Quantum Advantage in Tolerant Junta Testing

Avishay Tal
University of California, Berkeley, CA, USA

Weiqiang Yuan
EPFL, Lausanne, Switzerland

Abstract

We establish the first super-polynomial quantum advantage for the tolerant junta testing problem in the adaptive setting. Specifically, we show that within a certain parameter regime, tolerant k-junta testing with high precision can be solved using $\mathrm{poly}(k)$ quantum queries, whereas any classical algorithm requires at least $k^{\Omega(\log k)}$ queries.

The problem of tolerant k-junta testing is as follows: given parameters $(k, \varepsilon_1, \varepsilon_2)$, with $0 \leq \varepsilon_1 < \varepsilon_2 \leq 1/2$, and black-box access to a Boolean function f (defined on n variables), distinguish whether f is ε_1-close to some k-junta or ε_2-far from every k-junta.

We show the quantum advantage for a range of parameters close to $1/2$, for example, $\varepsilon_1 = 1/2 - 1/k$ and $\varepsilon_2 = 1/2 - 1/(2k^2)$. (As such, the problem is more naturally captured using the notion of correlation with closest k-junta.) The (non-adaptive) quantum tester we use was given by a recent work of Bao, Liu, Yao, Ye, and Zhang (SOSA 2026). We slightly adapt their analysis to show that it holds in the above parameter regime. On the other hand, our classical lower bound requires substantial new ideas. Inspired by the lower bound techniques of Chen and Patel (FOCS 2023), we introduce a new hard distribution of "yes" instances (i.e., instances with distance at most ε_1 to k-juntas) that is based on planting an "approximate-junta" as follows: we randomly pick k out of n coordinates, and for each fixing of the k coordinates, the 2^{n-k} values in the restricted subcube are drawn randomly except for an error-correcting code on which we place the same random bit. We show that this distribution is much closer to k-juntas than the uniform distribution, but on the other hand, they are indistinguishable with respect to any classical algorithm making $k^{o(\log k)}$ queries.

2012 ACM Subject Classification Theory of computation → Oracles and decision trees; Theory of computation → Quantum complexity theory

Keywords and phrases Property testing, Quantum query complexity, Error-correcting code, Boolean function analysis

Digital Object Identifier 10.4230/LIPIcs.CCC.2026.34

Funding *Avishay Tal*: Supported by the National Science Foundation CAREER Award CCF-2145474.
Weiqiang Yuan: Supported by the Swiss State Secretariat for Education, Research and Innovation (SERI) under contract number MB22.00026.

1 Introduction

A Boolean function $f : \{0,1\}^n \to \{\pm 1\}$ is said to be a k-junta if it depends on only k of n input variables. Despite representing a structurally simple class of functions, juntas play a foundational role across various areas of theoretical computer science, including computational learning theory [31, 21, 37], hardness amplification [9, 28], and property testing (see references below), which is the focus of this study.

In the standard junta testing problem, we are given oracle access to a hidden Boolean function $f : \{0,1\}^n \to \{\pm 1\}$. The objective is to determine whether f is a k-junta or ε-far from all k-juntas. To be precise, a *randomized* algorithm $\mathcal{A}^f$ is a (k, ε)-junta tester if it accepts every k-junta f with probability at least $2/3$, while rejecting with probability at least

2/3 every f that differs from every k-junta on at least an ε-fraction of inputs. We measure the tester's efficiency by its query complexity, namely, the number of oracle calls to f (while the running time is not considered). Ideally, we want the tester's query complexity to be as small as possible in terms of k and ε, but independent of n (the input length of f).

In the seminal work of Fischer, Kindler, Ron, Safra, and Samorodnitsky [22], the authors proposed the first junta tester with query complexity depending only polynomially on k and $1/\varepsilon$. Since then, a long line of research [19, 25, 2, 6, 7, 36, 1, 35, 18, 8, 30, 29, 4, 10, 20, 27, 33, 15, 13, 32, 14, 16, 17] has been dedicated to developing more efficient testers and sharper lower bounds, leading to a comprehensive understanding of the problem and its variants. In the standard model, Blais [7] provided a tester using $O(k \log k + k/\varepsilon)$ queries, which was later shown to be optimal for constant ε [35]. On the other hand, [18] shows that any non-adaptive tester must make $\tilde{\Omega}(k^{3/2}/\varepsilon)$ queries, nearly matching the $\tilde{O}(k^{3/2}/\varepsilon)$ upper bound established by Blais [6].

Beyond the classical setting, quantum algorithms have been shown to achieve a quadratic speedup: [1] provides a quantum tester making only $\tilde{O}(\sqrt{k/\varepsilon})$ queries, which is nearly optimal [11]. Lastly, recent work has extended these results to more general models, such as distribution-free testing [30, 10] and relative-error testing [17], demonstrating that junta testing remains easy even under these broader distance measures. However, this efficiency relies heavily on the assumption of perfect, noise-free inputs.

Tolerant Testing

A significant limitation of the standard testing model is that there is no guarantee on the performance of the tester on functions close to satisfying the property. A more realistic scenario is that given a black-box function f, one seeks to test whether there are k features that explain the output *in most cases*, or f is far from it. This is because data in the wild are often noisy and almost never perfect. To address this, Parnas, Ron, and Rubinfeld [34] introduce a natural relaxation of the standard testing model called *tolerant testing*.

In the context of junta testing, a $(k, \varepsilon_1, \varepsilon_2)$-tolerant junta tester must decide whether a Boolean function f is ε_1-close to some k-junta or ε_2-far from every k-junta by making oracle calls to f. Standard junta testing naturally emerges as a special case where $\varepsilon_1 = 0$.

In sharp contrast to the standard model, tolerant junta testing appears to be much harder. In the non-adaptive setting, Chen et al. [14] proved that any classical tolerant junta tester requires $2^{\Omega(\sqrt{k})}$ queries. This lower bound was subsequently shown to be tight by Nadimpalli and Patel [32] (building on [20, 27]), who provided a matching non-adaptive upper bound of $2^{\tilde{O}(\sqrt{k \log(1/(\varepsilon_2 - \varepsilon_1))})}$.

