

Learning junta distributions and quantum junta states, and QAC⁰ circuits

Francisco Escudero Gutiérrez*

Abstract

In this work we consider the problems of learning junta distributions, their quantum counterpart, quantum junta states, and QAC⁰ circuits, which we show to be juntas.

Junta distributions. A probability distribution $p : \{-1, 1\}^n \rightarrow [0, 1]$ is a k -junta if it only depends on k variables. We show that they can be learned with to error ε in total variation distance from $O(2^k \log(n)/\varepsilon^2)$ samples, which quadratically improves the upper bound of Aliakbarpour et al. (COLT'16) and matches their lower bound in every parameter.

Junta states. We initiate the study of n -qubit states that are k -juntas, those that are the tensor product of a k -qubit state and an $(n - k)$ -qubit maximally mixed state. We show that these states can be learned with error ε in trace distance with $O(12^k \log(n)/\varepsilon^2)$ single copies. We also prove a lower bound of $\Omega((4^k + \log(n))/\varepsilon^2)$ copies. Along the way, we give a new proof of the optimal performance of Classical Shadows based on Pauli analysis.

QAC⁰ circuits. Nadimpalli et al. (STOC'24) recently showed that the Pauli spectrum of QAC⁰ circuits (with not too many auxiliary qubits) is concentrated on low-degree. We remark that they showed something stronger, namely that the Choi states of those circuits are close to be juntas. As a consequence, we show that n -qubit QAC⁰ circuits with size s , depth d and a auxiliary qubits can be learned from $2^{O(\log(s^2 2^a)^d)} \log(n)$ copies of the Choi state, improving the $n^{O(\log(s^2 2^a)^d)}$ by Nadimpalli et al. In addition, we use this remark to improve on the lower bounds against QAC⁰ circuits to compute the address function.

1 Introduction

One of the main questions of computational learning theory is how efficiently can we learn an unknown object that is promised to have some structure. Two of the most studied structured objects are juntas, which are multi-bit or multi-qubit objects where only a few of the bits or qubits are relevant, and constant-depth circuits. There is plenty of literature about learning junta objects, such as Boolean juntas, junta distributions, quantum junta unitaries and quantum junta channels [AS07, ABR16, CJLW21, CNY23, BY23]. Two celebrated models of constant-depth circuits that have been studied from the point of view of learning are AC⁰ circuits and their quantum analogue, QAC⁰ circuits [LMN93, EIS22, NPVY23].

We continue this line of research by improving the upper bounds on learning classical junta distributions and QAC⁰ circuits, and by proving the first results on learning junta states. All of our upper bounds exploit that the considered objects satisfy two properties: their Fourier/Pauli expansions are close to be supported on a few low-degree sets. In other words, they are low-degree

*Qusoft and CWI feg@cw.nl

and sparse. In the case of juntas these two properties follow from definition, while for QAC⁰ circuits these properties are implicit in previous work and we uncover them here.

1.1 Summary of our results

We summarize our learning upper bounds in the following Table 1, where n is the number of bits or qubits, k stands for the number of relevant variables of a junta, s is the size and d the depth of a circuit, and ε is the error parameter with respect to metrics that we will specify later. In the case of classical objects, the complexity measure we consider is the sample complexity, while in the quantum case we consider the copy complexity.

	<i>Classical</i>	<i>Quantum</i>	
	Junta distributions	Junta states	QAC⁰ circuits
Previous best	$2^{2k} \log(n)/\varepsilon^4$ [ABR16]	—	$n^{\log(s/\varepsilon)^d}$ [NPVY23]
Our result	$2^k \log(n)/\varepsilon^2$	$12^k \log(n)/\varepsilon^2$	$2^{\log(s/\varepsilon)^d} \log(n)$

Table 1: Summary of our upper bounds.

Before we discuss our results in more detail, we make a few remarks about our main results.

- (i) **QAC⁰ circuits.** For constant d , s and ε , our result exponentially improves previous work. However, in the usual regime where $s = \text{poly}(n)$ and d, ε are constants, it yields a quasi-polynomial number of samples, which was already attained in previous work [NPVY23].
- (ii) **Junta states.** Recent works study the junta-learning problem of unitaries and quantum channels [CNY23, BY23], but there seems to be no previous work about quantum junta states. Hence, our result for junta states fills a gap in the literature. We also provide a $\Omega((4^k + \log(n))/\varepsilon^2)$ lower bound that shows that our upper bound cannot be improved by much.
- (iv) **Junta distributions.** Our upper bound is essentially optimal, as it matches the lower bound $\Omega(2^k/\varepsilon^2 + k \log(n)/\varepsilon)$ of [ABR16] in every parameter.

1.2 Learning junta distributions

Learning in the presence of irrelevant information, such as the *dummy* variables appearing in juntas, is one of the most famous yet open problems of classical computational learning theory since the 90's [Blu94, BHL95, BL97]. There are numerous works that consider the problems of learning and testing junta Boolean functions and distributions [MOS03, Val12, ABR16, CJLW21, CDL⁺24, NP24]. In particular, Aliakbarpour, Blais, and Rubinfeld considered the problem of learning a junta distribution $p : \{-1, 1\}^n \rightarrow [0, 1]$ from samples $x \sim p(x)$. They showed that in order to estimate p up to error ε in total variation distance, $O(2^{2k} k \log(n)/\varepsilon^4)$ samples suffice and $\Omega(2^k/\varepsilon^2 + k \log(n)/\varepsilon)$ are needed [ABR16]. We quadratically improve their upper bound matching their lower bound in every parameter.

Theorem 1. Let $p : \{-1, 1\}^n \rightarrow [0, 1]$ be a k -junta distribution. The distribution can be learned with error ε in total variation distance and success probability $\geq 1 - \delta$ with

$$O\left(\frac{2^k k \log(n/\delta)}{\varepsilon^2}\right)$$

samples.

As the upper bound of Aliakbarpour et al., ours exploits the fact that k -juntas have Fourier degree at most k . Our improvement comes from also using that their Fourier spectrum is concentrated on at most 2^k sets.

1.3 Learning junta states

In quantum learning theory the most commonly studied objects are states, unitaries and channels [OW15, HHJ⁺17, HKOT23]. By contrast, the problems of learning k -junta unitaries and channels were recently studied [CJLW21, BY23], but to the best of our knowledge no one has explored the version for states.

Definition 2. An n -qubit state ρ is said to be a k -junta state if there are a set $K \subseteq [n]$ of size k and a state ρ_K defined on K such that

$$\rho = \rho_K \otimes \frac{I_{[n]-K}}{2^{n-k}}.$$

In other words, ρ is a k -junta state if it is the tensor product of a k -qubit state and the maximally mixed state on the rest of the qubits.

Note that k -junta states are the quantum generalization of k -junta distributions, so the problem of learning them is the quantum analogue of the problem considered by Aliakbarpour et al. [ABR16]. We prove a nearly-optimal result for this problem in terms of copy complexity.

Theorem 3. Let ρ be a n -qubit k -junta quantum state. Then, ρ can be learned with error ε in trace distance and success probability $\geq 1 - \delta$ using

$$O\left(\frac{12^k \log(n/\delta)}{\varepsilon^2}\right)$$

copies of ρ , and $\Omega((\log(n) + 4^k)/\varepsilon^2)$ are necessary for this task. Furthermore, the algorithm just does Pauli measurements on single copies of the state.

