguishers indicates that all such distinguishers fail to distinguish (and thus learn) with any $> 50\%$ constant probability of success.

Information-theoretic bounds. The central object for understanding how large m must be before detection is possible is the *likelihood ratio*. Throughout this work, we will assume that D_u is absolutely continuous relative to $D_\emptyset$ for all u . We can then define the single- and m -sample likelihood ratios by

$$\overline{D}_u(x) = D_u(x)/D_\emptyset(x) \quad \text{and} \quad \overline{D}_u^{\otimes m}(\mathbf{x}) = D_u^{\otimes m}(\mathbf{x})/D_\emptyset^{\otimes m}(\mathbf{x}).$$

The classic Neyman-Pearson lemma implies the following:

Fact 15 *The likelihood ratio test $f(\mathbf{x}) \triangleq \mathbb{1}[\mathbb{E}_u \overline{D}_u^{\otimes m}(\mathbf{x}) > 1]$ achieves η -detection if and only if $d_{TV}(\mathbb{E}_u D_u^{\otimes m}(\mathbf{x}), D_\emptyset^{\otimes m}(\mathbf{x})) > \eta$.*

To motivate the low-degree likelihood ratio, we can consider the following “ L_2 ” formulation of the likelihood ratio test. Instead of considering L^∞ -bounded distinguishers $f : (\mathbb{R}^n)^m \rightarrow \{0, 1\}$, let us consider L^2 -bounded distinguishers $p : (\mathbb{R}^n)^m \rightarrow \mathbb{R}$, i.e. test functions for which $\|p(\mathbf{x})\|_{D_\emptyset^{\otimes m}} \leq 1$. One can readily verify that the choice of p which maximizes the *distinguishing advantage*, i.e. which solves

$$\sup_{\|p\|_{D_\emptyset^{\otimes m}}=1} \left| \mathbb{E}_{D_\emptyset^{\otimes m}}[p(\mathbf{x})] - \mathbb{E}_{u \sim \mu} \mathbb{E}_{D_u^{\otimes m}}[p(\mathbf{x})] \right| \quad (1)$$

is given by the shifted and normalized likelihood ratio:

$$p^*(\mathbf{x}) = \frac{\mathbb{E}_u \overline{D}_u^{\otimes m}(\mathbf{x}) - 1}{\left\| \mathbb{E}_u \overline{D}_u^{\otimes m} - \mathbf{1} \right\|_{D_\emptyset}}$$

where $\mathbf{1}$ denotes the constant 1 function. This achieves distinguishing advantage $\left\| \mathbb{E}_u \overline{D}_u^{\otimes m} - \mathbf{1} \right\|_{D_\emptyset}$; in this work, this quantity will always be finite as we only consider distributions $D_u, D_\emptyset$ over discrete domains.

By Jensen’s inequality, this optimal distinguishing advantage upper bounds $d_{TV}(\mathbb{E}_u D_u^{\otimes m}, D_\emptyset^{\otimes m})$, so by Fact 15, we obtain the following condition for information-theoretic hardness of detection:

Lemma 16 *If the optimal distinguishing advantage in Eq. (1) is $o_n(1)$, then weak detection is not achievable by any distinguisher.*

Computational bounds. The low-degree likelihood framework offers a computational analogue of the above reasoning. Instead of considering arbitrary L^2 -bounded distinguishers, we restrict our attention to distinguishers $p : (\mathbb{R}^n)^m \rightarrow \mathbb{R}$ which are given by polynomials of some degree k . Note that any such polynomial can be evaluated in time $(nm)^{O(k)}$.

The motivation for this first arose in the context of establishing sum-of-squares lower bounds for statistical estimation (Barak et al., 2019; Hopkins and Steurer, 2017; Hopkins et al., 2017), and more generally it is supported by the fact that many of the most powerful computationally efficient algorithms for hypothesis testing, e.g. spectral methods, message passing, and method of moments, can be implemented as low-degree polynomial estimators (see the survey (Kunisky et al., 2019) for an overview).

Let $(\chi_\alpha)_{|\alpha| \leq k}$ denote an orthonormal basis for the space $\mathcal{V}^{\leq k}$ of polynomials of degree at most k in $L^2(D_{\otimes}^{\otimes m})$. Given $f \in L^2(D_{\otimes}^{\otimes m})$, we denote by $f^{\leq k}$ the projection of f to this subspace. We will focus almost exclusively (with the exception of Section G) on $D_{\otimes}$ which is uniform over the Boolean cube, in which case $(\chi_\alpha)_{|\alpha| \leq k}$ are simply the Fourier characters of degree at most k and $f^{\leq k}$ is simply the degree- $\leq k$ Fourier truncation of f . We can now consider the following modification to the optimization problem in Eq. (1):

$$\sup_{p \in \mathcal{V}^{\leq k} : \|p\|_{D_{\otimes}^{\otimes m}} = 1} \left| \mathbb{E}_{D_{\otimes}^{\otimes m}}[p(\mathbf{x})] - \mathbb{E}_{u \sim \mu} \mathbb{E}_{D_{\mu}^{\otimes m}}[p(\mathbf{x})] \right|. \quad (2)$$

The optimal p is given, up to normalization and shifting, by the likelihood ratio *truncated to degree at most k* :

$$p^*(\mathbf{x}) = \frac{\mathbb{E}_u(\overline{D}_u^{\otimes m})^{\leq k}(\mathbf{x}) - 1}{\left\| \mathbb{E}_u(\overline{D}_u^{\otimes m})^{\leq k} - \mathbf{1} \right\|_{D_{\otimes}}}. \quad (3)$$

We refer to $\mathbb{E}_u(\overline{D}_u^{\otimes m})^{\leq k}$ as the *(m -sample) low-degree likelihood ratio*.

p^* achieves distinguishing advantage $\left\| \mathbb{E}_u(\overline{D}_u^{\otimes m})^{\leq k} - \mathbf{1} \right\|_{D_{\otimes}}$. In particular, if the likelihood ratio has the Fourier expansion $\mathbb{E}_u \overline{D}_u^{\otimes m}(\mathbf{x}) = \sum_{\alpha} \widehat{D}[\alpha] \chi_{\alpha}(\mathbf{x})$, then the optimal distinguishing advantage achieved by any degree- $\leq k$ distinguisher is given by $\chi_{\leq k}$ for

$$\chi_{\leq k}^2 \triangleq \sum_{0 < |\alpha| \leq k} \widehat{D}[\alpha]^2 \quad \text{for} \quad \widehat{D}[\alpha] \triangleq \mathbb{E}_{D_{\otimes}^{\otimes m}}[\mathbb{E}_u \overline{D}_u^{\otimes m}(\mathbf{x}) \chi_{\alpha}(\mathbf{x})].$$

We will refer to $\chi_{\leq k}$ as the *degree- k advantage*.

A long line of work in the classical literature on complexity of statistical inference has established bounds on this quantity for a wide range of tasks. Barring hypothesis testing problems with exceptional algebraic structure, the recurring finding is that boundedness of the low-degree likelihood ratio appears to coincide with the failure of computationally efficient distinguishers, giving rise to the following informal conjecture (see Conjecture 2.2.4 in Hopkins (2018) for a formalization).

Conjecture 17 (Low-degree conjecture, informal) *Let $k = k_n$ be some growing function of n . For “natural” distinguishing problems, there is no $(nm)^{\tilde{O}(k)}$ -time distinguisher achieving weak detection between $\mathbb{E}_u D_{\otimes}^{\otimes m}$ and $D_{\otimes}^{\otimes m}$ if $\left\| \mathbb{E}_u(\overline{D}_u^{\otimes m})^{\leq k} - \mathbf{1} \right\|_{D_{\otimes}} = o(1)$.*

In particular, if $m = \text{poly}(n)$ and $k = \omega(\log n)$, then boundedness of the low-degree likelihood ratio coincides with hardness for all $\text{poly}(n)$ -time distinguishers.

Sample-wise degree. In some parts of this work, we will consider an alternative notion of degree that allows for a wider range of efficient distinguishers.

We say that a polynomial $p : (\mathbb{R}^n)^m \rightarrow \mathbb{R}$ has *sample-wise degree* $(\mathcal{D}, k)$ if $p(x_1, \dots, x_m)$ is a linear combination of functions which each have degree at most $\mathcal{D}$ in each x_i and nonzero degree in at most k of the x_i ’s (the set of k indices can vary across functions in the linear combination). Note that any such p must have total degree at most $\mathcal{D}k$, and conversely any polynomial p with total degree at most k has sample-wise degree (k, k) . We will denote the orthogonal projection (with respect to $L^2(D_{\otimes}^{\otimes m})$) of any distinguisher p to the subspace of polynomials of sample-wise degree

$(\mathcal{D}, k)$ by $p^{\leq \mathcal{D}, k}$. The above discussion on optimal distinguishing advantage for low-degree polynomial distinguishers carries over immediately to this alternative setting by replacing all appearances of $\leq k$ with $\leq \mathcal{D}, k$. In particular, the relevant quantity to bound here would be $\mathbb{E}_u(\overline{D}_u^{\otimes m})^{\leq \mathcal{D}, k}$. We can also define the *degree- $(\mathcal{D}, k)$ advantage* $\chi_{\leq (\mathcal{D}, k)}$ accordingly.

Appendix B. Discussion and related work

Below we discuss relevant prior work in an effort to situate our contributions within the broader literature on the complexity of quantum and classical learning.

B.1. Information-theoretic bounds

Recent works have precisely pinned down the sample complexity for a number of fundamental quantum learning tasks (Chen et al., 2022a, 2023b, 2022c, 2024c,d,f,g; Gong et al., 2024; O’Donnell and Wright, 2015, 2016; Haah et al., 2016; Anshu et al., 2020). A proper survey of this literature is beyond the scope of this paper. We note that a large number of these works have focused on the question of distinguishing whether a state is maximally mixed or from some ensemble of states that are far from maximally mixed, and we too will focus on this problem, albeit from the perspective of computational efficiency.

In our setting, the ensembles of states we consider are typically very structured in the sense that each state can be described by a polynomial number of parameters. For states from such ensembles, we make the (folklore) observation that learning is information-theoretically tractable by a straightforward application of known results:

Remark 18 *We note that given any class $\mathcal{C}$ of n -qubit pure states with $\exp(\text{poly}(n))$ covering number, it is always information-theoretically possible to learn states in $\mathcal{C}$ by performing simple single-copy measurements on polynomially many copies. All of the classes mentioned on the first page, e.g. Gibbs states of local Hamiltonians, shallow circuit states, stabilizer states, and matrix product states, satisfy this property.*

The reason this problem is information-theoretically possible is that we can perform brute force over a suitable net $\{|\psi\rangle\}$ over $\mathcal{C}$. Techniques like classical shadows (Huang et al., 2020) allow for the estimation of the fidelity of the given state with all $|\psi\rangle$ in the net using only $O(\log |\mathcal{C}|) = \text{poly}(n)$ copies. Furthermore, the measurements needed are efficiently preparable: recent work from Schuster et al. (2024) proves that shadow estimation for fidelities can be performed using only measurements prepared by $\log(n)$ -depth circuits in a 1D nearest-neighbor layout. Consequently, polynomially many of these shallow measurements are sufficient to learn the circuit description of state $|\psi\rangle$ from the net that has high fidelity with the unknown state.

However, this strategy is computationally inefficient as it requires brute force enumeration and estimation of all $\exp(\text{poly}(n))$ fidelities in classical post-processing.

B.2. Classical low-degree method

The low-degree method emerged from Barak et al. (2019); Hopkins and Steurer (2017); Hopkins et al. (2017); Hopkins (2018) and has since been used extensively to obtain sharp predictions for a wide range of problems in classical statistical inference (see Kunisky et al. (2019) for a survey from 2019). It is impossible for us to do justice to this vast literature in this section, so here we highlight the parts of this literature most relevant to this work.

In [Brennan et al. \(2021\)](#), the authors established a formal connection between the low-degree method and the statistical query model (see below). This allowed them to transfer lower bounds for one framework to lower bounds for the other, recovering sharp predictions in both models for various problems. Most relevant to the present work are their results on planted clique and planted dense subgraph, in particular their bipartite analogues, which recover the tight statistical query lower bounds of [Feldman et al. \(2015\)](#). In addition to drawing upon this result in Section E.4 when we establish a similar connection in the quantum setting, in Section 2.3 we discuss how our lower bound for quantum planted biclique builds upon calculations performed in the classical setting. We also note the work of [Dhawan et al. \(2025\)](#) which provides classical low-degree bounds for a hypergraph version of this problem, as well as prior works of [Ma and Wu \(2015\)](#); [Bresler and Jiang \(2023\)](#); [Brennan et al. \(2018\)](#); [Brennan and Bresler \(2020\)](#); [Hajek et al. \(2015\)](#) which studied this same question from the perspective of average-case reductions rather than low-degree hardness.

Our results in Section H.2 also touch upon the well-studied problem of tensor PCA ([Montanari and Richard, 2014](#); [Lesieur et al., 2017](#)), where the goal is to distinguish between a Gaussian tensor and a Gaussian tensor plus a rank-1 spike. A sharp prediction for hardness for this problem via the low-degree method can be found in the note of [Kunisky et al. \(2019\)](#), and [Potechin and Rajendran \(2022\)](#) has also proven a sum-of-squares lower bound for this problem; these hardness results are matched by low-degree algorithms via sum-of-squares ([Hopkins et al., 2015](#)) and via message passing ([Wein et al., 2019](#)). Tensor PCA has also been studied in quantum contexts, with prior work giving quartic speedups over the best known classical algorithms for this problem ([Hastings, 2020](#); [Schmidhuber et al., 2025](#)).

B.3. Hardness of quantum learning from pseudorandomness

Another popular method to prove computational lower bounds on quantum learning tasks relies on recent constructions of pseudorandom quantum states ([Ji et al., 2018](#)) and unitaries ([Ma and Huang, 2024](#)). These are ensembles over states or unitaries that cannot be distinguished by efficient algorithms from Haar-random states or unitaries conditional on mild cryptographic assumptions. The fact that these ensembles cannot even be distinguished from their Haar-random counterparts implies that they also cannot be learned by efficient algorithms. Crucially, because these states or unitaries require only polynomially-sized circuits or even, in specialized architectures, circuits with subextensive lightcones ([Schuster et al., 2024](#)) to generate, this then implies that learning even extremely shallow circuits from measurements is computationally hard.

The existence of quantum pseudorandomness has had profound implications for physics. One key application is in “spoofing” quantum properties such as entanglement ([Aaronson et al., 2024](#)) and magic ([Gu et al., 2024a](#)) – thought to be measures of the “quantumness” of a state – with polynomial-sized circuits. This immediately places lower bounds on our ability to manipulate the amounts of these resources in a given quantum state by efficient algorithms. Our work immediately raises the question of whether these bounds could be strengthened by adopting a ‘low-degree’ notion of efficiency.

B.4. Quantum statistical query model

Within the quantum literature, perhaps most relevant to the present work is the *quantum statistical query (QSQ)* model, introduced in [Arunachalam et al. \(2020\)](#), which offers a complementary perspective on predicting information-computation gaps for quantum learning.

The model is motivated by the *statistical query* (SQ) model (Kearns, 1998) from the classical literature, another popular framework for proving hardness under a restricted model of computation. Roughly, the SQ considers a model where, instead of directly observing samples, the learning algorithm can only query population-level statistics of the data distribution to some level of precision. This framework has been used to obtain accurate predictions for a range of problems like Gaussian mixtures (Diakonikolas et al., 2017), learning neural networks (Chen et al., 2022b,d; Goel et al., 2020b,a; Diakonikolas et al., 2017, 2020), and robust linear classification (Diakonikolas and Kane, 2022). As with the low-degree method, the literature here is too vast to do justice to here.

Analogous to SQ, QSQ is a model where, instead of performing arbitrary multi-outcome measurements on the copies of the unknown state ρ , the learning algorithm can only query two-outcome observables of ρ of their choice up to some level of precision, called *tolerance*. Below, we briefly discuss the relative merits of these two models.

One advantage of QSQ is that it is possible to prove lower bounds even when the queries can be adaptively chosen, whereas in our low-degree framework, it is necessary to limit the adaptivity to avoid various “cheating” algorithms (see Section E.5).

But the QSQ model is limiting in some respects. For one, it assumes that the two-outcome observable values are given up to small *adversarial error*, which can be overly pessimistic (Abbe et al., 2021). In addition, the notion of sample complexity is somewhat conflated with the notion of tolerance: in this model, if observable values are given to error τ , the sample complexity of the algorithm is effectively regarded as the number of queries times $1/\tau^2$. But this imposes a rather rigid constraint on how samples get used in QSQ algorithms, which can sometimes lead to incorrect predictions for information-computation gaps; this is discussed at length in Brennan et al. (2021).

Perhaps the most important drawback of the QSQ model is that focusing on protocols that use two-outcome measurements can be overly restrictive. As the following example demonstrates, there can even be cases where the QSQ complexity of a problem exponentially overshoots both the true complexity and what is predicted by the low-degree method.

Example 1 *Consider the task of distinguishing between the maximally mixed state and a state which is a Haar-random pure state in the first $\log_2 n$ qubits and maximally mixed elsewhere. It is known that by measuring $O(\sqrt{n})$ copies in the computational basis, one can computationally efficiently distinguish between these two scenarios. The idea is that the distribution over readouts from measuring the first $\log_2 n$ qubits under the second scenario is the uniform distribution over $\{0, 1\}^{\log_2 n}$, whereas the distribution in the first scenario is $1/\sqrt{n}$ -far from uniform in L_2 distance. By counting the number of collisions between samples (Chan et al., 2014), one can distinguish between these two cases. Counting collisions among samples from a distribution over $\{0, 1\}^{\log_2 n}$ can be implemented with a degree- $O(\log n)$ estimator, meaning that the above task is tractable under the low-degree framework. In contrast, Arunachalam et al. (2023) showed that for QSQ algorithms, $\exp(\Omega(n))$ many QSQ queries are needed for this problem.*

Nevertheless, QSQ offers a fruitful complementary viewpoint on hardness of quantum learning. For instance, in Arunachalam et al. (2023), it was shown that ensembles that are approximate state 2-designs are hard to learn under QSQ, strengthening prior work of Hinsche et al. (2023); Nietner et al. (2023) which focused on *diagonal* two-outcome measurements. This mirrors our findings in the low-degree framework and serves as additional evidence in favor of the conjecture that random quantum circuits are pseudorandom. Even more interestingly, the mechanism by which 2-designs

manifest in the analysis for QSQ is quite different from how they appear in our low-degree arguments. In their case, the 2-design property is crucial for relating the variance of the response to a QSQ (as a random variable in $\rho \sim \mathcal{E}$) to the variance if ρ were Haar-random. In our case, the 2-design property is crucial for establishing local indistinguishability of copy-wise rotations of the quantum data.

Finally, as shown by [Brennan et al. \(2021\)](#) in the classical case, under certain conditions there is even a formal equivalence between the two models. In [Section E.4](#), we extend this equivalence to the quantum setting by observing that the conditions of [Brennan et al. \(2021\)](#) are satisfied by the distribution over readouts provided the measurements are corrupted by some stochastic noise.

Finally, note that [Arunachalam et al. \(2023\)](#) also studied a quantum generalization of the planted biclique problem, but despite sharing the same name, their version and motivation is entirely different: they take the classical distribution over right-neighbors of a random left-vertex of the bipartite graph and consider a coherent encoding of this. Their main finding is that if the size of the planted biclique is $\lambda \geq \Omega(\log n)$, then there is an algorithm for their problem that makes $n^{O(\lambda)}$ QSQs to precision $\sqrt{\lambda/n}$, whereas any classical algorithm requires precision λ/n .

B.5. Quantum learning algorithms as low-degree estimators

Observable averages. As a trivial first example of a low-degree estimator, consider a protocol that estimates an observable O by measuring multiple copies each with the two-outcome POVM $\{O, I - O\}$. The low-degree algorithm in this case is simply the proportion of copies that resulted in the former outcome, which is a degree-1 estimator.

Shallow circuit learning. It was recently shown in [Landau and Liu \(2024\)](#); [Kim et al. \(2024\)](#); [Huang et al. \(2024\)](#) that states prepared by quantum circuits of constant depth can be learned efficiently in the following sense: Given a state prepared by a constant depth circuit and the ability to perform measurements, using only estimations of $O(1)$ -local marginals allows one can find another constant-depth circuit that prepares a high-fidelity state. The first step of learning the local marginals can be done by learning the coefficients of all local Pauli operators, which is a degree $O(1)$ estimator. Then, in [Landau and Liu \(2024\)](#) a local inversion V is found that rotates the state to one of the form $|0^l\rangle \otimes |\psi^l\rangle$, for $l = O(1)$ qubits by inverting the local gates in the lightcone of the l qubits. As unitaries are inverted by taking the adjoint, this too is a low-degree polynomial map. These local inversions are combined with fixed SWAP operations, which results in an output that is low-degree in the measurement readouts.

Learning Hamiltonians via polynomial system solving. In [Haah et al. \(2022\)](#), the authors give a sample- and time-optimal algorithm for learning high-temperature Hamiltonians from Gibbs state access. The protocol first performs measurements to obtain local Pauli observables of the Gibbs state, and then post-processes these by solving a suitable polynomial system using the Newton-Raphson method. This latter algorithm proceeds by repeatedly applying low-degree polynomial maps to the iterate followed by projection, and can because the algorithm converges rapidly, the overall output can be well-approximated by a low-degree polynomial in the readouts.

The recent breakthrough of [Bakshi et al. \(2024b\)](#) gave an algorithm for learning *arbitrary-temperature* local Hamiltonians from Gibbs state access. They devised a sum-of-squares algorithm for post-processing readouts which were again given by local Pauli measurements. Roughly speaking, they formulated a certain polynomial system in the unknown coefficients of the Hamiltonian

and showed that a pseudodistribution of sufficient degree over solutions to this system could be rounded to a good approximation to the true coefficients. While it is unclear how to rigorously formulate this as a low-degree polynomial estimator and consequently how to rule out such algorithms when trying to prove information-computation gaps, in the classical literature, low-degree hardness is typically regarded as initial evidence for sum-of-squares hardness. Indeed, the low-degree advantage $\|\overline{D}^{\leq k} - 1\|^2$ emerges naturally in pseudo-calibration (Barak et al., 2019), and showing it is small is a necessary first step in proving a sum-of-squares lower bound.

Agnostic tomography of product states. Recent work (Bakshi et al., 2024a; Chen et al., 2024e) shows how to obtain, for any given pure state, a (discrete) product state that is ε -close to the optimal fidelity. For agnostic tomography of discrete product states, the algorithm proposed by Chen et al. (2024e) is adaptive and low-degree by looking for the high-correlated *single-qubit* project for each qubit via adaptively chosen local measurements followed by an adaptive local optimization procedure employing the measurement outcomes as a low-degree function. For agnostic tomography of product states when the optimal fidelity is larger than $5/6$, Bakshi et al. (2024a) gave the faster agnostic tomography algorithm of product states. The algorithm first uses a divide-and-conquer strategy and sews the local tomography of each qubit, and then performs a local product state optimization that only uses local unitary transformations and low-degree post-processing. The resulting learning algorithm as a whole is thus low-degree in the local measurement results.

For agnostic tomography of product states with any optimal fidelity, although the algorithm proposed in Bakshi et al. (2024b) is highly adaptive, the intuition and most intermediate subroutines are captured by low-degree functions. In particular, the algorithm maintains a cover over all product states that have large fidelity and only optimize fidelity in a low-degree way over extensions of good product states over a small subsystem.

B.6. Corner cases

Yet, there are instances where the classical low-degree framework, by limiting its attention to low-degree algorithms, makes incorrect predictions about the computational hardness of a problem. One notable example is k -XOR-SAT, a special case of learning parities. The low-degree method predicts that the problem should be computationally hard, even though it is in fact efficiently solvable by the high-degree algorithm of Gaussian elimination.

Our low-degree framework has the same issue. In Appendix I, we show low-degree hardness for learning quantum stabilizer states even though there exist time-efficient high-degree algorithms for the task that effectively perform a Gaussian elimination on measurement readouts (Montanaro, 2017).

In general, the low-degree method is particularly susceptible to such failures on problems (whether quantum or classical) with a high degree of algebraic structure (Holmgren and Wein, 2020). Nevertheless, algorithms that exploit such algebraic structure to run efficiently tend to be especially brittle – as exemplified by the famous contrast between the efficiency of learning parities, and the intractability of learning parities with noise. As with the classical low-degree method, we expect that our framework correctly captures the limits of learning quantum states in the presence of noise. While the current agnostic tomography algorithms (Grewal et al., 2024b; Chen et al., 2024e) of learning stabilizer states provide computational efficient protocols even when the fidelity between the underlying (noise-corrupted) state and any stabilizer states drop to $o(1)$, these algorithms depend on adaptive post-selection and Bell difference sampling that requires two-copy measurements. To

the best of our knowledge, there is no noise-robust algorithm based on subroutines such as computational difference sampling that only employs single-copy measurements. This can be evidence that our framework may correctly predict the vulnerability of high-degree protocols with single-copy measurements against noise.

Finally, we remark that very recently, [Buhai et al. \(2025\)](#) constructed a distinguishing problem for which the degree- $\Omega(n^{1-O(\varepsilon)})$ likelihood ratio vanishes, yet there is a *quasipolynomial*-time algorithm. Furthermore, unlike Gaussian elimination, their algorithm is noise-robust, based on a list-decoding algorithm for noisy polynomial interpolation. Notably, unlike a previous construction which also leveraged a connection to error correction ([Holmgren and Wein, 2020](#)), the distinguishing problem of [Buhai et al. \(2025\)](#) has permutational symmetry and thus yields a counterexample to a certain formalization of Conjecture 17 in [Hopkins \(2018\)](#). While their specialized construction offers a note of caution in interpreting predictions made by the low-degree method, the method remains useful for obtaining initial evidence of computational hardness and initial estimates for the thresholds at which hardness manifests. Furthermore, in cases where the best known existing algorithms fall under the umbrella of polynomial distinguishers, like the examples documented in Section B.5, low-degree hardness articulates a natural barrier faced by contemporary algorithmic techniques.

Appendix C. Outlook

Further refinements are possible to our framework and applications. Below, we mention some concrete future directions.

How to model full adaptivity. A key contribution of this work was to establish, for the first time, a framework for low-degree hardness against adaptive measurement strategies. Yet, here we can handle only a limited notion of adaptivity, where the adaptive choice of measurement basis depends only on a partial history of previous measurement outcomes. As we show in Section E.5, a variety of natural ways of modeling adaptivity run into fundamental issues or are as hard as proving circuit lower bounds. We leave as an important open question how to satisfactorily model adaptivity within the low-degree framework.

Other quantum applications. Finding further applications of our framework in different quantum tasks is an immediate open question. The candidate learning problems that one could hope to prove information-computation gaps for include > 1 -dimensional tensor network states, ground states of local Hamiltonians, shadow tomography with efficiently preparable observables, and agnostically learning shallow circuits.

Protocols with arbitrary POVMs. For non-adaptive measurement strategies, our framework falls a hair short of encompassing algorithms that use fully general POVMs, as we can only deal with ancilla-assisted PVMs with at most $\lesssim n/11$ ancillas – which do not suffice to simulate a general POVM.

Quantum hardness and advantage from encoding classical problems. In this work, we encode classical problems that are known to be average-case hard in the classical low-degree framework into quantum data to show average-case hardness in the quantum low-degree framework. It is natural to wonder if we can find a general reduction from classical average-case hardness to quantum average-case hardness in the low-degree setting. Our work also opens up the exciting possibility of finding

low-degree quantum advantage; that is, a classical problem, embedded into a quantum state, on which a low-degree quantum algorithm can perform better than a low-degree classical algorithm. It is also interesting to see if there are classical problems that are hard using *any* quantum encoding, indicating their hardness for any quantum algorithms.

Mixture vs. mixture hypothesis testing problems. The general framework established in this work focuses on a point vs. mixture distinguishing task. It is natural to ask if we can prove low-degree hardness for distinguishing between two random ensembles of quantum states, i.e. mixture vs. mixture state distinguishing tasks. This is of particular interest if one wants to extend our framework beyond single-copy measurements: as soon as one considers $t = 2$ -copy measurements, SWAP test provides another example of a low-degree learning algorithm. Low-degree lower bounds for mixture vs. mixture testing have been proven in classical settings, e.g. in [Rush et al. \(2023\)](#).

Distinguishing and learning quantum channels. While our general framework focuses on hypothesis testing tasks among quantum states, it is interesting to see if our framework can be generalized to hypothesis testing tasks among quantum channels, where we should also take into account the computational complexity of preparing input quantum states.

Existence of an information-computation gap for quantum error mitigation. In this work, we show that quantum error mitigation is low-degree hard for 1D geometrically local noisy circuit of super-logarithmic depth and inverse polynomially small noise rate and protocols using single-qubit measurements. While there exists a statistically efficient algorithm using joint measurements for the hypothesis testing problem we construct for our lower bound ([Buadescu and O’Donnell, 2021](#)), it is unknown whether there is a sample-efficient algorithm using single-qubit measurements. It thus remains open whether there is an actual information-computation gap in this setting.

Other quantum generalizations of classical inference problems. The problem of quantum planted biclique that we proposed in this work is just one of many possible “quantizations” of classical inference problems. One could have also considered a quantum version of (non-bipartite) planted clique, a quantum version of distinguishing random graphs from random geometric graphs ([Bangachev and Bresler, 2024](#); [Liu et al., 2022](#)), or any of a number of other models with planted sparse structure. Even for the quantum planted biclique question, it remains unclear what the full complexity landscape of the problem is when one allows general single-copy measurements. Given the extent to which such problems have shaped our understanding of the complexity of classical learning, we expect such problems to be a fruitful testbed for future work in the quantum learning community.

Appendix D. The general framework for non-adaptive measurements

In this section we present our low-degree framework for quantum learning, specialized to the case of non-adaptive measurement strategies. As mentioned in Section A.2, we will focus on *hypothesis testing* problems. More specifically, for most of this work we will focus on one of the most pervasive classes of hypothesis testing problems within the quantum learning literature ([O’Donnell and Wright, 2015](#); [Chen et al., 2022a](#)): given access to copies of an unknown n -qubit state ρ and given an ensemble $\mathcal{E}$ of “structured states,” distinguish between

- H_0 (null case): ρ is equal to the maximally mixed state $\rho_{\text{mm}} = I/2^n$.
- H_1 (alternative case): ρ was sampled at the outset from $\mathcal{E}$

Our notion of low-degree hardness depends on the choice of a class of measurement strategies. In general, one could consider measurement strategies which in each round process t copies of ρ by performing a projective measurement on those copies and n' ancilla qubits – by Naimark’s theorem, we can implement arbitrary $2^{n'}$ -outcome POVMs using a PVM with n' ancillary qubits. Each such measurement outputs a $(tn + n')$ -bit string, and we perform m such rounds of measurement and collect $m(tn + n')$ bits. A *degree- k* estimator in our setting is then any polynomial $f : \{0, 1\}^{m(tn+n')} \rightarrow \mathbb{R}$ whose variance under the null case is normalized to 1, such that each monomial in f depends on at most k bits. For any such f , we can consider the *distinguishing advantage* achieved by f . Recall from Eq. (2) that this is the absolute difference between the expectation of f under the null case and the expectation under the alternative case, and that the optimal distinguishing advantage is given by the variance of the projection of the *likelihood ratio* between alternative case and null case to the space of degree- $\leq k$ polynomials (we will review this calculation in the warm-up below).

For a given class of measurements, specified by t and n' and miscellaneous additional constraints on what kinds of measurements are allowed, we say that the hypothesis testing problem is *degree- k hard* for that class (throughout, we will think of k as $\omega(\log n)$) if *for any sequence of (non-adaptively chosen) measurements from the class*, the optimal distinguishing advantage $\chi_{\leq k}$ achieved by any degree- k estimator is $o(1)$. Note that the classical low-degree model considered in Appendix A.2, specialized to distributions over the Boolean cube and null case given by the uniform distribution, is the special case of the above framework when $\mathcal{E}$ only consists of states with diagonal density matrices. Indeed, the key challenge in the quantum setting is the additional quantification over which measurements are selected, whereas in the classical setting, the only measurements to consider are computational basis measurements.

Throughout this work, we will focus on the simple but practically relevant setting of single-copy projective measurements, i.e., where $t = 1$ (see Figure 1 for an illustration). Below, we first provide some warm-up calculations in the special case of local (i.e., single-qubit) measurements (Appendix D.1). We then turn to the main result of this section: if $\mathcal{E}$ forms an approximate 2-design, then hypothesis testing is low-degree hard for general single-copy measurements that use up to $O(n)$ ancillary qubits (Appendix D.2).