In the classical adaptive setting, despite significant effort, our understanding of tolerant junta testing remains limited. The best-known adaptive $(k, \varepsilon_1, \varepsilon_2)$-tolerant junta tester has query complexity $2^{\tilde{O}(k^{1/3})}$ when $\varepsilon_2 - \varepsilon_1$ is constant [16], while the best-known lower bound is $k^{\Omega(\log(1/(\varepsilon_2 - \varepsilon_1)))}$ when $\varepsilon_1, \varepsilon_2$ are bounded away from 0 and 1/2 [15]. In short, even in the parameter regime where $\varepsilon_1 = 0.01$ and $\varepsilon_2 = 0.49$, neither a tester with query complexity $\exp(o(k^{1/3}))$ nor a superpolynomial lower bound is known.

In the quantum world, however, the landscape changes dramatically. Recent work by Bao, Liu, Yao, Ye, and Zhang [3] introduces an elegant, non-adaptive quantum tester with query complexity $O(k \log k)$, provided that $4\varepsilon_1 < \varepsilon_2$. To evaluate this advantage, Bao et al. [3] complement their algorithm with a classical non-adaptive lower bound, demonstrating a surprising exponential quantum speedup over classical *non-adaptive* tolerant junta testers.

This provable advantage highlights the power of quantum in tolerant junta testing. However, [3]'s results do not establish any super-polynomial quantum speedup over classical *adaptive* tolerant junta testers, since the quantum tester in [3] works only in a parameter regime that is mutually exclusive of that of the superpolynomial lower bound by [15].[1] This leads to the following natural question:

Do adaptive quantum tolerant junta testers significantly outperform their classical counterparts?

1.1 Main Result

In this paper, we give an affirmative answer to the above question by showing a superpolynomial quantum speedup against classical adaptive tolerant junta testers.

▶ **Theorem 1** (Main Result). *For every $k \in \mathbb{N}$ and $\varepsilon > 2^{-k^{.99}}$, there exist $0 < \varepsilon_1 < \varepsilon_2 < 1/2$ such that $\varepsilon = \varepsilon_2 - \varepsilon_1$ and the following holds:*

- *There exists a* non-adaptive quantum *$(k, \varepsilon_1, \varepsilon_2)$-tolerant junta tester with query complexity $O(\mathrm{poly}(k, 1/\varepsilon))$.*
- *Any* adaptive classical *$(k, \varepsilon_1, \varepsilon_2)$-tolerant junta tester must make at least $k^{\Omega(\log(1/\varepsilon))}$ queries.*

The key to the proof of Theorem 1 is a stronger classical adaptive lower bound, inspired by [15], which applies to a wider range of parameters for which the quantum tester in [3] is valid. The quantum advantage is perhaps better captured by the notion of correlation with the closest k-junta, which is equivalent to the distance up to a linear transformation. In this notation, we show that for any $\tau > 2^{-k^{.99}}$,

Quantum Tester (Theorem 23) There exists a non-adaptive quantum algorithm, making $\mathrm{poly}(k/\tau)$ queries, that distinguishes between functions with at least $\tau_1 = \tau$ correlation and at most $\tau_2 = \tau^2/2$ correlation to k-juntas.

Classical Lower Bound (Theorem 16) On the other hand, we show that distinguishing $\tau_1 = \tau$ from $\tau_2 = \tau^2/2$ correlation with k-juntas (and even $\tau_1 = \tau$ versus $\tau_2 = \tau^c$ for any constant $c > 1$) requires $k^{\Omega(\log(1/\tau))}$ classical queries, even adaptively.

By picking $\tau = 1/\mathrm{poly}(k)$, we obtain a quantum algorithm with $\mathrm{poly}(k)$ queries, compared to a quasi-polynomial classical lower bound of $k^{\Omega(\log k)}$ queries.

Prior to our work, superpolynomial lower bound for tolerant junta testing was only known for much closer parameters τ_1, τ_2, i.e., with $\tau_1 = \tau_2 \cdot (1 + o(1))$ [15].

We view this as a quantum advantage to a natural tolerant property testing problem: distinguish whether there exist k features that explain at least τ or at most $\tau^2/2$ fraction of the outputs.

1.2 Proof Overview

1.2.1 Quantum Tester

Our quantum tolerant junta tester is a simple adaptation of the one provided in [3].

[1] More precisely, the quantum tester is efficient if either the gap $\varepsilon_2 - \varepsilon_1$ is constant, or both $\varepsilon_1, \varepsilon_2$ are close to 0, while the only known super-polynomial classical adaptive lower bound [15] requires a sub-constant gap between $\varepsilon_1, \varepsilon_2$ and that both $\varepsilon_1, \varepsilon_2$ are bounded away from 0 and 1/2, which we will discuss in more detail in Section 1.2.2.

▶ **Theorem 2** ([3, Theorem 1.3]). *Let $0 < \varepsilon_1 < \varepsilon_2 < 1/2$ be constants such that $4\varepsilon_1 < \varepsilon_2$. Then, there exists a non-adaptive quantum $(k, \varepsilon_1, \varepsilon_2)$-junta tester with query complexity $O(k \log k)$.*

To demonstrate this, Bao et al. [3] first approximate the minimum distance of a Boolean function f and T-juntas by $\mathrm{Inf}_{\overline{T}}[f]$, where $\mathrm{Inf}_{\overline{T}}(f)$ is expressed in terms of the Fourier expansion of f as $\sum_{S\subseteq[n]:S\cap T\neq\emptyset} \widehat{f}_S^2$. Specifically, they show

$$\mathrm{Inf}_{\overline{T}}[f]/4 \leq \mathrm{dist}(f, \mathcal{J}_T) \leq \mathrm{Inf}_{\overline{T}}[f],$$

where $\mathrm{dist}(f, \mathcal{J}_T)$ denotes the minimum distance of a Boolean function f from $\mathcal{J}_T$, the class of juntas on T. Then, they design an "influence tester" that distinguishes between the following two cases:

- Yes: $\mathrm{Inf}_{\overline{T}}[f] \leq \gamma_1$ for some $T \subseteq [n]$ with size $|T| = k$
- No: $\mathrm{Inf}_{\overline{T}}[f] \geq \gamma_2$ for all $T \subseteq [n]$ with size $|T| = k$

for any $\gamma_1 < \gamma_2$ with query complexity $\mathrm{poly}(k, (\gamma_2 - \gamma_1)^{-1})$.