For the upper bound we perform Classical Shadow tomography with Pauli measurements [HKP20, EFH⁺22]. Furthermore, we include a novel proof of the rigorous guarantees of the Classical Shadows algorithm based on Pauli analysis that might be of independent interest (see Theorem 8). The lower bound $\Omega(4^k/\varepsilon^2)$ follows from the lower bound by Haah et al. to learn k -qubit states [HHJ⁺17], and for the lower bound $\Omega(\log(n)/\varepsilon^2)$ we show that there are n states that 1-junta and difficult to distinguish.

1.4 Learning QAC⁰ circuits

QAC⁰ circuits were proposed by Moore as the quantum analogue of AC⁰ circuits [Moo99]. In that work Moore asked whether QAC⁰ circuits can compute parity, and despite various efforts the question remains open [FFG⁺03, PFGT20, Ros20, NPVY23, ADOY24]. In a recent work Nadimpalli, Parham, Vasconcelos and Yuen made progress in this direction, by showing that the Pauli spectrum of the Choi state of a QAC⁰ circuit with not too many auxiliary qubits is concentrated on low-degree. However, we note that they proved something stronger, and is that the Pauli spectrum of the Choi state of QAC⁰ circuits is not only concentrated on low-degree, but also the Choi state is close to be a junta (see Theorem 10). Using this, alongside the algorithm of Theorem 3, we can prove the following result.

Theorem 4. *Let ρ be the Choi state of a n -qubit QAC⁰ circuit with size s , depth d , and a auxiliary qubits. Then, using*

$$2^{O((\log(s^2 2^a/\varepsilon))^d)} \log(n/\delta)$$

singles copies of ρ , one can output a ρ' such that

$$2^n \|\rho - \rho'\|_F^2 \leq \varepsilon.$$

Furthermore, the algorithm just does Pauli measurements on single copies of the state.

The only previous result on learning QAC⁰ circuits was [NPVY23, Theorem 39], and our Theorem 4 improves it from $n^{O((\log(s^2 2^a/\varepsilon))^d)}$ copies to $2^{O((\log(s^2 2^a/\varepsilon))^d)} \log(n)$ copies. At this point it might be unclear why we have chosen to learn the Choi state of the circuit in the 2^n -Frobenius norm.¹ The reason why this is a good figure of merit for this learning task is explained in Section 4.1.1.²

In addition, in Section 4.1.3 we use that QAC⁰ are close to juntas, and not only to low-degree, to show new lower bounds for computing the address function, which is the canonical example of a low-degree function that depends on many variables.

1.5 Our algorithms in a nutshell

All of our algorithms are refinements of the *low-degree algorithm* of Linial, Mansour and Nisan [LMN93]. To sketch them, for simplicity, we will consider functions $f : \{-1, 1\}^n \rightarrow [-1, 1]$. Assume that we are promised that the Fourier spectrum of f is supported on L monomials of degree at most d , i.e.

$$f(x) = \sum_{s \in [L]} \hat{f}(S_s) \prod_{i \in S_s} x_i$$

for some $S_s \subseteq [n]$ with $|S_s| \leq d$. First, we will see how the low-degree algorithm would perform to learn f from samples $(x, f(x))$ where x is uniformly picked from $\{-1, 1\}^n$.

Low-degree algorithm

Step 1. For every $|S| \leq d$, obtain $\hat{f}'(S)$ that approximates $\hat{f}(S)$ up to error $\sqrt{\varepsilon/n^d}$.

Output. We output $f'(x) = \sum_{|S| \leq d} \hat{f}'(S) \prod_{i \in S} x_i$.

¹In a soon-to-appear work, Huang and Vasconcelos extend this result to recover not only the Choi-state, but also the unitary defined by the circuit [VH24].

²This is the same figure of merit as the one considered in [NPVY23], but the authors of that work use a slightly different notation.

It is well-known that with $(1/\alpha^2) \cdot \log(M)$ samples one can estimate M Fourier coefficients of f up to error α , so the low-degree algorithm requires $(n^d/\varepsilon^2) \cdot \log(n^d)$ samples. Now, f' is close to f , because

$$\sum_{|S| \leq d} |\widehat{f}(S) - \widehat{f}'(S)|^2 \leq \sum_{|S| \leq d} \frac{\varepsilon}{n^d} \leq \varepsilon,$$

where in the first inequality we have used the guarantees of Step 1, and in the second that $|\{|S| \leq d\}| \leq n^d$. In particular, this implies that $\Pr[f(x) \neq \text{sign}(f'(x))] \leq \varepsilon$.

However, note that the low-degree algorithm does not use that f is supported on L monomials out of the $\sim n^d$ low-degree monomials. Using that, one can improve on the low-degree algorithm.

Low-degree and sparse algorithm

Step 1. For every $|S| \leq d$, obtain $\widehat{f}'(S)$ that approximates $\widehat{f}(S)$ up to error $\sqrt{\varepsilon/4L}$.

Step 2. For every $|S| \leq d$, if $|\widehat{f}''(S)| \leq \sqrt{\varepsilon/4L}$, set $\widehat{f}'(S) = 0$, otherwise set $\widehat{f}''(S) = \widehat{f}'(S)$.

Output. We output $f''(x) = \sum_{|S| \leq d} \widehat{f}''(S) \prod_{i \in S} x_i$.

Note that Step 1 now just requires $(L/\varepsilon^2) \cdot \log(n^d)$ samples, considerably less than the $(n^d/\varepsilon^2) \cdot \log(n^d)$ samples of the low-degree algorithm. Also, notice that by adding the rounding of Step 2 we make sure that $\widehat{f}''(S) = 0$ for $S \notin \{S_1, \dots, S_L\}$, and every $S \in \{S_1, \dots, S_L\}$ satisfies that $|\widehat{f}(S) - \widehat{f}''(S)| \leq \sqrt{\varepsilon/L}$. Hence, we still have that

$$\sum_{|S| \leq d} |\widehat{f}(S) - \widehat{f}'(S)|^2 = \sum_{|S| \in \{S_1, \dots, S_L\}} |\widehat{f}(S) - \widehat{f}'(S)|^2 \leq \sum_{|S| \in \{S_1, \dots, S_L\}} \frac{\varepsilon}{L} = \varepsilon,$$

where in the first equality we have used that $\widehat{f}'(S) = 0$ for every $S \notin \{S_1, \dots, S_L\}$.

In conclusion, if we are promised that the Fourier or Pauli spectrum of our object is supported on $L \ll n^d$ coefficients of degree at most d , one should add a simple rounding step in the low-degree algorithm. This remark was already exploited by Eskenazis, Ivanišvili and Streck to learn Boolean functions [EIS22, Theorem 2], and we extend it to other contexts.

2 Preliminaries

Some notation. We will use ℓ_q to denote the q -norm with the counting measure, and L_q to denote the q -norm with the uniform probability measure. All expectations are taken with respect to the uniform probability measure unless otherwise stated. All logarithms are in base 2.

Concentration inequalities. We state a few concentration inequalities that we use often.