D.1. Warm-up: Low-degree hardness for local measurements

We start with protocols that only use local (single-qubit) PVMs with zero ancilla, i.e. PVMs of the form $\{|\phi_{s_i}\rangle\langle\phi_{s_i}|\}_{s_i}$ on the i -th copy, where $|\phi_{s_i}\rangle = \bigotimes_{r=1}^n |\phi_{s_{i,r}}\rangle$ and $s_i = (s_{i,1}, \dots, s_{i,n}) \in \{0, 1\}^n$. For any outcome s_i , the likelihood of observing s_i in the null case is always the same:

$$D_{\emptyset}(s_i) = \langle\phi_{s_i}|\rho_{\text{mm}}|\phi_{s_i}\rangle = 2^{-n}.$$

The likelihood ratio between observing s_i under the alternative case versus under the null case, for a specific choice of ρ sampled from the ensemble $\mathcal{E}$, is thus given by

$$\overline{D}_{\rho}(s_i) = 2^n \text{tr} \left(\rho \bigotimes_{r=1}^n |\phi_{s_{i,r}}\rangle\langle\phi_{s_{i,r}}| \right),$$

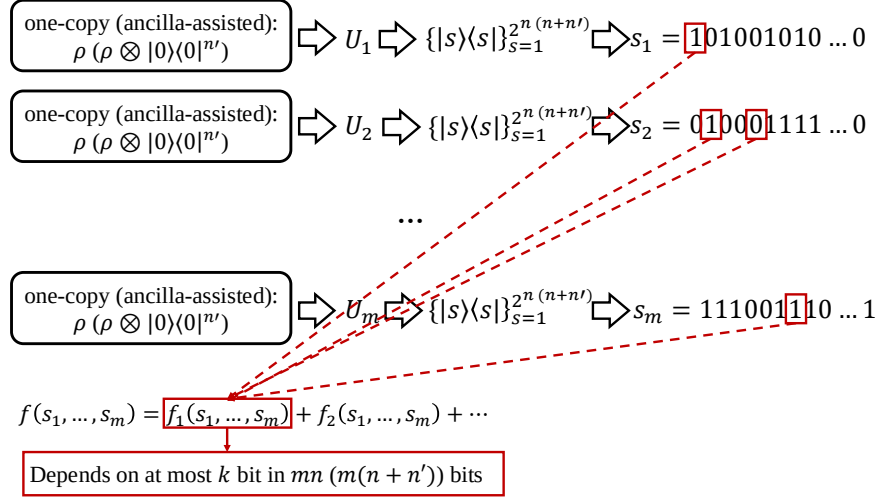


Figure 1: Non-adaptive single-copy PVM on one copy of ρ (expressions in parentheses correspond to the case where the PVM involves n' ancilla qubits).

and the likelihood ratio for observing a history $\mathbf{s} = (s_1, \dots, s_m) \in \{0, 1\}^{mn}$ after measuring m copies of ρ is given by

$$\overline{D}_{\rho, m}(\mathbf{s}) = 2^{mn} \prod_{i=1}^m \text{tr} \left(\rho \bigotimes_{r=1}^n |\phi_{s_{i,r}}\rangle \langle \phi_{s_{i,r}}| \right).$$

Let $\chi_T(\mathbf{s})$ denote the Fourier basis function $\chi_T(\mathbf{s}) = (-1)^{\sum_{(i,r) \in T} s_{i,r}}$, noting that because the distribution over histories under the null case is uniform, the functions (χ_T) form an orthonormal basis. We can then decompose the likelihood ratio for the whole history as

$$\overline{D}_{\rho, m}(\mathbf{s}) = \sum_{T \subseteq [mn]} \alpha_T^\rho \chi_T(\mathbf{s}), \quad \text{for } \alpha_T^\rho = \mathbb{E}_{\mathbf{s}} \overline{D}_{\rho, m}(\mathbf{s}) \chi_T(\mathbf{s}).$$

Now we can consider the low-degree likelihood ratio $\overline{D}_{\rho, m}^{\leq k}(\mathbf{s})$, recalling from Section A.2 that this is the degree- k Fourier truncation $\overline{D}_{\rho, m}^{\leq k}(\mathbf{s}) \triangleq \sum_{|T| \leq k} \alpha_T^\rho \chi_T(\mathbf{s})$ and that the optimal distinguishing advantage achieved by any degree- k estimator is given by:

$$\chi_{\leq k}^2 \triangleq \left\| \mathbb{E}_{\rho \sim \mathcal{E}} \overline{D}_{\rho, m}^{\leq k} - \mathbf{1} \right\|^2 = \sum_{|T| \leq k, T \neq \emptyset} (\mathbb{E}_{\rho} [\alpha_T^\rho])^2.$$

In the following, we consider the Fourier coefficient $\hat{D}[T] \triangleq \mathbb{E}_{\rho} [\alpha_T^\rho]$ for a particular $T \subseteq [mn]$ and $|T| \leq k$. We denote $\mathbf{s}[T]$ to be the set of $s_{i,r}$ with $(i, r) \in T$. Note that the parts not touched by T get traced out when we average over histories, yielding:

$$\hat{D}[T] = \mathbb{E}_{\rho} [\alpha_T^\rho] = \mathbb{E}_{\rho \sim \mathcal{E}} \mathbb{E}_{\mathbf{s}} \overline{D}_{\rho, m}(\mathbf{s}) \chi_T(\mathbf{s}) = \mathbb{E}_{\rho \sim \mathcal{E}} \mathbb{E}_{\mathbf{s}[T]} 2^{|T|} \text{tr} \left(\text{tr}_{[mn] \setminus T} (\rho^{\otimes m}) \bigotimes_{(i,r) \in T} |\phi_{s_{i,r}}\rangle \langle \phi_{s_{i,r}}| \right). \quad (3)$$

To show low-degree hardness, we wish to show that the sum of squares of these quantities over $|T| \leq k$ is small. Motivated by this, we consider the following:

Condition 19 (Bounded k -local likelihood ratio) *For a hypothesis testing problem ρ_{mm} vs. $\rho \sim \mathcal{E}$, define the k -local likelihood ratio as*

$$k\text{-LLR} := \max_{\substack{T \in [mn]: |T|=k \\ \Lambda \in \text{single-qubit PVMs}}} \left\| \mathbb{E}_{\rho \sim \mathcal{E}} \bar{D}_{\text{tr}_{[mn] \setminus T}(\rho^{\otimes m})} - \mathbf{1} \right\|^2,$$

where we maximize over all size- T subsets of the total mn qubits, and over all single-qubit PVM measurement strategies Λ (such strategies apply the PVM $\{|\phi_{s_i}\rangle\langle\phi_{s_i}|\}_{s_i \in \{0,1\}^n}$ on the i -th copy), and where the quantity $\bar{D}_{\text{tr}_{[mn] \setminus T}(\rho^{\otimes m})}$ is given by

$$\bar{D}_{\text{tr}_{[mn] \setminus T}(\rho^{\otimes m})} \triangleq 2^{|T|} \text{tr}_{[mn] \setminus T}(\rho^{\otimes m}) \bigotimes_{(i,r) \in T} |\phi_{s_{i,r}}\rangle\langle\phi_{s_{i,r}}|.$$

Often it is more convenient to check that the following stronger quantity, which we call k -local indistinguishability, is bounded:

$$d_{\text{tr}} \left(\mathbb{E}_{\rho \sim \mathcal{E}} \text{tr}_{[mn] \setminus T}(\rho^{\otimes m}), \frac{I}{2^{|T|}} \right), \quad (4)$$

i.e. any k -qubit subsystem of the m copies is close to maximally mixed. This is the natural quantum analogue of k -wise independence. It is easy to show that a bound on k -local distinguishability also implies a bound on k -LLR (see Appendix J.1 for the proof):

Corollary 20 (k -local indistinguishability implies k -LLR condition) *If the hypothesis testing problem ρ_{mm} vs. $\rho \sim \mathcal{E}$ satisfies $d_{\text{tr}} \left(\mathbb{E}_{\rho \sim \mathcal{E}} \text{tr}_{[mn] \setminus T}(\rho^{\otimes m}), \frac{I}{2^{|T|}} \right) \leq \varepsilon$ for any $T \in [mn]$ and $|T| \leq k$, then $k\text{-LLR} \leq \varepsilon^2 2^{2k}$.*

We can now conclude this warm-up section with the basic result that bounded k -LLR implies low-degree hardness. Later we will give examples of natural problems that satisfy this condition.

Proposition 21 (Low-degree hardness for single-qubit measurements) *For a hypothesis testing problem ρ_{mm} vs. $\rho \sim \mathcal{E}$ whose k -LLR is bounded by $2^{-\tilde{\Omega}(k \log n)}$, weak detection is degree- k hard for protocols that make non-adaptive single-qubit measurements on $\text{poly}(n)$ copies of ρ .*

Proof Suppose the k -LLR is bounded by ε^2 . Then

$$\mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{s}, \mathbf{s}'} \bar{D}_{\rho, m}(\mathbf{s}) \bar{D}_{\rho', m}(\mathbf{s}') \chi_T(\mathbf{s}) \chi_T(\mathbf{s}') \leq \left\| \mathbb{E}_{\rho \sim \mathcal{E}} \bar{D}_{\text{tr}_{[mn] \setminus T}(\rho^{\otimes m})} - \mathbf{1} \right\|^2 < \varepsilon^2.$$

Therefore, we can bound the square degree- k advantage by

$$\begin{aligned} \left\| \mathbb{E}_{\rho \sim \mathcal{E}} \bar{D}_{\rho, m}^{\leq k} - \mathbf{1} \right\|^2 &= \sum_{|T| \leq k, T \neq \emptyset} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{s}, \mathbf{s}'} \bar{D}_{\rho, m}(\mathbf{s}) \bar{D}_{\rho', m}(\mathbf{s}') \chi_T(\mathbf{s}) \chi_T(\mathbf{s}') \\ &= \varepsilon^2 \sum_{i=1}^k \binom{mn}{i} \leq \varepsilon^2 k (mn)^k. \end{aligned}$$

Choosing $\varepsilon = 2^{-\tilde{\Omega}(k \log n)}$ we can finish the proof. $\blacksquare$

By Theorem 20, the same conclusion (degree- k hardness of weak detection) follows when the k -local distinguishability measure in Eq. (4) is bounded by ε^2 . In addition, we are able to extend this result beyond local measurements to arbitrary measurements prepared by bounded-depth circuits using light cone arguments:

Corollary 22 (low-degree hardness for measurements with bounded depth) *For the hypothesis testing problem ρ_{mm} vs. $\rho \sim \mathcal{E}$ satisfying*

$$d_{\text{tr}} \left(\mathbb{E}_{\rho \sim \mathcal{E}} \text{tr}_{[mn] \setminus T} (\rho^{\otimes m}), \frac{I}{2^{|T|}} \right) \leq 2^{-\tilde{\Omega}(k \log n)}$$

for any $T \in [mn]$ and $|T| \leq 2^L k$ ($|T| \leq (2L + k)^{\mathcal{D}}$), weak detection is degree- k hard for any protocol that makes polynomially many non-adaptive measurements prepared by ($\mathcal{D}$ -dimensional geometrically-local) circuits of depth at most L .

We defer the details of this to Appendix J.2:

Remark 23 *The reader might wonder how the low-degree framework for local measurements differs from a more naive setting where one simply gets access to $\text{poly}(n)$ many size- $\log(n)$ patches of $\rho^{\otimes m}$, performs full tomography on those patches in polynomial time, and post-processes the resulting classical information in polynomial time. In the latter model, it is trivial to show that local indistinguishability implies hardness: if the patches are $1/\text{poly}(n)$ -close to maximally mixed, then no classical post-processing, no matter how powerful, can extract information about the underlying state.*

We clarify that our low-degree model is strictly more powerful than this latter model. For instance, in the latter model, if one performs tomography on a deterministic choice of patches and only those patches are indistinguishable whereas other patches of $\rho^{\otimes m}$ are not, then low-degree protocols would be able to detect them but the weaker model would not. On the other hand, even if algorithms in the weaker model could choose patches randomly, the resulting estimator will still correspond to a convex combination of polynomial functions that are $\text{poly}(n)$ -sparse, degree- $\log(n)$, and variance-1, which need not encompass all degree- $\log(n)$, variance-1 polynomial functions that would be considered in our low-degree model.

D.2. 2-design implies low-degree hardness for single-copy measurements

We now present the main result of this section: a sufficient condition for low-degree hardness for the most general non-adaptive setting where one can perform arbitrary single-copy PVMs with some number of ancillas, as depicted in Figure 1. Unlike in the previous subsection, each PVM can now arbitrarily entangle different qubits in a system of size $n + n'$.

More formally, we consider protocols using $m = \text{poly}(n)$ copies of ρ and applying ancilla-assisted PVMs on $\rho \otimes |0\rangle\langle 0|^{\otimes n'}$. We abuse the notation slightly and denote the measurement for each round as $\{|\phi_{s_i}\rangle\langle \phi_{s_i}| \}_{s_i}$ for $i = 1, \dots, m$, where the s_i 's are $n + n'$ -bit strings. Given an output s_i , the likelihoods under the null case and the alternative case for a fixed ρ from the ensemble are given by

$$D_{\emptyset}(s_i) = \text{tr}(\rho_{\text{mm}} \otimes |0\rangle\langle 0|^{\otimes n'} \cdot |\phi_{s_i}\rangle\langle \phi_{s_i}|), \quad D_{\rho}(s_i) = \text{tr}(\rho \otimes |0\rangle\langle 0|^{\otimes n'} \cdot |\phi_{s_i}\rangle\langle \phi_{s_i}|).$$

Therefore, the likelihood ratio for each copy is given by

$$\overline{D}_\rho(s_i) = \frac{\text{tr}(\rho \otimes |0\rangle\langle 0|^{\otimes n'} \cdot |\phi_{s_i}\rangle\langle \phi_{s_i}|)}{\text{tr}(\rho_{\text{mm}} \otimes |0\rangle\langle 0|^{\otimes n'} \cdot |\phi_{s_i}\rangle\langle \phi_{s_i}|)}$$

for $\rho \sim \mathcal{E}$. Given a history $\mathbf{s} = (s_1, \dots, s_m)$, the likelihood ratio for the whole history is given as

$$\overline{D}_{\rho, m}(\mathbf{s}) = \prod_{i=1}^m \frac{\text{tr}(\rho \otimes |0\rangle\langle 0|^{\otimes n'} \cdot |\phi_{s_i}\rangle\langle \phi_{s_i}|)}{\text{tr}(\rho_{\text{mm}} \otimes |0\rangle\langle 0|^{\otimes n'} \cdot |\phi_{s_i}\rangle\langle \phi_{s_i}|)}.$$

We first show that a generalization of the notion of local indistinguishability from the previous section implies low-degree hardness for ancilla-assisted PVMs:

Theorem 24 (Low-degree hardness for ancilla-assisted PVMs) *For the hypothesis testing problem ρ_{mm} vs. $\rho \sim \mathcal{E}$ satisfying*

$$d_{\text{tr}} \left(\mathbb{E}_{\rho \sim \mathcal{E}} \text{tr}_{[m(n+n')]\setminus T} \left(\bigotimes_{i=1}^m U_i \left(\rho \otimes |0\rangle\langle 0|^{\otimes n'} \right) U_i^\dagger \right), \text{tr}_{[m(n+n')]\setminus T} \left(\bigotimes_{i=1}^m U_i \left(\rho_{\text{mm}} \otimes |0\rangle\langle 0|^{\otimes n'} \right) U_i^\dagger \right) \right) \leq 2^{-\tilde{\Omega}(k \log n)},$$

for some $0 < n' \leq n$, any $T \in [m(n+n')]$, and $|T| \leq k$, weak detection is degree- k hard for any protocol that makes polynomially many non-adaptive PVMs $\{\{U_i | x\rangle \langle x| U_i^\dagger\}_{x=1}^{2^{n+n'}}\}_{i=1}^m$ on a single copy of ρ and n' ancillary qubits.

The proof is similar to Theorem 21, see Appendix J.3 for details.

With this in hand, we now show the main result of this section, namely that when $\rho \sim \mathcal{E}$ forms an approximate state 2-design, the hypothesis testing problem satisfies the condition in Theorem 24 and is thus low-degree hard for protocols with non-adaptive PVMs using a bounded number n' of ancillary qubits.

Theorem 25 (Approximate state 2-design indicates low-degree hardness for single-copy measurements)

For the hypothesis testing problem ρ_{mm} vs. $\rho \sim \mathcal{E}$ with $\mathcal{E}$ a ε -approximate state 2-design for $\varepsilon \leq 2^{-\tilde{\Omega}(k \log n)}$, weak detection is degree- k hard for any protocol that makes polynomially many non-adaptive PVMs $\{\{U_i | x\rangle \langle x| U_i^\dagger\}_{x=1}^{2^{n+n'}}\}_{i=1}^m$ acting on a single copy and at most $n' = \Theta(n)$ ancillary qubits.

Proof We consider each single copy with n' ancillary qubits, for each fixed unitary $U \in SU(2^n)$ we have

$$\begin{aligned}
& \mathbb{E}_\psi \left\| \text{tr}_{[n+n'] \setminus A} \left[U \left(|\psi\rangle\langle\psi| \otimes |0\rangle\langle 0|^{\otimes n'} \right) U^\dagger \right] - \text{tr}_{[n+n'] \setminus A} \left[U \left(\frac{I}{2^n} \otimes |0\rangle\langle 0|^{\otimes n'} \right) U^\dagger \right] \right\|_2^2 \\
& \leq \mathbb{E}_\psi \text{tr}_{[n+n'] \setminus A}^2 \left[U \left(|\psi\rangle\langle\psi| \otimes |0\rangle\langle 0|^{\otimes n'} \right) U^\dagger \right] + \text{tr}_{[n+n'] \setminus A}^2 \left[U \left(\frac{I}{2^n} \otimes |0\rangle\langle 0|^{\otimes n'} \right) U^\dagger \right] - \\
& \quad 2\mathbb{E}_\psi \text{tr}_{[n+n'] \setminus A} \left[U \left(|\psi\rangle\langle\psi| \otimes |0\rangle\langle 0|^{\otimes n'} \right) U^\dagger \right] \text{tr}_{[n+n'] \setminus A} \left[U \left(\frac{I}{2^n} \otimes |0\rangle\langle 0|^{\otimes n'} \right) U^\dagger \right] + 3\varepsilon \\
& = \mathbb{E}_\psi \text{tr}_{[n+n'] \setminus A}^2 \left[U \left(|\psi\rangle\langle\psi| \otimes |0\rangle\langle 0|^{\otimes n'} \right) U^\dagger \right] - \text{tr}_{[n+n'] \setminus A}^2 \left[U \left(\frac{I}{2^n} \otimes |0\rangle\langle 0|^{\otimes n'} \right) U^\dagger \right] + 3\varepsilon \\
& \leq \text{tr} \left[U \otimes U \left(\left(\frac{\text{SWAP}_{12}}{2^n(2^n+1)} - \frac{I_{2n}}{4^n(2^n+1)} \right) \otimes |0\rangle\langle 0|^{\otimes 2n'} \right) U^\dagger \otimes U^\dagger \cdot I_{[n+n'] \setminus T}^{\otimes 2} \otimes \text{SWAP}_T \right] + 3\varepsilon \\
& \leq \text{tr} \left[U \otimes U \left(\left(\frac{\text{SWAP}_{12}}{2^n(2^n+1)} \right) \otimes |0\rangle\langle 0|^{\otimes 2n'} \right) U^\dagger \otimes U^\dagger \cdot I_{[n+n'] \setminus T}^{\otimes 2} \otimes \text{SWAP}_T \right] + 2^{-n} + 3\varepsilon,
\end{aligned}$$

where the second inequality follows from the fact that the first moment of Haar random states is the same as the maximally mixed state.

Note that we can always decompose any n -qubit unitary U as:

$$U = \sum_i \alpha_i V_i \otimes W_i$$

with at most constant terms, $|\alpha_i| \leq 1$, and $V_i \in SU(2)$ and $W_i \in SU(2^{n-1})$ are single- and $n-1$ -qubit unitaries. Some examples along this line include the quantum Shannon decomposition and the Cosine-Sine Decomposition (Shende et al., 2005), which provides this constant as 8. We can repeat this procedure for n' times and have

$$U = \sum_i \alpha_i V_i \otimes W_i$$

with at most $2^{3n'}$ terms, $|\alpha_i| \leq 1$, and $V_i \in SU(2^{n'})$ and $W_i \in SU(2^{n-n'})$ are n' - and $n-n'$ -qubit unitaries. We thus have

$$\begin{aligned}
& \text{tr} \left[U \otimes U \left(\left(\frac{\text{SWAP}_{12}}{2^n(2^n+1)} \right) \otimes |0\rangle\langle 0|^{\otimes 2n'} \right) U^\dagger \otimes U^\dagger \cdot I_{[n+n'] \setminus T}^{\otimes 2} \otimes \text{SWAP}_T \right] \\
& = \sum_{ijkl} \alpha_i \alpha_j \alpha_k^* \alpha_l^* \text{tr} \left[W_i \otimes W_j \left(\frac{\text{SWAP}_{12}}{2^n(2^n+1)} \right) W_k^\dagger \otimes W_l^\dagger \otimes V_i \otimes V_j \left(|0\rangle\langle 0|^{\otimes 2n'} \right) V_k^\dagger \otimes V_l^\dagger \cdot \right. \\
& \quad \left. I_{[n+n'] \setminus T}^{\otimes 2} \otimes \text{SWAP}_T \right] \\
& \leq \sum_{ijkl} \alpha_i \alpha_j \alpha_k^* \alpha_l^* 2^{|T|-n-n'} \leq 2^{|T|+11n'-n}.
\end{aligned}$$

Thus, we have

$$\begin{aligned}
& \mathbb{E}_\psi \left\| \text{tr}_{[n+n'] \setminus A} \left[U \left(|\psi\rangle\langle\psi| \otimes |0\rangle\langle 0|^{\otimes n'} \right) U^\dagger \right] - \text{tr}_{[n+n'] \setminus A} \left[U \left(\frac{I}{2^n} \otimes |0\rangle\langle 0|^{\otimes n'} \right) U^\dagger \right] \right\|_2^2 \\
& \leq 2 \cdot 2^{|A|+11n'-n} + 3\varepsilon.
\end{aligned}$$

Then by Markov's inequality, we have

$$\begin{aligned} & \Pr \left[\left\| \text{tr}_{[n+n'] \setminus A} \left[U \left(|\psi\rangle\langle\psi| \otimes |0\rangle\langle 0|^{\otimes n'} \right) U^\dagger \right] - \text{tr}_{[n+n'] \setminus A} \left[U \left(\frac{I}{2^n} \otimes |0\rangle\langle 0|^{\otimes n'} \right) U^\dagger \right] \right\|_2 \right] \geq \left(2 \cdot 2^{|A|+11n'-n} + 3\varepsilon \right)^{1/4} \\ & \leq \left(2 \cdot 2^{|A|+11n'-n} + 3\varepsilon \right)^{1/2}. \end{aligned}$$

By the norm inequality, we have

$$\begin{aligned} & \Pr \left[d_{\text{tr}} \left(\text{tr}_{[n+n'] \setminus A} \left[U \left(|\psi\rangle\langle\psi| \otimes |0\rangle\langle 0|^{\otimes n'} \right) U^\dagger \right], \text{tr}_{[n+n'] \setminus A} \left[U \left(\frac{I}{2^n} \otimes |0\rangle\langle 0|^{\otimes n'} \right) U^\dagger \right] \right) \geq 2^{|A|/2} \left(2 \cdot 2^{|A|+11n'-n} + 3\varepsilon \right)^{1/4} \right] \\ & \leq \left(2 \cdot 2^{|A|+11n'-n} + 3\varepsilon \right)^{1/2}, \end{aligned}$$

for any fixed U .

We can now consider the full scenario on m copies of $[n]$ and denote $T = T_1 \cup \dots \cup T_m$, where T_i denotes the elements of T that are contained within the i th copy. We always have $|T_i| \leq |T|$. Consider matrices A, B, C, D with $\|A\|_1 = \|D\|_1 = 1$. Then,

$$\|A \otimes B - C \otimes D\|_1 \leq \|A \otimes (B - D) + (A - C) \otimes D\|_1 \leq \|A\|_1 \|B - D\|_1 + \|A - C\|_1 \|D\|_1 \leq \|B - D\|_1 + \|A - C\|_1.$$

Applying this bound m times and then use the union bound and the triangle inequality, we obtain:

$$\begin{aligned} & d_{\text{tr}} \left(\mathbb{E}_\psi \text{tr}_{[m(n+n')] \setminus T} \left(\bigotimes_{i=1}^m U_i \left(|\psi\rangle\langle\psi| \otimes |0\rangle\langle 0|^{\otimes n'} \right) U_i^\dagger \right), \text{tr}_{[m(n+n')] \setminus T} \left(\bigotimes_{i=1}^m U_i \left(\frac{I}{2^n} \otimes |0\rangle\langle 0|^{\otimes n'} \right) U_i^\dagger \right) \right) \\ & \leq m \cdot 2^{|T|/2} \left(2 \cdot 2^{|T|+11n'-n} + 3\varepsilon \right)^{1/4}, \end{aligned} \tag{5}$$

with probability at least $1 - m \cdot \left(2 \cdot 2^{|T|+11n'-n} + 3\varepsilon \right)^{1/2}$. Finally, note that $|T| \leq k$, we only need to choose $n' \leq (n - \tilde{\Omega}(k \log n))/11 = \Theta(n)$ and $\varepsilon = 2^{-\tilde{\Omega}(k \log n)}$, which will guarantee that

$$d_{\text{tr}} \left(\mathbb{E}_\psi \text{tr}_{[m(n+n')] \setminus T} \left(\bigotimes_{i=1}^m U_i \left(|\psi\rangle\langle\psi| \otimes |0\rangle\langle 0|^{\otimes n'} \right) U_i^\dagger \right), \text{tr}_{[m(n+n')] \setminus T} \left(\bigotimes_{i=1}^m U_i \left(\frac{I}{2^n} \otimes |0\rangle\langle 0|^{\otimes n'} \right) U_i^\dagger \right) \right)$$

is bounded by $2^{-\tilde{\Omega}(k \log n)}$ and finishes the proof for Theorem 25. $\blacksquare$

Appendix E. Frameworks for adaptive measurements

A feature which is largely unique to the quantum setting is the ability for the learner to perform *adaptively chosen* measurements. This turns out to be quite subtle to model within the low-degree framework; indeed, in Section E.5 we catalog a number of no-go results for natural attempts. We remark that in classical settings where the low-degree method has been applied and where there is the possibility of adaptive choice (Mardia et al., 2024; Bangachev and Bresler, 2024), it is also open how to model adaptivity.

In this section, we provide three frameworks for reasoning about measurement strategies with some nontrivial level of adaptivity. To do this, we first need to turn to a stronger *copy-wise* notion of low-degree, which is the quantum analogue of the classical sample-wise notion mentioned at the end of Section A.2. We provide an overview of this stronger notion in Appendix E.1 for details (see Figure 2(a)); roughly, low-degree protocols in this category compute functions that are linear combination of terms that have degree at most $\mathcal{D}$ in the bits of each copy, while each depending on at most k copies.

Under this stronger copy-wise model, we provide two *round-based* models of adaptivity, which divide all m samples into m_1 rounds each of size m_0 . In particular, we consider two kinds of round-based models. The first model has adaptivity with each block of size $m_0 = \text{polylog}(n)$ (see Appendix E.2 for details), and the second model has adaptivity among $m_1 = \text{polylog}(n)$ blocks (see Appendix E.3 for details). Finally, we also consider protocols based on statistical query (SQ) post-processing of readouts from non-adaptive measurements. In this model, while the measurements are still non-adaptive, the statistical queries, which can be thought of as an (arbitrarily high-degree) coarsening of our measurements to two-outcome observables, are chosen adaptively. We will show a general equivalence between SQ models and non-adaptive copy-wise low-degree models in the presence of readout noise in Appendix E.4, building on the classical result of Brennan et al. (2021).

Finally, in Appendix E.5, we explain how several natural notions of adaptivity in the low-degree method cannot be used to prove meaningful lower bounds.

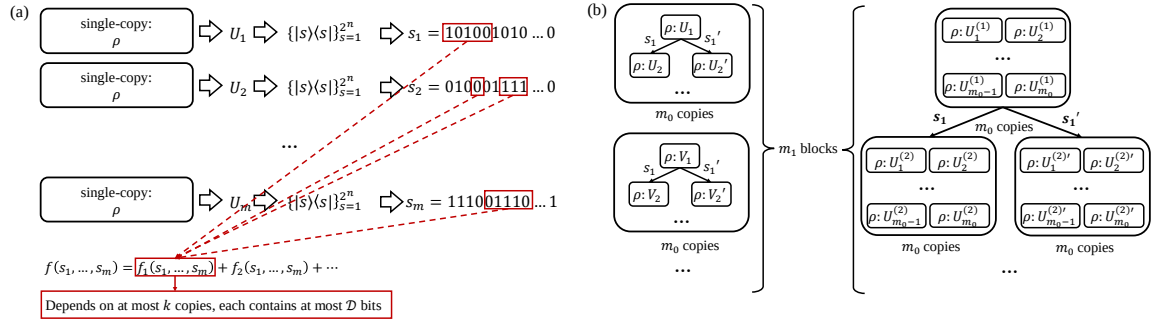


Figure 2: (a) The non-adaptive copy-wise low-degree model where each term of the final function depends on at most k copies each of degree at most $\mathcal{D}$. (b) Protocols using single-copy PVMs with restrictive adaptivity (round-based models), which can be decomposed into m_1 blocks each of m_0 copies. In the left model, we allow adaptivity within each block. In the right model, we allow adaptivity among blocks.

E.1. Copy-wise low-degree model

First, we introduce the copy-wise low-degree model. Recall that we consider protocols using $m = \text{poly}(n)$ copies of ρ and applying a single-copy PVM on each copy. Without loss of generality, we can denote the measurement for each round as $\{|\phi_{s_i}\rangle\langle\phi_{s_i}|\}_{s_i}$ for $i = 1, \dots, m$. We will assume that s_i can be any n -bit string by Appendix A.1. Given an output s_i , the likelihood ratio for each copy

is given by

$$\overline{D}_\rho(s_i) = 2^n \text{tr}(\rho |\phi_{s_i}\rangle \langle \phi_{s_i}|)$$

for $\rho \sim \mathcal{E}$. Given a history $\mathbf{s} = (s_1, \dots, s_m)$, the likelihood ratio for the whole history is given as

$$\overline{D}_{\rho,m}(\mathbf{s}) = 2^{mn} \prod_{i=1}^m \text{tr}(\rho |\phi_{s_i}\rangle \langle \phi_{s_i}|).$$

Here, we consider copy-wise $(\mathcal{D}, k)$ -low-degree likelihood ratio $D_{\rho,m}^{\leq \mathcal{D},k}$ instead of the low-degree likelihood ratio $D_{\rho,m}^{\leq k}$ considered in the previous parts. Recall from the end of Appendix A.2 that this is given by restricting the Fourier expansion of $D_{\rho,m}^{\leq \mathcal{D},k}$ to the subspace spanned by Fourier characters which each have degree at most $\mathcal{D}$ in each bit and nonzero degree in at most k of the bits. The square degree- $(\mathcal{D}, k)$ advantage can be computed as

$$\begin{aligned} \left\| \mathbb{E}_{\rho \sim \mathcal{E}} \overline{D}_{\rho,m}^{\leq \mathcal{D},k} - \mathbf{1} \right\|^2 &= \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left\langle \overline{D}_{\rho,m}^{\leq \mathcal{D},k}, \overline{D}_{\rho',m}^{\leq \mathcal{D},k} \right\rangle - 1 \\ &= \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left\langle \left(\bigotimes_{i=1}^m \overline{D}_\rho^{(i)} \right)^{\leq \mathcal{D},k}, \left(\bigotimes_{i=1}^m \overline{D}_{\rho'}^{(i)} \right)^{\leq \mathcal{D},k} \right\rangle - 1 \\ &= \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \sum_{T \in [m], |T| \leq k, T \neq \emptyset} \prod_{i \in T} \left(\left\langle \overline{D}_\rho^{(i) \leq \mathcal{D}}, \overline{D}_{\rho'}^{(i) \leq \mathcal{D}} \right\rangle - 1 \right) \\ &\leq \sum_{T \in [m], |T| \leq k, T \neq \emptyset} \left(\prod_{i \in T} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\left\langle \overline{D}_\rho^{(i) \leq \mathcal{D}}, \overline{D}_{\rho'}^{(i) \leq \mathcal{D}} \right\rangle - 1 \right)^{|T|} \right)^{1/|T|} \\ &\leq \sum_{T \in [m], |T| \leq k, T \neq \emptyset} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\left\langle \overline{D}_\rho^{\leq \mathcal{D}}, \overline{D}_{\rho'}^{\leq \mathcal{D}} \right\rangle - 1 \right)^{|T|}, \end{aligned} \tag{6}$$

where $\overline{D}_\rho^{(i)}$ in the second line is the likelihood ratio for the i -th sample, the third equality follows almost identically to the proof of Claim 3.3 in Brennan et al. (2021), the fourth line follows from Hölder's inequality, and $\overline{D}_\rho$ in the last line is the likelihood ratio for any single-copy measurements.

Similar to Theorem 19, we also identify a condition under which the hypothesis testing problem is low-degree hard under the copy-wise model.