The quantum tester in Theorem 1 follows directly from combining their influence tester with the following improved characterization of $\mathrm{corr}(f, \mathcal{J}_T)$ in terms of $\inf_{\overline{T}}[f]$:

$$1 - \mathrm{Inf}_{\overline{T}}(f) \leq \mathrm{corr}(f, \mathcal{J}_T) \leq \sqrt{1 - \mathrm{Inf}_{\overline{T}}(f)}.$$

We emphasize that this characterization is not new: the upper bound was shown in [3, Lemma 29] as an intermediate step, while the lower bound was provided in [22].

In what follows, we briefly discuss the main idea of the quantum influence tester in [3]. Its efficiency stems from the fact that a single quantum query suffices to sample $S \subseteq [n]$ from the Fourier distribution of f. With a sufficient number of such samples, we are able to estimate $\mathrm{Inf}_{\overline{T}}[f]$ for any $T \subseteq [n]$ up to some small error. We emphasize here that while the query complexity is $\mathrm{poly}(k)$, the running time of the algorithm is exponential in k – although this is not the focus of this paper.

As our goal is to determine if $\min_{T:|T|=k} \mathrm{Inf}_{\overline{T}}[f]$ is smaller than γ_1 or greater than γ_2, we may want a good approximation of $\mathrm{Inf}_{\overline{T}}[f]$ for all T with $|T| = k$. However, since there are $\binom{n}{k}$ choices of subsets of $[n]$ with size k, a naive union bound over all these subsets will incur a $\log n$ factor to the number of queries, which is intolerable.

De, Mossel, and Neeman [20] provide a solution by showing that the tolerant junta testing problem for functions with an arbitrarily large input length n can be reduced to the case where $n = \mathrm{poly}(k)$. Subsequently, Iyer, Tal, and Whitmeyer [27] further reduce n to $O(k)$, while barely changing the distance to the closest k-junta. Bao et al. [3] use Fourier samples to get an even simpler and more efficient reduction, using only $O(k \log k)$ quantum queries.

1.2.2 Classical Lower Bound

Our classical lower bound is inspired by the work of Chen and Patel [15], where they give the first superpolynomial adaptive lower bound for tolerant junta testing. Specifically, they prove the following.

▶ **Theorem 3** (Theorem 1 in [15]). *Let $0.01 \leq \varepsilon_1 < \varepsilon_2 \leq 0.49$ be two parameters such that $\varepsilon_2 - \varepsilon_1 \geq 2^{-O(k^{.99})}$. Then any $(k, \varepsilon_1, \varepsilon_2)$-tolerant junta tester requires $k^{\Omega(\log(1/(\varepsilon_2-\varepsilon_1)))}$ queries.*

Unfortunately, their result together with the quantum tester from [3] does not directly imply Theorem 1. When $\varepsilon_2 - \varepsilon_1 \to 0$, the quantum tester requires that $\frac{1}{2} - \varepsilon < \varepsilon_1 < \varepsilon_2$; while [15]'s lower bound requires $\varepsilon_1, \varepsilon_2$ to be bounded away from $1/2$. To further clarify why the lower bound in [15] does not apply to our setting, we provide a brief overview of the key ideas of their proof – some of which also play important roles in our proof.

Chen-Patel's Proof

The key of [15]'s proof is the construction of two hard distributions $\mathcal{D}_Y$ and $\mathcal{D}_N$, which are *mostly* supported on yes and no instances, respectively. By Yao's minimax lemma, it suffices to show that any deterministic tester making a polynomial number of queries cannot distinguish between the two distributions. Below, we will only focus on the core of the construction in [15], which corresponds to the parameter regime where

$$n = k + \ell, \qquad \ell < k^{0.99}, \qquad \tau_2 = \tau_0 + 2^{-4\ell}, \qquad \tau_1 = \tau_0 + 2^{-2\ell} - 2^{-4\ell},$$

where $\tau_1 \coloneqq 1 - 2\varepsilon_1, \tau_2 \coloneqq 1 - 2\varepsilon_2$ are the correlations between $\mathcal{D}_Y, \mathcal{D}_N$ and the closest k-junta, and $\tau_0 \coloneqq \mathbb{E}_{\boldsymbol{a} \sim \{\pm 1\}^{2^\ell}}[|\sum_{i=1}^{2^\ell} \boldsymbol{a}_i|/2^\ell] \approx \sqrt{2/\pi} \cdot 2^{-\ell/2}$.

The hard distributions $\mathcal{D}_Y$ and $\mathcal{D}_N$ are specified as follows:

- $\mathcal{D}_N$: The uniform distribution over all n-bit Boolean functions.
- $\mathcal{D}_Y$: Sample $\boldsymbol{S} \sim \binom{[n]}{k}$. Then, for each $y \in \{0,1\}^{\boldsymbol{S}}$, sample the values $\{\boldsymbol{f}(y,z)\}_{z \in \{0,1\}^{\overline{\boldsymbol{S}}}}$ uniformly at random from $\{\pm 1\}$, conditioned on the product of these values being -1:

$$\prod_{z \in \{0,1\}^{\overline{\boldsymbol{S}}}} \boldsymbol{f}(y,z) = -1.$$

A straightforward application of Hoeffding's inequality shows $\mathrm{corr}(\boldsymbol{g}, \mathcal{J}_k)$ is concentrated around τ_0 over $\boldsymbol{g} \sim \mathcal{D}_N$, hence at most τ_2 with high probability. On the other hand, by a coupling argument, [15] shows that

$$\mathbb{E}_{\boldsymbol{a}}[|\sum_{i=1}^{2^\ell} \boldsymbol{a}_i|/2^\ell] \geq \tau_0 + 2^{-2\ell},$$

where $\boldsymbol{a}$ is uniformly sampled from $\{\pm 1\}^{2^\ell}$ conditioned on $\prod_{i=1}^{2^\ell} \boldsymbol{a}_i = -1$. Using the same reasoning, we confirm that $\mathrm{corr}(\boldsymbol{f}, \mathcal{J}_k)$ is at least τ_1 with high probability over $\boldsymbol{f} \sim \mathcal{D}_Y$.[2]