Lemma 5 (Hoeffding bound). *Let X_1, \dots, X_m be independent-random variables that satisfy $-a_i \leq X_i \leq a_i$ for some $a_i > 0$. Then, for any $\tau > 0$, we have*

$$\Pr \left[\left| \sum_{i \in [m]} X_i - \sum_{i \in [m]} \mathbb{E}[X_i] \right| > \tau \right] \leq 2 \exp \left(-\frac{\tau^2}{2(a_1^2 + \dots + a_m^2)} \right).$$

Lemma 6 (Bernstein inequality). *Let X_1, \dots, X_m be independent-random variables with $|X_i| \leq M$ for some $M > 0$. Then,*

$$\Pr \left[\left| \sum_{i \in [m]} X_i - \sum_{i \in [m]} \mathbb{E}[X_i] \right| > \tau \right] \leq 2 \exp \left(- \frac{\tau^2/2}{\sum_{i \in [m]} \text{Var}[X_i] + \tau M/3} \right).$$

Lower bounds for state learning. The standard way of showing lower bounds for state learning is via an argument of Holevo. To state it, we first introduce the Holevo information of a set of states $\{\rho_i\}$, which is given by

$$\chi(\{\rho_i\}) = S \left(\frac{1}{n} \sum_{i \in [n]} \rho_i \right) - \frac{1}{n} \sum_{i \in [n]} S(\rho_i),$$

where $S(\rho) = -\text{Tr}[\rho \log(\rho)]$ is the von Neumann entropy. Now we are ready to write the precise statement we will use to show a lower bound for learning k -junta states. For a proof see [MMB⁺24, Lemma S14].

Lemma 7. *Let $\{\rho_i\}_{i \in [M]}$ be a family of M states that satisfy $\|\rho_i - \rho_j\|_{\text{tr}} \geq \varepsilon$ for every $i \neq j$. Assume that T copies are sufficient to learn this family of states with probability $\geq 2/3$. Then,*

$$\chi(\{\rho_i^{\otimes T}\}) = \Omega(\log(M)).$$

Additionally, we will need two facts about von Neumann entropy. The first is its additivity under tensor product,

$$S(\rho_A \otimes \rho_B) = S(\rho_A) + S(\rho_B), \tag{1}$$

and the second is subadditivity

$$S(\rho_{AB}) \leq S(\rho_A) + S(\rho_B). \tag{2}$$

3 Learning junta distributions

In this section we prove Theorem 1. We begin by recalling what is the usual model for learning distributions. Given a distribution $p : \{-1, 1\}^n \rightarrow [0, 1]$, one can access it by sampling $x \in \{-1, 1\}^n$ with probability $p(x)$. The goal of the learner is to use a few samples to output another distribution $p' : \{-1, 1\}^n \rightarrow [0, 1]$ that is ε -close to p in total variation distance, which is given by

$$d_{\text{TV}}(p, p') = \frac{1}{2} \|p - p'\|_{\ell_1} = \frac{1}{2} \sum_{x \in \{-1, 1\}^n} |p(x) - p'(x)|.$$

If $p : \{-1, 1\}^n \rightarrow [0, 1]$ is a k -junta depending on the variables of a set $K \subseteq [n]$ of size k , then it can be written as

$$p(x) = \sum_{S \subseteq K} \widehat{p}(S) \prod_{i \in S} x_i,$$

where $\widehat{p}(S) = \mathbb{E}_{x \in \{-1, 1\}^n} p(x) \prod_{i \in S} x_i$ are the Fourier coefficients of p . Note that all non-zero Fourier coefficients of a k -junta correspond to monomials of degree $\leq k$ and there is at most 2^k of them. We use this to show a nearly-optimal algorithm to learn k -junta distributions.

Proof of Theorem 1: Let $T = O\left(\frac{2^k}{\varepsilon^2} k \log\left(\frac{n}{\delta}\right)\right)$ be the number of samples (x^1, \dots, x^T) we take. For every $S \subseteq [n]$ with $|S| \leq k$ we define the empirical Fourier coefficient

$$\hat{p}'(S) = \frac{1}{2^{nT}} \sum_{s \in [T]} \prod_{i \in S} x_i^s.$$

Then, $\mathbb{E}[\hat{p}'(S)] = \hat{p}(S)$. Moreover, by a Hoeffding bound (Lemma 5) and a union bound over the at most n^k sets of size at most k , we have that with probability $\geq 1 - \delta$

$$|\hat{p}'(S) - \hat{p}(S)| \leq \frac{\varepsilon}{2 \cdot 2^n \sqrt{2^k}} \quad \text{for every } |S| \leq k. \quad (3)$$

For every $|S| \leq k$, we define

$$\hat{p}''(S) = \begin{cases} 0 & \text{if } |\hat{p}'(S)| \leq \varepsilon / (2 \cdot 2^n \cdot \sqrt{2^k}), \\ \hat{p}'(S) & \text{otherwise.} \end{cases} \quad (4)$$

Now, from Eq. (3) it follows that if

$$\hat{p}(S) = 0, \text{ then } \hat{p}''(S) = 0, \quad (5)$$

so in particular if K is the set of (at most k) variables that p depends on n , then

$$S \not\subseteq K, \text{ then } \hat{p}''(S) = 0. \quad (6)$$

In addition, we have that for every S with $|S| \leq k$

$$|\hat{p}''(S) - \hat{p}(S)| \leq \frac{\varepsilon}{2^n \sqrt{2^k}}. \quad (7)$$

We define $p''(x) = \sum_{|S| \leq k} \hat{p}''(S) \prod_{i \in S} x_i$ and claim that is close to p . Indeed,

$$\begin{aligned} \|p - p'\|_{L_2}^2 &= \sum_{S \subseteq K} |\hat{p}(S) - \hat{p}''(S)|^2 + \sum_{S \not\subseteq K} |\hat{p}''(S)|^2 \\ &= \sum_{S \subseteq K} |\hat{p}(S) - \hat{p}''(S)|^2 \\ &= 2^k \frac{\varepsilon^2}{2^{2n} 2^k} \\ &= \frac{\varepsilon^2}{2^{2n}}, \end{aligned}$$

where in the first line we have used Parseval's identity; in the second line we have used Eq. (6) and in the third Eq. (7) and that $|\mathcal{P}(K)| \leq 2^k$. Hence, $\|p - p'\|_{L_2} \leq \varepsilon/2^n$. Finally, as $\|\cdot\|_{\ell_1} \leq 2^n \|\cdot\|_{L_2}$, the result follows. \square

4 Learning quantum junta states and QAC⁰ circuits

In this section we prove Theorems 3 and 4. We begin by recalling the usual learning model for quantum states. We are given copies of ρ on which we can measure. We will only perform measurements on single copies of ρ , and the measurements may be different for different copies of

ρ . The goal is to use the outcomes of these measurements to output another state ρ' that is ε -close in trace distance to ρ , meaning that

$$d_{\text{tr}}(\rho, \rho') = \|\rho - \rho'\|_{\text{tr}} \leq \text{Tr}[\|\rho - \rho'\|] \leq \varepsilon.$$

Note that an n -qubit state ρ is a k -junta state if and only if it can be written as

$$\rho = \sum_{\substack{P \in \{I, X, Y, Z\}^{\otimes n} \\ \text{supp}(P) \subseteq K}} \widehat{\rho}(P) P,$$

for some $K \subseteq [n]$ of size k , where $\widehat{\rho}(P) = \text{Tr}[\rho P]/2^n$ are the Pauli coefficients, $\text{supp}(\otimes_{i \in [n]} P_i) = \{i \in [n] : P_i \neq I\}$,

$$I = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad X = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}, \quad Y = \begin{pmatrix} 0 & -i \\ i & 0 \end{pmatrix}, \quad \text{and} \quad Z = \begin{pmatrix} 1 & 0 \\ 0 & -1 \end{pmatrix}.$$

We stress that a quantum state is a generalization of a probability distribution, and that this generalization extends to the Pauli spectrum several notions related to the Fourier spectrum. Indeed, given a probability distribution $p : \{-1, 1\}^n \rightarrow [0, 1]$, it defines a n -qubit quantum state

$$\rho_p = \sum_{x \in \{-1, 1\}^n} p(x) |x\rangle \langle x|,$$

that satisfies $\widehat{\rho}_p(P) = \widehat{p}(\text{supp}(P))$ if $P \in \{I, Z\}^{\otimes n}$, and $\widehat{\rho}_p(P) = 0$ otherwise. Similarly, the biggest size of the support of a P such that $\widehat{\rho}(P) \neq 0$ generalizes the notion of degree. Furthermore, p is a k -junta distribution if and only if ρ_p is a k -junta state.