Condition 26 (Bounded $(\mathcal{D}, k)$ -likelihood ratio for each copy) *The hypothesis testing problem ρ_{mm} vs. $\rho \sim \mathcal{E}$ has an ε -bounded $(\mathcal{D}, k)$ -likelihood ratio for each copy if for any single-copy measurement strategy, we have*

$$\mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\left\langle \overline{D}_\rho^{\leq \mathcal{D}}, \overline{D}_{\rho'}^{\leq \mathcal{D}} \right\rangle - 1 \right)^k \leq \varepsilon.$$

Theorem 27 (Copy-wise low-degree hardness for single-copy measurements) *For a hypothesis testing problem ρ_{mm} vs. $\rho \sim \mathcal{E}$ satisfying Theorem 26 at $\varepsilon = 2^{-\tilde{\Omega}(k \log n)}$, i.e.,*

$$\mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\left\langle \overline{D}_\rho^{\leq \mathcal{D}}, \overline{D}_{\rho'}^{\leq \mathcal{D}} \right\rangle - 1 \right)^k \leq 2^{-\tilde{\Omega}(k \log n)}$$

for any single-copy measurement, weak detection is degree- $(\mathcal{D}, k)$ hard for any protocol that uses polynomially many non-adaptive single-copy PVMs $\{\{U_i | x\rangle \langle x| U_i^\dagger\}_{x=1}^{2^n}\}_{i=1}^m$.

We note that the hardness results in the above theorem is the same as Theorem 24 while Theorem 26 is stronger than approximate state 2-designs. Therefore, Theorem 27 is actually a weaker theorem than Theorem 24.

Proof We assume hypothesis testing problem ρ_{mm} vs. $\rho \sim \mathcal{E}$ satisfies $\varepsilon(\mathcal{D}, k)$ -likelihood ratio for each copy in Theorem 26, i.e.,

$$\mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\left\langle \overline{D}_\rho^{\leq \mathcal{D}}, \overline{D}_{\rho'}^{\leq \mathcal{D}} \right\rangle - 1 \right)^k \leq \varepsilon$$

for any single-copy measurement. We have the square degree- $(\mathcal{D}, k)$ advantage:

$$\begin{aligned} \left\| \mathbb{E}_{\rho \sim \mathcal{E}} \overline{D}_{\rho, m}^{\leq \mathcal{D}, k} - \mathbf{1} \right\|^2 &\leq \sum_{T \in [m], |T| \leq k, T \neq \emptyset} \left(\prod_{i \in T} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\left\langle \overline{D}_\rho^{(i) \leq \mathcal{D}}, \overline{D}_{\rho'}^{(i) \leq \mathcal{D}} \right\rangle - 1 \right)^{|T|} \right)^{1/|T|} \\ &\leq \sum_{t=1}^k \binom{m}{t} \varepsilon \\ &\leq km^k \varepsilon. \end{aligned}$$

Choosing $\varepsilon = 2^{-\tilde{\Omega}(k \log n)}$ we can finish the proof. ■

A priori it is unclear when Theorem 26 is satisfied. In the following, we show that approximate state k -designs provide one such example:

Lemma 28 (Approximate state k -design satisfies Theorem 26) *Given a hypothesis testing problem ρ_{mm} vs. $\rho \sim \mathcal{E}$ with $\mathcal{E}$ being an ε -approximate state k -design with $0 \leq \varepsilon \leq 1/2$, we have*

$$\mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\left\langle \overline{D}_\rho^{\leq \infty}, \overline{D}_{\rho'}^{\leq \infty} \right\rangle - 1 \right)^k \leq O(k^2 2^{k \log_2 k} (\varepsilon + 2^{-n})).$$

Proof We consider the expression

$$\begin{aligned} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\left\langle \overline{D}_\rho^{\leq \infty}, \overline{D}_{\rho'}^{\leq \infty} \right\rangle - 1 \right)^k &= \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\langle \overline{D}_\rho, \overline{D}_{\rho'} \rangle - 1 \right)^k \\ &= \sum_{j=0}^k \binom{k}{j} (-1)^j \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} 2^{jn} \left(\sum_{s \in \{0,1\}^n} D_\rho(s) D_{\rho'}(s) \right)^j. \end{aligned} \quad (7)$$

Let's further look at the expression

$$2^{jn} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\sum_{s \in \{0,1\}^n} D_\rho(s) D_{\rho'}(s) \right)^j = 2^{jn} \sum_{s_1, \dots, s_j} \mathbb{E}_\rho D_\rho(s_1) \cdots D_\rho(s_j) \mathbb{E}_{\rho'} D_{\rho'}(s_1) \cdots D_{\rho'}(s_j). \quad (8)$$

For a given list $s_1, \dots, s_j$, we denote by $\#y(s_1, \dots, s_j)$ the number of times the bit string y shows up in the list. We split the sum in (8) into j parts summing over the number r of distinct bit strings in the tuples $(s_1, \dots, s_j)$. In the following, $B(j)$ denotes the j th Bell number, which counts the number of partitions of a set of size j and $S(j, r)$ denotes the Stirling number of second kind, which

counts the number of partitions of a set of size j into r non-empty subsets. Suppose ρ and ρ' are drawn from the exact Haar measure, we have

$$\begin{aligned}
 & 2^{jn} \sum_{s_1, \dots, s_j} \mathbb{E}_\rho D_\rho(s_1) \cdots D_\rho(s_j) \mathbb{E}_{\rho'} D_{\rho'}(s_1) \cdots D_{\rho'}(s_j) \\
 &= \frac{2^{jn}}{d^2(d+1)^2 \cdots (d+j-1)^2} \left(\sum_{s_1, \dots, s_j} \prod_{y \in \{0,1\}^n} (\#y(s_1, \dots, s_j))!^2 \right) \\
 &\leq \frac{2^{jn}}{d^2(d+1)^2 \cdots (d+j-1)^2} \left(\sum_{r=1}^j (j-r)!^2 S(j, r) d(d-1) \cdots (d-r+1) \right),
 \end{aligned}$$

where the first equality follows from Theorem 11. The first inequality follows from organizing the sum by the number r of distinct bitstrings that appear, and then by the number of ways j bitstrings can have r distinct bitstrings. Here, $(j-r)!^2$ is the worst case upper bound on $\prod_{y \in \{0,1\}^n} (\#y(s_1, \dots, s_j))!^2$ when $(s_1, \dots, s_j)$ contains r distinct strings. Moreover, $S(j, r) d(d-1) \cdots (d-r+1)$ is a loose upper bound on the number of all such lists $(s_1, \dots, s_j)$. Using this, we can now further expand this sum

$$\begin{aligned}
 &\leq \frac{2^{jn}}{d^2(d+1)^2 \cdots (d+j-1)^2} \left(d \cdots (d-j+1) + \sum_{r=1}^{j-1} (j-r)!^2 B(j) d(d-1) \cdots (d-r+1) \right) \\
 &\leq 1 + B(j) \left(2j^2 2^{-n} + O \left(2^{-n} \sum_{l=0}^j 2^{2l \log_2(l) - ln} \right) \right) \\
 &= 1 + 2^{j \log(j)} \left(2j^2 2^{j \log(j)} 2^{-n} + O \left(2^{-n} \frac{1}{1 - 2^{2 \log_2(j) - n}} \right) \right) \\
 &= 1 + O(j^2 2^{j \log_2 j} 2^{-n}),
 \end{aligned}$$

Here we used that $S(j, j) = 1$ and a simple upper bound $B(j) \leq j!$. This is because any partition can be achieved as the cycle structure of a permutation.

However, ρ, ρ' are drawn from an ε -approximate state k -design here instead. Therefore, we have

$$(1 - 2\varepsilon)(1 + O(j^2 2^{-n})) \leq (8) \leq (1 + 2\varepsilon)(1 + O(j^2 2^{-n})),$$

where we used

$$\begin{aligned}
 1 - 2\varepsilon &\leq 1/(1 + \varepsilon) \\
 1/(1 - \varepsilon) &\leq 1 + 2\varepsilon.
 \end{aligned}$$

for $0 < \varepsilon \leq 1/2$. Plugging this equation into Eq. (7) yields

$$\begin{aligned}
\mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\left\langle \overline{D}_\rho^{\leq \infty}, \overline{D}_{\rho'}^{\leq \infty} \right\rangle - 1 \right)^k &= \sum_{j=0}^k \binom{k}{j} (-1)^j \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} 2^{jn} \left(\sum_{s \in \{0,1\}^n} D_\rho(s) D_{\rho'}(s) \right)^j \\
&\leq \sum_{j=0}^k \binom{k}{j} (-1)^j (1 + (-1)^j 2\varepsilon) (1 + O(j^2 2^j \log_2 j 2^{-n})) \\
&\leq \sum_{j=0}^k \binom{k}{j} (-1)^j + O(k^2 2^{-n} + \varepsilon) \sum_{j=0}^k \binom{k}{j} \\
&\leq O(k^2 2^{-n} + \varepsilon) \sum_{j=0}^k \binom{k}{j} \\
&\leq O(2^k (k^2 2^{k \log_2 k} 2^{-n} + \varepsilon)),
\end{aligned}$$

which finishes the proof for Theorem 28. ■

Combining Theorem 28 and Theorem 27, we have the following corollary:

Corollary 29 (Approximate state k -design indicates copy-wise low-degree hardness) *For the hypothesis testing problem ρ_{mm} vs. $\rho \sim \mathcal{E}$ with $\mathcal{E}$ a $2^{-\tilde{\Omega}(k \log n)}$ -approximate state k -design, weak detection is degree- $(\mathcal{D}, k)$ hard for any protocol that measures a polynomial number of non-adaptive single-copy PVMs $\{\{U_i | x\rangle \langle x| U_i^\dagger\}_{x=1}^{2^n}\}_{i=1}^m$.*

E.2. Round-based models with adaptivity within blocks

As shown in Figure 2(b), we consider a framework for learning with single-copy PVMs and restricted adaptivity called *round-based models*. In this model, we divide the m samples into m_1 blocks, each of m_0 samples.

The first model we consider is when we have $m_1 = \text{poly}(n)$ blocks, each consisting of $m_0 = \text{polylog}(n)$ copies that can be measured adaptively. While adaptivity is allowed *within* the blocks, no adaptivity is allowed *between blocks* (see LHS of Figure 2(b) for illustration). Roughly speaking, in this model the measurements within each block are allowed to be chosen arbitrarily adaptively based on readouts from earlier in the same block, but at the start of each new block the adaptivity “resets.” We consider a different notion of adaptivity in the next section. We will focus on the $(\mathcal{D}, k)$ copy-wise low-degree model throughout this subsection for technical convenience. We use $\overline{\Pi}$ between $\overline{D}$ to denote adaptive multiplication, and $\otimes$ to denote non-adaptive product.

We analyze adaptivity using the learning tree framework introduced in Aharonov et al. (2022); Bubeck et al. (2020); Chen et al. (2022a, 2023a); Chen and Gong (2023); Chen et al. (2024d). Roughly speaking, in this approach one models the learning protocol as a decision tree. Starting from the root node, we pick a PVM and measure a copy of the unknown quantum state using the picked PVM. Based on different measurement outcomes, we pick different PVMs in the different child nodes. For adaptive protocols with m measurements, we repeat this procedure for m times, resulting in a decision tree of depth m . As we choose a PVM with 2^n outcomes per measurement, we can map all possible histories to a 2^{mn} bit string. This fact is our crucial interface with the Boolean analytic techniques used in the previous section.

Formally, we define the protocol as below

Definition 30 (Round-based model with adaptivity within blocks) *Given $m = \text{poly}(n)$ copies of unknown n -qubit state ρ , a round-based model with adaptivity within blocks contains $m_1 = \text{poly}(n)$ blocks each of $m_0 = \text{polylog}(n)$ copies. The blocks are independent of each other, and no adaptivity is allowed among blocks. Each block can be represented as a rooted tree $\mathcal{T}_i, i \in [m_1]$ of depth m_0 with each node on the tree recording the measurement outcomes of the block so far. Each tree $\mathcal{T}_i$ has the following properties:*

- We assign a probability $p^\rho(u_i)$ for any node u_i on the tree $\mathcal{T}_i$.
- The probability assigned to the root r_i of the tree $\mathcal{T}_i$ is $p^\rho(r_i) = 1$.
- At each non-leaf node u_i , we measure ρ using an adaptively chosen PVM $\{U_{u_i}|s\rangle\langle s|U_{u_i}^\dagger\}$, which results in a classical outcome s . Each child node v_i corresponding to the classical outcome s of the node u_i is connected through the edge $e_{u_i,s}$.
- If v_i is the child of u_i through the edge $e_{u_i,s}$, the probability to traverse this edge is $p^\rho(s|u_i) = \langle s|U_{u_i}^\dagger\rho U_{u_i}|s\rangle$. Then

$$p^\rho(v_i) = p^\rho(u_i) \cdot p^\rho(s|u_i) = p^\rho(u_i) \cdot \langle s|U_{u_i}^\dagger\rho U_{u_i}|s\rangle.$$
- Each root-to-leaf path is of length m_0 . Note that for a leaf node ℓ_i , $p^\rho(\ell_i)$ is the probability for the classical memory to be in state ℓ_i at the end of the learning protocol. Denote the set of leaves of $\mathcal{T}_i$ by $\text{leaf}(\mathcal{T}_i)$

We will need the following condition:

Condition 31 ($(M, \varepsilon, \mathcal{D}, k)$ -condition for restricted adaptivity within blocks) *Consider the hypothesis testing problem ρ_{mm} vs. $\rho \sim \mathcal{E}$. We denote the output history as $\mathbf{s} = (s_1, \dots, s_{m_0}) \in \{0, 1\}^{m_0 n}$ with each $s_i \in \{0, 1\}^n$ for $i \in [m_0]$. We assume that for any composite block of km_0 samples (with history $\mathbf{s} = (s_1, \dots, s_{km_0})$), we have*

$$\mathbb{E}_{\mathbf{s}_{1:km_0-1}} \mathbb{E}_{\rho \sim \mathcal{E}} \left[\left(\prod_{i=1}^{km_0-1} \overline{D}_\rho^{\leq \mathcal{D}}(s_i) \right)^{2(km_0-1)} \right] \leq M.$$

We further assume that for any single-copy measurement, we have

$$\mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\left\langle \overline{D}_\rho^{\leq \mathcal{D}}, \overline{D}_{\rho'}^{\leq \mathcal{D}} \right\rangle - 1 \right)^{2km_0} \leq \varepsilon.$$

Given the above condition, we can prove the following result:

Theorem 32 (Low-degree hardness for round-based model with adaptivity within blocks) *For the hypothesis testing problem ρ_{mm} vs. $\rho \sim \mathcal{E}$ satisfying an $(M, 2^{-\Omega(k^3 m_0 \log k \log m_1 \log m_0 \log M)}, \mathcal{D}, k)$ -condition for restricted adaptivity within blocks (see Theorem 31), i.e.,*

$$\begin{aligned} \mathbb{E}_{\mathbf{s}_{1:km_0-1}} \mathbb{E}_{\rho \sim \mathcal{E}} \left[\left(\prod_{i=1}^{km_0-1} \overline{D}_\rho^{\leq \mathcal{D}}(s_i) \right)^{2(km_0-1)} \right] &\leq M, \quad \text{and} \\ \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\left\langle \overline{D}_\rho^{\leq \mathcal{D}}, \overline{D}_{\rho'}^{\leq \mathcal{D}} \right\rangle - 1 \right)^{2km_0} &\leq 2^{-\Omega(k^3 m_0 \log k \log m_1 \log m_0 \log M)} \end{aligned}$$

for each block of m_0 samples and any single-copy measurement, weak detection is degree- $(\mathcal{D}, k)$ hard for any adaptivity within round-based model protocol that makes polynomially many PVMs $\{\{U_i | x\rangle \langle x| U_i^\dagger\}_{x=1}^{2^n}\}_{i=1}^m$. These protocols use $m_1 = \text{poly}(n)$ blocks each with $m_0 = m/m_1 = \text{polylog}(n)$ samples, and adaptivity is only allowed in each block.

Proof We assume the $(M, \varepsilon, \mathcal{D}, k)$ -assumption for restricted adaptivity within blocks in Theorem 31 holds. We start with a single block of size m_0 . Recall that

$$\overline{D}_\rho^{(i)}(s_i) = 2^n \text{tr}(\rho |\phi_{s_i}\rangle \langle \phi_{s_i}|), \quad \overline{D}_{\rho, m_0}(\mathbf{s}) = 2^{m_0 n} \prod_{i=1}^{m_0} \text{tr}(\rho |\phi_{s_i}\rangle \langle \phi_{s_i}|).$$

Our first step is to write $\overline{D}_{\rho, m_0}$ as a product of functions with degree $\mathcal{D}$ on each copy:

$$\overline{D}_{\rho, m_0}^{\leq \mathcal{D}, \infty}(\mathbf{s}) = \prod_{i=1}^{m_0} \overline{D}_{\rho'}^{(i) \leq \mathcal{D}}(s_i).$$

We have

$$\begin{aligned} \left\| \mathbb{E}_{\rho \sim \mathcal{E}} \overline{D}_{\rho, m_0}^{\leq \mathcal{D}, k} - \mathbf{1} \right\|^2 &= \mathbb{E}_{\mathbf{s}} (\mathbb{E}_{\rho \sim \mathcal{E}} \overline{D}_{\rho, m_0}^{\leq \mathcal{D}, k} - 1)^2 \\ &= \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left\langle \overline{D}_{\rho, m_0}^{\leq \mathcal{D}, k}, \overline{D}_{\rho', m_0}^{\leq \mathcal{D}, k} \right\rangle - 1 \\ &= \mathbb{E}_{\mathbf{s}} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \overline{D}_{\rho, m_0}^{\leq \mathcal{D}, k}(\mathbf{s}) \overline{D}_{\rho', m_0}^{\leq \mathcal{D}, k}(\mathbf{s}) - 1. \end{aligned}$$

Instead of choosing $\chi_T(\mathbf{s})$ in the qubit-wise, we now choose $\chi_T(\mathbf{s})$ in the copy-wise. We denote

$$\overline{D}_{\rho, m_0}^{\leq \mathcal{D}}(\mathbf{s}) = \sum_{T \subseteq [m_0]} \alpha_T^\rho \chi_T(\mathbf{s}), \quad \alpha_T^\rho = \frac{1}{2^{nm_0}} \sum_{\mathbf{s}} \overline{D}_{\rho, m_0}^{\leq \mathcal{D}}(\mathbf{s}) \chi_T(\mathbf{s}) = \mathbb{E}_{\mathbf{s}} \overline{D}_{\rho, m_0}^{\leq \mathcal{D}}(\mathbf{s}) \chi_T(\mathbf{s}).$$

Again, we focus on the Fourier basis in this work. The low-degree likelihood ratio satisfies:

$$\overline{D}_{\rho, m_0}^{\leq \mathcal{D}, k}(\mathbf{s}) = \sum_{|T| \leq k} \alpha_T^\rho \chi_T(\mathbf{s}).$$

We compute the square degree- $(\mathcal{D}, k)$ advantage as

$$\begin{aligned} \left\| \mathbb{E}_{\rho \sim \mathcal{E}} \overline{D}_{\rho, m_0}^{\leq \mathcal{D}, k} - \mathbf{1} \right\|^2 &= \mathbb{E}_{\mathbf{s}} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \overline{D}_{\rho, m_0}^{\leq \mathcal{D}, k}(\mathbf{s}) \overline{D}_{\rho', m_0}^{\leq \mathcal{D}, k}(\mathbf{s}) - 1 \\ &= \mathbb{E}_{\mathbf{s}} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \sum_{|T|, |T'| \leq k} \alpha_T^\rho \alpha_{T'}^{\rho'} \chi_T(\mathbf{s}) \chi_{T'}(\mathbf{s}) - 1 \\ &= \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \sum_{|T| \leq k} \alpha_T^\rho \alpha_T^{\rho'} - 1 \\ &= \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{s}, \mathbf{s}'} \sum_{|T| \leq k} \overline{D}_{\rho, m_0}^{\leq \mathcal{D}}(\mathbf{s}) \overline{D}_{\rho', m_0}^{\leq \mathcal{D}}(\mathbf{s}') \chi_T(\mathbf{s}) \chi_T(\mathbf{s}') - 1 \\ &= \sum_{|T| \leq k, T \neq \emptyset} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{s}, \mathbf{s}'} \overline{D}_{\rho, m_0}^{\leq \mathcal{D}}(\mathbf{s}) \overline{D}_{\rho', m_0}^{\leq \mathcal{D}}(\mathbf{s}') \chi_T(\mathbf{s}) \chi_T(\mathbf{s}'), \end{aligned}$$

where the fourth line follows from the fact that $\langle \chi_T(\mathbf{s}), \chi_{T'}(\mathbf{s}) \rangle = \delta_{T,T'}$. However, we can not use the previous Hölder inequality directly due to adaptivity. To this end, we have to use the recursive trick. We denote

$$A_{m_0,k} = \left\| \mathbb{E}_{\rho \sim \mathcal{E}} \bar{D}_{\rho,m_0}^{\leq \mathcal{D},k} - \mathbf{1} \right\|^2$$

for convenience. We thus have:

$$A_{m_0,k} = A_{m_0-1,k} + \sum_{|T'| \leq k-1, T' \neq \emptyset} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{s}, \mathbf{s}'} \bar{D}_{\rho,m_0}^{\leq \mathcal{D}}(\mathbf{s}) \bar{D}_{\rho',m_0}^{\leq \mathcal{D}}(\mathbf{s}') \chi_{T' \cup [m_0]}(\mathbf{s}) \chi_{T' \cup [m_0]}(\mathbf{s}'). \quad (9)$$

We now consider the second term. We have (we always assume $T' \neq \emptyset$ in the following)

$$\begin{aligned} & \sum_{|T'| \leq k-1} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{s}, \mathbf{s}'} \bar{D}_{\rho,m_0}^{\leq \mathcal{D}}(\mathbf{s}) \bar{D}_{\rho',m_0}^{\leq \mathcal{D}}(\mathbf{s}') \chi_{T' \cup [m_0]}(\mathbf{s}) \chi_{T' \cup [m_0]}(\mathbf{s}') \\ = & \sum_{|T'| \leq k-1} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{s}_{1:m_0-1}, \mathbf{s}'_{1:m_0-1}} \bar{D}_{\rho,m_0-1}^{\leq \mathcal{D}}(\mathbf{s}_{1:m_0-1}) \bar{D}_{\rho',m_0-1}^{\leq \mathcal{D}}(\mathbf{s}'_{1:m_0-1}) \chi_{T'}(\mathbf{s}_{1:m_0-1}) \chi_{T'}(\mathbf{s}'_{1:m_0-1}) \\ & \mathbb{E}_{\mathbf{s}_{m_0}, \mathbf{s}'_{m_0}} [\bar{D}_{\rho}^{(m_0) \leq k}(s_{m_0}) \bar{D}_{\rho'}^{(m_0) \leq k}(s'_{m_0}) \chi_{[s_{m_0}]}(s_{m_0}) \chi_{[s'_{m_0}]}(s'_{m_0})] \\ \leq & \left(\mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{s}_{1:m_0-1}, \mathbf{s}'_{1:m_0-1}} \left(\sum_{|T'| \leq k-1} \bar{D}_{\rho,m_0-1}^{\leq \mathcal{D}}(\mathbf{s}_{1:m_0-1}) \bar{D}_{\rho',m_0-1}^{\leq \mathcal{D}}(\mathbf{s}'_{1:m_0-1}) \chi_{T'}(\mathbf{s}_{1:m_0-1}) \chi_{T'}(\mathbf{s}'_{1:m_0-1}) \right)^2 \right)^{\frac{1}{2}} \\ & \cdot \varepsilon^{\frac{1}{2km_0}} \\ \leq & km_0^k \varepsilon^{\frac{1}{2km_0}} \mathbb{E}_{\rho \sim \mathcal{E}} \mathbb{E}_{\mathbf{s}_{1:m_0-1}} \left[\left(\prod_{i=1}^{m-1} \bar{D}_{\rho}^{(i) \leq \mathcal{D}}(s_i) \right)^2 \right] \\ \leq & km_0^k \varepsilon^{\frac{1}{2km_0}} \mathbb{E}_{\mathbf{s}_{1:m_0-1}} \mathbb{E}_{\rho \sim \mathcal{E}} \left[\left(\bar{D}_{\rho}^{(i) \leq \mathcal{D}}(s_i) \right)^{2(m_0-1)} \right], \end{aligned} \quad (10)$$

where the second step follows from implementing Hölder's inequality twice and $(M, \varepsilon, \mathcal{D}, k)$ -assumption for restricted adaptivity within blocks in Theorem 31, and the fact that

$$\left(\sum_{|T'| \leq k-1} \chi_{T'}(\mathbf{s}_{1:m_0-1}) \chi_{T'}(\mathbf{s}'_{1:m_0-1}) \right)^2 \leq \left(\sum_{t=1}^{k-1} \binom{m_0}{k} \right)^2 \leq k^2 m_0^k$$

and the third step follows from the Cauchy-Schwarz inequality. This indicates that

$$\left\| \mathbb{E}_{\rho \sim \mathcal{E}} \bar{D}_{\rho,m_0}^{\leq \mathcal{D},k} - \mathbf{1} \right\|^2 = A_{m_0,k} = km_0^{k+1} \varepsilon^{\frac{1}{2km_0}} \mathbb{E}_{\mathbf{s}_{1:m_0-1}} \mathbb{E}_{\rho \sim \mathcal{E}} \left[\left(\bar{D}_{\rho}^{(i) \leq \mathcal{D}}(s_i) \right)^{2(m_0-1)} \right]. \quad (11)$$

We then consider the full likelihood ratio $\overline{D}_{\rho,m}(\mathbf{S})$ for the complete bit string output $\mathbf{S} = (s_1, \dots, s_{m_1}) \in \{0, 1\}^{m_0 m_1}$ for $m = m_0 m_1$ copies using the non-adaptive result:

$$\begin{aligned}
\left\| \mathbb{E}_{\rho \sim \mathcal{E}} \overline{D}_{\rho,m}^{\leq \mathcal{D},k} - \mathbf{1} \right\|^2 &= \mathbb{E}_{\mathbf{S}} (\mathbb{E}_{\rho \sim \mathcal{E}} \overline{D}_{\rho,m}^{\leq \mathcal{D},k} - 1)^2 = \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left\langle \overline{D}_{\rho,m}^{\leq \mathcal{D},k}, \overline{D}_{\rho',m}^{\leq \mathcal{D},k} \right\rangle - 1 \\
&= \mathbb{E}_{\mathbf{S}} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \overline{D}_{\rho,m}^{\leq \mathcal{D},k}(\mathbf{S}) \overline{D}_{\rho',m}^{\leq \mathcal{D},k}(\mathbf{S}) - 1 \\
&\leq \sum_{t=1}^k \binom{m}{t} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\left\langle \overline{D}_{\rho,m_0}^{\leq \mathcal{D},k}, \overline{D}_{\rho',m_0}^{\leq \mathcal{D},k} \right\rangle - 1 \right)^t \\
&\leq k m_1^k \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\left\langle \overline{D}_{\rho,m_0}^{\leq \mathcal{D},k}, \overline{D}_{\rho',m_0}^{\leq \mathcal{D},k} \right\rangle - 1 \right)^k \\
&\leq k m_1^k \sum_{j=0}^k \binom{k}{j} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\mathbb{E}_{\mathbf{s} \in \{0,1\}^{m_0 n}} \overline{D}_{\rho,m_0}^{\leq \mathcal{D},k}(\mathbf{s}) \overline{D}_{\rho',m_0}^{\leq \mathcal{D},k}(\mathbf{s}) \right)^j.
\end{aligned}$$

We note that

$$\begin{aligned}
\mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\mathbb{E}_{\mathbf{s} \in \{0,1\}^{m_0 n}} \overline{D}_{\rho,m_0}^{\leq \mathcal{D},k}(\mathbf{s}) \overline{D}_{\rho',m_0}^{\leq \mathcal{D},k}(\mathbf{s}) \right)^j &= \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{s}_1, \dots, \mathbf{s}_j \in \{0,1\}^{m_0 n}} \overline{D}_{\rho,m_0}^{\leq \mathcal{D},k}(\mathbf{s}_1) \dots \overline{D}_{\rho,m_0}^{\leq \mathcal{D},k}(\mathbf{s}_j) \overline{D}_{\rho',m_0}^{\leq \mathcal{D},k}(\mathbf{s}_1) \dots \overline{D}_{\rho',m_0}^{\leq \mathcal{D},k}(\mathbf{s}_j) \\
&\leq \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{s} \in \{0,1\}^n} \overline{D}_{\rho}^{\leq \mathcal{D},jk}(\mathbf{s}) \overline{D}_{\rho'}^{\leq \mathcal{D},jk}(\mathbf{s}) \\
&\leq 1 + A_{km_0, jk}
\end{aligned}$$

as choosing $\mathbf{s}_1, \dots, \mathbf{s}_j$ is a special case for adaptively choosing all jm_0 copies and there can be at most copy-wise degree of $(\mathcal{D}, jk)$. We thus have

$$\begin{aligned}
\left\| \mathbb{E}_{\rho \sim \mathcal{E}} \overline{D}_{\rho,m}^{\leq \mathcal{D},k} - \mathbf{1} \right\|^2 &\leq k m_1^k \sum_{j=0}^k \binom{k}{j} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\mathbb{E}_{\mathbf{s} \in \{0,1\}^{m_0 n}} \overline{D}_{\rho,m_0}^{\leq \mathcal{D},k}(\mathbf{s}) \overline{D}_{\rho',m_0}^{\leq \mathcal{D},k}(\mathbf{s}) \right)^j \\
&\leq k m_1^k 2^k A_{km_0, k^2} \\
&\leq k 2^k m_1^k \cdot k^2 (km_0)^{k^2} \varepsilon^{\frac{1}{2km_0}} \mathbb{E}_{\mathbf{s}_{1:km_0-1}} \mathbb{E}_{\rho \sim \mathcal{E}} \left[\left(\overline{D}_{\rho}^{(i) \leq \mathcal{D}}(s_i) \right)^{2(km_0-1)} \right].
\end{aligned}$$

Recall that $(M, \varepsilon, \mathcal{D}, k)$ -assumption for restricted adaptivity within blocks in Theorem 31 indicates that

$$\mathbb{E}_{\mathbf{s}_{1:km_0-1}} \mathbb{E}_{\rho \sim \mathcal{E}} \left[\left(\prod_{i=1}^{km_0-1} \overline{D}_{\rho}^{\leq \mathcal{D}}(s_i) \right)^{2(km_0-1)} \right] \leq M.$$

We can thus compute the overall square degree- $(\mathcal{D}, k)$ advantage by

$$\begin{aligned}
\left\| \mathbb{E}_{\rho \sim \mathcal{E}} \overline{D}_{\rho,m}^{\leq \mathcal{D},k} - \mathbf{1} \right\|^2 &\leq k 2^k m_1^k \cdot k^2 (km_0)^{k^2} \varepsilon^{\frac{1}{2km_0}} \mathbb{E}_{\mathbf{s}_{1:km_0-1}} \mathbb{E}_{\rho \sim \mathcal{E}} \left[\left(\overline{D}_{\rho}^{(i) \leq \mathcal{D}}(s_i) \right)^{2(km_0-1)} \right] \\
&\leq k 2^k m_1^k \cdot k^2 (km_0)^{k^2} \varepsilon^{\frac{1}{2km_0}} M \\
&= 2^{k^2 \log k \log m_0 \log m_1} \varepsilon^{\frac{1}{2km_0}} M
\end{aligned}$$

By choosing

$$\varepsilon = 2^{-\Omega(k^3 m_0 \log k \log m_1 \log m_0 \log M)},$$

we can ensure the square degree- $(\mathcal{D}, k)$ advantage below $o(1)$, which finishes the proof for Theorem 32. $\blacksquare$

At a first glance, the $(M, \varepsilon, \mathcal{D}, k)$ -assumption for restricted adaptivity within blocks in Theorem 31 in Theorem 32 seems complicated. In the following, we show that approximate state designs satisfy such an assumption. Formally, we have the following corollary (see Appendix K.1 for the proof):

Corollary 33 (Approximate state $2km_0$ -design indicates hardness for round-based model with adaptivity within blocks)

For the hypothesis testing problem ρ_{mm} vs. $\rho \sim \mathcal{E}$ with $\mathcal{E}$ a $2^{-\Omega(k^4 m_0^2 \log^2 k \log m_1 \log^2 m_0)}$ -approximate state $2km_0$ -design, weak detection is degree- $(\mathcal{D}, k)$ hard for any adaptivity within round-based model protocol that makes polynomially many PVMs $\{\{U_i | x\rangle \langle x| U_i^\dagger\}_{x=1}^{2^n}\}_{i=1}^m$. These protocols use $m_1 = \text{poly}(n)$ blocks each with $m_0 = m/m_1 = \text{polylog}(n)$ samples, and adaptivity is only allowed in each block.