To show the two distributions are indistinguishable, Chen and Patel [15] observe the following: conditioned on any fixed $\boldsymbol{S} = S \in \binom{n}{k}$, an algorithm can distinguish the two distributions only if for some $y \in \{0,1\}^S$, all the points in $\{y,z\}_{z \in \{0,1\}^{\overline{S}}}$ are queried. In particular, after querying $x^{(1)}, \ldots, x^{(m)}$, the algorithm can achieve non-zero advantage only if there exists a pair of indices $1 \leq i < j \leq m$, such that $x^{(i)} = (x^{(j)})^{\oplus \overline{S}}$. Since the probability that $x^{(i)} = (x^{(j)})^{\oplus \overline{S}}$ for random $\boldsymbol{S} \sim \binom{[n]}{k}$ is at most $1/\binom{n}{k}$, by union bound, the algorithm distinguishes two distributions with advantage at most $m^2/\binom{n}{k}$. Therefore, any tester that distinguishes between the two distributions with constant advantage requires $\Omega\left(\sqrt{\binom{n}{k}}\right)$ queries.

[2] In [15], the authors did not manage to optimize the bound for the gap $\tau_1 - \tau_2$ under this construction. Indeed, we conjecture that the right answer here is $\Theta(2^{-3\ell/2})$. However, since we can transform the odd-parity distribution into the uniform distribution by flipping at most one bit, the gap cannot exceed $2^{-\ell}$, thus, does not suffice for our purpose.

In the above construction, $\mathbb{E}_{\boldsymbol{g}\sim\mathcal{D}_N}[\mathrm{corr}(\boldsymbol{g},\mathcal{J}_k)]$ equals $(1-o(1))\cdot\mathbb{E}_{\boldsymbol{f}\sim\mathcal{D}_Y}[\mathrm{corr}(\boldsymbol{f},\mathcal{J}_k)]$, whereas we require the former to be quadratic in the latter. Although [15] extends their construction to a wider class of parameters by padding and planting biased output bits, unfortunately, any straightforward adaptation of their proof remains incompatible with the parameter regime required for our quantum tester.

An Unsuccessful Attempt

Let $\mathcal{D}_N$ remain the uniform distribution. Our objective is to find a "pseudorandom" distribution $\mathcal{D}_Y$ with correlation at least $2^{-\ell/4}$ to k-juntas. To achieve this, we still first sample $\boldsymbol{S}\sim\binom{[n]}{k}$, which is presumed to be the support of the nearest k-junta. Then, for any fixed assignment $y\in\{0,1\}^S$, we aim to ensure that $\mathbb{E}[|\sum_{z\in\{0,1\}^{\overline{S}}}\boldsymbol{f}(y,z)|]\geq 2^{3\ell/4}$. This allows us to apply Hoeffding's inequality to show that $\mathrm{corr}(f,\mathcal{J}_{\boldsymbol{S}})=\Omega(2^{-\ell/4})$ holds with high probability. A natural way to implement this is to first sample a uniform sign $\boldsymbol{b}_y\sim\{\pm1\}$. Then generate $\{\boldsymbol{f}(y,z)\}_{z\in\{0,1\}^{\overline{S}}}$ from an independent biased distribution towards $\boldsymbol{b}_y$, that is, for each $z\in\{0,1\}^{\overline{S}}$, let $\boldsymbol{f}(y,z)=\boldsymbol{b}_y$ with probability $1/2+2^{-\ell/4}$, and $\boldsymbol{f}(y,z)=-\boldsymbol{b}_y$ otherwise.

Unfortunately, this strategy is vulnerable to a simple, efficient distinguisher based on estimating influences. While a function $\boldsymbol{g}\sim\mathcal{D}_N$ satisfies $\mathrm{Inf}_i[\boldsymbol{g}]\geq 1/2-\Theta(2^{-n/2}\cdot\sqrt{\log n})$ for all $i\in[n]$ with high probability, a function $\boldsymbol{f}\sim\mathcal{D}_Y$ exhibits significantly lower influence: for any $i\in\overline{\boldsymbol{S}}$, $\mathrm{Inf}_i[\boldsymbol{f}]\leq 1/2-\Theta(2^{-\ell/2})$ with high probability. Consequently, an algorithm can distinguish the two distributions by estimating $\mathrm{Inf}_i[\boldsymbol{f}]$ for each $i\in[n]$ with precision $O(2^{-\ell/2})$ using $\Theta(2^\ell)=\Theta(1/\varepsilon)$ samples, and then simply checking if the minimum of the estimations over $i\in[n]$ is below a threshold of $1/2-\Theta(2^{-\ell/2})$.

Our Strategy

The fundamental flaw in the previous construction is that it introduces local correlations between the values of $\boldsymbol{f}$ at adjacent points. More generally, $\mathcal{D}_Y$ remains distinguishable from $\mathcal{D}_N$ if the marginal distribution of $\boldsymbol{f}$ on any Hamming ball of constant radius deviates from being uniform. To circumvent this, we propose a construction based on error-correcting codes. By planting the bias only on the codewords of a high-distance code, we ensure that correlations only emerge between inputs with large Hamming distance; hence, $\boldsymbol{f}$ is locally uniform.

Formally, our "yes" distribution $\mathcal{D}_Y$ is generated as follows:

1. Sample a support $\boldsymbol{S}\sim\binom{[n]}{k}$.
2. Let $C\subseteq\{0,1\}^{\overline{S}}$ be an arbitrary error-correcting code with rate $R=3/4$ and relative distance $\delta=0.01$, as guaranteed by the Gilbert-Varshamov bound.
3. For each assignment $y\in\{0,1\}^S$, sample a uniform sign $\boldsymbol{b}_y\sim\{\pm1\}$.
4. For each $z\in\{0,1\}^{\overline{S}}$, let

$$\boldsymbol{f}(y,z)\begin{cases}:=\boldsymbol{b}_y & \text{if } z\in C,\\ \sim\{\pm1\} & \text{if } z\notin C.\end{cases}$$

Since $\mathbb{E}_{\boldsymbol{f}}[|\sum_{z\in\{0,1\}^{\overline{S}}}\boldsymbol{f}(y,z)|]\geq|C|-2^{\ell/2}=\Omega(2^{3\ell/4})$, a standard application of Hoeffding's inequality confirms that $\mathrm{corr}(\boldsymbol{f},\mathcal{J}_k)\geq\Omega(2^{-\ell/4})$ happens with high probability.