As in the classical case, the non-zero Pauli coefficients of a k -junta state correspond to *low-degree* Pauli operators, those with small support, and they are at most 4^k . Using this we could learn k -junta states in a similar way that we used to learn k -junta distributions if we had an mechanism way of learning the low-degree Pauli coefficients. Such a mechanism is the Classical Shadows algorithm by Huang, Kueng and Preskill [HKP20], which was later improved by Elben et al. [EFH⁺22, Sec.II.B.].

Theorem 8 ([HKP20, EFH⁺22]). *Let ρ be a n -qubit state. Then, by performing Pauli measurements on*

$$O \left(\frac{3^k \log((3n)^k / \delta)}{2^{2n} \varepsilon^2} \right)^3$$

single copies of ρ one can output estimates $\widehat{\rho}'(P)$ such that with success probability $\geq 1 - \delta$ satisfy

$$|\widehat{\rho}(P) - \widehat{\rho}'(P)| \leq \varepsilon$$

for every $P \in \{I, X, Y, Z\}^{\otimes n}$ with $|\text{supp}(P)| \leq k$.

We include a proof of Theorem 8 that uses a novel Pauli analytic approach inspired on the proof of the non-commutative Bohnenblust-Hille inequality by Volberg and Zhang [VZ23].

³The factor 2^{2n} in the denominator appears because the Pauli coefficients are the expectations of the Pauli observables over 2^n .

Proof of Theorem 8: We will make use of $T = O(3^k \log((3n)^k/\delta)/(2^{2n}\varepsilon^2))$ copies of ρ . Let B_Q be a basis that diagonalizes $Q \in \{X, Y, Z\}^{\otimes n}$. For every $s \in [T]$, we will pick $Q^s \in \{X, Y, Z\}^{\otimes n}$ independently uniformly at random and measure ρ in the basis B_{Q^s} . For every $i \in [n]$, let $x_i^s = \pm 1$ if the outcome of the s -th measurement on the i -th qubit is the ± 1 eigen-space of Q_i^s . Then, for every $P \in \{I, X, Y, Z\}^{\otimes n}$ we define an empirical estimator of $\widehat{\rho}(P)$ via

$$\widehat{\rho}'(P) = \frac{3^{|\text{supp}(P)|}}{2^{nT}} \sum_{s \in [T]} \prod_{i \in \text{supp}(P)} x_i^s \delta_{P_i=Q_i^s}.$$

We claim that $\widehat{\rho}'(P)$ equals $\widehat{\rho}(P)$ on expectation. Indeed,

$$\begin{aligned} \mathbb{E}\widehat{\rho}'(P) &= \frac{3^{|\text{supp}(P)|}}{2^n} \mathbb{E}_{Q \in \{X, Y, Z\}^{\otimes n}} \prod_{i \in \text{supp}(P)} \sum_{x_i \in \{-1, 1\}} \Pr_{\rho, B_{Q_i}}[x_i] x_i \delta_{P_i=Q_i} \\ &= \frac{3^{|\text{supp}(P)|}}{2^n} \mathbb{E}_{Q \in \{X, Y, Z\}^{\otimes \text{supp}(P)}} \prod_{i \in \text{supp}(P)} \sum_{x_i \in \{-1, 1\}} x_i \Pr_{\rho, B_{Q_i}}[x_i] \delta_{P_i=Q_i} \\ &= \frac{1}{2^n} \prod_{i \in \text{supp}(P)} \sum_{x_i \in \{-1, 1\}} x_i \Pr_{\rho, P_i}[x_i] \\ &= \frac{1}{2^n} \prod_{i \in \text{supp}(P)} \text{Tr}[\rho P_i] = \frac{1}{2^n} \text{Tr}[\rho P] = \widehat{\rho}(P), \end{aligned}$$

the first line is true because the expectation of $\widehat{\rho}'(P)$ does not change if T changes; the second line follows from the fact that inside \mathbb{E}_Q there is no dependence on the variables outside $\text{supp}(P)$; the third line is true because the term inside \mathbb{E}_Q is 0 unless $Q_i = P_i$ for every $i \in \text{supp}(P)$; and fourth line is true because $\Pr_{\rho, B_{P_i}}[x_i] = \text{Tr}[\rho |P_i(x_i)\rangle \langle P_i(x_i)|]$ where $|P_i(x_i)\rangle$ is a unit eigenvector of P_i with eigenvalue x_i .

In addition, the second moment (and thus the variance) of $\widehat{\rho}'(P)$ for $T = 1$ is considerably smaller than the trivial upper bound $\mathbb{E}[|\widehat{\rho}'(P)|^2] \leq \|\widehat{\rho}'(P)\|_\infty^2 = 9^{|\text{supp}(P)|}/4^n$. Indeed, for $T = 1$ we have

$$\begin{aligned} \mathbb{E}(\widehat{\rho}'(P))^2 &= \frac{9^{|\text{supp}(P)|}}{4^n} \mathbb{E}_{Q \in \{X, Y, Z\}^{\otimes n}} \prod_{i \in \text{supp}(P)} \sum_{x_i \in \{-1, 1\}} \Pr_{\rho, B_{Q_i}}[x_i] (x_i \delta_{P_i=Q_i})^2 \\ &= \frac{9^{|\text{supp}(P)|}}{4^n} \mathbb{E}_{Q \in \{X, Y, Z\}^{\otimes \text{supp}(P)}} \prod_{i \in \text{supp}(P)} \sum_{x_i \in \{-1, 1\}} \Pr_{\rho, B_{Q_i}}[x_i] \delta_{P_i=Q_i} \\ &= \frac{9^{|\text{supp}(P)|}}{4^n} \mathbb{E}_{Q \in \{X, Y, Z\}^{\otimes \text{supp}(P)}} \prod_{i \in \text{supp}(P)} \delta_{P_i=Q_i} \\ &= \frac{3^{|\text{supp}(P)|}}{4^n}, \end{aligned}$$

where the second line follows from the fact that the quantity inside $\mathbb{E}_{Q \in \{X, Y, Z\}^{\otimes n}}$ does not depend on the variables outside of $\text{supp}(P)$ and the fact that $(x_i \delta_{P_i=Q_i})^2 = \delta_{P_i=Q_i}$; and the third line is true because $\sum_{x_i} \Pr_{\rho, B_{Q_i}}[x_i] = 1$.