E.3. Round-based models with adaptivity among blocks

The second model we consider is when we have $m_1 = \text{polylog}(n)$ adaptive blocks of $m_0 = \text{poly}(n)$ non-adaptive copies within each block. Roughly speaking, at the start of each block, we can prepare a fixed measurement strategy to apply to the next $\text{poly}(n)$ copies, where the strategy is chosen fully adaptively based on all readouts from the previous blocks. We formalize this as follows:

Definition 34 (Round-based model with adaptivity among blocks) *Given $m = \text{poly}(n)$ copies of unknown n -qubit state ρ , a round-based model with adaptivity within blocks contains $m_1 = \text{polylog}(n)$ blocks each of $m_0 = \text{poly}(n)$ copies. While no adaptivity is allowed within each block, how to measure a whole new block can be adaptively decided by previous history. The whole process can be as a rooted tree $\mathcal{T}$ of depth m_1 with each node on the tree recording the measurement outcomes of a block. It has the following properties:*

- We assign a probability $p^\rho(u)$ for any node u on the tree $\mathcal{T}$.
- The probability assigned to the root r of the tree $\mathcal{T}$ is $p^\rho(r) = 1$.
- At each non-leaf node u , we measure m_1 copies of ρ using an m_0 independent PVMs $\{U_{u,i} | s\rangle \langle s| U_{u,i}^\dagger, i \in [m_0]\}$, which results in a classical outcome $\mathbf{s}_u = s_u^{(1)}, \dots, s_u^{(m_0)}$. Each child node v corresponding to the classical outcome $\mathbf{s}_u$ of the node u is connected through the edge $e_{u,s}$.
- If v is the child of u through the edge $e_{u,s}$, the probability to traverse this edge is $p^\rho(s|u) = \prod_{i=1}^{m_0} \langle s| U_{u,i}^\dagger \rho U_{u,i} | s\rangle$. Then

$$p^\rho(v) = p^\rho(u) \cdot p^\rho(s|u) = p^\rho(u) \cdot \prod_{i=1}^{m_0} \langle s| U_{u,i}^\dagger \rho U_{u,i} | s\rangle.$$

- Each root-to-leaf path is of length m_1 . Note that for a leaf node ℓ , $p^\rho(\ell)$ is the probability for the classical memory to be in state ℓ at the end of the learning protocol. Denote the set of leaves of $\mathcal{T}$ by $\text{leaf}(\mathcal{T})$

We will need the following condition:

Condition 35 $((M', \varepsilon, \mathcal{D}, k)$ -condition for restricted adaptivity among blocks) *Consider the hypothesis testing problem ρ_{mm} vs. $\rho \sim \mathcal{E}$. We assume that for any single-copy measurement with output s , we have*

$$\mathbb{E}_{\rho \sim \mathcal{E}} \left(\overline{D}_{\rho}^{\leq \mathcal{D}}(s) - 1 \right)^{2(m_1-1)} \leq M'.$$

We further assume that for any single-copy measurement, we have

$$\mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\left\langle \overline{D}_{\rho}^{\leq \mathcal{D}}, \overline{D}_{\rho'}^{\leq \mathcal{D}} \right\rangle - 1 \right)^{2k} \leq \varepsilon.$$

Given the above condition, we can prove the following result:

Theorem 36 (Low-degree hardness for round-based model with adaptivity among blocks) *For the hypothesis testing problem ρ_{mm} vs. $\rho \sim \mathcal{E}$ satisfying a $(M', 2^{-\Omega(km_1 \log k \log m_0 \log M')}, \mathcal{D}, k)$ -condition for restricted adaptivity among blocks (see Theorem 35), i.e.,*

$$\begin{aligned} \mathbb{E}_{\rho \sim \mathcal{E}} \left(\overline{D}_{\rho}^{\leq \mathcal{D}}(s) - 1 \right)^{2(m_1-1)} &\leq M' \\ \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\left\langle \overline{D}_{\rho}^{\leq \mathcal{D}}, \overline{D}_{\rho'}^{\leq \mathcal{D}} \right\rangle - 1 \right)^{2k} &\leq 2^{-\Omega(km_1 \log k \log m_0 \log m_1 \log M')} \end{aligned}$$

for any single-copy PVM, weak detection is degree- $(\mathcal{D}, k)$ hard for any adaptivity among round-based model protocol that makes polynomially many single-copy PVMs $\{\{U_i | x\rangle \langle x| U_i^\dagger\}_{x=1}^{2^n}\}_{i=1}^m$. These protocols use $m_1 = \text{polylog}(n)$ blocks each with $m_0 = m/m_1 = \text{poly}(n)$ samples, and adaptivity is only allowed among blocks.

Intuitively, the proof of Theorem 36 also uses the recursive induction trick. However, instead of inducting on each single copy, we should induct on each block instead.

Proof The first step is to project the overall likelihood-ratio $\overline{D}_{\rho, m}^{\leq \mathcal{D}}(\mathbf{S})$ for $\mathbf{S} = (s_1, \dots, s_{m_1}) \in \{0, 1\}^{mn}$ into product of low-degree blocks as $\overline{D}_{\rho, m}^{\leq \mathcal{D}}(\mathbf{S}) = \prod_{i=1}^{m_1} \overline{D}_{\rho, m_0}^{\leq \mathcal{D}}(s_i)$, where each $s_i \in \{0, 1\}^{m_0 n}$ is the history in block $i = 1, \dots, m_1$. Note that in the copy-wise low-degree model, we always have $\overline{D}_{\rho, m}^{\leq \mathcal{D}, k}(\mathbf{S})$ included in $\prod_{i=1}^{m_1} \overline{D}_{\rho, m_0}^{\leq \mathcal{D}, k}(s_i)$. In the following, we abuse the notation a little and instead denote

$$\overline{D}_{\rho, m}^{\leq \mathcal{D}} = \prod_{i=1}^{m_1} \overline{D}_{\rho, m_0}^{\leq \mathcal{D}, k}(s_i).$$

We use the Fourier basis for each block $\chi_T(\mathbf{S})$ for $T \subseteq [m_1]$ here and decompose $\overline{D}_{\rho, m}^{\leq \mathcal{D}}(\mathbf{S})$ as

$$\overline{D}_{\rho, m}^{\leq \mathcal{D}}(\mathbf{S}) = \sum_{|T| \subseteq [m_1]} \alpha_T^\rho \chi_T(\mathbf{S}), \quad \alpha_T^\rho = \frac{1}{2^{nm}} \sum_{\mathbf{S}} \overline{D}_{\rho, m}^{\leq \mathcal{D}}(\mathbf{S}) \chi_T(\mathbf{S}) = \mathbb{E}_{\mathbf{S}} \overline{D}_{\rho, m}^{\leq \mathcal{D}}(\mathbf{S}) \chi_T(\mathbf{S}).$$

By the definition of copy-wise square degree- $(\mathcal{D}, k)$ advantage, we have

$$\left\| \mathbb{E}_{\rho \sim \mathcal{E}} \overline{D}_{\rho, m}^{\leq \mathcal{D}, k} - \mathbf{1} \right\|^2 \leq \sum_{|T| \leq k, T \neq \emptyset} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{S}, \mathbf{S}'} \overline{D}_{\rho, m}^{\leq \mathcal{D}}(\mathbf{S}) \overline{D}_{\rho', m}^{\leq \mathcal{D}}(\mathbf{S}') \chi_T(\mathbf{S}) \chi_T(\mathbf{S}').$$

We denote the above quantity as $B_{m_1,k}$. We have the following induction by dividing T into those with the last block or without:

$$B_{m_1,k} = B_{m_1-1,k} + \sum_{|T'| \leq k-1, T' \neq \emptyset} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{S}, \mathbf{S}'} \bar{D}_{\rho, m}^{\leq \mathcal{D}}(\mathbf{S}) \bar{D}_{\rho', m}^{\leq \mathcal{D}}(\mathbf{S}') \chi_{T' \cup [m_1]}(\mathbf{S}) \chi_{T' \cup [m_1]}(\mathbf{S}').$$

We now consider the second term. We have (we always assume $T' \neq \emptyset$ in the following)

$$\begin{aligned} & \sum_{|T'| \leq k-1, T' \neq \emptyset} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{S}, \mathbf{S}'} \bar{D}_{\rho, m}^{\leq \mathcal{D}}(\mathbf{S}) \bar{D}_{\rho', m}^{\leq \mathcal{D}}(\mathbf{S}') \chi_{T' \cup [m_1]}(\mathbf{S}) \chi_{T' \cup [m_1]}(\mathbf{S}') \\ = & \sum_{|T'| \leq k-1} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{S}_{1:m-m_0}, \mathbf{S}'_{1:m-m_0}} \bar{D}_{\rho, m-m_0}^{\leq \mathcal{D}}(\mathbf{S}_{1:m-m_0}) \bar{D}_{\rho', m-m_0}^{\leq \mathcal{D}}(\mathbf{S}'_{1:m-m_0}) \chi_{T'}(\mathbf{S}_{1:m-m_0}) \chi_{T'}(\mathbf{S}'_{1:m-m_0}) \\ & \mathbb{E}_{\mathbf{s}_{m_0}, \mathbf{s}'_{m_0}} [\bar{D}_{\rho, m_0}^{\leq \mathcal{D}, k}(\mathbf{s}_{m_0}) \bar{D}_{\rho', m_0}^{\leq \mathcal{D}, k}(\mathbf{s}'_{m_0}) \chi_{[s_{m_0}]}(\mathbf{s}_{m_0}) \chi_{[s'_{m_0}]}(\mathbf{s}'_{m_0})] \\ \leq & \left(\mathbb{E}_{\substack{\mathbf{S}_{1:m-m_0}, \mathbf{S}'_{1:m-m_0} \\ \rho, \rho' \sim \mathcal{E}}} \left(\sum_{|T'| \leq k-1} \bar{D}_{\rho, m-m_0}^{\leq \mathcal{D}}(\mathbf{S}_{1:m-m_0}) \bar{D}_{\rho', m-m_0}^{\leq \mathcal{D}}(\mathbf{S}'_{1:m-m_0}) \chi_{T'}(\mathbf{S}_{1:m-m_0}) \chi_{T'}(\mathbf{S}'_{1:m-m_0}) \right)^2 \right)^{\frac{1}{2}} \\ & \left(\mathbb{E}_{\substack{\mathbf{S}_{1:m-m_0}, \mathbf{S}'_{1:m-m_0} \\ \rho, \rho' \sim \mathcal{E}}} \left(\langle \bar{D}_{\rho, m_0}^{\leq \mathcal{D}, k}, \bar{D}_{\rho', m_0}^{\leq \mathcal{D}, k} \rangle - 1 \right)^2 \right)^{1/2} \\ \leq & km_1^k \mathbb{E}_{\mathbf{S}_{1:m-m_0}, \rho \sim \mathcal{E}} [\bar{D}_{\rho, m-m_0}^{\leq \mathcal{D}}(\mathbf{S}_{1:m-m_0})^2] \cdot \left(\mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\sum_{\substack{B \neq \emptyset, |B| \leq k \\ B \subseteq [m_0]}} \prod_{j \in B} (\langle \bar{D}_{\rho}^{(j) \leq \mathcal{D}}, \bar{D}_{\rho'}^{(j) \leq \mathcal{D}} \rangle - 1) \right)^2 \right)^{1/2} \\ \leq & k^2 m^k \mathbb{E}_{\mathbf{S}_{1:m-m_0}, \rho \sim \mathcal{E}} [\bar{D}_{\rho, m-m_0}^{\leq \mathcal{D}}(\mathbf{S}_{1:m-m_0})^2] \cdot \left(\mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\prod_{\substack{B \neq \emptyset, |B| \leq k \\ B \subseteq [m_0], j \in B}} (\langle \bar{D}_{\rho}^{(j) \leq \mathcal{D}}, \bar{D}_{\rho'}^{(j) \leq \mathcal{D}} \rangle - 1) \right)^2 \right)^{1/2} \\ \leq & k^2 m^k \mathbb{E}_{\mathbf{S}_{1:m-m_0}, \rho \sim \mathcal{E}} [\bar{D}_{\rho, m-m_0}^{\leq \mathcal{D}}(\mathbf{S}_{1:m-m_0})^2] \cdot \left(\mathbb{E}_{\rho, \rho' \sim \mathcal{E}} (\langle \bar{D}_{\rho}^{\leq \mathcal{D}}, \bar{D}_{\rho'}^{\leq \mathcal{D}} \rangle - 1)^{2k} \right)^{1/2}, \end{aligned}$$

where the second step follows from Cauchy-Swartz inequality, the third step uses the fact that

$$\sum_{|T'| \leq k-1} \chi_{T'}(\mathbf{S}_{1:m-m_0}) \chi_{T'}(\mathbf{S}'_{1:m-m_0}) \leq \sum_{t=1}^{k-1} \binom{m_1-1}{t} \leq km_1^k$$

and the non-adaptive result for each block, and the fifth step follows from Hölder's inequality. By the $(M', \varepsilon, \mathcal{D}, k)$ -assumption for restricted adaptivity among blocks, we further have

$$\begin{aligned} & \sum_{|T'| \leq k-1, T' \neq \emptyset} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{S}, \mathbf{S}'} \bar{D}_{\rho, m}^{\leq \mathcal{D}}(\mathbf{S}) \bar{D}_{\rho', m}^{\leq \mathcal{D}}(\mathbf{S}') \chi_{T' \cup [m_1]}(\mathbf{S}) \chi_{T' \cup [m_1]}(\mathbf{S}') \\ & \leq k^2 m^k \varepsilon^{1/2} \mathbb{E}_{\mathbf{S}_{1:m-m_0}, \rho \sim \mathcal{E}} [\bar{D}_{\rho, m-m_0}^{\leq \mathcal{D}}(\mathbf{S}_{1:m-m_0})^2]. \end{aligned}$$

Now, we consider the term $\mathbb{E}_{\mathbf{S}_{1:m-m_0}, \rho \sim \mathcal{E}} \left[\overline{D}_{\rho, m-m_0}^{\leq cD} (\mathbf{S}_{1:m-m_0})^2 \right]$. It can be computed by Hölder's inequality as:

$$\begin{aligned} \mathbb{E}_{\mathbf{S}_{1:m-m_0}, \rho \sim \mathcal{E}} \left[\overline{D}_{\rho, m-m_0}^{\leq \mathcal{D}} (\mathbf{S}_{1:m-m_0})^2 \right] &= \mathbb{E}_{\mathbf{S}_{1:m-m_0}} \mathbb{E}_{\rho \sim \mathcal{E}} \left[\prod_{i=1}^{m_1-1} \overline{D}_{\rho, m_0}^{\leq \mathcal{D}, k} (\mathbf{s}_i)^2 \right] \\ &\leq \mathbb{E}_{\mathbf{S}_{1:m-m_0}} \prod_{i=1}^{m_1-1} \left(\mathbb{E}_{\rho \sim \mathcal{E}} \left[D_{\rho, m_0}^{\leq \mathcal{D}} (\mathbf{s}_i)^{2(m_1-1)} \right] \right)^{1/(m_1-1)} \end{aligned}$$

Here, we consider the maximal value of $\mathbb{E}_{\rho \sim \mathcal{E}} \left[D_{\rho, m_0}^{\leq \mathcal{D}, k} (\mathbf{s}_i)^{2(m_1-1)} \right]$. Given a term $\overline{D}_{\rho, m_0}^{\leq \mathcal{D}, k} (\mathbf{s}_i)^l$, fix $B_i^1, \dots, B_i^l \subseteq [m_0]$ with size $\leq k$, we can denote $C_0, \dots, C_l$ as the maximal subset of $[m_0]$ that is covered by exactly 0, ..., l sets within $B_i^1, \dots, B_i^l$. We thus have

$$\begin{aligned} &\mathbb{E}_{\rho \sim \mathcal{E}} \left[D_{\rho, m_0}^{\leq \mathcal{D}, k} (\mathbf{s}_i)^{2(m_1-1)} \right] \\ &\leq \mathbb{E}_{\rho \sim \mathcal{E}} \left[\sum_{B_i^1, \dots, B_i^{m_1-1}} 1^{\otimes C_0} \otimes \bigotimes_{j \in C_1} (\overline{D}_{\rho}^{\leq \mathcal{D}} (s_{i,j}) - 1) \otimes \dots \otimes \bigotimes_{j \in C_{2(m_1-1)}} (\overline{D}_{\rho}^{\leq \mathcal{D}} (s_{i,j}) - 1)^{2(m_1-1)} \right] \\ &\leq k^{2(m_1-1)} m_0^{2k(m_1-1)} \mathbb{E}_{\rho \sim \mathcal{E}} \left[\bigotimes_{j \in \bigcup_{u=0}^{2(m_1-1)} C_u} (\overline{D}_{\rho}^{\leq \mathcal{D}} (s_{i,j}) - 1)^{2(m_1-1)} \right], \end{aligned}$$

where $s_{i,j} \in \{0, 1\}^n$ is the output bit string of the j -th copy in the i -th block. By the $(M', \varepsilon, \mathcal{D}, k)$ -assumption for restricted adaptivity among blocks, we further have

$$\mathbb{E}_{\rho \sim \mathcal{E}} \left[D_{\rho, m_0}^{\leq \mathcal{D}, k} (\mathbf{s}_i)^{2(m_1-1)} \right] \leq M'^{km_1} k^{2(m_1-1)} m_0^{2k(m_1-1)}.$$

We thus have

$$\begin{aligned} B_{m_1, k} &= \left\| \mathbb{E}_{\rho \sim \mathcal{E}} \overline{D}_{\rho, m}^{\leq \mathcal{D}, k} - 1 \right\|^2 \\ &\leq m_1 k^2 m^k M'^{km_1} k^{2(m_1-1)} m_0^{2k(m_1-1)} \varepsilon^{1/2} \\ &= 2^{O(km_1 \log M' \log m_0 \log m_1 \log k)} \varepsilon^{1/2}. \end{aligned}$$

By choosing

$$\varepsilon = 2^{-\Omega(km_1 \log k \log m_1 \log m_0 \log M')},$$

we can ensure the square degree- $(\mathcal{D}, k)$ advantage below $o(1)$, which finishes the proof for Theorem 36. $\blacksquare$

Again, we show that approximate state designs satisfy $(M', \varepsilon, \mathcal{D}, k)$ -assumption for restricted adaptivity among blocks in Theorem 35. Formally, we have the following corollary (see Appendix K.2 for the proof):

Corollary 37 (Approximate state $2k$ -design indicates hardness for round-based model among adaptivity within b

For the hypothesis testing problem ρ_{mm} vs. $\rho \sim \mathcal{E}$ with $\mathcal{E}$ a $2^{-\Omega(km_1^2 \log^2 m_1 \log m_0 \log k)}$ -approximate state $2k$ -design, weak detection is degree- $(\mathcal{D}, k)$ hard for any adaptivity among round-based model protocol that makes polynomially many single-copy PVMs $\{\{U_i | x\rangle \langle x| U_i^\dagger\}_{x=1}^{2^n}\}_{i=1}^m$. These protocols use $m_1 = \text{polylog}(n)$ blocks each with $m_0 = m/m_1 = \text{poly}(n)$ samples, and adaptivity is only allowed among blocks.

E.4. The statistical query model and its equivalence with the copy-wise low-degree model

In this section, we make the simple observation that the copy-wise low-degree model that we consider is equivalent to the *statistical query* model applied to readouts from non-adaptive measurements, provided the readouts in the low-degree model are noisy. The conclusion follows in a straightforward fashion from an analogous result previously proved in the classical setting (Brennan et al., 2021), and the main contribution of this section is conceptual: to identify a physically motivated scenario under which the conditions for their result hold in the quantum setting.

E.4.1. REVIEW OF STATISTICAL QUERY MODEL

Here, we introduce another computational model of *statistical query* (SQ) model. The SQ model was first proposed to design and analyze noise-tolerant algorithms (Kearns, 1998). It is further developed to be a popular restricted model of computation for studying information-computation tradeoffs (Diakonikolas et al., 2017; Feldman et al., 2017, 2015). In this work, we focus on the VSTAT(m) algorithm, which access a distribution p over $\mathbb{R}^n$ via queries $\phi : \mathbb{R}^n \rightarrow [0, 1]$ to an oracle. For an query ϕ , it returns to a value $\mathbb{E}_{x \sim p}[\phi(x)] + \xi$ with an adversarial chosen $\xi \in \mathbb{R}$ with $|\xi| \leq \max \left\{ \frac{1}{m}, \sqrt{\frac{\mathbb{E}[\phi](1-\mathbb{E}[\phi])}{m}} \right\}$. The approximation of $\mathbb{E}[\phi]$ is the same as an m -sample empirical estimate under the guarantees of concentration inequality. An algorithm that makes q queries to VSTAT(m) is roughly a proxy for an algorithm running in time q on m samples despite that ϕ need not to be polynomial-time computable but have must be a function of only a single sample.

To prove lower bounds against SQ algorithms, we introduce a complexity measure on hypothesis testing problems called *statistical dimension*. We will use the following definition of statistical dimension from Brennan et al. (2021):

Definition 38 (Definition 1.2 of Brennan et al. (2021)) Given the hypothesis testing problem $D_\emptyset$ vs. $\mathcal{S} = \{D_u\}_{u \sim \mu}$, the statistical dimension $SDA(\mathcal{S}, \mu, m)$ is defined to be:

$$SDA(\mathcal{S}, \mu, m) := \max \left\{ N \in \mathbb{N} : \mathbb{E}_{u,v \sim \mu} [|\langle D_u, D_v \rangle - 1| | A] \leq \frac{1}{m}, \forall \text{ events } A \text{ s.t. } Pr_{u,v \sim \mu}(A) \geq \frac{1}{N^2} \right\}. \quad (12)$$

The intuition for the above definition might be obscure at first glance. Recall that $\langle D_u, D_v \rangle - 1$ is the centered average of the likelihood ratio $\overline{D}_v$ over samples from D_u . When this quantity is bounded below by δ , there can be a common event such that one can simultaneously distinguish D_u and D_v from $D_\emptyset$ with probability at least some δ' depending only on δ . The statistical dimension in (12) thus quantifies the measure of pairs of distributions with no such common events according to the prior distribution μ . It is shown that the statistical dimension provides a lower bound on the SQ complexity of hypothesis testing with a VSTAT oracle (Feldman et al., 2017; Brennan et al., 2021):

Lemma 39 (Theorem 2.7 of [Feldman et al. \(2017\)](#) and Theorem A.5 of [Brennan et al. \(2021\)](#)) *Given the hypothesis testing problem $D_\emptyset$ vs. $\mathcal{S} = \{D_u\}_{u \sim \mu}$. Any (possibly) randomized statistical query algorithm which solves the hypothesis testing problem with probability at least $1 - \delta$ requires at least $(1 - \delta)SDA(\mathcal{S}, \mu, m)$ queries to $VSTAT(m/3)$.*

Remark 40 *We remark that the statistical dimension definition we used here is mildly stronger than the standard statistical dimension defined in [Feldman et al. \(2017\)](#) and the statistical query model used in previous results in learning distributions of quantum circuits ([Nietner et al., 2023](#); [Hinsche et al., 2021](#)). In other words, the lower bound proved in our work automatically holds for both models mentioned above.*

E.4.2. EQUIVALENCE

We now show the equivalence between copy-wise low-degree models introduced in the previous part and statistical query models in Appendix E.4.1 following the framework in [Brennan et al. \(2021\)](#).

To begin with, we go back to the classical distinguishing task $D_\emptyset$ vs. $\mathcal{S} = \{D_u\}_{u \sim \mu}$ on $\mathbb{R}^n$ as in Appendix A.2. We now introduce the concept of niceness:

Definition 41 ((δ, k) -nice ([Brennan et al., 2021](#))) *Given a null distribution $D_\emptyset$ on $\mathbb{R}^n$, a polynomial $p : (\mathbb{R}^n)^k \rightarrow \mathbb{R}$ of k samples $x_1, \dots, x_k \in \mathbb{R}^n$ is called k purely high degree if it is orthogonal to any function $q : (\mathbb{R}^n)^k \rightarrow \mathbb{R}$ that has degree at most k in each sample $x_1, \dots, x_k$, i.e., $\mathbb{E}_{x_1, \dots, x_k \sim D_\emptyset} p(x_1, \dots, x_k) q(x_1, \dots, x_k) = 0$ for all function $q : (\mathbb{R}^n)^k \rightarrow \mathbb{R}$ that has degree at most k in each sample $x_1, \dots, x_k$. The hypothesis testing problem $D_\emptyset$ vs. $\mathcal{S} = \{D_u\}_{u \sim \mu}$ is called (δ, k) -nice if no k -purely high-degree function of k samples can distinguish between the two cases with probability δ .*

Theorem 41 might look strange at first glance. However, the intuition here is that it rules out testing problems that can be solved using very few samples as usually we choose $k = \text{polylog}(n)$ as degrees of functions. It is shown that classical random noise can be a sufficient condition for ensuring niceness:

Lemma 42 (Theorem 5.2 of [Brennan et al. \(2021\)](#)) *Consider a hypothesis testing problem $D_\emptyset$ vs. $\mathcal{S} = \{D_u\}_{u \sim \mu}$ and suppose that $D_\emptyset = D_{\emptyset, \text{ind}}^{\otimes n}$ is a product distribution. Suppose that there is no k -sample algorithm that can distinguish between the two cases with probability δ_0 . We construct $\mathcal{S}' = \{D'_u\}_{u \sim \mu}$ by sampling $x'_i \sim D'_u, x'_i \in \mathbb{R}^n$ using a sample $x \sim D_u, x \in \mathbb{R}^n$ and replacing each coordinate x_j independently by a fresh sample from $D_{\emptyset, \text{ind}}$ with probability κ . Then the distinguishing task $D_\emptyset$ vs. $\mathcal{S}'$ is $(\delta_0(1 - \kappa)^{k^2}, k)$ -nice.*

Given the niceness condition, [Brennan et al. \(2021\)](#) shows that

Lemma 43 (Theorem 1.6 of [Brennan et al. \(2021\)](#)) *Consider a $(m^{-k/2}/4, k)$ -nice hypothesis testing problem $D_\emptyset$ vs. $\mathcal{S} = \{D_u\}_{u \sim \mu}$, we have:*

- *If there exists some $0 \leq m' \leq m$ such that $SDA(\mathcal{S}, m', \mu) \leq \left(\frac{2m}{m'}\right)^k$, then we have the $4mk$ -sample square degree- (∞, k^2) advantage bounded by $o(1)$.*
- *If the m -sample square degree- (∞, k) advantage is bounded by $o(1)$, then there exists $0 \leq m' \leq m$ such that $SDA(\mathcal{S}, m', \mu) \geq \left(\frac{2m}{m'}\right)^{\Omega(k)}$.*

Now, we go to the quantum hypothesis testing problem between ρ_{mm} vs. $\rho \sim \mathcal{E}$. As we only consider single-copy PVMs here, the null distribution $D_\emptyset$ after measuring ρ_{mm} is always the uniform distribution, which is a product distribution of $(1/2, 1/2)$ on each bit. We can obtain a quantum version of Theorem 42 using the local depolarization noise $\Lambda_\kappa^{\otimes n}$ with an identical depolarization channel Λ_κ with noise rate κ on each qubit, which is a common noise model considered on near-term noisy quantum devices (Chen et al., 2023a). Here, Λ_κ is given by

$$\Lambda_\kappa(\rho) \rightarrow (1 - \kappa)\rho + \kappa \frac{I_2}{2}, \quad (13)$$

where ρ is a single-qubit state. In real experiments, we usually have to implement a PVM $\{U|x\rangle\langle x|U^\dagger\}_{x=1}^{2^n}$ by implementing the unitary U first and perform a computational basis measurement. If we assume a readout error before the computational basis measurement in the form of local depolarization noise $\Lambda_\kappa^{\otimes n}$, it will naturally replace every bit of the readout information with $(1/2, 1/2)$ with probability κ , which is exactly $D_{\emptyset, \text{ind}}$. We thus have the following observations:

Observation 44 (*Hypothesis testing with local depolarization noise as readout error is nice*) Consider a quantum state hypothesis testing problem ρ_{mm} vs. $\rho \sim \mathcal{E}$ and corresponding protocols using PVMs. Suppose that we suffer from a readout error before the computational basis measurement in the form of local depolarization noise $\Lambda_\kappa^{\otimes n}$. If there is no k -sample algorithm that can distinguish between the two cases with probability δ_0 . Then the noisy distinguishing task $D_\emptyset$ vs. $\mathcal{S}$ is $(\delta_0(1 - \kappa)^{k^2}, k)$ -nice.

In the low-degree framework, we usually assume that $k = \text{polylog}(n)$ and $k = \omega(\log n)$. In this case, we can always find some **constant** threshold κ_0 such that $\delta_0(1 - \kappa)^{k^2} \leq m^{-k/2}/4$ for any $m = \text{poly}(n)$ and constant δ_0 . We thus have the following equivalence between the statistical query model and the copy-wise low-degree model when we have local depolarization noise as readout error:

Theorem 45 (Equivalence between statistical query model and the copy-wise low-degree model under noise)

Consider solving a quantum state hypothesis testing problem ρ_{mm} vs. $\rho \sim \mathcal{E}$ using protocols with $m = \text{poly}(n)$ single-copy PVMs. Assume that there is no k -sample algorithm for $k = \omega(\log n)$ that can distinguish between the two cases with constant probability δ_0 . Then there exists some constant threshold κ_0 such that if there is a readout error in the form of local depolarization noise $D_\kappa^{\otimes n}$ with $\kappa > \kappa_0$, we have:

- If there exists some $0 \leq m' \leq m$ such that $\text{SDA}(\mathcal{E}, m', \mu) \leq \left(\frac{2m}{m'}\right)^k$, then the square degree- (∞, k^2) advantage $\chi_{\leq(\infty, k^2)}^2$ (see Appendix A.2 for the formal definition) is bounded by $o(1)$.
- If the optimal distinguishing advantage achieved by any $4mk$ -sample copy-wise degree- $\leq(\infty, k)$ distinguished (i.e. the square degree- (∞, k) advantage $\chi_{\leq(\infty, k)}^2$) is bounded by $o(1)$, then there exists $0 \leq m' \leq m$ such that $\text{SDA}(\mathcal{E}, m', \mu) \geq \left(\frac{2m}{m'}\right)^{\Omega(k)}$.

E.5. No-go results for other models of adaptivity

It is natural to ask what would happen if we considered protocols with a greater level of adaptivity. The difficulty stems mainly from the fact that our model focuses on the complexity of post-processing and does not take into account the computational complexity of making adaptive choices of bases to measure in. As such, it is possible to “hide” unbounded computation in the adaptive choices.

E.5.1. k -GRAM COPY-WISE ADAPTIVITY

For starters, one might consider a k -gram adaptivity model. In this model, we choose the measurement bases for each copy based on the output of the previous $k - 1$ copies. Unfortunately, one can show that if the hypothesis testing task is *information-theoretically* possible using any number of single-copy measurements, possibly arbitrarily adaptive, there is always a 2-gram adaptive sequence of measurements for which the classical post-processing is low-degree tractable, in fact, trivial.

The construction for the cheating algorithm is given as follows. Denote the adaptive sequence of PVMs that would information-theoretically suffice to solve the problem as $\{\{U_i | x\rangle \langle x| U_i^\dagger\}_{x=1}^{2^n}\}_{i=1}^m$. Note that for each U_i , we can decompose it into constant terms in the form of a direct product of the first $n - 1$ qubits and the last qubit, so we can simulate this protocol using a polynomial number of PVMs on $n - 1$ qubits instead of n qubits. We can then measure $\text{poly}(n)$ copies of the state and record the whole history in a Markovian way in the last qubit. The reason is that we can encode arbitrarily many bits of information into $SU(2)$, so we can always record the history of all results of the protocol being simulated up to a given point into the choice of the measurement for the last qubit of the next copy being measured. After the simulation is finished, we adaptively choose one last basis to measure in: we perform an inefficient computation to post-process the information encoded in the last qubit of the previous qubit and determine the answer to the hypothesis testing problem, and then we measure the last qubit of the final copy in either the X or Z basis depending on the answer. Having hidden all of this computation in the adaptive choice of bases, we can trivially post-process the readouts to determine the answer by looking at what basis we measured the last qubit of the final copy in.

E.5.2. ARBITRARY SINGLE-QUBIT MEASUREMENT WITH LOW-DEGREE ADAPTIVITY

The above suggests that merely restricting the adaptivity to depend on the readouts for the previous copy does not suffice, because too much computation can be hidden even into such adaptive choices.

Another natural model to consider is a low-degree adaptivity model. In this model, we are restricted to using m single-qubit measurements, and the choice of the basis on each qubit in the total mn qubits can depend on at most the outcomes of the previous $k \geq 3$ qubits. For this model, we also show that it is possible to hide exponential computational complexity in computing the adaptivity if we assume that the task is sample efficient for local measurements.

The construction for the algorithm is given as follows. Note that the overall probability distribution based on performing a local measurement on a copy at once should be the same as performing it qubit by qubit, so we consider the latter one.

We keep a classical counter. We just measure each qubit on the computational basis first and add the counter by one after each qubit. When the counter goes to k , we store the history of the previous $k - 1$ qubits on the choice of basis we measure on the k -th qubit (recall that $SU(2)$ can store infinitely many bits of information). Next, we set the counter to 0. We again measure each qubit on the computational basis and add the counter by one after each qubit. When the counter again goes to k , we store the previous history of all $2k$ qubits using the k -th qubit basis and the results from $(k + 1)$ -th to $(2k - 1)$ -th qubit. We repeat the above procedure for $\text{poly}(n)$ time and, as in the previous section, we can collect the information from all readouts for the protocol being simulated into the choice of a basis for one qubit in $SU(2)$.

E.5.3. DISCRETE SINGLE-QUBIT BASIS MEASUREMENT WITH LOW-DEGREE ADAPTIVITY

Both the above algorithms stem from the fact that the low-degree model is a poor proxy for the complexity of adaptivity in situations where one can use arbitrarily large bit complexity. It is natural to hope that if we work with a discrete set of measurement bases for each qubit, then these pathologies might go away.