Indistinguishability

Next, we present the main idea of showing the indistinguishability of $\mathcal{D}_Y$ and $\mathcal{D}_N$ generated as described above. Let $\mathcal{A}^f$ be an arbitrary algorithm making m queries to f, and consider any path $\mathcal{P}$ in its decision tree representation. Let $x^{(1)}, x^{(2)}, \ldots, x^{(m)}$ denote the set of points queried along $\mathcal{P}$. Under the uniform distribution $\boldsymbol{g} \sim \mathcal{D}_N$, $\mathcal{A}^{\boldsymbol{g}}$ follows $\mathcal{P}$ with probability exactly 2^{-m}. We show that under $\boldsymbol{f} \sim \mathcal{D}_Y$, this probability remains $(1 \pm o(1))2^{-m}$ as long as $m = k^{O(\ell)}$. The indistinguishability of $\mathcal{D}_Y$ and $\mathcal{D}_N$ follows from a standard coupling argument. Below, we briefly discuss how to show $\mathcal{A}^{\boldsymbol{f}}$ follows $\mathcal{P}$ with probability close to 2^{-m} over $\boldsymbol{f} \sim \mathcal{D}_Y$.

We say $S \in \binom{[n]}{k}$ is good if for every distinct pair of query points $x^{(i)}, x^{(j)}$, at least one of the following holds:

- $x^{(i)}$ and $x^{(j)}$ are inconsistent on at least one of the coordinates in S,
- $x^{(i)}$ and $x^{(j)}$ has Hamming distance less than $\ell/100$.

We claim that conditioned on $\boldsymbol{S} = S$ for a good S, the queried values $\boldsymbol{f}(x^{(1)}), \ldots, \boldsymbol{f}(x^{(m)})$ are independent and uniform. To see this, fix any $y \in \{0,1\}^S$. The "good" property ensures that the subcube $\{x \in \{0,1\}^n \mid x_S = y\}$ contains at most one queried point $x^{(i)}$ whose projection on $\overline{S}$ is a codeword in C, because if there were two such points, they must agree on S while having a Hamming distance of at least $\ell/100$, which contradicts the assumption that S is good. Since each query hits at most one codeword per subcube, and the signs $(\boldsymbol{b}_y)_{y \in \{0,1\}^S}$ associated with the codewords are independent and uniform, we deduce that $\boldsymbol{f}(x^{(1)}), \ldots, \boldsymbol{f}(x^{(m)})$ are independent and uniform. Therefore, conditioned on $\boldsymbol{S}$ being good, $\mathcal{A}^{\boldsymbol{f}}$ traverses $\mathcal{P}$ with probability exactly 2^{-m}.

It remains to show that a random $\boldsymbol{S} \sim \binom{[n]}{k}$ is good with high probability. For any two queried points $x^{(i)}, x^{(j)}$ with Hamming distance at least $\ell/100$, the probability that they coincide on coordinates in $\boldsymbol{S}$ is at most $\binom{n-\ell/100}{k} / \binom{n}{k} = k^{-\Omega(\ell)}$. By applying a union bound over all m^2 pairs of queries, we conclude that $\boldsymbol{S}$ is good with probability at least $1 - m^2 \cdot k^{-\Omega(\ell)}$.

1.3 Organization

The remainder of this paper is structured as follows: In Section 2, we provide the necessary notation and recall the standard tools in Boolean function analysis, coding theory, and concentration inequalities. In Section 3, we establish our new classical lower bound. Finally, in Section 4, we describe our quantum tester and analyze its query complexity and correctness.

2 Preliminaries

Throughout the paper, we use log to denote the base-2 logarithm, and ln to denote the natural logarithm. For any distribution $\mathcal{D}$ over a finite space Ω, we use $\mathcal{D}(x)$ to denote $\Pr_{\boldsymbol{x} \sim \mathcal{D}}[\boldsymbol{x} = x]$ for every $x \in \Omega$. Given two distributions μ, ν over the same space, we use $d_{\mathrm{TV}}(\mu, \nu)$ to denote the *total variation distance* between μ and ν. Given $S \subseteq [n]$, we use $\overline{S} := [n] \setminus S$ to denote the complement of S when n is clear from the context. Given $x \in \{0,1\}^n$, we use $x^{\oplus S} := x \oplus 1^S 0^{\overline{S}}$ to denote the string obtained by flipping all the coordinates in S of x.

Boolean Functions

Every n-bit Boolean function $f : \{0,1\}^n \to \{\pm 1\}$ can be uniquely written as a polynomial in the Fourier basis: $f(x) = \sum_{S \subseteq [n]} \widehat{f}_S \chi_S(x)$, where $\chi_S(x) := (-1)^{\sum_{i \in S} x_i}$, and $\{\widehat{f}_S\}_{S \subseteq [n]}$ are the Fourier coefficients of f.

Parseval's identity states that the sum of the squares of the Fourier coefficients of any Boolean function is exactly 1, that is, $\sum_{S\subseteq[n]} \widehat{f}_S^2 = 1$. Plancherel's theorem generalizes Parseval's identity. It states that for any two functions $f, g : \{0,1\}^n \to \mathbb{R}$, $\mathbb{E}_{\boldsymbol{x}}[f(\boldsymbol{x})g(\boldsymbol{x})] = \sum_{S\subseteq[n]} \widehat{f}_S\widehat{g}_S$.

We recall the standard notion of influence for Boolean functions.