Now, the claimed result follows from the Bernstein inequality and an union bound over the at most $(3n)^k$ Pauli operators of degree lower than k . \square

Our algorithm to learn k -junta states is robust, in the sense that it also applies in the case of the Pauli spectrum of the state is $(\varepsilon^2/2^{2n})$ -concentrated on the Pauli coefficients corresponding to k -qubits, which is the case where it exists $K \subseteq [n]$ of size k such that

$$\sum_{\text{supp}(P) \not\subseteq K} |\widehat{\rho}(P)|^2 \leq \frac{\varepsilon^2}{2^{2n}}.$$

Theorem 9. *Let ρ be a n -qubit state whose Pauli spectrum is $(\varepsilon^2/2^{2n})$ -concentrated on a set of k qubits. Then, using*

$$O\left(\frac{12^k \log((3n)^k/\delta)}{\varepsilon^2}\right)$$

copies of ρ one can output ρ' such that with success probability $\geq 1 - \delta$ satisfies

$$\sum_{P \in \{I, X, Y, Z\}^{\otimes n}} |\widehat{\rho}'(P) - \widehat{\rho}(P)|^2 \leq \frac{2\varepsilon^2}{2^{2n}}.$$

In particular, $\|\rho' - \rho\|_{\text{tr}} \leq \sqrt{2}\varepsilon$. Furthermore, the algorithm just does Pauli measurements on single copies of the state.

Proof of Theorem 9: Similarly to the the proof of classical case Theorem 1, we use $T = O\left(\frac{12^k \log((3n)^k/\delta)}{\varepsilon^2}\right)$ copies of the state obtain an estimate $\widehat{\rho}'(P)$ for every P with $|\text{supp}(P)| \leq k$ such that

$$|\widehat{\rho}(P) - \widehat{\rho}'(P)| \leq \frac{\varepsilon}{4\sqrt{4^k}2^n}. \quad (8)$$

This can be done via Classical Shadows (see Theorem 8). Now, for every $P \in \{I, X, Y, Z\}^{\otimes n}$ we define

$$\widehat{\rho}''(P) = \begin{cases} 0 & |\text{supp}(P)| > k, \\ 0 & |\widehat{\rho}'(P)| \leq \varepsilon/(2 \cdot 2^n \cdot \sqrt{4^k}) \text{ and } |\text{supp}(P)| \leq k, \\ \widehat{\rho}'(P) & \text{otherwise.} \end{cases}$$

In particular, from Eq. (8) it follows that for every S with $|\text{supp}(S)| \leq k$ we have that

$$|\widehat{\rho}(P) - \widehat{\rho}''(P)| \leq \frac{\varepsilon}{2^n \sqrt{4^k}}. \quad (9)$$

In addition, we claim that for every $P \in \{I, X, Y, Z\}^{\otimes n}$

$$|\widehat{\rho}(P) - \widehat{\rho}''(P)| \leq |\widehat{\rho}(P)|. \quad (10)$$

Indeed, the only non-trivial case of Eq. (10) corresponds to P with $|\text{supp}(P)| \leq k$ and $|\widehat{\rho}'(P)| \geq \varepsilon/(2 \cdot 2^n \cdot \sqrt{4^k})$. In that case, we have that

$$\begin{aligned} |\widehat{\rho}(P)| &\geq |\widehat{\rho}'(P)| - |\widehat{\rho}(P) - \widehat{\rho}'(P)| \\ &\geq \varepsilon/(4 \cdot 2^n \cdot \sqrt{4^k}) \\ &\geq |\widehat{\rho}(P) - \widehat{\rho}'(P)| \\ &= |\widehat{\rho}(P) - \widehat{\rho}''(P)|, \end{aligned}$$

where the first is due to triangle inequality; the second line is true because of Eq. (8) and the hypothesis on P ; the third line again follows from Eq. (8); and the fourth line is true because of the choice of P and the definition of $\widehat{\rho}''(P)$.

Finally, we claim that $\rho'' = \sum_P \hat{\rho}''(P)P$ is a good approximation to ρ . Indeed, let $K \subseteq [n]$ be the subset of qubits where the spectrum of ρ is concentrated on, then

$$\begin{aligned} \sum_{P \in \{I, X, Y, Z\}^{\otimes n}} |\hat{\rho}(P) - \hat{\rho}''(P)|^2 &= \sum_{P \in \{I, X, Y, Z\}^{\otimes K}} |\hat{\rho}(P) - \hat{\rho}''(P)|^2 + \sum_{P \notin \{I, X, Y, Z\}^{\otimes K}} |\hat{\rho}(P) - \hat{\rho}''(P)|^2 \\ &\leq \sum_{P \in \{I, X, Y, Z\}^{\otimes K}} \frac{\varepsilon^2}{2^{2n} 4^k} + \sum_{P \notin \{I, X, Y, Z\}^{\otimes K}} |\hat{\rho}(P)|^2 \\ &\leq 2 \frac{\varepsilon^2}{2^{2n}}, \end{aligned}$$

where in the second line we have used Eqs. (9) and (10); and in the third line that $|\{I, X, Y, Z\}^{\otimes K}| = 4^k$ and that the spectrum of ρ is $(\varepsilon^2/2^{2n})$ -concentrated on K . \square

Now we are ready to prove our learning result for k -junta states.

Proof of Theorem 3: The upper bound follows from Theorem 9. The lower bound $\Omega(4^k/\varepsilon^2)$ follows from the fact that k -qubit states are k -juntas and the lower bound for learning k -qubit states of Haah et al. [HHJ⁺17]. For the lower bound $\Omega(\log(n)/\varepsilon^2)$ we will provide a set of n states $\{\rho_i\}_{i \in [n]}$ of n qubits that are 1-junta, and satisfy

$$\|\rho_i - \rho_j\|_{\text{tr}} \geq \varepsilon \text{ if } i \neq j, \quad (11)$$

$$\chi(\{\rho_i^{\otimes T}\}) \leq T\varepsilon^2 \text{ for every } T \in \mathbb{N}. \quad (12)$$

From Eqs. (11) and (12) the lower bound $\Omega(\log(n)/\varepsilon^2)$ follows from Lemma 7. For every $\varepsilon \in (0, 1/2)$ we define

$$\rho_\varepsilon = \frac{1}{2} \begin{pmatrix} 1 + \varepsilon & 0 \\ 0 & 1 - \varepsilon \end{pmatrix}.$$

For $i \in [n]$, we define

$$\rho_i = \frac{I}{2} \otimes \cdots \otimes \underbrace{\rho_\varepsilon}_{i\text{-th qubit}} \otimes \cdots \otimes \frac{I}{2}.$$

Eq. (11) holds because if $i \neq j$, then

$$\|\rho_i - \rho_j\|_{\text{tr}} = \left\| \rho_\varepsilon \otimes \frac{I}{2} - \frac{I}{2} \otimes \rho_\varepsilon \right\|_{\text{tr}} = \left\| \frac{1}{2} \begin{pmatrix} 0 & \varepsilon \\ \varepsilon & -\varepsilon \\ & & 0 \end{pmatrix} \right\|_{\text{tr}} = \varepsilon.$$

Proving Eq. (12) requires just a bit more work. We begin by noting that

$$|S(\rho_\varepsilon) - 1| \leq O(\varepsilon^2) \quad (13)$$

for every $\varepsilon < 1/2$. Indeed,

$$\begin{aligned} |S(\rho_\varepsilon) - 1| &= \left| - \sum_{x \in \{\pm 1\}} \frac{1 + x\varepsilon}{2} \log \left(\frac{1 + x\varepsilon}{2} \right) - 1 \right| = \left| \sum_{x \in \{\pm 1\}} \frac{1 + x\varepsilon}{2} \log(1 + x\varepsilon) \right| \\ &= \left| \sum_{x \in \{\pm 1\}} \frac{1 + x\varepsilon}{2} (x\varepsilon + O(\varepsilon^2)) \right| = O(\varepsilon^2), \end{aligned}$$

where in the second line we have applied Taylor's theorem. We recall that the Holevo information is given by

$$\chi(\{\rho_i^{\otimes T}\}) = S \underbrace{\left(\frac{1}{n} \sum_{i \in [n]} \rho_i^{\otimes T} \right)}_{(*)} - \underbrace{\frac{1}{n} \sum_{i \in [n]} S(\rho_i^{\otimes T})}_{(**)}.$$