One exemplary model is when one can perform m single-qubit measurements. However, the choice of the basis on each qubit is a low-degree function of degree k on all previous history, and the possible choices are within the single-qubit Pauli basis $\{|0\rangle\langle 0|, |1\rangle\langle 1|, \{|+\rangle\langle +|, |-\rangle\langle -|, \{|+i\rangle\langle +i|, |-i\rangle\langle -i|\}$. Quantitatively, we suppose that every adaptive choice of single-qubit Pauli basis to measure is a low-degree junta $\{0, 1, +, -, +i, -i\} \rightarrow \{\sigma_x, \sigma_y, \sigma_z\}$. In the following, we show that even if the adaptivity choice function is a function of $O(1)$ previous single-qubit measurement outcomes, proving a hardness result is as hard as proving a circuit lower bound if the problem can be information-theoretically solved using $m_{\text{info}} = \text{poly}(n)$ single-qubit measurements.

Without loss of generality, we suppose the problem can be solved in information-theoretically solved using $m_{\text{info}} = \text{poly}(n)$ computational basis measurements. To show a lower bound, we at least need to rule out the following family of strategies: for copies $i = m_{\text{info}} + 1, \dots, m_{\text{info}} + \text{poly}(n)$, the basis in which the first qubit of copy i is measured depends on either $O(1)$ single-qubit outcomes from the first m_{info} copies or $O(1)$ of the bases in which the first qubits of copies $m_{\text{info}} + 1, \dots, m_{\text{info}} + i - 1$ are measured. At the end of the protocol, one looks at the basis in which the first qubit of the last copy was measured and outputs a Boolean predicate of that.

Note that we can think of the first qubits of these copies as effectively computing intermediate gates of a Boolean circuit whose inputs are the bits from measuring the first m_{info} copies. To rule out the possibility that there is some protocol of the above form, we would have to prove there is no polynomial-sized circuit that processes the information from the first m_{info} computational basis measurements and solves the detection problem.

Appendix F. Applications of the general framework

F.1. Low-degree hardness of error mitigation

In this section, we will apply the low-degree framework to show an information-computation gap for error mitigation. Previously, [Quek et al. \(2024\)](#) proved that quantum error mitigation is *statistically* hard for 1D geometrically local noisy quantum circuits of depth $\omega(\log n)$ at *constant* noise rate ([Quek et al., 2024](#)). Here we prove that even with significantly less noise, in fact at a rate inverse polynomial in the system size and with only a *constant* number of layers of noise, error mitigation is low-degree hard against protocols with a polynomial number of single-qubit measurements, even though it is statistically tractable for protocols with polynomial number of noisy data samples.

In Section [F.1.1](#) we formally define the problem and lower bound instance, and in Section [F.1.2](#) we formally state and prove our hardness result.

F.1.1. ERROR MITIGATION PRELIMINARIES

For any circuit consisting of l layers, $\mathcal{C} = \mathcal{U}_l \circ \dots \circ \mathcal{U}_1$ where $\mathcal{U}_i(\cdot) := U_i(\cdot)U_i^\dagger$, let $\tilde{\Phi}_{\mathcal{C}}$ denote the noisy version of $\mathcal{C}$, that is, a channel acting on $X \in \mathcal{H}_n$ as $X \mapsto \mathcal{D}_{\kappa}^{\otimes n} \circ \mathcal{U}_l \circ \dots \circ \mathcal{D}_{\kappa}^{\otimes n} \circ \mathcal{U}_1(X)$ where each $\mathcal{D}_{\kappa}^{\otimes n}$ is a layer of single-qubit depolarizing noise.

Error mitigation is defined as the following problem:

Definition 46 (Error mitigation) Upon input of:

1. a classical description of a noiseless circuit $\mathcal{C}$ and a finite set $\mathcal{M} = \{O_i\}$ of observables such that $\max_i \|O_i\| < 1$;
2. copies of the output state σ' of the noisy circuit $\tilde{\Phi}_{\mathcal{C}}(\rho_{\text{in}})$, and the ability to perform collective measurements,

output $1/\text{poly}(n)$ -precise estimates of the expectation values $\text{tr}(O_i \mathcal{C}(\rho_{\text{in}}))$ for each $O_i \in \mathcal{M}$. The number m of copies of σ' needed is the sample complexity of error mitigation.

To show a hardness result for this problem, we will consider the following hypothesis testing task:

Problem 47 (Noisy circuit hypothesis testing) Given m copies of some state ρ and the description of some circuit $\mathcal{C} \sim \mathcal{E}$, decide if ρ is

- Null hypothesis: $I/2^n = \tilde{\Phi}_{\mathcal{C}} \circ \mathcal{C}^\dagger(I/2^n)$
- Alternative hypothesis: $\tilde{\Phi}_{\mathcal{C}} \circ \mathcal{C}^\dagger(|\bar{0}\rangle\langle\bar{0}|)$

promised that one of the two is the case.

Hardness for this task implies hardness for error mitigation as follows. By using just a single observable, $\mathcal{M} = \{|\bar{0}\rangle\langle\bar{0}|\}$ and setting $\rho_{\text{in}} = \mathcal{C}^\dagger(I/2^n)$ (in the null hypothesis) or $\rho_{\text{in}} = \mathcal{C}^\dagger(|\bar{0}\rangle\langle\bar{0}|)$ (in the alternative hypothesis), an algorithm that solves error mitigation can certainly solve Theorem 47.

Note that we have used a different reduction from that in [Quek et al. \(2024\)](#). The upshot is that to prove low-degree hardness of error mitigation, it suffices to prove the low-degree hardness of the hypothesis testing problem, which we will do in the rest of this section. While the hard instance in [Quek et al. \(2024\)](#) assumes that there is a layer of single-qubit depolarization noise after each layer of the circuit, we consider the ensemble of quantum circuits that interleaves a constant number of random circuits of depth $\omega(\log(n))$ that form approximate unitary 2-designs and local depolarization noise of noise rate $O(1/n^{1-\delta'})$ here. Therefore, the total layer of noise in our model is only constant while the depth of the circuit can be up to $\omega(\log n)$.

F.1.2. LOWER BOUND

Theorem 48 (Extremely low-noise error mitigation is low-degree hard) For any $\delta, \delta' < 1$, there exists an ensemble of noisy circuits on geometrically-local architectures in any dimension, with only constantly-many layers of single-qubit depolarizing noise at rate $\kappa = \Omega(1/n^{1-\delta'})$, such that error mitigation is degree- k hard for algorithms that make single-qubit measurements on polynomially-many copies of the noisy state, or polynomially-many non-adaptive measurements prepared by depth k geometrically local circuits. That is, the hardness of error mitigation persists even in this extremely low-noise setting.

Even at constant depolarizing noise rate (higher than what we consider in the above theorem), error mitigation is only known to become statistically hard for a geometrically local circuit after $\omega(\log(n))$ layers of noise have acted. This is because every additional layer of noise in the circuit

results in a multiplicative contraction in relative entropy to the maximally mixed state². Since we only have a *constant* number of layers of noise at an extremely low rate $1/n^{1-\delta'}$ in the above setting, that is far from enough to render the problems statistically hard. Thus, Theorem 48 suggests an information-computation gap for this setting.

Fact 49 *For a circuit $\mathcal{C} = \mathcal{U}_l \circ \dots \circ \mathcal{U}_1$, let $\rho^{(l)} := \tilde{\Phi}_{\mathcal{C}}(\rho_{in})$. If each U_i is sampled from an exact n -qubit 2-design,*

$$\mathbb{E}_{\mathcal{U}_1^l} \text{tr}(\rho^{(l)})^2 \leq c^{nl}(1 - 2^{-n}) + \frac{1}{2^n}$$

where c is a noise-dependent parameter,

$$c = \frac{1 + 3(1 - \kappa)^2}{4}.$$

This fact is proven in Quek et al. (2024). When $\kappa = \frac{1}{n^{1-\delta'}}$ for $0 < \delta' < 1$,

$$\lim_{n \rightarrow \infty} c^{nl} = e^{-\frac{3}{2}ln^{\delta'}}.$$

To see this, let $y = c^{nl}$, and consider that:

$$\begin{aligned} \log(y) &= nl \log \left(\frac{1 + 3(1 - \frac{1}{n^{1-\delta'}})^2}{4} \right) \\ &= nl \log \left(1 - \frac{3}{2n^{1-\delta'}} + \frac{3}{4n^{2(1-\delta')}} + O\left(\frac{1}{n^{3(1-\delta')}}\right) \right) \\ &= nl \left[-\frac{3}{2n^{1-\delta'}} + \frac{3}{4n^{2(1-\delta')}} + \frac{9}{8n^{2(1-\delta')}} + O\left(\frac{1}{n^{3(1-\delta')}}\right) \right] \end{aligned}$$

and therefore

$$\lim_{n \rightarrow \infty} \log(y) = -\frac{3}{2}ln^{\delta'}$$

as desired. This computation illustrates that in the expression

$$c^{nl}(1 - 2^{-n}) + l\varepsilon$$

which appears in the next corollary, the term $l\varepsilon$ asymptotically (in n) dominates the term c^{nl} at noise rate $\kappa = 1/n^{1-\delta'}$ for constant ε .

2. It was proven in Quek et al. (2024) that the hardness of error mitigation kicks in at $\text{poly log log}(n)$ depth on circuits with all-to-all connectivity. When restricted to $\mathcal{D}$ -dimensional geometrically local connectivity, Quek et al. (2024) proved that quantum error mitigation becomes *statistically* hard only after $\min(O(\log(n)^{2/\mathcal{D}} \text{poly log log}(n), \log(n)))$ layers of constant rate noise – while we only require constantly many layers of extremely low noise.

Corollary 50 For any $a, l \in \mathbb{Z}_+$ and $\varepsilon, \varepsilon^* \in \mathbb{R}_+$ let

$$R(a) := 2^{a-n} \left[c^{nl} (1 - 2^{-n}) + l\varepsilon \right] + \varepsilon^*.$$

Define the random state $\rho^{(l)} := \tilde{\Phi}_{\mathcal{C}}(\rho_{in})$ where $\mathcal{C} = \mathcal{U}_l \circ \dots \circ \mathcal{U}_1$, each $\mathcal{U}_i$ is sampled from an ε -approximate 2-design for $i = 1, \dots, l$ and $\mathcal{U}_l$ is sampled from an ε^* -approximate 2-design.

For any $A \subseteq [n]$, and input state ρ_{in} , we have:

$$\Pr_{\mathcal{U}_1, \dots, \mathcal{U}_l} \left[\left\| \text{tr}_{[n] \setminus A} [\rho^{(l)}] - \frac{I_A}{2^{|A|}} \right\|_1 \geq 2^{|A|/2} R(|A|)^{1/4} \right] \leq R(|A|)^{1/2}.$$

Proof Suppose instead the last block $\mathcal{U}_l$ were sampled from an exact 2-design $\mu(2)$, instead of an approximate one. Let us first consider the expected value (over this alternative distribution) of the quantity

$$\begin{aligned} \mathbb{E}_{\mathcal{U}_l \sim \mu(2)} \left[\left\| \text{tr}_{[n] \setminus A} [\rho^{(l)}] - \frac{I_A}{2^{|A|}} \right\|_2^2 \right] &= \mathbb{E}_{\mathcal{U}_l \sim \mu(2)} \left[\text{tr} \left(\left(\text{tr}_{[n] \setminus A} [\rho^{(l)}] \right)^2 \right) - 2 \text{tr} \left(\frac{\text{tr}_{[n] \setminus A} [\rho^{(l)}]}{2^{|A|}} \right) + \text{tr} \left(\frac{I_A}{2^{2|A|}} \right) \right] \\ &= \mathbb{E}_{\mathcal{U}_l \sim \mu(2)} \left[\text{tr} \left(\left(\text{tr}_{[n] \setminus A} [\rho^{(l)}] \right)^2 \right) \right] - 2^{-|A|} \\ &= 2^{|A|-n} \left[\text{tr} \left((\rho^{(l-1)})^2 \right) - \frac{1}{2^n} \right]. \end{aligned} \quad (14)$$

Here we have integrated over the Haar measure in Equation (14). However, since the actual distribution from which $\mathcal{U}_l$ is sampled is an ε^* -approximate 2-design, the expectation over the actual distribution satisfies

$$\mathbb{E}_{\mathcal{U}_l} \left[\left\| \text{tr}_{[n] \setminus A} [\rho^{(l)}] - \frac{I_A}{2^{|A|}} \right\|_2^2 \right] \leq 2^{|A|-n} \left[\text{tr} \left((\rho^{(l-1)})^2 \right) - \frac{1}{2^n} \right] + \varepsilon^*.$$

Taking expectation over the remaining $l - 1$ blocks:

$$\begin{aligned} \mathbb{E}_{\mathcal{U}_1, \dots, \mathcal{U}_l} \left[\left\| \text{tr}_{[n] \setminus A} [\rho^{(l)}] - \frac{I_A}{2^{|A|}} \right\|_2^2 \right] &= 2^{|A|-n} \left[\mathbb{E}_{\mathcal{U}_1, \dots, \mathcal{U}_{l-1}} \left[\text{tr} \left((\rho^{(l-1)})^2 \right) \right] - \frac{1}{2^n} \right] + \varepsilon^* \\ &\leq 2^{|A|-n} \left[c^{nl} (1 - 2^{-n}) + \frac{1}{2^n} + l\varepsilon - \frac{1}{2^n} \right] + \varepsilon^* \\ &= 2^{|A|-n} \left[c^{nl} (1 - 2^{-n}) + l\varepsilon \right] + \varepsilon^*. \end{aligned}$$

where in the inequality we have used Theorem 49.

By Markov's inequality:

$$\Pr_{\mathcal{U}_1, \dots, \mathcal{U}_l} \left[\left\| \text{tr}_{[n] \setminus A} [\rho^{(l)}] - \frac{I_A}{2^{|A|}} \right\|_2 \geq R^{1/4} \right] = \Pr \left[\|\dots\|_2^2 \geq R^{1/2} \right] \leq R^{1/2}.$$

Converting ℓ_2 to ℓ_1 via Cauchy-Schwarz:

$$\Pr_{\mathcal{U}_1, \dots, \mathcal{U}_l} \left[\left\| \text{tr}_{[n] \setminus A} [\rho^{(l)}] - \frac{I_A}{2^{|A|}} \right\|_1 \geq 2^{|A|/2} R^{1/4} \right] \leq R^{1/2}.$$

■

We now consider the full scenario on m copies of the system, each copy being output by the same circuit of n qubits that is sampled once at the start. Let $T \subseteq [mn]$. We may write any such T as a union over the elements of T in each of the copies: $T = T_1 \cup \dots \cup T_m$, meaning that the overall product state can also be decomposed as:

$$\text{tr}_{[mn] \setminus T}[(\rho^{(l)})^{\otimes m}] = \bigotimes_{i=1}^m \text{tr}_{[n] \setminus T_i}[\rho^{(l)}]$$

By a union bound over all m copies, one then has that with probability at least

$$1 - \sum_{i=1}^m R(|T_i|)^{1/2} \geq 1 - mR(|T|)^{1/2}$$

$$\left\| \text{tr}_{[n] \setminus T_i}[\rho^{(l)}] - \frac{I_{T_i}}{2^{|T_i|}} \right\|_1 \leq 2^{|T_i|/2} R(|T_i|)^{1/4} \quad \forall i \in [m].$$

Again using the identity $\|A \otimes B - C \otimes D\|_1 \leq \|B - D\|_1 + \|A - C\|_1$ m times, we obtain that

$$\begin{aligned} \Pr_{\mathcal{U}_1, \dots, \mathcal{U}_\ell} \left[\left\| \text{tr}_{[mn] \setminus T}[(\rho^{(l)})^{\otimes m}] - \frac{I_T}{2^{|T|}} \right\|_1 \leq \sum_{i=1}^m 2^{|T_i|/2} R(|T_i|)^{1/4} \right] \\ \geq \Pr_{\mathcal{U}_1, \dots, \mathcal{U}_\ell} \left[\left\| \text{tr}_{[mn] \setminus T}[(\rho^{(l)})^{\otimes m}] - \frac{I_T}{2^{|T|}} \right\|_1 \leq m 2^{|T|/2} R(|T|)^{1/4} \right] \\ \geq 1 - mR(|T|)^{1/2}. \end{aligned}$$

Therefore

$$\begin{aligned} \mathbb{E}_{\mathcal{U}_1, \dots, \mathcal{U}_\ell} \left\| \text{tr}_{[mn] \setminus T}[(\rho^{(l)})^{\otimes m}] - \frac{I_T}{2^{|T|}} \right\|_1 &\leq m 2^{|T|/2} R(|T|)^{1/4} + mR(|T|)^{1/2} \\ &\leq m 2^{|T|/2} R(|T|)^{1/4} \end{aligned}$$

By the strong convexity of trace distance,

$$\left\| \mathbb{E}_{\mathcal{U}_1, \dots, \mathcal{U}_\ell} \left(\text{tr}_{[mn] \setminus T}[(\rho^{(l)})^{\otimes m}] \right) - \frac{I_T}{2^{|T|}} \right\|_1 \leq m 2^{|T|/2} R(|T|)^{1/4}$$

$\rho_k^{(l)}$ is a state output by l ε -approximate 2-designs alternating with a layer of depolarizing noise. So far, we have been agnostic as to how we generate the ε -approximate 2 designs, but as we want to obtain the shallowest depth at which error mitigation fails, now we prescribe that we use the shallowest ε -approximate 2-design generators we know of. We use the prescription of [Schuster et al. \(2024\)](#) to implement all the blocks with geometrically-local circuits on any architecture, in depth $\log(n/\varepsilon)$ for the first $n - 1$ blocks and depth $\log(n/\varepsilon^*)$ for the last.

By the calculations below Theorem 49, for the extremely low noise rate we consider,

$$R(|T|) := 2^{|T|-n} \left[c^{nl} (1 - 2^{-n}) + l\varepsilon \right] + \varepsilon^* \approx 2^{|T|-n} l\varepsilon + \varepsilon^*.$$

Then, if we choose $k = \log(n)$, $\varepsilon^* = 2^{-\Omega(k) \log(n)}$, it is easy to check that $\varepsilon^* > 2^{|T|-n} l \varepsilon$ for any $|T| < k$, and so for this choice of parameters,

$$\left\| \mathbb{E}_{\mathcal{U}_1, \dots, \mathcal{U}_\ell} \left(\text{tr}_{[mn] \setminus T} [(\rho^{(l)})^{\otimes m}] \right) - \frac{I_T}{2^{|T|}} \right\|_1 \leq m 2^{|T|/2} O(\varepsilon^*)^{1/4} \quad (15)$$

To bound the performance of low-degree algorithms using single-qubit measurements, since we only need to consider $|T| < k$ and $m = \text{poly}(n)$, we may check that the right hand side of Equation (15) becomes $2^{-\Omega(k \log n)}$, which allows us to apply Theorem 21. A similar argument and application of Theorem 22 allows us to prove the statement for measurements of bounded depth $\Omega(k)$.

F.2. Low-degree hardness of learning and testing random circuits

We now turn to another application of our framework, learning and testing quantum circuits of bounded depth (Huang et al., 2024; Landau and Liu, 2024; Zhao et al., 2024; Nietner et al., 2023; Vasconcelos and Huang, 2024). In Nietner et al. (2023), it is shown that learning the distribution of measuring a state prepared by a quantum circuit of depth $\omega(\log n)$ is hard for the statistical query model. For shallow circuits of constant depth, however, Huang et al. (2024); Landau and Liu (2024) provides computational- and sample-efficient algorithms to learn unitaries and states. It is natural to wonder if this task is hard under the low-degree framework after a certain depth threshold.

F.2.1. RANDOM QUANTUM CIRCUITS AND CIRCUIT MOMENT BOUNDS

Throughout this work, we consider geometrically-local quantum circuits, which contain only local single- and two-qubit gates on neighboring two qubits. More concretely, we focus on the brickwork architecture defined as follows:

Definition 51 *A n -qubit brickwork quantum circuit of depth L with a periodic boundary condition is a quantum circuit that is of the form*

$$U = (U_{2,3}^{(L)} \otimes \dots \otimes U_{n,1}^{(L)}) (U_{1,2}^{(L-1)} \otimes \dots \otimes U_{n-1,n}^{(L-1)}) \dots (U_{2,3}^{(2)} \otimes \dots \otimes U_{n,1}^{(2)}) (U_{1,2}^{(1)} \otimes \dots \otimes U_{n-1,n}^{(1)}),$$

where $U_{i,j}^{(l)}$ is the a two-qubit gate from $U(4)$ acting on the i -th and the j -th qubit in the l -th layer. Here, without loss of generality, we assume L and n are even.

We consider two kinds of random quantum circuit ensembles throughout this work. The first kind of ensemble we consider is called *brickwork random quantum circuits*, which is a random brickwork quantum circuit with each two-qubit gate chosen Haar randomly.

The second kind of ensemble we consider is the *coarse-grained random quantum circuit*. Consider a probability distribution over random circuits consisting of two steps: First, we apply Haar random unitaries to n/n' blocks each consisting of n' qubits. Second, we shift the layers such that the gates of the two layers overlap on $n'/2$ and apply Haar random unitaries to the new n/n' blocks. Moreover, we denote by $\nu_{n',n,L}$ the ensemble as above where, instead of Haar random unitaries, we draw the unitary on the blocks of n' qubits from a brickwork random quantum circuits of depth L .

For these ensembles of random quantum circuits, we summarize the results from Chen et al. (2024b); Brandao et al. (2016); Haferkamp (2022); Schuster et al. (2024):

Lemma 52 (Combination of results in Haferkamp (2022); Schuster et al. (2024)) *We have that:*

- *Brickwork random quantum circuits are a ε -approximate unitary t -design in depth $O(nt \log^7(t) + \log(1/\varepsilon))$.*
- *Coarse-grained random quantum circuits $\nu_{C \log(nt/\varepsilon), n, Ct \log(nt/\varepsilon)}$ is a ε -approximate unitary t -design for a sufficiently large constant $C > 0$.*

Combining these two results together, we consider the ensemble $\mu_{n', n, L}$ that is similar to $\nu_{n', n, L}$ but replacing the n' -qubit Haar random unitary with a brickwork random quantum circuits with large enough depth. We have $\mu_{C \log(nt/\varepsilon), n, Ct \log(nt/\varepsilon)}$ is a ε -approximate unitary t -design for a sufficiently large constant $C > 0$.

Based on the general framework, the low-degree hardness for learning and testing quantum circuits follows directly.

Here, we consider the following hypothesis testing problem:

- H_0 (null case): the state is the maximally mixed state $\rho_{\text{mm}} = I/2^n$.
- H_1 (alternative case): the state ρ is randomly chosen from a random geometrically local circuit of depth L according to $\mu_{n', n, L}$ (see Appendix F.2.1 for the definition).

Here, we focus on states prepared by circuits of depth L (Huang et al., 2024). Suppose we have an algorithm that can learn a state prepared by a circuit of depth L using single-copy PVMs, which given copies of $U |0^n\rangle$ promised that U is prepared by a quantum circuit of depth L , outputs a classical description of the circuit that implements unitary U' such that $d_{\text{tr}}(U |0^n\rangle, U' |0^n\rangle) \leq \varepsilon$. We can then solve the hypothesis testing using single-copy ancilla-free measurements via the following procedure:

1. We first implement the algorithm and output a unitary U' .
2. We then implement the measurement $\{U'^{\dagger} |x\rangle \langle x| U'\}_{x=1}^{2^n}$, which is a unitary circuit $U'^{\dagger}$ followed by a computational basis measurement.

In the null case H_0 , we are sampling from a uniform distribution. However, in the alternative case H_1 , we are sampling from a state that is ε -close to $|0\rangle$.

F.2.2. LOW-DEGREE HARDNESS FOR NON-ADAPTIVE SINGLE-COPY MEASUREMENTS

In this subsection, we consider non-adaptive protocols with single-copy PVMs. By Theorem 25, given $k \leq O(n)$ and an ensemble $\mathcal{E}$ that is $2^{-\tilde{\Omega}(k \log n)}$ -approximate state 2-design, weak detection is degree- k hard for any protocol between ρ_{mm} vs. $\mathcal{E}$ using polynomially many non-adaptive PVMs $\{\{U_i |x\rangle \langle x| U_i^{\dagger}\}_{x=1}^{2^{n+n'}}\}_{i=1}^m$ with up to $n' = \Theta(n)$ ancillary qubits. By Theorem 52, $\mu_{C \log(nt/\varepsilon), n, Ct \log(nt/\varepsilon)}$ is a ε -approximate unitary t -design for a sufficiently large constant $C > 0$. To form a $2^{-\tilde{\Omega}(k \log n)}$ -approximate state 2-design, we only need circuit depth of scaling

$$O(\log(nt/\varepsilon)) = \tilde{O}(k \log n),$$

where we use $t = 2$.

Therefore, combining these two results together, we obtain the following low-degree hardness results for non-adaptive protocols with single-copy measurements and bounded ancillary qubits:

Corollary 53 *For the hypothesis testing problem ρ_{mm} vs. $\rho \sim \mu_{\tilde{O}(k \log n), n, \tilde{O}(k \log n)}$ (i.e. an random circuit ensemble of depth $\tilde{O}(k \log n)$), weak detection is degree- k hard for any protocol that makes polynomially many non-adaptive single-copy PVMs $\{\{U_i | x\rangle \langle x| U_i^\dagger\}_{x=1}^{2^{n+n'}}\}_{i=1}^m$ with up to $n' = \Theta(n)$ ancillary qubits.*

By the correspondence between the hypothesis testing and the learning problem, we have the following corollary.

Corollary 54 *Given any $k = O(n)$, learning quantum circuits of depth $\tilde{O}(k \log n)$ is degree- k hard for degree- k protocols with a polynomial number of non-adaptive single-copy PVMs with up to $n' = \Theta(n)$ ancillary qubits.*

F.2.3. LOW-DEGREE HARDNESS FOR MODELS WITH RESTRICTED ADAPTIVITY

We then consider adaptive protocols. We first consider the model of restricted adaptivity within blocks, where we have $m_1 = \text{poly}(n)$ blocks, each of $m_0 = \text{polylog}(n)$ adaptive copies. Yet the blocks are independent from one another. (see Appendix E.2). By Theorem 33, if $\mathcal{E}$ forms a $2^{-\Omega(k^4 m_0^2 \log^2 k \log m_1 \log^2 m_0)}$ -approximate state $2km_0$ -design, weak detection between $\mathcal{E}$ and ρ_{mm} is degree- $(\mathcal{D}, k)$ hard for any adaptivity within round-based model protocol that makes polynomially many single-copy PVMs $\{\{U_i | x\rangle \langle x| U_i^\dagger\}_{x=1}^{2^n}\}_{i=1}^m$. These protocols use $m_1 = \text{poly}(n)$ blocks each with $m_0 = m/m_1 = \text{polylog}(n)$ samples, and adaptivity is only allowed in each block. By Theorem 52, $\mu_{C \log(nt/\varepsilon), n, Ct \log(nt/\varepsilon)}$ is a ε -approximate unitary t -design for a sufficiently large constant $C > 0$. To form a $2^{-\Omega(k^4 m_0^2 \log^2 k \log m_1 \log^2 m_0)}$ -approximate state $2km_0$ -design, we only need circuit depth of scaling

$$O(\log(nt/\varepsilon)) = \tilde{O}(k^4 m_0^2 \log m_1),$$

where we use $t = 2km_0$. We thus have the following result:

Corollary 55 *For the hypothesis testing problem ρ_{mm} vs. $\rho \sim \mu_{\tilde{O}(k^4 m_0^2 \log m_1), n, \tilde{O}(k^4 m_0^2 \log m_1)}$ (i.e. an random circuit ensemble of depth $\tilde{O}(k^4 m_0^2 \log m_1)$), weak detection is degree- $(\mathcal{D}, k)$ hard for any adaptivity within round-based model protocol making polynomially many single-copy PVMs $\{\{U_i | x\rangle \langle x| U_i^\dagger\}_{x=1}^{2^n}\}_{i=1}^m$. These protocols use $m_1 = \text{poly}(n)$ blocks each with $m_0 = m/m_1 = \text{polylog}(n)$ samples, and adaptivity is only allowed in each block. Moreover, this indicates that learning quantum circuits of depth $\tilde{O}(k^4 m_0^2 \log m_1)$ is hard any adaptivity within round-based model protocol with a degree at most $(\mathcal{D}, k)$ and $m = \text{poly}(n)$ single-copy PVMs.*

We then consider the model of restricted adaptivity among blocks, where we have $m_1 = \text{polylog}(n)$ adaptive blocks of $m_0 = \text{poly}(n)$ non-adaptive copies within each block. By Theorem 37, if $\mathcal{E}$ forms a $2^{-\Omega(km_1^2 \log^2 m_1 \log m_0 \log k)}$ -approximate state $2k$ -design, weak detection is degree- $(\mathcal{D}, k)$ hard for any adaptivity among round-based model protocol that makes polynomially many single-copy PVMs $\{\{U_i | x\rangle \langle x| U_i^\dagger\}_{x=1}^{2^n}\}_{i=1}^m$. These protocols use $m_1 = \text{polylog}(n)$ blocks each with $m_0 = m/m_1 = \text{poly}(n)$ samples, and adaptivity is only allowed among blocks. By Theorem 52, $\mu_{C \log(nt/\varepsilon), n, Ct \log(nt/\varepsilon)}$ is a ε -approximate unitary t -design for a sufficiently large constant $C > 0$. To form a $2^{-\Omega(km_1^2 \log^2 m_1 \log m_0 \log k)}$ -approximate state $2k$ -design, we only need circuit depth of scaling

$$O(\log(nt/\varepsilon)) = \tilde{O}(km_1^2 \log m_0),$$

where we use $t = 2k$. We thus have the following corollary:

Corollary 56 *For the hypothesis testing problem ρ_{mm} vs. $\rho \sim \mu_{\tilde{O}(km_1^2 \log m_0), n, \tilde{O}(km_1^2 \log m_0)}$ (i.e. an random circuit ensemble of depth $\tilde{O}(km_1^2 \log m_0)$), weak detection is degree- $(\mathcal{D}, k)$ hard for any adaptivity among round-based model protocol that makes polynomially many single-copy PVMs $\{\{U_i | x\rangle \langle x| U_i^\dagger\}_{x=1}^{2^n}\}_{i=1}^m$. These protocols use $m_1 = \text{polylog}(n)$ blocks each with $m_0 = m/m_1 = \text{poly}(n)$ samples, and adaptivity is only allowed among blocks. Moreover, this indicates that learning quantum circuit of depth $\tilde{O}(km_1^2 \log m_0)$ is hard any adaptivity among round-based model protocol with a degree at most k and $m = \text{poly}(n)$ single-copy PVMs.*

F.2.4. CRYPTOGRAPHIC APPLICATIONS

In this part, we provide some discussion on cryptographic applications of the low-degree hardness proved in our work. A recent work from Fefferman, Ghosh, Sinha, and Yuen (Fefferman et al., 2024) showed that *assumptions* about learning the output state of random quantum circuits could be used as the foundation for secure *one-way state generators* in quantum cryptography. A one-way state generator is an efficient algorithm that takes as input a classical key key , and outputs a quantum state $|\psi_{\text{key}}\rangle$ from that key. Anyone with the key key can efficiently verify that $|\psi_{\text{key}}\rangle$ is the correct output. However, given polynomially many copies of the $|\psi_{\text{key}}\rangle$, it should be computationally hard for an adversary to produce a key key' with the corresponding state $|\psi_{\text{key}'}\rangle$ close to $|\psi_{\text{key}}\rangle$.

In their setting, they have to conjecture that random quantum circuits of depth $L \geq \log^2 n$ are computationally hard to learn in the average case as there is no solid proof for the hardness. Here, however, we are able to provide a rigorous hardness result in the low-degree model based on Theorem 53:

Observation 57 (Computational low-degree $\delta = 2^{-\tilde{\Theta}(k \log n)}$ -no-learning observation) *Let $\mathcal{C}_{n, \tilde{\Theta}(k \log n)}$ be the class of coarse-grained random quantum circuit (see Appendix F.2.1 for definition). For all low-degree algorithms with degree at most k and single-copy PVMs with up to $\Theta(n)$ ancillary qubits, given polynomially-many copies of $C |0^n\rangle$ for $C \in \mathcal{C}_{n, \tilde{\Theta}(k \log n)}$ and $C \sim \mu_{\tilde{\Theta}(k \log n), n, \tilde{\Theta}(k \log n)}$, the probability of the algorithm outputting a circuit D such that $|\langle 0^n | C^\dagger D | 0^n \rangle| \geq \delta$ is bounded by $\delta \leq 2^{-\tilde{\Theta}(k \log n)}$.*

Based on this observation, we can use the following protocol to create a *one-way state generators against low-degree adversaries*, which is an efficient algorithm that takes as input a classical key key and outputs a quantum state $|\psi_{\text{key}}\rangle$ from that key such that

- Anyone with the key key can efficiently verify that $|\psi_{\text{key}}\rangle$ is the correct output.
- Given polynomially many copies of the $|\psi_{\text{key}}\rangle$, it should be computationally hard for any adversary that can perform single-copy PVMs with up to $\Theta(n)$ ancillary qubits and low-degree post-processing with degree at most k on the data to produce a key key' with the corresponding state $|\psi_{\text{key}'}\rangle$ close to $|\psi_{\text{key}}\rangle$.