▶ **Definition 4** (Influence). *Let $f : \{0,1\}^n \to \{\pm 1\}$ be a Boolean function and $i \in [n]$. The influence of i on f is defined as*

$$\mathrm{Inf}_i[f] \coloneqq \Pr_{\boldsymbol{x}\sim\{0,1\}^n}[f(\boldsymbol{x}) \neq f(\boldsymbol{x}^{\oplus\{i\}})] = \sum_{S\ni i} \widehat{f}_S^2.$$

In this paper, we also work with the following two variants of influence.

▶ **Definition 5** (Level-k Influence). *Let $f : \{0,1\}^n \to \{\pm 1\}$ be a Boolean function and $k, i \in [n]$. The level-k influence of i on f is defined as*

$$\mathrm{Inf}_i^{\leq k}[f] \coloneqq \sum_{S\ni i : |S|\leq k} \widehat{f}_S^2.$$

▶ **Definition 6** (Generalized Influence). *Let $f : \{0,1\}^n \to \{\pm 1\}$ be a Boolean function and $T \subseteq [n]$ be an arbitrary set. The influence of T on f is defined as*

$$\mathrm{Inf}_T(f) \coloneqq \sum_{S : S\cap T\neq\varnothing} \widehat{f}_S^2 = 1 - \sum_{S : S\subseteq\overline{T}} \widehat{f}_S^2.$$

Juntas

For any set $S \subseteq [n]$, we say a Boolean function $f : \{0,1\}^n \to \{\pm 1\}$ is an S-junta if it only depends on variables in S. We denote the set of S-juntas on n variables by $\mathcal{J}_S^n$. Moreover, for any integer $k \in [n]$, we say f is a k-junta if it depends on at most k variables. We denote the set of k-juntas on n variables by $\mathcal{J}_k^n$. Note that $\mathcal{J}_k^n = \bigcup_{S\subseteq[n]:|S|=k} \mathcal{J}_S^n$. We will write $\mathcal{J}_S^n$ and $\mathcal{J}_k^n$ simply as $\mathcal{J}_S$ and $\mathcal{J}_k$ when n is clear from the context.

For any two Boolean functions $f, g : \{0,1\}^n \to \{\pm 1\}$, we use $\mathrm{dist}(f,g) \coloneqq \Pr_{\boldsymbol{x}}[f(\boldsymbol{x}) \neq g(\boldsymbol{x})]$ to denote the distance between f and g. We also define $\mathrm{corr}(f,g) \coloneqq \mathbb{E}_{\boldsymbol{x}}[f(\boldsymbol{x})g(\boldsymbol{x})] = 1 - 2\mathrm{dist}(f,g)$ to denote the correlation between f and g. Let $\mathcal{F}$ be a set of n-bit Boolean functions. We use $\mathrm{dist}(f,\mathcal{F}) \coloneqq \min_{g\in\mathcal{F}} \mathrm{dist}(f,g)$ to denote the distance between f and the closest function in $\mathcal{F}$. Similarly, we use $\mathrm{corr}(f,\mathcal{F}) \coloneqq \max_{g\in\mathcal{F}} \mathrm{corr}(f,g) = 1 - 2\mathrm{dist}(f,\mathcal{F})$ to denote the maximum correlation between f and any function in $\mathcal{F}$.

Given $T \subseteq [n]$, the correlation between f and the closest T-junta can be calculated as follows.

▷ Claim 7 ([27]). Let $f : \{0,1\}^n \to \{\pm 1\}$ be a Boolean function and $T \subseteq [n]$. Define $f_{\mathrm{avg},T} : \{0,1\}^n \to [-1,1]$ where

$$f_{\mathrm{avg},T}(x) \coloneqq \mathbb{E}_{\boldsymbol{y}\sim\{0,1\}^n}[f(\boldsymbol{y}) \mid \boldsymbol{y}_T = \boldsymbol{x}_T].$$

It holds that

$$\mathrm{corr}(f, \mathcal{J}_T) = \mathbb{E}_{\boldsymbol{x}\sim\{0,1\}^n}[|f_{\mathrm{avg},T}(\boldsymbol{x})|].$$

The following lemma states that $\mathrm{corr}(f, \mathcal{J}_k)$ can be approximated by the maximum correlation between f and k-juntas supported only on coordinates with high level-k influence.

▶ **Lemma 8** ([27]). *Let $f : \{0,1\}^n \to \{\pm 1\}$ be any Boolean function and $\tau > 0$ be any parameter. Let $C \coloneqq \{i \in [n] \mid \mathrm{Inf}_i^{\leq k}[f] \geq \tau^2/k\}$. Then $|C| \leq k^2/\tau^2$. Moreover,*

$$\max_{S \subseteq C : |S| \leq k} \mathrm{corr}(f, \mathcal{J}_S) \geq \mathrm{corr}(f, \mathcal{J}_k) - \tau.$$

Quantum Query Algorithms

Let $f : \{0,1\}^n \to \{\pm 1\}$ be a Boolean function with Fourier representation $f(x) = \sum_{S \subseteq [n]} \widehat{f}_S \chi_S(x)$. By Parseval's theorem, one can define a natural distribution $\mathcal{S}_f$ over $2^{[n]}$ where $\mathcal{S}_f(S) = \widehat{f}_S^2$. The following folklore states that $\mathcal{S}_f$ can be sampled by an algorithm that makes a single quantum query to f.

▷ Claim 9 ([5]). There is an algorithm $\textsc{FourierSample}^f$ that makes one quantum query to $f : \{0,1\}^n \to \{\pm 1\}$, and for every $S \subseteq [n]$, it outputs S with probability $\widehat{f}_S^2$.

Error-correcting Codes

Given two n-bit strings $x, y \in \{0,1\}^n$, let $\Delta(x,y) \coloneqq \{i \mid x_i \neq y_i\}$ denote the Hamming distance between x and y. Moreover, the relative distance between x and y is defined as $\delta(x,y) \coloneqq \Delta(x,y)/n$.

▶ **Definition 10.** *A binary code of block length n is a subset $C \subseteq \{0,1\}^n$. The rate of C is defined as $\log|C|/n$, and the relative distance of C is defined as $\min_{c_1 \neq c_2 \in C} \delta(c_1, c_2)$. We say C is an (n, R, δ)-code if it has block length n, rate R, and relative distance δ.*

The following textbook result states that as long as $R + H_2(\delta) \leq 1$, there exists an (n, R, δ)-code, where $H_2(x) \coloneqq x\log(1/x) + (1-x)\log(1/(1-x))$ is the binary entropy function, extended continuously to the boundaries so that $H_2(0) = H_2(1) = 0$.