We will analyze the terms (*) and (**) separately. We begin with (**):

$$(**) = S(\rho_1^{\otimes T}) = T[S(\rho_\varepsilon) + (n-1)] \geq Tn - O(T\varepsilon^2),$$

where we have applied additivity of the entropy under tensor product (see Eq. (1)) and Eq. (13). The analysis of the term (*) is a bit more involved:

$$\begin{aligned} (*) &= S \left(\frac{1}{n} \left\{ \rho_\varepsilon^{\otimes T} \otimes \left(\frac{I}{2} \right)^{\otimes T} \otimes \cdots \otimes \left(\frac{I}{2} \right)^{\otimes T} + \cdots + \left(\frac{I}{2} \right)^{\otimes T} \otimes \cdots \otimes \rho_\varepsilon^{\otimes T} \right\} \right) \\ &\leq nS \left(\frac{1}{n} \left\{ \rho_\varepsilon^{\otimes T} + (n-1) \left(\frac{I}{2} \right)^{\otimes T} \right\} \right) \\ &\leq nTS \left(\frac{1}{n} \left\{ \rho_\varepsilon + (n-1) \frac{I}{2} \right\} \right) \\ &= nTS \left(\rho_{\frac{\varepsilon}{n}} \right) \\ &\leq nT + TO(\varepsilon^2/n^2), \end{aligned}$$

where in the lines 2 and 3 we have applied subadditivity of the entropy (see Eq. (2)), and in the last line Eq. (13). Putting the analysis for terms (*) and (**) together, Eq. (12) follows. \square

4.1 QAC⁰ circuits

4.1.1 Very brief introduction to QAC⁰ circuits

A QAC⁰ is a circuit composed by single-qubit gates and Toffoli gates, which are the unitaries defined via

$$|x_1, \dots, x_l, b\rangle \rightarrow |x_1, \dots, x_l, b \cdot \text{AND}(x_1, \dots, x_l)\rangle,$$

where here $x_1, \dots, x_l, b \in \{-1, 1\}$ and $\text{AND}(x_1, \dots, x_l) = -1$ if and only if $x_1 = \dots = x_l = -1$. Given a $(n+a+1)$ -qubit QAC⁰ circuit one should think of the first n qubits as input qubits, of the next a qubits as auxiliary qubits and of the last qubit as an output qubit. Also, the last $a+1$ qubits are initialized in a fixed state σ . Hence, a QAC⁰ circuit defines an n -to-1 qubit channel via

$$\Phi_\sigma(\rho) = \text{Tr}_{[n+a]}[U(\rho \otimes \sigma)U^\dagger],$$

where U is the unitary implemented by the circuit and $\text{Tr}_{[n+a]}$ is the trace with respect to the input and auxiliary qubits. The Choi state of a QAC⁰ circuit is the Choi state of its correspondent channel, namely the $(n+1)$ -qubit state

$$\rho_{\Phi_\sigma} = \Phi_\sigma \otimes \text{Id}_n(|\text{EPR}_n\rangle \langle \text{EPR}_n|),$$

where $|\text{EPR}_n\rangle$ is the tensor product of n EPR states.

The original motivation when Moore introduced of QAC⁰ circuits was to use them to approximate Boolean functions $f : \{-1, 1\}^n \rightarrow \{-1, 1\}$ [Moo99], namely to approximate n -to-1 qubit channels like

$$\Phi_f(\rho) = \sum_{x \in \{-1, 1\}^n} \langle x | \rho | x \rangle |f(x)\rangle \langle f(x)|.$$

It is easy to check that the Choi state of these channels is given by

$$\rho_f = \frac{1}{2^n} \left(I^{\otimes(n+1)} + \sum_{S \subseteq [n]} \widehat{f}(S) Z_S \otimes Z \right), \quad (14)$$

where $Z_S = \otimes_{i \in n} Z^{\delta_{i \in S}}$. Hence, for $f, g : \{-1, 1\}^n \rightarrow \{-1, 1\}$, we have that

$$2^n \|\rho_f - \rho_g\|_F^2 = 2^{2n} \sum_{P \in \{I, X, Y, Z\}^{\otimes(n+1)}} |\widehat{\rho}_f(P) - \widehat{\rho}_g(P)|^2 = \sum_{S \subseteq [n]} |\widehat{f}(S) - \widehat{g}(S)|^2 = \Pr[f(x) \neq g(x)], \quad (15)$$

where in the first equality we have used Parseval's identity, in the second Eq. (14) and the last equality is elementary. From Eq. (14) follows that learning the Choi state of a QAC⁰ circuit in the 2^n -Frobenius norm is a pretty natural problem.

4.1.2 Learning QAC⁰ circuits

Nadimpalli et al. showed that, for fixed depth and size, the Pauli spectrum of the Choi state of a QAC⁰ circuit is concentrated on low-degree coefficients [NPVY23, Theorem 18]. However, we have noticed that they proved something stronger, namely that these states are close to be juntas.

Theorem 10. *Let ρ be the Choi state of $(n + a + 1)$ QAC⁰ circuit of depth d and size s and let $\varepsilon > 0$. Then, there exists a set $K \subseteq [n + 1]$ with $|K| \leq (\log(2^a s^2 / \varepsilon))^d$ such that*

$$\sum_{\text{supp}(P) \not\subseteq K} |\widehat{\rho}(P)|^2 \leq \frac{\varepsilon}{2^{2n}}.$$

Note that Theorem 4 quickly follows from Theorems 9 and 10 and using that for two n -qubit states ρ and ρ' we have that by Parseval's identity

$$2^{2n} \sum_{P \in \{I, X, Y, Z\}^{\otimes n}} |\widehat{\rho}(P) - \widehat{\rho}'(P)|^2 = 2^n \|\rho - \rho'\|_F^2.$$

As Theorem 10 was not explicitly stated in [NPVY23], we prove it here. To do that, we just have to borrow a few lemmas from [NPVY23] and apply them in a careful way. We note that in that work results are stated for the Choi representation of a channel and in our work we use the Choi state of the channel. Both are easily related, as the Choi state is obtained by dividing the Choi representation by the dimension of the space.

Let U be the unitary implemented by a $(n + a + 1)$ QAC⁰ circuit. Then, it defines a $(n + a + 1)$ -to-1 qubit channel via

$$\Phi(\cdot) = \text{Tr}_{[n+a]}[U \cdot U^\dagger].$$

The first lemma we need states that the Choi state of this $(n + a + 1)$ -to-1 qubit channel does not change much if one removes from the circuit a few Toffoli gates acting on many qubits [NPVY23, Lemma 23].

Lemma 11 ([NPVY23]). *Let Φ be the $(n+a+1)$ -to-1 channel defined by an $(n+a+1)$ -qubit QAC⁰ circuit. Let Φ' be the $(n+a+1)$ -to-1 channel obtained by removing from the circuit m Toffoli gates acting on at least l qubits each. Then, the Choi states satisfy*

$$\sum_{P \in \{I, X, Y, Z\}^{\otimes (n+a+2)}} |\widehat{\rho}_{\Phi}(P) - \widehat{\rho}_{\Phi'}(P)|^2 = O\left(\frac{m^2}{2^l 2^{2(n+a+2)}}\right).$$

Recall that $(n+a+1)$ -qubit QAC⁰ circuit also defines an n -to-1 qubit channel when the auxiliary register is initialized on a fixed $(a+1)$ -qubit state σ , namely

$$\Phi_{\sigma}(\rho \otimes \sigma) = \Phi(\rho \otimes \sigma).$$

The second lemma we need relates the Pauli spectrum of the Choi states of Φ_{σ} and Φ [NPVY23, Proposition 28].