Moreover, by the fact that $k = \omega(\log(n))$ in the usual low-degree model, δ is thus inversely superpolynomially small. According to Fefferman et al. (2024), the one-way state generator against low-degree adversaries can be made noise robust against any noisy model via parallel repetition as long as the fidelity of implementing the circuit in the noisy case is at least $1/(\text{poly}(n)L)$ where L is the circuit depth.

Generation algorithm: Given input key $\text{key} \in \{0, 1\}^{r(n)}$ for some polynomial $r(n)$, we interpret it as a description of an n -qubit circuit C from the ensemble $\mathcal{C}_{n, \tilde{\Theta}(k \log n)}$. Output $C|0\rangle$.

Verification procedure: Given input key $\text{key} \in \{0, 1\}^{r(n)}$ and a state $D|0\rangle$, apply $C^\dagger$ to the state and measure in the computational basis. Accept if the result is all zeroes, and reject otherwise.

Algorithm 1: Random circuit one-way state generators against low-degree adversaries

F.3. Low-degree hardness of learning sparse Hamiltonian from Gibbs states

In this section, we introduce the second application of our framework, learning Hamiltonian from Gibbs states (Anshu et al., 2020, 2021; Haah et al., 2022; Bakshi et al., 2024b; Wodecki and Marecek, 2024; Narayanan, 2024). In this task, we are asked to estimate a Hamiltonian $H = \sum_{P_a \in \mathcal{P}_n} \lambda_a P_a$ containing J terms given access to the Gibbs state of the Hamiltonian under temperature parameter $\beta = 1/kT$:

$$\rho_{\beta, H} = \frac{e^{\beta H}}{\text{tr}(e^{\beta H})}.$$

The algorithm has to output each λ_a within additive error ε without prior knowledge of the structure of the Hamiltonian.

All the known algorithms for this task focus on the case when the Hamiltonian is ς -local. In Anshu et al. (2020), the first polynomial sample complexity (but computationally inefficient) bound for this task is given as $O\left(\frac{2^{\text{poly}(\beta)} J^2 \log J}{\beta^{O(1)} \varepsilon^2}\right)$. Anshu et al. (2021) then shows that in the case when all P_a are commuting the algorithm can be made computationally efficient. In the high-temperature regime where $\beta \leq \beta_c = (100e^6 2^{O(\varsigma)})^{-1}$, Haah et al. (2022) provides an algorithm that achieves $O\left(\frac{2^{O(\beta)} \log J}{\beta^2 \varepsilon^2}\right)$ sample complexity and $O\left(\frac{2^{O(\beta)} J \log J}{\beta^2 \varepsilon^2}\right)$ time complexity. A recent result (Bakshi et al., 2024b) provided an algorithm that works for all β with sample complexity $\text{poly}(J, 1/\beta, (1/\varepsilon)^{\max\{1, O(\beta)\}})$ and time complexity $\text{poly}(J, 1/\beta, (1/\varepsilon)^{2^{O(\beta)}})$. Follow-up works improved the double exponential scaling in β to single exponential scaling (Wodecki and Marecek, 2024; Narayanan, 2024).

It turns out that in the regime where $\varsigma = O(1)$, $\varepsilon = \Theta(1)$ and $\beta = \Theta(1)$, statistical and computational efficient algorithms are available. From the lower bound side, the best-known sample complexity lower bound is provided as $\Omega\left(\frac{e^\beta}{\beta^2 \varepsilon^2} \cdot \log n\right)$. It is natural to ask if we can computationally efficiently learn non-local (sparse) Hamiltonian (i.e., $J = \text{poly}(n)$ but ς can be as large as n) from Gibbs state when $\beta = \Theta(1)$ and $\varepsilon = \Theta(1)$.

In this section, we prove that learning even sparse but non-local Hamiltonian from the Gibbs state at $\beta = \Theta(1)$ is low-degree hard for non-adaptive single-copy measurements.

F.3.1. BASIC RESULTS IN RANDOM MATRIX THEORY

We will also need some results from the random matrix theory. In this work, we consider *Wigner's Gaussian unitary ensemble (GUE)* (Wigner, 1958). A $d \times d$ GUE is a family of complex Hermitian

random matrices specified by

$$H_{jj} = \frac{g_{jj}}{\sqrt{d}},$$

$$H_{jk} = \frac{g_{jk} + ig'_{jk}}{\sqrt{2d}}, \text{ for } k > j,$$

where g_{jj}, g_{jk}, g'_{jk} are independent standard Gaussian $\mathcal{N}(0, 1)$. The definition we consider here follows from [Chen et al. \(2024a\)](#) and has an additional $1/\sqrt{d}$ normalization factor from the standard definition of Gaussian unitary ensemble. In the following, we denote a random Hamiltonian chosen from GUE as H_{GUE} . We can also write a H_{GUE} in the Pauli basis as:

$$H_{\text{GUE}} = \sum_{P \in \mathcal{P}_n} \frac{g_P}{d} P$$

with each g_P an independent standard Gaussian, and $d = 2^n$. The GUE is invariant under any unitary transformation. Therefore, any subspace of low-energy eigenvectors will be Haar random. We will also need the following fact

Fact 58 *With probability at least $1 - \exp(-\Theta(n))$ for random Hamiltonian H_{GUE} from GUE, we have $\|H_{\text{GUE}}\| \leq 3$.*

Proof We define normalized q -norms to be $\|B\|_q = (\overline{\text{tr}}|B|^q)^{1/q}$ where $\overline{\text{tr}} = \frac{1}{N} \cdot \text{tr}$. The average-case normalized q -norms is defined as $|||B|||_q = (\mathbb{E}\overline{\text{tr}}|B|^q)^{1/q}$. By the random matrix theory ([Tropp et al., 2015](#); [Tropp, 2018](#); [Bandeira et al., 2023](#); [Brailovskaya and van Handel, 2024](#); [Anderson et al., 2010](#)), we have

$$|||H_{\text{GUE}}|||_p \leq 2 \left(1 + \frac{(p/2)^{3/4}}{\sqrt{d}} \right).$$

By Markov's inequality, we have

$$\Pr[|||H_{\text{GUE}}||| \geq t] \leq \frac{\mathbb{E}_{H_{\text{GUE}}} [|||H_{\text{GUE}}|||^p]}{t^p} \leq d \frac{|||H_{\text{GUE}}|||_p^p}{t^p} \leq \left(\frac{2^{\log d/p}}{t} \cdot 2 \left(1 + \frac{(p/2)^{3/4}}{\sqrt{d}} \right) \right)^p.$$

Recall that $\log d = n$, we can choose $p = C \log n / \varepsilon$ for some large enough C and $t = 2(1 + \varepsilon)$. Then we have

$$\Pr[|||H_{\text{GUE}}||| \geq 2(1 + \varepsilon)] \leq \left(\frac{2^{\log d/p}}{t} \cdot 2 \left(1 + \frac{(p/2)^{3/4}}{\sqrt{d}} \right) \right)^p \leq \left(\frac{1 + \varepsilon/2}{1 + \varepsilon} \right)^p \leq \exp(-cn),$$

for some constant c . By choosing $\varepsilon = 1/2$ we finish the proof. ■

F.3.2. LOW-DEGREE HARDNESS FOR LEARNING AN ARBITRARY HAMILTONIAN

Note that if we allow J to be arbitrarily large, we will probably have to output an exponential number of λ_a when J is exponentially large. To avoid this output size barrier that prevents the possible computationally efficient algorithms, we consider the following distinguishing algorithm. In particular, the Hamiltonian is chosen either of the following two ensembles:

- The Hamiltonian is a random Hamiltonian of form $H_{\text{GUE}} = \sum_{P_a \in \mathcal{P}_n} \frac{g_a}{d} P_a$ chosen from the Wigner's Gaussian unitary ensemble (GUE) (see Appendix F.3.1). Here, each g_a is chosen as an IID standard Gaussian. H can also be a random complex matrix with $H_{ij} = H_{ji}^* = \frac{g_{ij} + i g'_{ij}}{\sqrt{2d}}$ and $H_{ii} = \frac{g_{ii}}{\sqrt{d}}$ where g_{ii}, g_{ij}, g'_{ij} are independent standard Gaussian.
- The Hamiltonian is $H = 0$, which yields a maximally mixed state as the Gibbs state.

The goal is to decide which is the case using copies of Gibbs state of the Hamiltonian. The task is again to distinguish between the random Gibbs state $\rho_{\beta, H} = \frac{e^{-\beta H}}{\text{tr}(e^{-\beta H})}$, where H is chosen from the GUE ensemble H_{GUE} , from the maximally mixed state. Therefore, we can exploit our standard framework. In the following, we stick to the case when $\beta = 1$ (low-temperature) and we denote such state as ρ_H .

We have the following result showing the low-degree hardness of distinguishing between ρ_{mm} vs. $\rho_{H_{\text{GUE}}}$, indicating the low-degree hardness of learning arbitrary Hamiltonian from its Gibbs state at $\beta = 1$:

Theorem 59 *For the hypothesis testing problem ρ_{mm} vs. $\rho \sim \rho_{H_{\text{GUE}}}$ and $k \leq O(n)$, weak detection is degree- k hard for any protocol that makes polynomially many non-adaptive single-copy PVMs with up to $n' = \Theta(n)$ ancillary qubits. Moreover, this indicates that learning arbitrary Hamiltonian from its Gibbs state at $\beta = 1$ is hard for any non-adaptive protocol with a degree at most $k \leq O(n)$ and $m = \text{poly}(n)$ single-copy PVMs with up to $n' = \Theta(n)$ ancillary qubits.*

Proof According to the general framework for non-adaptive ancilla-assisted single-copy PVMs in Theorem 24, we only need to consider the term

$$d_{\text{tr}} \left(\mathbb{E}_{H_{\text{GUE}}} \text{tr}_{[mn] \setminus T} \left[\bigotimes_{i=1}^m U_i \rho_H U_i^\dagger \right], \frac{I_T}{2^{|T|}} \right)$$

for any choice of unitary sequence $\{U_i\}_{i=1}^m$, and any $T \in [mn]$ and $|T| = k \leq O(n)$. Recall from Theorem 58, with probability $1 - \exp(-\Theta(n))$ for random Hamiltonian H_{GUE} from GUE, we have $\|H_{\text{GUE}}\| \leq 3$. As H_{GUE} is invariant from any unitary, we thus have the eigenstates of H_{GUE} chosen Haar randomly. Given any random $H_{\text{GUE}} = \sum_{j=1}^{2^n} \lambda_j |H_j\rangle \langle H_j|$ where λ_i is the i -th eigenvalue and $|H_j\rangle$ is the corresponding eigenstate. With probability $1 - \exp(-\Theta(n))$, we have $-3 \leq \lambda_j \leq 3$ by Theorem 58, and each $|H_j\rangle$ chosen from Haar random ensemble. By Theorem 25, as $|H_j\rangle$ is chosen from a unitary 2-design ensemble (actually exact Haar randomly), for any subset $T \subset [mn]$ with $|T| \leq n$, any m , and any fixed sequence of unitaries $\{U_i\}_{i=1}^m$, we have

$$d_{\text{tr}} \left(\text{tr}_{[mn] \setminus T} \left[\bigotimes_{i=1}^m U_i |H_j\rangle \langle H_j| U_i^\dagger \right], \frac{I_T}{2^{|T|}} \right) \leq m \cdot 2^{|T|/2} \cdot (2^{|T|-n})^{1/4} = m \cdot 2^{3|T|/4 - n/4}$$

with probability $1 - m \cdot (2^{|T|-n})^{1/2} = 1 - m \cdot 2^{|T|/2 - n/2}$ over the choice of $|H_j\rangle$.

We then consider the Gibbs state $\rho_{H_{\text{GUE}}} = \frac{e^{-H_{\text{GUE}}}}{\text{tr}(e^{-H_{\text{GUE}}})} = \frac{1}{\text{tr}(e^{-H_{\text{GUE}}})} \sum_{i=1}^{2^n} e^{-\lambda_j} |H_j\rangle \langle H_j|$. We focus on the case when $\|H_{\text{GUE}}\| \leq 3$. In this case, we always have $e^{-\lambda_j} = \Theta(2^{-n})$. We thus have

$$d_{\text{tr}} \left(\mathbb{E}_{H_{\text{GUE}}} \text{tr}_{[mn] \setminus T} \left[\bigotimes_{i=1}^m U_i \rho_H U_i^\dagger \right], \frac{I_T}{2^{|T|}} \right) \leq O \left(m \cdot 2^{3|T|/4 - n/4} \right).$$

We then consider PVMs with n' ancillary qubits as in Appendix D.2. By Theorem 24, we need to compute the following term:

$$d_{\text{tr}} \left(\mathbb{E}_{\rho_{H_{\text{GUE}}}} \text{tr}_{[m(n+n')] \setminus T} \left(\bigotimes_{i=1}^m U_i \left(\rho_{H_{\text{GUE}}} \otimes |0\rangle\langle 0|^{\otimes n'} \right) U_i^\dagger \right), \text{tr}_{[m(n+n')] \setminus T} \left(\bigotimes_{i=1}^m U_i \left(\rho_{\text{mm}} \otimes |0\rangle\langle 0|^{\otimes n'} \right) U_i^\dagger \right) \right).$$

Recall that $\rho_{H_{\text{GUE}}} = \frac{e^{-H_{\text{GUE}}}}{\text{tr}(e^{-H_{\text{GUE}}})} = \frac{1}{\text{tr}(e^{-H_{\text{GUE}}})} \sum_{i=1}^{2^n} e^{-\lambda_j} |H_j\rangle \langle H_j|$ has each $|H_j\rangle$ chosen from state 2-design (actually true Haar random) and the coefficient is of the scaling $O(2^{-n})$, we can use the triangle inequality with (5) to obtain

$$\begin{aligned} & d_{\text{tr}} \left(\mathbb{E}_{\rho_{H_{\text{GUE}}}} \text{tr}_{[m(n+n')] \setminus T} \left(\bigotimes_{i=1}^m U_i \left(\rho_{H_{\text{GUE}}} \otimes |0\rangle\langle 0|^{\otimes n'} \right) U_i^\dagger \right), \text{tr}_{[m(n+n')] \setminus T} \left(\bigotimes_{i=1}^m U_i \left(\rho_{\text{mm}} \otimes |0\rangle\langle 0|^{\otimes n'} \right) U_i^\dagger \right) \right) \\ & \leq 2m \cdot 2^{|T|/2} \left(2 \cdot 2^{|T|+11n'-n} + 3\varepsilon \right)^{1/4}. \end{aligned}$$

■

F.3.3. LOW-DEGREE HARDNESS FOR LEARNING A SPARSE HAMILTONIAN

In the above section, we consider learning an arbitrary Hamiltonian from its Gibbs state at temperature $\beta = 1$. Now, we show that even for sparse Hamiltonian, which has $J = \text{poly}(n)$ Pauli terms, is low-degree hard to learn from its Gibbs states at temperature $\beta = 1$.

We consider the following distinguishing task between two ensembles:

- The Hamiltonian is a random sparse Pauli string (RSPS) Hamiltonian ensemble H_{RSPS} where each Hamiltonian is an independent sum of J random Pauli strings with random sign coefficients (Chen et al., 2024a):

$$H = \sum_{a=1}^J \frac{r_a}{\sqrt{J}} P_a,$$

where P_a is IID chosen from $\mathcal{P}_n$ and r_a is chosen IID uniformly in $\{+1, -1\}$.

- The Hamiltonian is $H = 0$, which yields a maximally mixed state as the Gibbs state.

The goal is to decide which is the case using copies of the Gibbs state of the Hamiltonian at $\beta = 1$. For this distinguishing task, we have the following hardness result.

Corollary 60 *For the hypothesis testing problem ρ_{mm} vs. $\rho \sim \rho_{H_{\text{RSPS}}}$ at $J = O(n^3)$ and any $k \leq O(n)$, weak detection is degree- k hard any protocol that makes polynomially many non-adaptive single-copy PVMs with up to $n' = \Theta(n)$ ancillary qubits. Moreover, this indicates that learning a sparse Hamiltonian (containing polynomial terms in the Pauli basis) from its Gibbs state at $\beta = 1$ is hard for any non-adaptive protocol with a degree at most $k \leq O(n)$ and $m = \text{poly}(n)$ single-copy PVMs with up to $n' = \Theta(n)$ ancillary qubits.*

Proof The first observation is that the RSPS ensemble matches the GUE to the third moment, i.e.,

$$\begin{aligned}\mathbb{E}_{H_{\text{GUE}}}[H_{\text{GUE}}] &= \mathbb{E}_{H_{\text{RSPS}}}[H_{\text{RSPS}}] = 0, \\ \mathbb{E}_{H_{\text{GUE}}}[H_{\text{GUE}} \otimes H_{\text{GUE}}] &= \mathbb{E}_{H_{\text{RSPS}}}[H_{\text{RSPS}} \otimes H_{\text{RSPS}}] = \frac{1}{d^2} \sum_{P_a \in \mathcal{P}_n} P_a \otimes P_a, \\ \mathbb{E}_{H_{\text{GUE}}}[H_{\text{GUE}} \otimes H_{\text{GUE}} \otimes H_{\text{GUE}}] &= \mathbb{E}_{H_{\text{RSPS}}}[H_{\text{RSPS}} \otimes H_{\text{RSPS}} \otimes H_{\text{RSPS}}] = 0.\end{aligned}$$

It is invariant under random Clifford unitary, which forms a unitary 3-design (Zhu, 2017). We can thus conclude that the eigenstates of H_{RSPS} form a unitary 3-design.

Chen et al. (2024a) shows that when $J = \text{poly}(n)$, the behavior of H_{RSPS} is close to H_{GUE} . In particular, by Theorem III.1 of Chen et al. (2024a), when $J \geq O(n^3/\varepsilon^4)$ ensures that

$$\Pr_{H \sim H_{\text{RSPS}}}[\|H\| \geq 2(1 + \varepsilon)] \leq \exp(-\Theta(n)). \quad (16)$$

As the eigenstates of H_{RSPS} form the state 3-design and with $1 - \exp(-\Theta(n))$ the eigenvalue of H_{RSPS} is bounded within $[-3, 3]$. Note that for learning arbitrary Hamiltonian in Theorem 59, we only use the fact that the eigenvalue is bounded with high probability and the eigenstates are chosen from state 2-design. We thus again have

$$\begin{aligned}d_{\text{tr}} \left(\mathbb{E}_{\rho_{\text{RSPS}}} \left[\text{tr}_{[m(n+n')]\setminus T} \left(\bigotimes_{i=1}^m U_i \left(\rho_{\text{RSPS}} \otimes |0\rangle\langle 0|^{\otimes n'} \right) U_i^\dagger \right) \right], \text{tr}_{[m(n+n')]\setminus T} \left(\bigotimes_{i=1}^m U_i \left(\rho_{\text{mm}} \otimes |0\rangle\langle 0|^{\otimes n'} \right) U_i^\dagger \right) \right) \\ \leq 2m \cdot 2^{|T|/2} \left(2 \cdot 2^{|T|+11n'-n} + 3\varepsilon \right)^{1/4}.\end{aligned}$$

We can then obtain the claimed result via Theorem 24. ■

Appendix G. Quantum planted biclique

Here, we consider a new quantum learning problem that can be understood as the quantum analogue of the planted biclique problem. In this problem, we will focus on n -qudit quantum states with each qudit of d dimensions. We will abuse the notation by using d to represent the local qudit dimension instead of the dimension of a single copy of the quantum state only throughout this subsection.

Before formulating the task, we will have to define a few related notations. Given operators $A, B \in \mathbb{C}^{d \times d}$ and a subset $T \subseteq [n]$, define

$$A \otimes_T B = \bigotimes_{i=1}^n (A \cdot \mathbb{1}[i \notin T] + B \cdot \mathbb{1}[i \in T]).$$

We suppose $\lambda, d, m, n \in \mathbb{N}$, and without loss of generality, d is prime. Given $S \subseteq [n]$, let Λ_S denote the n -qudit channel that acts as identity on the qudits outside of S and as fully depolarizing noise on the qudits within S . Let Λ_κ denote the depolarizing channel on $(\mathbb{C}^d)^{\otimes n}$ with rate κ . We will take $\kappa = \lambda/n$ throughout, so we will omit the subscript.

We are now given either m copies of the following two cases:

- $\sigma = I/d^n$.
- $\sigma = \sigma_{\rho,S}$ which is defined as follows. Nature samples Haar-random, d -dimensional pure state ρ , and a subset $S \subseteq [n]$ of expected size λ by including each element of $[n]$ independently with probability λ (ρ and S are the same across all m copies of $\sigma_{\rho,S}$). The state $\sigma_{\rho,S}$ is given by

$$\sigma_{\rho,S} \triangleq (\Lambda_\kappa \circ \Lambda_S)[\rho^{\otimes n}].$$

The goal is to distinguish between these alternatives.

G.1. Lower bound for local measurements

We first prove the hardness for protocols of bounded degree making local measurements, i.e., single-qudit measurements. Here, we suppose that the learner performs (non-adaptive) local measurements. In particular, suppose the j -th qudit of the i -th copy of $\sigma_{\rho,S}$ is measured using the basis vectors $\{|\psi_1^{i,j}\rangle, \dots, |\psi_d^{i,j}\rangle\}$. Given $\mathbf{x} = (x_{i,j}) \in [d]^{m \times n}$, define $|\psi_{\mathbf{x}}\rangle \triangleq \bigotimes_{i,j} |\psi_{x_{i,j}}^{i,j}\rangle$.

Theorem 61 (Low-degree hardness for local measurements) *Weak detection for quantum planted biclique is degree- k hard for any protocol that makes m non-adaptive local measurements at $k = n^{o(1)}$ and $\lambda \leq o(nd^{1/4}m^{-1/2})$.*

Proof Note that if $\sigma = I/d^n$, then the distribution over outcomes after performing these measurements is uniform over $[d]^{m \times n}$. On the other hand, if $\sigma = \sigma_{\rho,S}$, the probability of observing $\mathbf{x} \in [d]^{m \times n}$ is given by $\langle \psi_{\mathbf{x}} | \mathbb{E}_{\rho,S} \sigma_{\rho,S}^{\otimes m} | \psi_{\mathbf{x}} \rangle$. For fixed ρ, S , denote the likelihood ratio by

$$\overline{D}_{\rho,S}(\mathbf{x}) \triangleq d^{mn} \langle \psi_{\mathbf{x}} | \sigma_{\rho,S}^{\otimes m} | \psi_{\mathbf{x}} \rangle,$$

and define $\overline{D} \triangleq \mathbb{E}_{\rho,S} \overline{D}_{\rho,S}$.

Lemma 62 *Given $t \in \mathbb{N}$ and density matrix $\rho \in \mathbb{C}^{d \times d}$,*

$$\mathbb{E}_S \sigma_{\rho,S}^{\otimes t} = \sum_{U \subseteq [t]} \kappa^{|U|} (I/d^n)^{\otimes U^c} \otimes \left(\sum_{T_1, \dots, T_{|U|} \neq \emptyset} \bigotimes_{\ell \in U} (I/d \otimes_{T_\ell} \Delta) \cdot \kappa^{|\cup_{\ell \in U} T_\ell|} \right),$$

where $\Delta \triangleq \rho - I/d$.

Proof Note that we can write

$$\begin{aligned} \sigma_{\rho,S} &= \kappa \left(\bigotimes_{i=1}^n (I/d + \mathbb{I}[i \in S] \Delta) \right) + (1 - \kappa) I/d^n \\ &= \kappa \sum_{T \subseteq [n]} (I/d \otimes_T \Delta) \cdot \mathbb{I}[T \subseteq S] + (1 - \kappa) I/d^n \\ &= I/d^n + \kappa \sum_{T \neq \emptyset} (I/d \otimes_T \Delta) \cdot \mathbb{I}[T \subseteq S], \end{aligned}$$

so

$$\sigma_{\rho,S}^{\otimes t} = \sum_{U \subseteq [t]} \kappa^{|U|} (I/d^n)^{\otimes U^c} \otimes \left(\sum_{T_1, \dots, T_{|U|} \neq \emptyset} \bigotimes_{\ell \in U} (I/d \otimes_{T_\ell} \Delta) \cdot \mathbb{K}[T_1 \cup \dots \cup T_\ell \subseteq S] \right).$$

The claim follows. ■

We will consider the standard Fourier basis over $[d]^{m \times n}$. Let $\xi \triangleq e^{\frac{2\pi i}{d}}$. Given a partial assignment $\mathbf{W} = (W, \alpha)$, where $W \subseteq [m] \times [n]$ and $\alpha : W \rightarrow \{1, \dots, d-1\}$, define the basis function

$$\chi_{\mathbf{W}}(\mathbf{x}) \triangleq \prod_{(i,j) \in W} \xi^{\alpha_{i,j} x_{i,j}},$$

and define the associated Fourier coefficient of a function $f : [d]^{m \times n} \rightarrow \mathbb{C}$ by $\hat{f}[\mathbf{W}] = \mathbb{E}[f(\mathbf{x}) \chi_{\mathbf{W}}(\mathbf{x})]$, where $\mathbf{x}$ is uniform over $[d]^{m \times n}$. To lighten notation, we will denote the Fourier transform of $\widehat{D}$ by $\hat{D}$ instead of $\widehat{\widehat{D}}$.

Let $\mathbf{W} = (W, \alpha)$ where $W = \sqcup_{i=1}^m W_i$ for $W_i \subseteq \{i\} \times [n]$ and $\alpha_i : W_i \rightarrow \{1, \dots, d-1\}$ is the restriction of α to W_i ; define $\mathbf{W}_i = (W_i, \alpha_i)$. Given $\mathbf{x} \in [d]^{m \times n}$, denote by $\mathbf{x}|_{W_i} \in [d]^n$ the restriction of $\mathbf{x}$ to the coordinates indexed by W_i , and define $|\psi_{\mathbf{x}|_{W_i}}\rangle \in (\mathbb{C}^d)^{\otimes |W_i|}$ in the natural way.

Then we have

$$\widehat{D}_{\rho,S}[\mathbf{W}] = d^{|W|} \prod_{(i,j) \in W} \mathbb{E}_{x_{i,j}} [\xi^{\alpha_{i,j} x_{i,j}} \langle \psi_{x_{i,j}}^{i,j} | \sigma_{\rho,S} | \psi_{x_{i,j}}^{i,j} \rangle]. \quad (17)$$

We can now compute the expectation of this over ρ, S by applying Lemma 62. Let $\text{supp}(\mathbf{W}) \triangleq \{i \in [m] : W_i \neq \emptyset\}$. Note that because $\mathbb{E}_x[\xi^{\alpha x}] = 0$ for $\alpha \in \{1, \dots, d-1\}$, the only terms in the expression for $\mathbb{E} \sigma_{\rho,S}^{\otimes W}$ which contribute to Eq. (17) are given by $U = \text{supp}(\mathbf{W})$ and $T_i = W_i$ for all $i \in \text{supp}(\mathbf{W})$. We thus obtain

$$\hat{D}[\mathbf{W}] = \kappa^{|\text{supp}(\mathbf{W})| + |\cup_{\ell} W_\ell|} \mathbb{E}_{\Delta} \left[\prod_{(i,j) \in W} \mathbb{E}_{x_{i,j}} [\xi^{\alpha_{i,j} x_{i,j}} \langle \psi_{x_{i,j}}^{i,j} | d\Delta | \psi_{x_{i,j}}^{i,j} \rangle] \right].$$

For any $W \subseteq [m] \times [n]$, we have that the Fourier mass

$$\mu_W \triangleq \sum_{\alpha: W \rightarrow \{1, \dots, d-1\}} |\hat{D}[(W, \alpha)]|^2$$

can be expressed as

$$\mu_W = \kappa^{2|\text{supp}(\mathbf{W})| + 2|\cup_{\ell} W_\ell|} \mathbb{E}_{\Delta, \Delta'} \prod_{(i,j) \in W} C_{i,j}(\Delta, \Delta'), \quad (18)$$

where

$$C_{i,j}(\Delta, \Delta') \triangleq \mathbb{E}_{x, x' \sim [d]} \left[\sum_{a \in \{1, \dots, d-1\}} \xi^{a(x-x')} \langle \psi_x^{i,j} | d\Delta | \psi_x^{i,j} \rangle \langle \psi_{x'}^{i,j} | d\Delta' | \psi_{x'}^{i,j} \rangle \right]. \quad (19)$$

Note that

$$\sum_{a \in \{1, \dots, d-1\}} \xi^{a(x-x')} = d \cdot \mathbb{1}[x = x'] - 1,$$

so

$$C_{i,j}(\Delta, \Delta') = \mathbb{E}_x[\langle \psi_x^{i,j} | d\Delta | \psi_x^{i,j} \rangle \cdot \langle \psi_x^{i,j} | d\Delta' | \psi_x^{i,j} \rangle]$$

To compute (19), we will appeal to the following moment calculation:

Lemma 63 *For $\Delta = \rho - I/d$ with ρ Haar-random,*

$$\mathbb{E}(d\Delta)^{\otimes t} = \sum_{\pi \in \mathcal{S}_t} \gamma_{|\text{fix}(\pi)|} P_\pi,$$

where $\text{fix}(\pi) \subseteq [t]$ denotes the set of fixed points of π , and P_π denotes the n -qudit permutation operator which acts via π on the qudits indexed by U and as identity elsewhere. Here, $\gamma_0, \dots, \gamma_t$ satisfy

$$|\gamma_0| = \frac{d^t}{d(d+1) \cdots (d+t-1)} \quad \text{and} \quad |\gamma_f| \leq O(t/\sqrt{d})^f \text{ for } f > 0$$

if $t = o(\sqrt{d})$.

Proof By a standard calculation,

$$\mathbb{E}(d\Delta)^{\otimes t} \triangleq \sum_{U \subseteq [t]} (-1)^{t-|U|} \beta_{d,|U|} \sum_{\pi \in \mathcal{S}_U} P_\pi \quad \text{for} \quad \beta_{d,s} \triangleq \frac{d^s}{d(d+1) \cdots (d+s-1)}, \quad (20)$$

where $\mathcal{S}_U$ denotes the symmetric group on the elements in U .

For any n -qudit permutation operator P_π , if $V = \text{fix}(\pi)$ has size f , then its coefficient in the above sum is given by

$$\sum_{U \supseteq V^c} (-1)^{t-|U|} \beta_{d,|U|} = \sum_{r=0}^f (-1)^r \binom{f}{r} \beta_{d,t-r} \triangleq \gamma_f.$$

The claimed expression for γ_0 is immediate.

It remains to establish the bound on $|\gamma_f|$ for $f > 0$. By Fact 74, $\beta_{d,t-r}$ admits the series expansion $1 + \sum_{a=1}^{\infty} c_a(t-r) \cdot d^{-a}$, where each c_a is a degree- $2a$ polynomial in $t-r$ (and thus in r) satisfying $|c_a(t-r)| \leq (1+t-r)^{2a} \leq (2t)^{2a}$. Note that for each $a \leq f/2$, we have

$$\sum_{r=0}^f (-1)^r \binom{f}{r} c_a(t-r) = 0.$$

We conclude that

$$|\gamma_f| = \left| \sum_{r=0}^f (-1)^r \binom{f}{r} \sum_{a=\lfloor f/2 \rfloor + 1}^{\infty} c_a(t-r) d^{-a} \right| \leq O(t/\sqrt{d})^f$$

as claimed. ■

Therefore,

$$\begin{aligned}
\mathbb{E}_{\Delta, \Delta'} \prod_{(i,j) \in W} C_{i,j}(\Delta, \Delta') &= \sum_{\pi, \pi' \in \mathcal{S}_W} \gamma_{|\text{fix}(\pi)|} \gamma_{|\text{fix}(\pi')|} \mathbb{E}_{\mathbf{x}} [\langle \psi_{\mathbf{x}|_W} | P_{\pi} | \psi_{\mathbf{x}|_W} \rangle \cdot \langle \psi_{\mathbf{x}|_W} | P_{\pi'} | \psi_{\mathbf{x}|_W} \rangle] \\
&\leq \left(\sum_{\pi \in \mathcal{S}_W} \gamma_{|\text{fix}(\pi)|} \mathbb{E}_{\mathbf{x}} [|\langle \psi_{\mathbf{x}|_W} | P_{\pi} | \psi_{\mathbf{x}|_W} \rangle|^2]^{1/2} \right)^2 \\
&\leq |W|!^2 \cdot O(1/d)^{|W|/2}, \tag{21}
\end{aligned}$$

where in the third step we used Lemma 63 and Lemma 75 below to conclude that for any permutation $\pi \in \mathcal{S}_W$,

$$\begin{aligned}
\gamma_{|\text{fix}(\pi)} \mathbb{E}_{\mathbf{x}} [|\langle \psi_{\mathbf{x}|_{W \setminus \text{fix}(\pi)}} | P_{\pi} | \psi_{\mathbf{x}|_{W \setminus \text{fix}(\pi)}} \rangle|^2]^{1/2} &\leq O(|W|/\sqrt{d})^{|\text{fix}(\pi)|/2} \cdot d^{-|W \setminus \text{fix}(\pi)|/4} \\
&\leq O(1/d)^{|W|/4} \\
&\leq O(|W|^4/d)^{|W|/2}.
\end{aligned}$$

Finally, we can substitute the bound in Eq. (21) back into Eq. (18) to obtain

$$\mu_W \leq \kappa^{2|\text{supp}(\mathbf{W})|+2|\cup_{\ell} W_{\ell}|} \cdot O(|W|^2/\sqrt{d})^{|W|}.$$

Therefore, for any fixed $Z \subseteq [n]$ and t , the sum over μ_W for which $|\text{supp}(W)| = t$, $\cup_{\ell} W_{\ell} = Z$, and $|W| \leq k$ is upper bounded by

$$\binom{m}{t} \kappa^{2t+2|Z|} \sum_{1 \leq |W_1|, \dots, |W_t| \leq k: |\cup_{\ell} W_{\ell}| = Z} O(k^2/\sqrt{d})^{\sum_{\ell} |W_{\ell}|}.$$

Note that the sum in the above equation can be upper bounded by

$$\begin{aligned}
\sum_{1 \leq |W_1|, \dots, |W_t| \leq k: W_{\ell} \subseteq Z \forall \ell} O(k^2/\sqrt{d})^{\sum_{\ell} |W_{\ell}|} &\leq ((1 + O(k^2/\sqrt{d}))^{|Z|} - 1)^t \\
&\leq O(k^3/\sqrt{d})^t.
\end{aligned}$$

Summing over all nonempty $Z \subseteq [n]$ of size at most k , we conclude that

$$\begin{aligned}
\sum_{W: 1 \leq |W| \leq k} \mu_W &\leq \sum_{t=1}^k \sum_{s=1}^k \binom{m}{t} \binom{n}{s} \kappa^{2s+2t} O(k^3/\sqrt{d})^t \\
&\leq \sum_{t=1}^D O\left(\frac{k^3 m \lambda^2}{n^2 \sqrt{d}}\right)^t \cdot \sum_{s=1}^k (\lambda^2/n)^s.
\end{aligned}$$

In particular, if $k = n^{o(1)}$, this is bounded by $o(1)$ provided

$$\lambda \leq \tilde{o}(nd^{1/4}m^{-1/2}).$$

■

G.2. Upper bound for single-copy measurements

Here, we demonstrate that for the detection problem $\{\sigma_{\rho,S}\}_{\rho \sim \mu_{\text{Haar}}, S}$ vs. I/d^m , there is a low-degree algorithm using (non-local) single-copy measurements that succeeds provided $\lambda \geq \omega(n^{2/3})$.