▶ **Theorem 11** (Gilbert–Varshamov bound [23, 38]). *Let $\delta \in (0, 1/2)$ and $R \in (0,1)$ be parameters such that $R + H_2(\delta) \leq 1$. Then there exists an (n, R, δ)-code for every $n \in \mathbb{N}$.*

We also use $H_2^{-1} : [0,1] \to [0,1/2]$ to denote the inverse binary entropy function, where given $x \in [0,1]$, $H_2^{-1}(x)$ is defined as the unique $y \in [0,1/2]$ such that $H_2(y) = x$.

▶ **Fact 12** ([12, Theorem 2.2]). *For every $x \in [0,1]$, it holds that $H_2^{-1}(x) \geq \frac{x}{2\log(6/x)}$.*

Concentration Bounds

We recall the following standard concentration inequalities.

▶ **Lemma 13** (Hoeffding's inequality [26]). *Let $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \in [0,1]$ be independent random variables and $\boldsymbol{S} = \sum_{i=1}^n \boldsymbol{X}_i$. Then for every $t > 0$,*

$$\Pr[|\boldsymbol{S} - \mathbb{E}[\boldsymbol{S}]| \geq t] \leq 2\exp(-2t^2/n).$$

▶ **Fact 14** ([24]). *Let n be even and $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n \sim \{\pm 1\}$ be n independent Rademacher variables. Then $\mathbb{E}_{\boldsymbol{a}}[|\sum_{i=1}^n \boldsymbol{a}_i|] \leq \sqrt{2n/\pi}$.*

We will also use the following simple inequality.

▶ **Fact 15.** *Let $0 \leq k, m \leq n$ be integers such that $m + k \leq n$. Then*

$$\binom{n-m}{k} \Big/ \binom{n}{k} \leq \left(\frac{n-k}{n}\right)^m.$$

Proof. Let $\ell := n - k$. Then,

$$\binom{n-m}{k}/\binom{n}{k} = \binom{n-m}{\ell-m}/\binom{n}{\ell} = \prod_{i=0}^{m-1}\frac{\ell-i}{n-i} \le (\ell/n)^m. \qquad \blacktriangleleft$$

3 Classical Lower Bound

In this section, we demonstrate our new classical lower bound for tolerant junta testing.

▶ **Theorem 16.** *Let $k \in \mathbb{N}$, $\tau_2 \in (2^{-k^{0.99}}, 1/2)$, and $\tau_2 < \tau_1 = \tau_2^c$ and for some $c \in (\frac{\log\log(1/\tau_2)}{\log(1/\tau_2)}, 1)$. Then any adaptive classical algorithm which*

- *accepts f with probability at least $2/3$ if f has correlation at least τ_1 to some k-junta;*
- *rejects f with probability at least $2/3$ if f has correlation at most τ_2 to every k-junta*

must make $k^{\Omega(\frac{\log(1/\tau_1)}{\log(2/c)})}$ queries to f.

To prove the theorem, we follow the standard approach for establishing testing lower bound: Using Yao's minimax lemma, it suffices to find two hard distributions $\mathcal{D}_Y$ and $\mathcal{D}_N$ – supported mostly on yes instances and no instances respectively – and prove that they are indistinguishable by any classical algorithm making a polynomial number of queries.

The rest of the section is organized as follows: In Section 3.1, we define the hard distributions $\mathcal{D}_Y$ and $\mathcal{D}_N$ and verify that they satisfy the required correlation properties. Then in Section 3.2, we show the indistinguishability of these two distributions. Finally, in Section 3.3 we put everything together to show Theorem 16.

3.1 Hard Distributions

Given any $k \in \mathbb{N}$, $\tau_2 \in (2^{-k^{0.99}}, 1/2)$, and $\tau_2 < \tau_1 = \tau_2^c$ and for some $c \in (\frac{\log\log(1/\tau_2)}{\log(1/\tau_2)}, 1)$, let $\ell := 2\log(1/\tau_2)$. We will focus only on the special case where $n = k + \ell$. The proof for arbitrary large n follows from a simple padding argument, namely, adding dummy input bits to f.

The two hard distributions $\mathcal{D}_N$ and $\mathcal{D}_Y$ are described as follows.

▶ **Definition 17.** *Let $R := 1 - c/2 + 2/\ell$. For every $S \in \binom{[n]}{k}$, let $C_{\overline{S}} \subseteq \{0,1\}^{\overline{S}}$ be a fixed arbitrary (ℓ, R, δ)-code with $\delta := H_2^{-1}(1-R)$. The distributions $\mathcal{D}_Y$ and $\mathcal{D}_N$ are defined as follows:*

- *$\mathcal{D}_N$: The uniform distribution over all n-bit Boolean functions.*
- *$\mathcal{D}_Y$: Draw $\boldsymbol{S} \sim \binom{[n]}{k}$. For each $y \in \{0,1\}^{\boldsymbol{S}}$, draw $\boldsymbol{b}_y \sim \{\pm 1\}$. Then for each $z \in \{0,1\}^{\overline{\boldsymbol{S}}}$, let*

$$\boldsymbol{f}(y,z) \begin{cases} := \boldsymbol{b}_y & z \in C_{\overline{\boldsymbol{S}}} \\ \sim \{\pm 1\} & z \notin C_{\overline{\boldsymbol{S}}} \end{cases}.$$

We first show $\mathcal{D}_N$ is mostly supported on functions with correlation at most τ_2 with every k-junta.

▷ Claim 18. Let $\boldsymbol{f} \sim \mathcal{D}_N$. With probability at least $1 - \exp(-2^{k/3})$, $\mathrm{corr}(\boldsymbol{f}, \mathcal{J}_k) \le 2^{-\ell/2} = \tau_2$.