Lemma 12 ([NPVY23]). *Let Φ and Φ_{σ} be the channels as above determined by a QAC⁰ circuit. Then, their Choi states satisfy*

$$\widehat{\rho}_{\Phi_{\sigma}}(P) = 2^{a+1} \sum_{Q \in \{I, X, Y, Z\}^{\otimes n}} \widehat{\rho}_{\Phi}(P \otimes Q) \text{Tr}[Q\sigma^T].$$

Now, we are ready to prove Theorem 10.

Proof: Let Φ be the $(n+a+1)$ -to-1 channel determined by $(n+a+1)$ -qubit QAC⁰ circuit of depth d a size s . Let $l \in \mathbb{N}$ to be fixed later. Let Φ' be the $(n+a+1)$ -to-1 channel obtained by removing from the circuit the Toffoli gates that act on more than l qubits. As there is at most s of them, by Lemma 11 we have that the Choi states satisfy

$$\sum_{P \in \{I, X, Y, Z\}^{\otimes (n+a+2)}} |\widehat{\rho}_{\Phi}(P) - \widehat{\rho}_{\Phi'}(P)|^2 = O\left(\frac{s^2}{2^l 2^{2(n+a+1)}}\right). \quad (16)$$

Now, by a light-cone argument, as the depth of the circuit without the *long* Toffoli gates is at most d , then at the end of the circuit the output qubit only depends on at most l^d other qubits. This implies that the $\rho_{\Phi'}$ is a k -junta state for $k = l^d + 1$. By Eq. (16), if $K \subseteq [n+a+2]$ is the set of k qubits on which $\rho_{\Phi'}$ depends on, then

$$\sum_{P \notin \{I, X, Y, Z\}^K} |\widehat{\rho}_{\Phi}(P)|^2 = O\left(\frac{s^2}{2^l 2^{2(n+a+2)}}\right). \quad (17)$$

Now, if $K' \subseteq [n+1]$ is the subset of non-auxiliary qubits of K , i.e., $K' = K \cap [n+1]$, then

$$\begin{aligned}
\sum_{\text{supp}(P) \subseteq [K']} |\widehat{\rho}_{\Phi'}(P)|^2 &= 2^{2(a+1)} \sum_{\text{supp}(P) \subseteq [K']} \left| \sum_{Q \in \{I, X, Y, Z\}^{\otimes(a+1)}} \widehat{\rho}_{\Phi}(P \otimes Q) \text{Tr}[Q\sigma^T] \right|^2 \\
&\leq 2^{2(a+1)} \sum_{\text{supp}(P) \subseteq [K']} \left(\sum_{Q \in \{I, X, Y, Z\}^{\otimes(a+1)}} |\widehat{\rho}_{\Phi}(P \otimes Q)|^2 \right) \cdot \left(\sum_{Q \in \{I, X, Y, Z\}^{\otimes(a+1)}} |\text{Tr}[Q\sigma^T]|^2 \right) \\
&= 2^{3(a+1)} \|\sigma^T\|_F^2 \sum_{\text{supp}(P) \subseteq [K']} \sum_{Q \in \{I, X, Y, Z\}^{\otimes(a+1)}} |\widehat{\rho}_{\Phi}(P \otimes Q)|^2 \\
&\leq 2^{3(a+1)} \sum_{\text{supp}(P) \not\subseteq [K]} |\widehat{\rho}_{\Phi}(P)|^2 \\
&= 2^{3(a+1)} O\left(\frac{s^2}{2^l 2^{2(n+a+2)}}\right) \\
&= O\left(\frac{s^2 2^{a+1}}{2^l 2^{2(n+1)}}\right),
\end{aligned}$$

where the first line is true by Lemma 12; in the second we apply Cauchy-Schwarz; in the third we use Parseval identity with σ^T ; in the fourth line we use that if $\text{supp}(P) \not\subseteq K'$, then $\text{supp}(P \otimes Q) \not\subseteq K$; and in the fifth line we use Eq. (17). Now, the result follows by taking $l = \log(s^2 2^{a+1} / \varepsilon^2)$. \square

4.1.3 New lower bounds on the computing power of QAC⁰ circuits

Finally, we show how to use Theorem 10 to improve on the lower bounds on the computing power of QAC⁰ circuits. To improve on the lower bounds that one would obtain with [NPVY23, Theorem 18], one should seek for functions of low-degree that are far from being juntas. With that purpose, we consider the address function, that is known to be the Boolean function of degree $D+1$ that depends on more variables [NS94]. To define it, let $\text{add}: \{-1, 1\}^D \rightarrow [2^D]$ be a bijection. The D -address function $f: \{-1, 1\}^D \times \{-1, 1\}^{2^D} \rightarrow \{-1, 1\}$ defined by

$$f(x, y) = \sum_{a \in \{-1, 1\}^D, y \in \{-1, 1\}^{2^D}} \left(\frac{x_1 a_1 + 1}{2}\right) \cdots \left(\frac{x_k a_k + 1}{2}\right) y_{\text{add}(a)} \quad (18)$$

for every $x \in \{-1, 1\}^D$ and $y \in \{-1, 1\}^{2^D}$. Note that f has degree $D+1$, but depends on $2^D + D$ variables. Moreover, we can show that f is far from every Boolean function that depends on less than 2^D variables.

Fact 13. *Let f be the D -address function. Let $k \in [2^D]$. Then, the degree of f is $D+1$ and f is $((2^D - k)/2^{D+1})$ -far from being a k -junta.*

Proof: Let $g: \{-1, 1\}^{D+2^D} \rightarrow \{-1, 1\}$ be a k -junta. The distance between g and f is

$$d(f, g) = \Pr_{x,y}[g(x, y) \neq f(x, y)] = 1 - \Pr_{x,y}[g(x, y) = f(x, y)] = 1 - \Pr_{x,y}[g(x, y) = y_{\text{add}(x)}],$$

where in the last equality we have used that $\left(\frac{x_1 a_1 + 1}{2}\right) \cdots \left(\frac{x_k a_k + 1}{2}\right) = \delta_{a,x}$. Now,

$$\Pr_{x,y}[g(x, y) = y_{\text{add}(x)}] = \frac{1}{2^D} \sum_{x \in \{-1, 1\}^D} \Pr_y[g(x, y) = y_{\text{add}(x)}] \leq \frac{1}{2^D} \left(k + \frac{1}{2}(2^D - k)\right),$$

where in the inequality we have used that g depends on at most k variables of y_1, \dots, y_{2^D} , and that if g does not depend on y_i , then $\Pr_y[g(x, y) = y_i] = 1/2$. Putting everything together follows that $d(f, g) \geq ((2^D - k)/2^{D+1})$. \square

Corollary 14. *In order to compute the D -address function with a depth d , size s QAC^0 circuit with a -auxiliary qubits up to error $1/4$, the parameters need to satisfy*

$$s^2 2^a = \Omega(2^{(2^D)^{1/d}}).$$

Proof: By Fact 13 it follows that the D -address function is $1/8$ -far from every $((3/4)2^D)$ -junta. On the other hand, by Theorem 10 it follows that the Choi-state of the QAC^0 that does not satisfy $\log(s^2 2^a)^d = \Omega(2^D)$ circuit is $1/8$ -close to a $((3/4)2^D)$ -junta. Putting both things together, the claimed result follows. \square

Remark 15. If one used [NPVY23, Theorem 18] instead of Theorem 10 in the proof of Corollary 14, one would obtain a weaker lower bound $s^2 2^a = \Omega(2^{D/d})$.