Theorem 64 (A low-degree efficient protocol with single-copy PVMs) *There is a constant-degree protocol for weak detection for quantum planted biclique that uses polynomially many non-adaptive single-copy PVMs, provided $\lambda \geq \omega(n^{2/3})$.*

Proof Let $\Pi \triangleq \frac{I + \text{SWAP}}{2} \in \mathbb{C}^{d^2 \times d^2}$. Suppose that n is even, and for every copy of the unknown n -qudit state, consider measuring the $2j$ -th and $(2j+1)$ -th qudits with the PVM $\{\Pi, I - \Pi\}$. Given $i \in [n], j \in [n/2]$, let $Z_{ij}(\sigma)$ denote the indicator random variable for whether the outcome of that measurement is Π when the i -th copy of the underlying state being measured is σ . For any i, j , we have

$$\mathbb{E}[Z_{ij}(I/d^m)] = \frac{d+1}{2d},$$

whereas for fixed S ,

$$\begin{aligned} \mathbb{E}[Z_{ij}(\sigma_{\rho,S})] &= \kappa \cdot \left\{ \frac{d+1}{2d} \cdot (1 - \mathbb{1}_j) + \mathbb{1}_j \right\} + (1 - \kappa) \cdot \frac{d+1}{2d} \\ &= \frac{d+1}{2d} + \frac{\kappa(d-1)}{2d} \cdot \mathbb{1}_j \triangleq \alpha_{S,j}, \end{aligned}$$

where $\mathbb{1}_j$ denotes the indicator for the event that $(2j, 2j+1) \in S \times S$. The expected value of the latter over the randomness of S is

$$\mathbb{E}_S[\alpha_{S,j}] = \frac{d+1}{2d} + \kappa^3 \cdot \frac{d-1}{2d} \triangleq \bar{\alpha}.$$

Let us now consider the test statistic $Z = \frac{2}{mn} \sum_{ij} Z_{ij}$.

As each $Z_{ij}(I/d^m)$ is an independent sample from $\text{Ber}(\frac{d+1}{2d})$, we have

$$\text{Var}[Z(I/d^m)] = \frac{2}{mn} \cdot \frac{d^2-1}{4d^2} \asymp \frac{1}{mn}.$$

We can also compute

$$\begin{aligned} \mathbb{E}_S \mathbb{E}[Z(\sigma_{\rho,S})^2] &= \mathbb{E}_S \left[\frac{4}{m^2 n^2} \left(\sum_{i,j} \alpha_{S,j} + \sum_{i \neq i'; j} \alpha_{S,j}^2 + \sum_{i, i'; j \neq j'} \alpha_{S,j} \alpha_{S,j'} \right) \right] \\ &= \frac{4}{m^2 n^2} \left(\frac{mn}{2} \bar{\alpha} + \frac{m(m-1)n}{2} \left(\kappa^2 \bar{\alpha}^2 + (1 - \kappa^2) \left(\frac{d+1}{2d} \right)^2 \right) + \frac{m^2 n(n-2)}{4} \bar{\alpha}^2 \right) \\ &= \bar{\alpha}^2 + \frac{2\bar{\alpha}}{mn} + \frac{4}{mn} \left(\frac{m-1}{2} \left(\kappa^2 \bar{\alpha}^2 + (1 - \kappa^2) \left(\frac{d+1}{2d} \right)^2 \right) - \frac{m}{2} \bar{\alpha}^2 \right) \\ &= \bar{\alpha}^2 + \frac{2}{mn} \left(\bar{\alpha} - \left(\frac{d+1}{2d} \right)^2 \right) + \frac{4}{mn} \left(\frac{m-1}{2} \kappa^2 - \frac{m}{2} \right) \left(\bar{\alpha}^2 - \left(\frac{d+1}{2d} \right)^2 \right) \\ &= \bar{\alpha}^2 + \Theta(1/mn + \kappa^3/n), \end{aligned}$$

so the variance of the test statistic $Z(\sigma_{\rho,S})$ is bounded by $\Theta(1/mn + \kappa^3/n)$. In particular, detection is possible provided that

$$\kappa^6 \gg 1/mn + \kappa^3/n,$$

which holds provided $\lambda \gg n^{2/3}$, as claimed. ■

G.3. Upper bounds for local measurements

Finally, we show that with computational basis measurements, there is a computationally inefficient algorithm that succeeds provided $\lambda \gg \min(d \log n, n^{1/2}d^{1/4})$ and a low-degree algorithm that succeeds provided $\lambda \gg n^{1/2}d^{1/4}$. In particular, the latter threshold matches the threshold below which we showed low-degree hardness in Section G.1.

Proposition 65 (Information-theoretic and computational tractability with local measurements)

Weak detection for quantum planted biclique is information-theoretically possible using measurements in the computational basis provided $\lambda \gg \min(d \log n, n^{1/2}d^{1/4})$. There is a constant-degree protocol for weak detection using measurements in the computational basis provided $\lambda \gg n^{1/2}d^{1/4}$.

Proof Consider the observable value $\text{tr}(Z\rho)$ for $Z = \text{diag}(1, \dots, -1, \dots)$. By standard anti-concentration of the Haar measure in d dimensions, with high probability $|\text{tr}(Z\rho)| = \Omega(1/\sqrt{d})$. In other words, if for every qudit of every copy, we measure in the d -dimensional computational basis and look at whether the measurement outcome is among the first $d/2$ computational basis vectors, this is distributed as $\text{Ber}(1/2 + \Theta(1/\sqrt{d}))$ if the qudit being measured is a copy of ρ , and as $\text{Ber}(1/2 + \Theta(1/\sqrt{d}))$ if the qudit being measured is a copy of I/d . In particular, with large constant probability, post-processing the readouts to solve quantum planted biclique now amounts to solving an instance of classical *bipartite planted dense subgraph* (Feldman et al., 2017). In particular, this entails distinguishing between a completely random $n \times n$ bipartite graph (with edge probabilities equal to $1/2$) versus one in which a random subgraph of expected size $\lambda \times \lambda$ has been planted whose edge connection probabilities are elevated to $1/2 + p$ for $p = \Theta(1/\sqrt{d})$.

First suppose that $\lambda \gg d \log n$. Note that for a random $t \times t'$ subgraph with edge probabilities equal to $1/2$, the number of edges is distributed as $\text{Bin}(tt', 1/2)$ and thus lies within $tt'/2 \pm O(\sqrt{tt'}\sqrt{\log 1/\delta})$ with probability $1 - \delta$. By a union bound over all $t \times t'$ subgraphs for $t, t' \leq O(\lambda)$ of a completely random $n \times n$ bipartite graph with edge probabilities equal to $1/2$, we conclude that under the null hypothesis, with high probability no such $t \times t'$ subgraph has more than $tt'/2 + O(t^{3/4}t'^{3/4}\sqrt{\log n})$ edges. On the other hand, in the planted scenario, the number of edges in the planted $t \times t'$ subgraph, where $\mathbb{E}[t] = \mathbb{E}[t'] = \lambda$, is distributed as $\text{Bin}(tt', 1/2 + p)$. Provided that $tt'p \gg t^{3/2}t'^{3/4}\sqrt{\log n}$, which happens with high probability over t, t' because $\lambda \gg d \log n$, one can thus perform weak detection for quantum planted biclique by brute force enumerating over all subsets of t copies of the state, and all subsets of t' qudits out of n qudits.

Next suppose that $\lambda \gg n^{1/2}d^{1/4}$. The total number of edges present in the full graph under the null hypothesis is distributed as $\text{Bin}(n^2, 1/2)$. Under the alternative hypothesis where there is a planted $t \times t'$ subgraph with elevated edge probabilities $1/2 + p$ for $p = \Theta(1/\sqrt{d})$, the total number of edges is instead distributed as $\text{Bin}(n^2 - tt', 1/2) + \text{Bin}(tt', 1/2 + p)$. The expected difference, over the randomness of t, t' , between the means of these random variables is $\lambda^2 p$, whereas both random variables have variance $O(n^2)$. So provided that $\lambda \gg \sqrt{n/p} \asymp n^{1/2}d^{1/4}$, there is a constant-degree algorithm for performing weak detection by simply counting the number of edges present in the bipartite graph given by the readouts. $\blacksquare$

Appendix H. Low-degree hardness for hybrid and classical problems.

In this section, we provide some extended applications of the quantum low-degree model in different problems with classical or hybrid input by encoding classical input into quantum data. In particular,

we embed a classical problem that is hard for classical low-degree algorithms into quantum data in a low-degree way to prove the hardness of quantum tasks under our quantum low-degree model. We first use a reduction from classical low-degree hardness of Learning Parity with Noise to show the low-degree hardness for Learning Stabilizer with Noise (Poremba et al., 2024) in Appendix H.1. We also show the low-degree hardness of agnostic learning product states by embedding tensor PCA, which is shown to be classically low-degree hard, into quantum data in Appendix H.2.

H.1. Low-degree hardness of Learning Stabilizers with Noise

We now consider the task of learning stabilizer with noise, which is another candidate for a quantum analog of learning parities with noise (Poremba et al., 2024) other than learning stabilizer states in Appendix I. Importantly, one has to be careful how to formulate the low-degree framework in this hybrid context. LPN is notably low-degree hard even in the absence of noise because Gaussian elimination, while polynomial-time, is not a low-degree algorithm, and LSN inherits this behavior as a generalization of LPN. In particular, if one does not constrain what kind of measurement is performed on the encoded state, there are simple Gaussian elimination-based ways to “cheat” and select a measurement which renders the subsequent classical post-processing tractable under low degree. We circumvent this issue by modeling the choice of measurement as low-degree in the classical input. Below, we provide the detailed description of our approach.

The *Learning Parity with Noise* (LPN) problem and its assumed hardness, are cornerstones on which much of modern classical and quantum cryptography is built (Blum et al., 1994). This is the problem of decoding a random linear code under Bernoulli noise. Formally, the input to the problem is

$$(A \sim \{0, 1\}^{n \times l}, Ax + e \pmod{2}),$$

where $A \sim \{0, 1\}^{n \times l}$ and $x \in \{0, 1\}^l$ are chosen uniformly at random, and $e = \text{Ber}_\kappa^{\otimes l}$ is a random Bernoulli error for some appropriate noise rate κ . The goal is to recover x . For $n = \text{poly}(l)$, LPN is believed to be hard for both classical and quantum algorithms running in time $\text{poly}(l)$ in the constant noise regime.

Note that the LPN decoding problem is an *average-case* problem, where the success probability of an algorithm is measured on average over the random choices of A , x , and e . A prominent piece of evidence for the hardness of LPN is a worst-to-average-case reduction (Brakerski et al., 2019; Yu and Zhang, 2021). Such a reduction means that an algorithm that succeeds with high probability over a random instance would even suffice to solve the worst-case instance of LPN. That is to say, most instances within that class are hard. Brakerski et al. (2019); Yu and Zhang (2021) are able to provide such a reduction from the nearest codeword problem (Brakerski et al., 2019; Yu and Zhang, 2021), a worst-case promise version of LPN.

Regarding the low-degree hardness of LPN-related problems, an exemplary result is on a special case of learning parity (without noise), known as k -XORSAT. Here, the variables are l $\{\pm 1\}$ -valued $x_1, \dots, x_l$ and we are told n random constraints of the form $x_{i_1,s} x_{i_2,s} \dots x_{i_k,s} = b_s$ with $b_s = \pm 1$, for $s \in [n]$. The goal is to determine whether there is an assignment $x_1, \dots, x_l \in \{\pm 1\}$ that satisfies all the constraints. This problem at $l \ll n \ll l^{3/2}$ is computationally hard for low-degree methods, i.e., it is hard to distinguish between a random b (which is unsatisfiable with high probability) and a formula with a planted assignment (Kunisky et al., 2019). Although this problem can be solved in polynomial time using Gaussian elimination over the finite field, these intrinsic high-degree algorithms tend not to be robust to even a tiny amount of noise.

Poremba et al. (2024) provides a potential natural quantum analog of LPN, *Learning Stabilizer with Noise* (LSN), and gave evidence for its hardness by showing that LSN too has a worst-to-average-case reduction. It is natural to ask if LSN replicates the above feature of LPN, i.e. that even its noiseless version is low-degree hard. Here, we answer in the affirmative, where the low-degree hardness we prove is qubit-wise low-degree hardness.

The LSN problem is to decode a random quantum stabilizer code under local depolarizing noise. Let $S \leq \mathcal{P}_n$ be an abelian subgroup of the Pauli group which has $n - l$ generators

$$S = \langle g_1, \dots, g_{n-l} \rangle, \quad g_i \in \pm\{I, X, Y, Z\}^n.$$

This is already sufficient to specify an error correcting code, whose codespace $C(S) \in (\mathbb{C}^2)^{\otimes n}$ is the simultaneous $+1$ eigenspace of all elements of S , i.e.

$$C(S) = \{|\psi\rangle : g|\psi\rangle = |\psi\rangle, \forall g \in S\}.$$

If S has $n - l$ generators, then $C(S)$ has 2^l many codewords. We use $\text{Stab}(n, l)$ to denote the set of all stabilizer subgroups of $\mathcal{P}_n$ with $n - l$ generators. Formally, the goal for LSN is to find the underlying random string $x \in \{0, 1\}^l$ given as input a sample

$$(S \sim \text{Stab}(n, l), E|\psi_{S,x}\rangle), \quad \text{with} \quad |\psi_{S,x}\rangle = U_{\text{Enc},S}(|0^{n-l}\rangle \otimes |x\rangle),$$

where $S \sim \text{Stab}(n, l)$ is a uniformly random stabilizer subgroup, $E = \Lambda_\kappa^{\otimes n}$ is a Pauli error from a local depolarizing channel with noise rate κ , and $U_{\text{Enc},S}$ is some canonical encoding circuit for the stabilizer code associated with S .

Poremba et al. (2024) implicitly show that for every instance of LPN, of the form $(A \in \mathbb{Z}_2^{n \times l}, A \cdot s + e \pmod{2})$, there exists a corresponding instance of LSN:

$$(U_A, |A \cdot s + e \pmod{2}\rangle) \tag{22}$$

where U_A is the Clifford encoding circuit³ encoding operation

$$U_A : |0^{n-l}\rangle \otimes |x\rangle \rightarrow |A \cdot x \pmod{2}\rangle \tag{23}$$

which is an injective matrix multiplication for any vector $x \in \mathbb{Z}_2^l$. These instances correspond to each other in the sense that a solution for the LSN problem (which takes the form of an output bit string $s \in \{0, 1\}^l$) is correct if and only if s is correct for the corresponding LPN problem.

In this work, we prove the average-case hardness of LSN in the low-degree framework. Before, we have to clarify what ‘low-degree’ means in the context of a problem with classical-quantum hybrid input such as LSN. This input consists of a classical matrix $S \sim \text{Stab}(n, l)$ containing $n - l$ stabilizers and a (noisy) quantum encoded state $E|\psi_{S,x}\rangle$. In the noiseless case, a general decoding algorithm could apply the inverse $U_{\text{Enc},S}^\dagger$ of the encoding unitary $U_{\text{Enc},S}$ and measure in the computational basis, which solves the problem exactly. While the measurement itself is a low-degree algorithm, the decision of which unitary to apply is high-degree as it depends on all $O(n^2)$ bits of S . This indicates that the definition of the algorithm’s degree must take into account the degree of the function that chooses the unitary to apply. This would rule out the above ‘cheating’ algorithm, as it should – because the above algorithm could also solve the classical parities problem

3. Note that U_A can be described in terms of CNOT gates (Patel et al., 2008), which are themselves Clifford gates.

by performing Gaussian elimination to choose which unitary to apply. Yet it is known that parities is low-degree hard.

To address this issue, we will define a low-degree algorithm in two ways, depending on how the hybrid quantum-classical input is supplied to the algorithm:

- **LSN with purely quantum input.** In this model, we are given m -copies of purely quantum input consisting of two parts. For each copy, the first $n \times l$ qubits are just a computational basis state indicating which Stabilizer code was applied (at most $O(n^2)$ bits are required to specify its stabilizer group). The last l qubits are the (noisy) encoded state. In this case, what we can do is to perform a joint PVM (a global fixed unitary acting on all qubits and a computational basis measurement) on the joint state. Suppose the post-processing after the measurement is a function of degree k , then the overall protocol is equivalent to a classical algorithm of degree $2k$ as the output distribution is at most quadratic in each input qubit.
- **LSN with hybrid quantum-classical input with low-degree quantum encoding of classical input.** In this case, we are given the classical matrix S and m copies of l -qubit (noisy) encoded state $|y\rangle$. We can then perform a measurement on each copy of code state by implementing a quantum unitary $U(S)$ encoding all classical information on the unitary and then implementing a computational basis measurement. The output distribution is given by $p(z) = \{|\langle z|U(S)|y\rangle|^2\}_z$, where z is a l bit string. In this case, the definition of a low-degree algorithm is one where $p(z)$ is a low-degree function of degree at most k on each bit of S . This eliminates the above-mentioned ‘cheating’ algorithm.

Given these two types of LSN, we have the following low-degree hardness result.

Theorem 66 (Average-case LSN is low-degree hard, even without noise) *There are no non-adaptive low-degree algorithms of degree $\text{polylog}(l)$ solving the average-case problem LSN, even without noise, with probability at least $2/3$ at $n = \text{poly}(l)$.*

One drawback of our result however is that it applies even when there is no noise. This is an inherent limitation of the low-degree method, as LPN and other problems with significant algebraic structure are low-degree hard despite being solvable in the absence of noise using brittle algebraic methods like Gaussian elimination (Holmgren and Wein, 2020). Nevertheless, given that these algorithms are “high-degree,” the low-degree method provides evidence for the hardness of these problems in the presence of noise.

Proof We first recall the classical proof of worst-case low-degree hardness for r -XORSAT, a special case of learning parity (without noise). Let $\mathcal{H}$ be a r -uniform hypergraph with l vertices and n hyperedges $e_1, \dots, e_n$. We consider the task of distinguishing between the following two distributions over $b \in \{\pm 1\}^n$:

- $b \sim \text{Unif}(\{\pm 1\}^n)$ (null).
- Sample a random $x \in \{\pm 1\}^l$ and for $i \in [n]$ set $b_i = \prod_{j \in C_i} x_j$ (alternative planted).

We would like to show that these two distributions are low-degree indistinguishable (for some fixed $\mathcal{H}$).

An even cover is a set of hyperedges such that each vertex appears an even number of times. Equivalently, we can write $\mathcal{H} \in \mathbb{Z}_2^{n \times l}$ as a matrix (equivalent to A in the definition of LPN) where

the i -th row C_i is the indicator vector of the hyperedge e_i . Then an even cover is a set of rows that sum up to 0.

We will show that if $\mathcal{H}$ has no even cover of size $\leq k$, then no degree- k polynomials can distinguish between the two cases. Since the null case is just the uniform distribution over $\{\pm 1\}^n$, the natural basis is the monomial basis $\{b_S = \prod_{i \in S} b_i\}_{S \subseteq [n]}$. The degree- k advantage is thus

$$\chi_{\leq k}^2 = \sum_{S \subseteq [n], 0 < |S| \leq k} \mathbb{E}_b[b_S]^2.$$

Note that the expectation is taken over the unknown random solution x , for any $S \subseteq [n]$ of size at most k , we have $\bigoplus_{i=1}^n C_i \neq \emptyset$ as $\mathcal{H}$ has no even cover of size $\leq k$. Therefore, we have

$$\mathbb{E}_b[b_S] = \mathbb{E}_{x \sim \{\pm 1\}^n} \prod_{i \in S} x_{C_i} = \mathbb{E}_{x \sim \{\pm 1\}^n} x_{\bigoplus_{i=1}^n C_i \neq \emptyset} = 0.$$

Via the above argument, we have shown that any low-degree protocol of degree $\text{polylog}(l)$ cannot solve worst-case Learning Parities (without noise) with even negligible probability. By the mapping given by (22) and (23), there is a worst-case Learning Stabilizers (without noise) problem corresponding to every Learning Parities (without noise) problem. Furthermore, a non-adaptive degree- k quantum algorithm for the latter problem implies a degree- $2k$ classical algorithm for the former problem. This is because, although a quantum algorithm is allowed to perform quantum measurements (assumed to be single-copy PVMs) in an arbitrary basis $\{U |x\rangle\langle x| U^\dagger\}_{x=1}^{2^n}$, U is a linear transformation which doesn't increase degree and the probability distribution is at most quadratic in each input qubits. Therefore, the nonexistence of a low-degree classical algorithm for Learning Parities in the worst-case implies the nonexistence of a low-degree quantum algorithm for Learning Stabilizers in the worst-case.

We further argue that there is also no low-degree quantum algorithm for Learning Stabilizers (without noise) problem in the average-case, by performing a worst-to-average-case reduction that preserves low-degrees. (Note that we perform this reduction *exactly*, while Poremba et al. (2024) were only able to perform a worst-to-average-case reduction to a nonstandard definition of average case where the distribution depends on the worst-case instance.) Given any worst-case instance of Learning Stabilizers (without noise), we can create an average-case instance by applying the following 're-randomizing unitary' to the input quantum state $U_{\text{Enc},S}(|0^{n-l}\rangle \otimes |x\rangle)$:

- a logical Pauli X^u for the known code S , where $u \sim \{0, 1\}^l$ is a random string
- a random n -qubit Clifford C to the received state $|\psi_{S,x}\rangle$.

This maps it to the exact average-case distribution over noiseless instances, as the logical Pauli maps the message x to a uniformly random one:

$$\overline{X^u} E |\overline{\psi_x}\rangle^S = E |\overline{\psi_{x \oplus u}}\rangle^S \quad \text{where } x \oplus u \sim \{0, 1\}^l$$

while the random Clifford C maps the code to a uniformly random code:

$$C \circ U_{\text{Enc},S} = U_{\text{Enc},S'} \quad \text{where } S' \sim \text{Stab}(n, l).$$

We can then feed the re-randomized quantum state to the low-degree LSN solver, which outputs the correct x' with high probability. We then output $x' \oplus u$. This amounts to a low-degree quantum algorithm for the worst-case LSN instance that succeeds with high probability as, again, applying the re-randomizing unitary described above does not increase the degree of the solution. $\blacksquare$

H.2. Low-degree hardness of agnostic tomography of product states

In this section, we observe that if one starts with a classical hypothesis testing problem which is classically low-degree hard, then if one encodes it into a quantum hypothesis testing problem using any low-degree mapping, the resulting quantum problem is quantum low-degree hard. The reasoning is along similar lines to the previous section: given a choice of measurements of the encoded state, the distribution over readouts is itself low-degree in the original classical input.

We carry out this observation formally for the special case of tensor PCA and use it to derive a new average-case hardness result for quantum learning.

Specifically, we consider the task of agnostic tomography (Grewal et al., 2024b,a; Chen et al., 2024e; Bakshi et al., 2024a). We are given copies of an unknown quantum state ρ that has fidelity at least τ with some state in a given class Ξ . Our goal is to find a state $\sigma \in \Xi$ such that the fidelity between σ and ρ is at least $\tau - \varepsilon$. When Ξ is the set of product states, Bakshi et al. (2024a) proposed a computational and sample efficient algorithm in the regime $\tau \geq O(1)$ and $\varepsilon = 1/\text{poly}(n)$. It is further shown that when $\tau = 1/\text{poly}(n)$, this task is NP-hard in the worst-case.

In this section, we show that agnostic tomography of product state when $\tau = 1/\text{poly}(n)$ is *average-case* low-degree hard for protocols with a polynomial number of single-copy PVMs. Quantitatively, we have:

Theorem 67 (Average-case hardness of agnostic tomography of product state at $\tau = 1/\text{poly}(n)$)

There exists a polynomial $p(n)$ and a non-trivial probability distribution μ over all pure states such that given $|\psi\rangle \sim \mu$, it has fidelity at least $\tau \geq 1/p(n)$ with the set Ξ of product state. However, finding a product state $\sigma \in \Xi$ such that the fidelity $\langle \psi | \sigma | \psi \rangle$ between $|\psi\rangle$ and σ is at least $\tau - \varepsilon$ for any $0 < \varepsilon \leq \tau/2$ is average-case hard for quantum algorithms with degree $k \leq o(n)$ and polynomially many single-copy measurements.

Remark 68 *The proof of Theorem 67 follows from a low-degree quantum encoding of the classical problem of tensor principal components analysis (PCA) into quantum data. A similar encoding can be extended to transfer any low-degree-hard additive Gaussian problem (Kunisky et al., 2019) into a quantum hypothesis testing problem, for which quantum low-degree hardness will follow from classical low-degree hardness.*

To prove Theorem 67, we first consider the following classical hypothesis testing task, which is known as the order-4 tensor principal components analysis (PCA) problem. Given signal parameter λ , we are given a classical tensor $T \in \mathbb{C}^{n \times n \times n \times n}$ which is sampled from the following two cases:

- $T = G$ where $G \in \mathbb{C}^{n \times n \times n \times n}$ is a Gaussian random tensor with each entry an IID standard Gaussian variable $\sim \mathcal{N}(0, 1)$ (null).
- $T = \lambda x^{\otimes 4} + G$ where $x \in \{\pm 1\}^n$ is chosen uniformly at random and $G \in \mathbb{C}^{n \times n \times n \times n}$ is a Gaussian random tensor (alternative).

Regarding this hypothesis testing problem, Kunisky et al. (2019) proved the following results on the classical low-degree likelihood ratio in the query model where one can query the entries T via a low-degree function of degree at most k .

Lemma 69 (Theorem 3.3 of Kunisky et al. (2019)) *Consider the above order-4 tensor PCA distinguishing problem with $\lambda \leq O(n^{-1}k^{-1/2})$, the square degree- k advantage for algorithms of degree at most k that query the tensor T is bounded by $o(1)$.*

In the following, we consider a modified version of this tensor PCA problem where each tensor is normalized. Given signal parameter λ , we are given a classical tensor $T \in \mathbb{C}^{n \times n \times n \times n}$ which is sampled from the following two cases:

- $T = G/\|G\|_F$ where $G \in \mathbb{C}^{n \times n \times n \times n}$ is a Gaussian random tensor with each entry an IID standard Gaussian variable $\sim \mathcal{N}(0, 1)$ (null).
- $T = (\lambda x^{\otimes 4} + G)/\|\lambda x^{\otimes 4} + G\|_F$ where $x \in \{\pm 1\}^n$ is chosen uniformly at random and $G \in \mathbb{C}^{n \times n \times n \times n}$ is a Gaussian random tensor (alternative).

Here, the Frobenius norm $\|T\|_F$ of a tensor T is defined as

$$\|T\|_F := \sqrt{\sum_{i,j,k,l=1}^n T_{ijkl}^2}.$$

We can show that this normalized order-4 tensor PCA distinguishing problem is also low-degree hard for algorithms with degree at most k under the same parameter regime with Theorem 69.

Lemma 70 *Consider the above normalized order-4 tensor PCA distinguishing problem with $\lambda \leq O(n^{-1}k^{-1/2})$, the square degree- k advantage for algorithms of degree at most k that query the tensor T is bounded by $o(1)$.*

Proof We first consider the null case when $T = G/\|G\|_F$ for a Gaussian random tensor $G \sim \mathbb{C}^{n \times n \times n \times n}$. Given a Gaussian random tensor G , the average value and variance of square Frobenius norm of G is given by

$$\begin{aligned} \mathbb{E}[\|G\|_F^2] &= \mathbb{E}\left[\sum_{i,j,k,l=1}^n T_{ijkl}^2\right] = n^4, \\ \text{Var}[\|G\|_F^2] &= \text{Var}\left[\sum_{i,j,k,l=1}^n T_{ijkl}^2\right] = \sum_{i,j,k,l=1}^n \text{Var}[T_{ijkl}^2] = 2n^4, \end{aligned}$$

where we use the fact that each entry of G is chosen from IID standard Gaussian distribution $\mathcal{N}(0, 1)$. Therefore, by the Chernoff bound, we have

$$(1 - 3\varepsilon)n^4 \leq \|G\|_F^2 \leq (1 + 3\varepsilon)n^4$$

with probability at least $1 - 2\exp(-\Theta(n^4\varepsilon^2))$ for $0 \leq \varepsilon \leq 1/2$. Note that $(1 - 3\varepsilon) \leq (1 - \varepsilon)^2$ and $(1 + \varepsilon)^2 \leq 1 + 3\varepsilon$ for $0 \leq \varepsilon \leq 1/2$, we thus have

$$(1 - \varepsilon)n^2 \leq \|G\|_F \leq (1 + \varepsilon)n^2$$

with probability at least $1 - 2\exp(-\Theta(n^4\varepsilon^2))$ for $0 \leq \varepsilon \leq 1/2$.

We then consider the alternative case when $T = (\lambda x^{\otimes 4} + G)/\|\lambda x^{\otimes 4} + G\|_F$. For a random choice of $\lambda x^{\otimes 4} + G$, the square Frobenius norm is computed as

$$\|\lambda x^{\otimes 4} + G\|_F^2 = \lambda^2 \|x^{\otimes 4}\|_F^2 + \lambda \langle x^{\otimes 4}, G \rangle + \|G\|_F^2.$$

The first term will always be $\lambda^2 n^4$ by definition, and the last term has mean value n^4 and variance $2n^4$. We thus consider the cross term $\lambda \langle x^{\otimes 4}, G \rangle$. As x is chosen uniformly at random from $\{\pm 1\}^n$, the mean value for this cross term is 0. The variance is given by

$$\begin{aligned} \text{Var} [\lambda \langle x^{\otimes 4}, G \rangle] &= \lambda \mathbb{E} [\langle x^{\otimes 4}, G \rangle^2] \\ &= \lambda \mathbb{E} \left[\left(\sum_{i,j,k,l=1}^n x_i x_j x_k x_l G_{ijkl} \right)^2 \right] \\ &= \lambda \mathbb{E} \left[\sum_{i,j,k,l=1}^n (x_i x_j x_k x_l)^2 G_{ijkl}^2 \right] \\ &= \lambda n^4 \end{aligned}$$

where we use the fact that G_{ijkl} are IID standard Gaussian variables and thus $\mathbb{E}[G_{ijkl} G_{i'j'k'l'}] = \delta_{ii'} \delta_{jj'} \delta_{kk'} \delta_{ll'}$. Summing all these together, we have

$$\begin{aligned} \mathbb{E}[\|\lambda x^{\otimes 4} + G\|_F^2] &= \mathbb{E} \left[\sum_{i,j,k,l=1}^n T_{ijkl}^2 \right] = (1 + \lambda^2) n^4, \\ \text{Var}[\|\lambda x^{\otimes 4} + G\|_F^2] &= \text{Var} \left[\sum_{i,j,k,l=1}^n T_{ijkl}^2 \right] = \sum_{i,j,k,l=1}^n \text{Var} [T_{ijkl}^2] = (2 + \lambda) n^4. \end{aligned}$$

Given that $\lambda \leq O(n^{-1} k^{-1/2}) = o(1)$, we conclude that

$$(1 - \varepsilon) \sqrt{1 + \lambda^2} n^2 \leq \|\lambda x^{\otimes 4} + G\|_F \leq (1 + \varepsilon) \sqrt{1 + \lambda^2} n^2$$

with probability at least $1 - 2\exp(-\Theta(n^4 \varepsilon^2))$ for $0 \leq \varepsilon \leq 1/2$.

If we add a fixed scaling factor to the original version of tensor PCA, the low-degree hardness result in Theorem 69 will not change. In the normalized tensor PCA, with probability $1 - 2\exp(-\Theta(n^4 \varepsilon^2))$ we will add a scaling factor within the range of $[(1 + \varepsilon)^{-1} n^{-2}, (1 - \varepsilon)^{-1} n^{-2}]$ for the null case compared to the standard case, and a scaling factor within the range of $[(1 + \varepsilon)^{-1} (\sqrt{1 + \lambda^2})^{-1} n^{-2}, (1 - \varepsilon)^{-1} (\sqrt{1 + \lambda^2})^{-1} n^{-2}]$ for the alternative case compared to the standard case. The deviation on the low-degree likelihood ratio is within the range $[(\sqrt{1 + \lambda^2})^{-1} \cdot \frac{1-\varepsilon}{1+\varepsilon}, (\sqrt{1 + \lambda^2})^{-1} \cdot \frac{1+\varepsilon}{1-\varepsilon}]$, the two sides of which are both of constant scaling. Thus, we can still obtain the same low-degree hardness result for the normalized tensor PCA. $\blacksquare$

We then consider the spectral norm of the two cases for the normalized tensor PCA. The spectral norm of an $n \times n \times n \times n$ tensor $T \in \mathbb{C}^{n \times n \times n \times n}$ is defined to be

$$\|T\|_{\text{spec}} = \max_{x,y,z,u \in \mathbb{C}^n} \frac{|\langle T, x \otimes y \otimes z \otimes u \rangle|}{\|x\|_2 \|y\|_2 \|z\|_2 \|u\|_2}.$$

It is shown in Bakshi et al. (2024a) that the existence of an algorithm for agnostic tomography of product states at parameter τ and $\varepsilon \lesssim \tau$ indicates the existence of an algorithm that estimates the spectral norm of a tensor $T \in \mathbb{C}^{n' \times n' \times n' \times n'}$. Here, we further show that such reduction is low-degree. Formally, we have the following lemma:

Lemma 71 *Suppose approximating the spectral norm of random tensors $T \in \mathbb{C}^{n' \times n' \times n' \times n'}$ with respect to some distribution corresponding to μ to additive error $\varepsilon' \|T\|_F$ is hard for any classical algorithm of degree $2k$. then agnostic tomography of product states with the unknown quantum state encoded from the random classical data drawn from μ in a particular way at some parameter $\tau = \text{poly}(\varepsilon')$ and $\varepsilon \lesssim \tau$ is hard for any quantum algorithm of degree k using polynomially many single-copy measurements.*

We will delay the proof of this lemma. Based on Theorem 71, we now provide the proof for the main result claimed in Theorem 67.

Proof [Proof of Theorem 67] We first consider the normalized tensor PCA. In the null case, the tensor is $T = G/\|G\|_F$ for a Gaussian random tensor $G \in \mathbb{C}^{n \times n \times n \times n}$. By the concentration inequality for Gaussian random tensor (Tomioka and Suzuki, 2014), we have

$$\|G\|_{\text{spec}} \leq \Theta(\sqrt{n})$$

with probability at least $1 - \exp(-\Theta(n))$. Therefore, we have

$$\left\| \frac{G}{\|G\|_F} \right\|_{\text{spec}} \leq \Theta(n^{-3/2})$$

with probability at least $1 - \exp(-\Theta(n))$.

In the alternative case, we observe that

$$\|\lambda x^{\otimes 4} + G\|_{\text{spec}} \geq \lambda n^2.$$

Therefore, we have

$$\left\| \frac{\lambda x^{\otimes 4} + G}{\|\lambda x^{\otimes 4} + G\|_F} \right\|_{\text{spec}} \geq \Theta(\lambda)$$

with probability at least $1 - \exp(-\Theta(n^2))$ when $\lambda = o(1)$.

Given $\lambda = \Theta(n^{-1}k^{-1/2})$ and $k = o(n)$, we have

$$\left\| \frac{\lambda x^{\otimes 4} + G}{\|\lambda x^{\otimes 4} + G\|_F} \right\|_{\text{spec}} \geq \omega(n^{-3/2}).$$

Note that the tensor in the normalized tensor PCA always has $\|T\|_F = 1$. If there is an algorithm that can estimate the spectral norm to $n^{-3/2}\|T\|_F$, we can use this algorithm to solve the normalized tensor PCA problem at any $k = o(n)$.

Since the normalized tensor PCA is low-degree hard by as shown by Theorem 70 for any classical algorithms of degree $k = o(n)$ at $\lambda \leq O(n^{-1}k^{-1/2})$, then estimating the spectral norm of these random tensors $T \in \mathbb{C}^{n \times n \times n \times n}$ to $n^{-3/2}\|T\|_F$ additive error is hard any classical algorithms of degree $k = o(n)$. By Theorem 71, when we encode these random tensors into random quantum states in (24), agnostic tomography of product states with these random quantum state is hard any quantum algorithms of degree $k/2 = o(n)$ making polynomially many single-copy measurements. ■

Finally, we provide the proof of Theorem 71.

Proof [Proof of Theorem 71] We will follow the reduction inspired by Bakshi et al. (2024a). Assume we are given a tensor $T \in \mathbb{C}^{n \times n \times n \times n}$ with $\|T\|_F = 1$, we construct a quantum state on $4n$ qubits as

$$|\psi_T\rangle = \sum_{i,j,k,l} T'_{ijkl} |e_i e_j e_k e_l\rangle, \quad (24)$$

where $|e_i e_j e_k e_l\rangle = |e_i\rangle \otimes |e_j\rangle \otimes |e_k\rangle \otimes |e_l\rangle$, where $|e_i\rangle$ is the product state that is $|1\rangle$ on the i -th qubit and $|0\rangle$ on the others. We note that this is a qubit encoding that preserves the Boolean properties. Moreover, this qubit encoding makes every qubit of the quantum state a degree-1 function on any classical input bit.

For $T \in \mathbb{C}^{n' \times n' \times n' \times n'}$ and a Haar random isometry $U \in \mathbb{C}^{n \times n'}$, as long as $n' > 10 \log n$, then with 0.99 probability over the randomness of U , we have $\max_{ijkl} |U^{\otimes 4} T| \leq \left(\frac{10n'}{n}\right)^2$ with $U^{\otimes 4} T \in \mathbb{C}^{n \times n \times n \times n}$ by Claim 7.5 in Bakshi et al. (2024a). Note that U is linear, $U^{\otimes 4} T$ is a degree-1 function on any classical input bit.

Given a tensor $T \in \mathbb{C}^{n \times n \times n \times n}$ with $\|T\|_F = 1$ and denote $M = \max_{ijkl} n^2 |T_{ijkl}|$, let OPT_T be the spectral norm of T and $\text{OPT}_{|\psi_T\rangle}$ to be defined as

$$\text{OPT}_{|\psi_T\rangle} = \max_{|\sigma\rangle = |\sigma_1\rangle \otimes \dots \otimes |\sigma_n\rangle} |\langle \sigma | \psi_T \rangle|.$$

Lemma 7.6 of Bakshi et al. (2024a) then shows that for large enough n , we have

$$e^{-2} \text{OPT}_T - \frac{10M}{n^{0.2}} \leq \text{OPT}_{|\psi_T\rangle} \leq e^{-2} \text{OPT}_T + \frac{10}{n^{0.1}} + \frac{10M}{n^{0.2}}.$$

In the following, we reduce approximating the norm of a tensor to agnostic learning product states to inverse polynomial accuracy. Given $\|T\|_F = 1$, we consider $U^{\otimes 4} T$ for U an ε -approximate Haar random isometry, and the corresponding quantum state $|\psi_{U^{\otimes 4} T}\rangle$. We thus have

$$e^{-2} \text{OPT}_{U^{\otimes 4} T} - \frac{10M}{n^{0.2}} \leq \text{OPT}_{|\psi_{U^{\otimes 4} T}\rangle} \leq e^{-2} \text{OPT}_{U^{\otimes 4} T} + \frac{10}{n^{0.1}} + \frac{10M}{n^{0.2}}.$$

By the claim, we can choose $M = (10m)^2$. As U is a isometry, we have $\text{OPT}_{U^{\otimes 4} T} = \text{OPT}_T$. We thus have

$$e^{-2} \text{OPT}_T - \frac{1000m^2}{n^{0.2}} \leq \text{OPT}_{|\psi_{U^{\otimes 4} T}\rangle} \leq e^{-2} \text{OPT}_T + \frac{10}{n^{0.1}} + \frac{1000m^2}{n^{0.2}}.$$

Choosing $\varepsilon < 0.01$ and $n = \Omega((m/\varepsilon)^{20})$, we have

$$e^{-2}(\text{OPT}_T - \varepsilon) \leq \text{OPT}_{|\psi_{U^{\otimes 4} T}\rangle} \leq e^{-2}(\text{OPT}_T + \varepsilon).$$

Therefore, finding the optimal product state fidelity to accuracy $0.01\varepsilon^2$ gives the spectral norm of the tensor to ε additive error.

Recall that both the encoding from classical tensor to the quantum state is degree 1, and the random rotation $U^{\otimes 4} T$ is also degree 1. When we measure each copy of the quantum state $|\psi_{U^{\otimes 4} T}\rangle$ using a single-copy PVM $\{V|x\rangle\langle x|V^\dagger\}_x$ for some fixed unitary V , the output distribution over $\{|\langle x|V|\psi_T\rangle|^2\}_x$ is a degree -2 function on each qubit as V is also linear. Combining all the above

results together, we conclude that, if there exists a low-degree quantum algorithm of degree k for agnostic tomography of product states such that for any n qubit state at parameter τ and $\varepsilon \lesssim \tau$ with sample complexity and running time $f(n, \tau)$ for some function f and high success probability. There is a classical algorithm of degree $2k$ that can approximate the spectral norm of any tensor $T \in \mathbb{C}^{n' \times n' \times n' \times n'}$ to additive error $\varepsilon' \|T\|_F$ with running time $f(\text{poly}(n'/\varepsilon'), \text{poly}(\varepsilon'))$ with high probability. This is the claim in Theorem 71. $\blacksquare$

Appendix I. The corner case: Low-degree hardness of learning stabilizer states

Here, we also provide a proof of low-degree hardness for learning stabilizer states, which is an exact quantum analogue of parity. We call this a corner case because, similar to parity, we can show low-degree hardness for learning stabilizer states even though there exist time efficient high-degree algorithms.

Let us first define stabilizer states. Let $\mathcal{P}_n$ be the n -qubit Pauli group, i.e. all tensor products of single-qubit Pauli matrices $\{I, X, Y, Z\}$ (including the global phases $\{\pm 1, \pm i\}$), where

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \sigma_x = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad \sigma_y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \sigma_z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

A subgroup $G \subseteq \mathcal{P}_n$ is called a *stabilizer group* if:

1. G is abelian (i.e. all elements commute),
2. $-I \notin G$, where I is the n -qubit identity operator,
3. $|G| = 2^r$ for some $r \leq n$.

A *stabilizer state* is any state that is the simultaneous $+1$ eigenstate of a maximal stabilizer group G , i.e. one with n generators. The task of learning stabilizer states is the following:

Definition 72 (Learning stabilizer states)

Input: m copies of $|\psi\rangle$, where $|\psi\rangle$ is an n -qubit stabilizer state.

Output: Classical description of $|\psi\rangle$.

We note that $|\psi\rangle$ has an efficient classical description – for example, its set of stabilizers, which take $O(n^2)$ bits to specify.

On the algorithmic front, Aaronson and Gottesman (Aaronson and Gottesman, 2008) provided polynomial-time learning algorithms using either $m = O(n^2)$ single-copy measurements or $m = O(n)$ joint measurements on multiple copies. In particular, the single-copy measurement approach relies on a procedure called computational difference sampling. Montanaro (Montanaro, 2017) then provided an improved algorithm using $O(n)$ two-copy measurements via a procedure called Bell difference sampling. Both computational difference sampling based on single-copy measurements and Bell difference sampling based on two-copy measurements rely on high-degree classical post-processing using the measurement data collected. The key subroutine of these sampling algorithms is to sample the Pauli observables (in a high-degree fashion) that stabilize the unknown stabilizer

states from the sampling. However, there are up to 2^n possible stabilizer states that share the same group of Pauli observables. It is even unknown if there are any low-degree algorithms that can find the underlying stabilizer state given a group of Pauli observables that stabilize it.

In general, it is natural to ask if there is a low-degree algorithm that learns stabilizer states also from a learning theory perspective. This is because learning stabilizer states can be one possible natural quantum analog of learning parity, which is shown to be low-degree hard in the classical framework (Kunisky et al., 2019). In the following, we show a negative answer for single-copy PVMs with up to $\Theta(n)$ ancillary qubits.

We simply consider distinguishing between random stabilizer states (alternative ensemble) and the maximally mixed state (null). As any low-degree algorithm that learns stabilizer states with high probability would imply a low-degree algorithm for this distinguishing problem, the following implies low-degree hardness for learning stabilizer states:

Corollary 73 *For the hypothesis testing problem ρ_{mm} vs. random stabilizer states, weak detection is degree- k hard for any protocol that makes polynomially many non-adaptive single-copy PVMs $\{\{U_i |x\rangle \langle x| U_i^\dagger\}_{x=1}^{2^{n+n'}}\}_{i=1}^m$ with up to $n' = \Theta(n)$ ancillary qubits. Moreover, the task of learning stabilizer states is hard for degree- k protocols with a polynomial number of non-adaptive single-copy PVMs with up to $n' = \Theta(n)$ ancillary qubits.*

Proof It is known that random stabilizer states form a state 3-design (Zhu, 2017). The claim thus follows immediately by Theorem 25. $\blacksquare$

Appendix J. Deferred proofs in the general framework for non-adaptive measurements

J.1. Proof of Theorem 20

For any subset $T \subseteq [mn]$ of size $\leq k$,

$$\begin{aligned} \left\| \mathbb{E}_{\rho \sim \mathcal{E}} \bar{D}_{\text{tr}_{[mn] \setminus T}(\rho^{\otimes m})} - \mathbf{1} \right\|^2 &= \mathbb{E}_{\mathbf{s}} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \bar{D}_{\text{tr}_{[mn] \setminus T}(\rho^{\otimes m})}(\mathbf{s}) \bar{D}_{\text{tr}_{[mn] \setminus T}(\rho'^{\otimes m})}(\mathbf{s}) - 1 \\ &= \mathbb{E}_{\mathbf{s}} 2^{2|T|} \left(2^{-|T|} + \zeta_{\mathbf{s}} \right) \left(2^{-|T|} + \zeta_{\mathbf{s}} \right) - 1, \quad \text{for some } \zeta_{\mathbf{s}} \in [-\varepsilon, \varepsilon] \\ &\leq \varepsilon^2 2^{2|T|}. \end{aligned}$$

where the second equality follows from the k -local distinguishability assumption.

J.2. Low-degree hardness for measurements prepared by bounded-depth circuits

We also consider other single-copy PVM class. Given any PVM $\{|\phi_{s_i}\rangle \langle \phi_{s_i}| \}_{s_i}$, we can always write it $\{U_i |s_i\rangle \langle s_i| U_i^\dagger\}_{s_i}$ for $s_i \in \{0, 1\}^n$ (see Appendix A.1). Recall that the square degree- k advantage can be computed as

$$\left\| \mathbb{E}_{\rho \sim \mathcal{E}} \bar{D}_{\rho, m}^{\leq k} - \mathbf{1} \right\|^2 = \sum_{|T| \leq k, T \neq \emptyset} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{s}, \mathbf{s}'} \bar{D}_{\rho, m}(\mathbf{s}) \bar{D}_{\rho', m}(\mathbf{s}') \chi_T(\mathbf{s}) \chi_T(\mathbf{s}'),$$

where

$$\overline{D}_{\rho,m}(\mathbf{s}) = 2^{mn} \prod_{i=1}^m \text{tr} \left(\rho U_i |s_i\rangle \langle s_i| U_i^\dagger \right) = 2^{mn} \text{tr} \left(\rho^{\otimes m} \bigotimes_{i=1}^m U_i |s_i\rangle \langle s_i| U_i^\dagger \right).$$

As a particular example, we consider the case when U_i 's can be efficiently prepared by ($\mathcal{D}$ -dimensional geometrically-local) circuits of depth at most L . In this case, we again fix a particular T with $|T| \leq k$. We have:

$$\begin{aligned} & \mathbb{E}_{\rho,\rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{s},\mathbf{s}'} \overline{D}_{\rho,m}(\mathbf{s}) \overline{D}_{\rho',m}(\mathbf{s}') \chi_T(\mathbf{s}) \chi_T(\mathbf{s}') \\ &= \mathbb{E}_{\rho,\rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{s},\mathbf{s}'} 2^{2mn} \text{tr} \left(\rho^{\otimes m} \bigotimes_{i=1}^m U_i |s_i\rangle \langle s_i| U_i^\dagger \right) \text{tr} \left(\rho'^{\otimes m} \bigotimes_{i=1}^m U_i |s'_i\rangle \langle s'_i| U_i^\dagger \right) \chi_T(\mathbf{s}) \chi_T(\mathbf{s}') \\ &= \left(\mathbb{E}_{\rho \sim \mathcal{E}} \mathbb{E}_{\mathbf{s}[T]} 2^{|T|} \text{tr} \left(\rho^{\otimes m} \bigotimes_{i=1}^m U_i \cdot |\mathbf{s}[T]\rangle \langle \mathbf{s}[T]| \cdot \bigotimes_{i=1}^m U_i^\dagger \right) \chi_T(\mathbf{s}) \right)^2. \end{aligned}$$

When U_i 's can be efficiently prepared by ($\mathcal{D}$ -dimensional geometrically-local) circuits of depth at most L , $\bigotimes_{i=1}^m U_i \cdot |\mathbf{s}[T]\rangle \langle \mathbf{s}[T]| \cdot \bigotimes_{i=1}^m U_i^\dagger$ is an observable of size at most $2^L k ((2L + k)^\mathcal{D})$ qubits. In this case, similar to the case for single-qubit measurements, we need local indistinguishability for subsystems of size $2^L k ((2L + k)^\mathcal{D})$.

We note that we prove Theorem 22 directly using the stronger assumption similar to Theorem 20 for convenience. It is possible to prove the same result based on a weaker assumption in Theorem 21.

J.3. Proof of Theorem 24

We now compute the square degree- k advantage, which is

$$\left\| \mathbb{E}_{\rho \sim \mathcal{E}} \overline{D}_{\rho,m}^{\leq k} - \mathbf{1} \right\|^2 = \mathbb{E}_{\mathbf{s}} \mathbb{E}_{\rho,\rho' \sim \mathcal{E}} \overline{D}_{\rho,m}^{\leq k}(\mathbf{s}) \overline{D}_{\rho',m}^{\leq k}(\mathbf{s}) - 1.$$

We can also decompose the likelihood ratio for the whole history as

$$\overline{D}_{\rho,m}(\mathbf{s}) = \sum_{T \subseteq [mn]} \alpha_T^\rho \chi_T(\mathbf{s}), \quad \alpha_T^\rho = \frac{1}{2^{nm}} \sum_{\mathbf{s}} \overline{D}_{\rho,m}(\mathbf{s}) \chi_T(\mathbf{s}) = \mathbb{E}_{\mathbf{s}} \overline{D}_{\rho,m}(\mathbf{s}).$$

The set $\chi_T(\mathbf{s})$ is again an orthonormal basis of functions. We note that here we cannot choose the Fourier basis as it is no longer orthogonal under $D_\emptyset$. But there exists a set of orthogonal bases $\chi_T(\mathbf{s})$.

Given that the square degree- k advantage satisfies:

$$\overline{D}_{\rho,m}^{\leq k}(\mathbf{s}) = \sum_{|T| \leq k} \alpha_T^\rho \chi_T(\mathbf{s}).$$

We can thus compute the square degree- k advantage as

$$\left\| \mathbb{E}_{\rho \sim \mathcal{E}} \overline{D}_{\rho,m}^{\leq k} - \mathbf{1} \right\|^2 = \sum_{|T| \leq k, T \neq \emptyset} \mathbb{E}_{\rho,\rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{s},\mathbf{s}'} \overline{D}_{\rho,m}(\mathbf{s}) \overline{D}_{\rho',m}(\mathbf{s}') \chi_T(\mathbf{s}) \chi_T(\mathbf{s}').$$

In the following, we consider the contribution for a fixed set $T \in [m(n + n')]$. We denote $\mathbf{s}[T]$ to be the set of $s_{i,r}$ with $(i, r) \in T$ for $i = 1, \dots, m$ and $r = 1, \dots, n + n'$. Recall that each PVM with n' ancilla qubits can be represented by $\{\{U_i | s_i\rangle \langle s_i| U_i^\dagger\}_{s_i=1}^{2^{n+n'}}\}_{i=1}^m$. We have:

$$\begin{aligned}
 & \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{s}, \mathbf{s}'} \overline{D}_{\rho, m}(\mathbf{s}) \overline{D}_{\rho', m}(\mathbf{s}') \chi_T(\mathbf{s}) \chi_T(\mathbf{s}') \\
 &= \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{s}, \mathbf{s}'} \prod_{i=1}^m \frac{\text{tr}(\rho \otimes |0\rangle\langle 0|^{\otimes n'} \cdot |\phi_{s_i}\rangle\langle \phi_{s_i}|)}{\text{tr}(\rho_{\text{mm}} \otimes |0\rangle\langle 0|^{\otimes n'} \cdot |\phi_{s_i}\rangle\langle \phi_{s_i}|)} \prod_{i=1}^m \frac{\text{tr}(\rho' \otimes |0\rangle\langle 0|^{\otimes n'} \cdot |\phi_{s'_i}\rangle\langle \phi_{s'_i}|)}{\text{tr}(\rho_{\text{mm}} \otimes |0\rangle\langle 0|^{\otimes n'} \cdot |\phi_{s'_i}\rangle\langle \phi_{s'_i}|)} \chi_T(\mathbf{s}) \chi_T(\mathbf{s}') \\
 &= \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \sum_{\mathbf{s}, \mathbf{s}'} \prod_{i=1}^m \text{tr}(\rho \otimes |0\rangle\langle 0|^{\otimes n'} \cdot |\phi_{s_i}\rangle\langle \phi_{s_i}|) \text{tr}(\rho' \otimes |0\rangle\langle 0|^{\otimes n'} \cdot |\phi_{s'_i}\rangle\langle \phi_{s'_i}|) \chi_T(\mathbf{s}) \chi_T(\mathbf{s}') \\
 &= \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \sum_{\mathbf{s}, \mathbf{s}'} \text{tr} \left(\bigotimes_{i=1}^m U_i \left(\rho \otimes |0\rangle\langle 0|^{\otimes n'} \right) U_i^\dagger \cdot |s_i\rangle\langle s_i| \right) \text{tr} \left(\bigotimes_{i=1}^m U_i \left(\rho' \otimes |0\rangle\langle 0|^{\otimes n'} \right) U_i^\dagger \cdot |s'_i\rangle\langle s'_i| \right) \chi_T(\mathbf{s}) \chi_T(\mathbf{s}') \\
 &= \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \sum_{\mathbf{s}[T], \mathbf{s}'[T]} \text{tr} \left(\text{tr}_{[m(n+n')]\setminus T} \left(\bigotimes_{i=1}^m U_i \left(\rho \otimes |0\rangle\langle 0|^{\otimes n'} \right) U_i^\dagger \right) \bigotimes_{(i,r) \in T} |s_{i,r}\rangle\langle s_{i,r}| \right) \cdot \\
 & \quad \text{tr} \left(\text{tr}_{[m(n+n')]\setminus T} \left(\bigotimes_{i=1}^m U_i \left(\rho' \otimes |0\rangle\langle 0|^{\otimes n'} \right) U_i^\dagger \right) \bigotimes_{(i,r) \in T} |s'_{i,r}\rangle\langle s'_{i,r}| \right) \chi_T(\mathbf{s}) \chi_T(\mathbf{s}')
 \end{aligned}$$

where the last line follows from the fact that $\chi_T(\mathbf{s}), \chi_T(\mathbf{s}')$ only depend on $\mathbf{s}[T], \mathbf{s}'[T]$.

Note that $\text{tr}_{[m(n+n')]\setminus T} \left(\bigotimes_{i=1}^m U_i \left(\rho' \otimes |0\rangle\langle 0|^{\otimes n'} \right) U_i^\dagger \right)$ is the RDM of $\bigotimes_{i=1}^m U_i \left(\rho' \otimes |0\rangle\langle 0|^{\otimes n'} \right) U_i^\dagger$ on the qubits in T . The above equation can be bounded if

$$d_{\text{tr}} \left(\mathbb{E}_{\rho \sim \mathcal{E}} \text{tr}_{[m(n+n')]\setminus T} \left(\bigotimes_{i=1}^m U_i \left(\rho \otimes |0\rangle\langle 0|^{\otimes n'} \right) U_i^\dagger \right), \text{tr}_{[m(n+n')]\setminus T} \left(\bigotimes_{i=1}^m U_i \left(\rho_{\text{mm}} \otimes |0\rangle\langle 0|^{\otimes n'} \right) U_i^\dagger \right) \right) \leq \varepsilon$$

for $\varepsilon \leq 2^{-\tilde{\Omega}(k \log n)}$. This is because suppose the above equation holds, according to the fact that χ_T are orthogonal under the null distribution, we have

$$\mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{s}, \mathbf{s}'} \overline{D}_{\rho, m}(\mathbf{s}) \overline{D}_{\rho', m}(\mathbf{s}') \chi_T(\mathbf{s}) \chi_T(\mathbf{s}') \leq \sum_{\mathbf{s}[T], \mathbf{s}'[T]} \varepsilon^2 \chi_T(\mathbf{s}) \chi_T(\mathbf{s}') \leq 2^{2|T|} \varepsilon^2$$

and thus

$$\left\| \mathbb{E}_{\rho \sim \mathcal{E}} \overline{D}_{\rho, m}^{\leq k} - \mathbf{1} \right\|^2 = \sum_{|T| \leq k, T \neq \emptyset} \mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \mathbb{E}_{\mathbf{s}, \mathbf{s}'} \overline{D}_{\rho, m}(\mathbf{s}) \overline{D}_{\rho', m}(\mathbf{s}') \chi_T(\mathbf{s}) \chi_T(\mathbf{s}') \leq k(m(n + n'))^k 2^k \varepsilon^2 \leq o(1)$$

as $n' \leq n$, which finishes the proof for Theorem 24.

Appendix K. Deferred proofs in frameworks for adaptive measurements

K.1. Proof of Theorem 33

By Theorem 28, given an ε' -approximate state $2km_0$ design with $0 \leq \varepsilon' \leq 1/2$, we have

$$\mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\left\langle \overline{D}_{\rho}^{\leq \infty}, \overline{D}_{\rho'}^{\leq \infty} \right\rangle - 1 \right)^{2km_0} \leq O((2km_0)^2 2^{2km_0} (\varepsilon' + 2^{-n})).$$

Given any $s_{1:km_0-1}$, we have

$$\begin{aligned}\mathbb{E}_{\rho \sim \mathcal{E}} \left[\left(\overline{D}_\rho^{\leq \infty}(s) \right)^{2(km_0-1)} \right] &\leq (1 + \varepsilon') \frac{2^{2(km_0-1)n} (2(km_0-1))!}{2^n (2^n + 1) \dots (2^n + 2(km_0-1) - 1)} \\ &\leq (1 + \varepsilon') (2(km_0-1))!.\end{aligned}$$

Therefore, we can choose

$$M = 2(2(km_0-1))! = 2^{O(km_0 \log k \log m_0)}.$$

By Theorem 32, we can thus choose

$$\begin{aligned}\varepsilon &= 2^{-\Omega(k^3 m_0 \log k \log m_1 \log m_0 \log M)} \\ &= 2^{-\Omega(k^4 m_0^2 \log^2 k \log m_1 \log^2 m_0)},\end{aligned}$$

which indicates that

$$\varepsilon' = 2^{-\Omega(k^4 m_0^2 \log^2 k \log m_1 \log^2 m_0)}.$$

K.2. Proof of Theorem 37

By Theorem 28, given an ε' -approximate state $2k$ design with $0 \leq \varepsilon' \leq 1/2$, we have

$$\mathbb{E}_{\rho, \rho' \sim \mathcal{E}} \left(\left\langle \overline{D}_\rho^{\leq \infty}, \overline{D}_{\rho'}^{\leq \infty} \right\rangle - 1 \right)^{2k} \leq O((2k)^2 2^{2k} (\varepsilon' + 2^{-n})).$$

We also have

$$\begin{aligned}\mathbb{E}_{\rho \sim \mathcal{E}} \left[\left(\overline{D}_\rho^{\leq \infty}(s) - 1 \right)^{2(m_1-1)} \right] &\leq (1 + \varepsilon') \frac{2^{2(m_1-1)n} (2m_1)!}{2^n (2^n + 1) \dots (2^n + 2(m_1-1) - 1)} \\ &\leq (1 + \varepsilon') (2m_1)!.\end{aligned}$$

Therefore, we can choose

$$M' = 2(2m_1)! = 2^{O(m_1 \log m_1)}.$$

By Theorem 36, we can thus choose

$$\begin{aligned}\varepsilon &= 2^{-\Omega(km_1 \log k \log m_0 \log M')} \\ &= 2^{-\Omega(km_1^2 \log^2 m_1 \log m_0 \log k)},\end{aligned}$$

which indicates that

$$\varepsilon' = 2^{-\Omega(km_1^2 \log^2 m_1 \log m_0 \log k)}.$$

Appendix L. Auxiliary lemmas for quantum planted biclique

Fact 74 Let $\beta_{d,s}$ be as defined in (20). Then

$$\beta_{d,s} = 1 + \sum_{a=1}^{\infty} c_a(s) \cdot d^{-a},$$

for coefficients $c_a(s)$ which are degree- $2a$ polynomials in s satisfying

$$|c_a(s)| \leq (1+s)^{2a}$$

for all $s \geq 0$.

Proof Taylor expanding $\log(\beta_{d,s})$ and exponentiating, we find that

$$c_a(s) = (-1)^a \sum_{\ell=1}^a \frac{1}{\ell!} \sum_{\substack{i_1+\dots+i_\ell=a+\ell \\ i_j \geq 2 \forall j}} \sigma_{i_1}(s) \cdots \sigma_{i_\ell}(s) \cdot \frac{1}{(i_1-1) \cdots (i_\ell-1)},$$

where $\sigma_i(s) \triangleq 1^{i-1} + \dots + (s-1)^{i-1}$. Note that $\sigma_i(s) \leq s^i$, so the inner sum above can be bounded by $s^{a+\ell} \cdot \binom{a-1}{\ell-1}$. Finally, $\sum_{\ell=1}^a \frac{1}{\ell!} \binom{a-1}{\ell-1} s^{a+\ell} \leq (1+s)^{2a}$, as claimed. $\blacksquare$

Lemma 75 Let $\pi \in \mathcal{S}_W$ be a permutation with no fixed points. Then

$$\mathbb{E}_{\mathbf{x}}[|\langle \psi_{\mathbf{x}|_W} | P_\pi | \psi_{\mathbf{x}|_W} \rangle|^2] \leq d^{-|W|/2}.$$

Proof Let $C = (\mathbf{i}, \pi(\mathbf{i}), \pi^2(\mathbf{i}), \dots, \pi^{p-1}(\mathbf{i}) = \mathbf{i})$ be any cycle in π . By assumption, $p > 1$. Then if the cycle length p is even,

$$\mathbb{E} \prod_{\ell=0}^{p-1} \langle \psi_{x_{\pi^\ell(\mathbf{i})}} | \psi_{x_{\pi^{\ell+1}(\mathbf{i})}} \rangle^2 \leq \mathbb{E} \prod_{\ell=0}^{p/2-1} \langle \psi_{x_{\pi^{2\ell}(\mathbf{i})}} | \psi_{x_{\pi^{2\ell+1}(\mathbf{i})}} \rangle^2 = (1/d)^{p/2},$$

where we used that for all $\mathbf{i} \in [m] \times [n]$, $\mathbb{E}[|\psi_{x_{\mathbf{i}}}\rangle \langle \psi_{x_{\mathbf{i}}}|] = I/d$. Similarly, if the cycle length p is odd, then because $p > 1$,

$$\mathbb{E} \prod_{\ell=0}^{p-1} \langle \psi_{x_{\pi^\ell(\mathbf{i})}} | \psi_{x_{\pi^{\ell+1}(\mathbf{i})}} \rangle^2 \leq \mathbb{E} \prod_{\ell=1}^{p/2-1/2} \langle \psi_{x_{\pi^{2\ell}(\mathbf{i})}} | \psi_{x_{\pi^{2\ell+1}(\mathbf{i})}} \rangle^2 \cdot \langle \psi_{x_{\pi^p(\mathbf{i})}} | \psi_{x_{\mathbf{i}}} \rangle^2 \quad (25)$$

$$= (1/d)^{(p+1)/2} \cdot \mathbb{E} \|\psi_{x_{\mathbf{i}}}\|^4 = (1/d)^{(p+1)/2} \leq (1/d)^{p/2}. \quad (26)$$

The claim follows as the sum of the lengths of the cycles in π is equal to $|W|$. $\blacksquare$