Acknowledgements. I thank Jop Briët, Alexandros Eskenazis, Jonas Helsen, Yuval Filmus and Francisca Vasconcelos for useful conversations. I thank Hausdorff Research Institute of Mathematics of Bonn, which hosted me during the Dual Trimester Program: “Boolean Analysis in Computer Science” where this paper was written. This research was supported by the European Union’s Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement no. 945045, and by the NWO Gravitation project NETWORKS under grant no. 024.002.003.

References

- [ABR16] Maryam Aliakbarpour, Eric Blais, and Ronitt Rubinfeld. Learning and testing junta distributions. In *Conference on Learning Theory*, pages 19–46. PMLR, 2016.
- [ADOY24] Anurag Anshu, Yangjing Dong, Fengning Ou, and Penghui Yao. On the computational power of qac0 with barely superlinear ancillae. *arXiv preprint arXiv:2410.06499*, 2024.
- [AS07] Alp Atıcı and Rocco A Servedio. Quantum algorithms for learning and testing juntas. *Quantum Information Processing*, 6(5):323–348, 2007.
- [BHL95] A. Blum, L. Hellerstein, and N. Littlestone. Learning in the presence of finitely or infinitely many irrelevant attributes. *Journal of Computer and System Sciences*, 50(1):32–40, 1995. URL: <https://www.sciencedirect.com/science/article/pii/S0022000085710045>, doi:10.1006/jcss.1995.1004.
- [BL97] Avrim L Blum and Pat Langley. Selection of relevant features and examples in machine learning. *Artificial intelligence*, 97(1-2):245–271, 1997.
- [Blu94] Avrim Blum. Relevant examples and relevant features: Thoughts from computational learning theory. In *AAAI Fall Symposium on ‘Relevance*, volume 5, page 1, 1994.

- [BY23] Zongbo Bao and Penghui Yao. On Testing and Learning Quantum Junta Channels. In Gergely Neu and Lorenzo Rosasco, editors, *Proceedings of Thirty Sixth Conference on Learning Theory*, volume 195 of *Proceedings of Machine Learning Research*, pages 1064–1094. PMLR, 12–15 Jul 2023. URL: <https://proceedings.mlr.press/v195/bao23b.html>.
- [CDL⁺24] Xi Chen, Anindya De, Yuhao Li, Shivam Nadimpalli, and Rocco A Servedio. Mildly exponential lower bounds on tolerant testers for monotonicity, unateness, and juntas. In *Proceedings of the 2024 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 4321–4337. SIAM, 2024.
- [CJLW21] Xi Chen, Rajesh Jayaram, Amit Levi, and Erik Waingarten. Learning and testing junta distributions with sub cube conditioning. In *Conference on Learning Theory*, pages 1060–1113. PMLR, 2021.
- [CNY23] Thomas Chen, Shivam Nadimpalli, and Henry Yuen. Testing and Learning Quantum Juntas Nearly Optimally. In *Proceedings of the 2023 Annual ACM-SIAM Symposium on Discrete Algorithms (SODA)*, pages 1163–1185. SIAM, 2023.
- [EFH⁺22] Andreas Elben, Steven T. Flammia, Hsin-Yuan Huang, Richard Kueng, John Preskill, Benoît Vermersch, and Peter Zoller. The randomized measurement toolbox. *Nature Reviews Physics*, 5:9 – 24, 2022. doi:10.1038/s42254-022-00535-2.
- [EIS22] Alexandros Eskenazis, Paata Ivanisvili, and Lauritz Streck. Low-degree learning and the metric entropy of polynomials. *arXiv preprint arXiv:2203.09659*, 2022.
- [FFG⁺03] Maosen Fang, Stephen Fenner, Frederic Green, Steven Homer, and Yong Zhang. Quantum lower bounds for fanout. *arXiv preprint quant-ph/0312208*, 2003.
- [HHJ⁺17] Jeongwan Haah, Aram W Harrow, Zhengfeng Ji, Xiaodi Wu, and Nengkun Yu. Sample-optimal tomography of quantum states. *IEEE Transactions on Information Theory*, 63(9):5628–5641, 2017.
- [HKOT23] Jeongwan Haah, Robin Kothari, Ryan O’Donnell, and Ewin Tang. Query-optimal estimation of unitary channels in diamond distance, 2023. arXiv:2302.14066.
- [HKP20] 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, 2020.
- [KS02] Guy Kindler and Shmuel Safra. Noise-resistant boolean functions are juntas. *preprint*, 5(7):19, 2002.
- [LMN93] Nathan Linial, Yishay Mansour, and Noam Nisan. Constant depth circuits, Fourier transform, and learnability. *Journal of the ACM (JACM)*, 40(3):607–620, 1993.
- [MMB⁺24] Francesco Anna Mele, Antonio Anna Mele, Lennart Bittel, Jens Eisert, Vittorio Giovannetti, Ludovico Lami, Lorenzo Leone, and Salvatore FE Oliviero. Learning quantum states of continuous variable systems. *arXiv preprint arXiv:2405.01431*, 2024.
- [Moo99] Cristopher Moore. Quantum circuits: Fanout, parity, and counting. *arXiv preprint quant-ph/9903046*, 1999.

- [MOS03] Elchanan Mossel, Ryan O’Donnell, and Rocco P Servedio. Learning juntas. In *Proceedings of the thirty-fifth annual ACM symposium on Theory of computing*, pages 206–212, 2003.
- [NP24] Shivam Nadimpalli and Shyamal Patel. Optimal non-adaptive tolerant junta testing via local estimators. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1039–1050, 2024.
- [NPVY23] Shivam Nadimpalli, Natalie Parham, Francisca Vasconcelos, and Henry Yuen. On the Pauli spectrum of QAC^0 . *arXiv preprint arXiv:2311.09631*, 2023.
- [NS94] Noam Nisan and Mario Szegedy. On the degree of boolean functions as real polynomials. *Computational complexity*, 4:301–313, 1994.
- [OW15] Ryan O’Donnell and John Wright. Quantum spectrum testing. In *Proceedings of the 47th annual ACM symposium on Theory of computing (STOC)*, pages 529–538, 2015. doi:10.1145/2746539.2746582.
- [PFGT20] Daniel Padé, Stephen Fenner, Daniel Grier, and Thomas Thierauf. Depth-2 QAC^0 circuits cannot simulate quantum parity. *arXiv preprint arXiv:2005.12169*, 2020.
- [Ros20] Gregory Rosenthal. Bounds on the QAC^0 complexity of approximating parity. *arXiv preprint arXiv:2008.07470*, 2020.
- [Val12] Gregory Valiant. Finding correlations in subquadratic time, with applications to learning parities and juntas. In *2012 IEEE 53rd Annual Symposium on Foundations of Computer Science*, pages 11–20. IEEE, 2012.
- [VH24] Francisca Vasconcelos and Hsin-Yuan Huang. Learning shallow quantum circuits with many-qubit gates. 2024.
- [VZ23] Alexander Volberg and Haonan Zhang. Noncommutative Bohnenblust–Hille inequalities. *Mathematische Annalen*, pages 1–20, 2023. doi:10.1007/s00208-023-02680-0.