Proof. Fix any $S \in \binom{[n]}{k}$. For any assignment $y \in \{0,1\}^S$ over coordinates in S, define $\boldsymbol{a}_y := 2^{-\ell} \cdot |\sum_{z \in \{0,1\}^{\overline{S}}} \boldsymbol{f}(y,z)|$. By Claim 7, it holds that $\mathrm{corr}(\boldsymbol{f}, \mathcal{J}_S) = 2^{-k} \cdot \sum_{y \in \{0,1\}^S} \boldsymbol{a}_y$.

We first bound the expected value of each individual $\boldsymbol{a}_y$: Since $\{\boldsymbol{f}(y,z)\}_{z\in\{0,1\}^{\overline{S}}}$ are independent Rademacher variables, by Fact 14, we have $\mathbb{E}[\boldsymbol{a}_y] \le \sqrt{2/\pi}\cdot 2^{-\ell/2}$. Then using Hoeffding's inequality, we have $\mathrm{corr}(\boldsymbol{f},\mathcal{J}_S) = 2^{-k}\cdot\sum_{y\in\{0,1\}^S}\boldsymbol{a}_y \ge 2^{-\ell/2} = \tau_2$ with probability $1-\exp(-2^{k/2})$. Finally, by applying union bound over all $S\in\binom{[n]}{k}$, we obtain

$$\Pr[\mathrm{corr}(\boldsymbol{f},\mathcal{J}_k)\ge\tau_2] \le \sum_{S:|S|=k}\Pr[\mathrm{corr}(\boldsymbol{f},\mathcal{J}_S)\ge\tau_2] \le \binom{n}{k}\cdot\exp(-2^{k/2}) \le \exp(-2^{k/3}). \quad\triangleleft$$

Then we demonstrate that $\mathcal{D}_Y$ is mostly supported on functions with correlation at least τ_1 with some k-juntas.

▷ Claim 19. Let $\boldsymbol{f}\sim\mathcal{D}_Y$. With probability at least $1-\exp(-2^{k/2})$, $\mathrm{corr}(\boldsymbol{f},\mathcal{J}_k)\ge\tau_1$.

Proof. Fix any $S\in\binom{[n]}{k}$. Sample a uniform $\boldsymbol{f}\sim\mathcal{D}_Y$ conditioned on $\boldsymbol{S}=S$. We similarly define $\boldsymbol{a}_y \coloneqq 2^{-\ell}\cdot|\sum_{z\in\{0,1\}^{\overline{S}}}\boldsymbol{f}(y,z)|$ for each assignment $y\in\{0,1\}^S$. By Claim 7, $\mathrm{corr}(\boldsymbol{f},\mathcal{J}_S) = 2^{-k}\cdot\sum_{y\in\{0,1\}^S}\boldsymbol{a}_y$. However, for this time, we have

$$\boldsymbol{a}_y = 2^{-\ell}\cdot\left|C_{\overline{S}}|\cdot\boldsymbol{b}_y + \sum_{z\notin C_{\overline{S}}}\boldsymbol{f}(y,z)\right|.$$

Using the fact that $\{\boldsymbol{f}(y,z)\}_{z\notin C_{\overline{S}}}$ are independent, an application of Fact 14 yields

$$\mathbb{E}[\boldsymbol{a}_y] \ge |C_{\overline{S}}|/2^{\ell} - 2^{-\ell/2} \ge 2^{-(1-R)\cdot\ell-1}.$$

Consequently, by Hoeffding's inequality, with probability at least $1-\exp(-2^{k/2})$, we have

$$\mathrm{corr}(\boldsymbol{f},\mathcal{J}_S) = 2^{-k}\cdot\sum_{y\in\{0,1\}^S}\boldsymbol{a}_y \ge 2^{-(1-R)\cdot\ell-2} = 2^{-c\cdot\ell/2} = \tau_2^c = \tau_1,$$

where the second equality follows from the definition of $R\coloneqq 1-c/2+2/\ell$. The desired claim follows by averaging over $\boldsymbol{S}\sim\binom{[n]}{k}$. $\triangleleft$

3.2 Indistinguishability

Next, we show that $\mathcal{D}_Y$ and $\mathcal{D}_N$ are indistinguishable by classical algorithms making a polynomial number of queries.

▶ **Lemma 20.** *Let $\mathcal{D}_Y$ and $\mathcal{D}_N$ be defined as in Definition 17. Then for any classical algorithm $\mathcal{A}^f$ that makes m queries to f,*

$$\left|\Pr_{\boldsymbol{f}\sim\mathcal{D}_Y}[\mathcal{A}^{\boldsymbol{f}}=1] - \Pr_{\boldsymbol{g}\sim\mathcal{D}_N}[\mathcal{A}^{\boldsymbol{g}}=1]\right| \le m^2\cdot k^{-\Omega(\delta\cdot\ell)}. \tag{1}$$

Proof of Lemma 20. Consider $\mathcal{A}^f$ as a depth-m decision tree. Without loss of generality, we can assume that every root-to-leaf path in the tree has length exactly m. Let $\mathcal{P}$ denote an arbitrary path. It is straightforward to observe that $\mathcal{A}^{\boldsymbol{f}}$ follows path $\mathcal{P}$ with probability exactly 2^{-m} when $\boldsymbol{f}$ is drawn from the uniform distribution $\mathcal{D}_N$. Our primary task is to show that when $\boldsymbol{f}\sim\mathcal{D}_Y$, $\mathcal{A}^{\boldsymbol{f}}$ follows $\mathcal{P}$ with probability close to 2^{-m}.

Let $x^{(1)},\dots,x^{(m)}$ denote the points queried on $\mathcal{P}$. We say $S\in\binom{[n]}{k}$ is *good* (w.r.t. $\mathcal{P}$) if for every $1\le i<j\le n$, either $x_S^{(i)}\ne x_S^{(j)}$, or $\Delta(x^{(i)},x^{(j)})<\delta\cdot\ell$. We first show that when S is good, the probability that $A^{\boldsymbol{f}}$ follows $\mathcal{P}$ is exactly 2^{-m} over $\boldsymbol{f}\sim\mathcal{D}_Y$ conditioned on $\boldsymbol{S}=S$.

▷ Claim 21. Let $r^{(1)}, \ldots, r^{(m)} \in \{\pm 1\}$ be m arbitrary signs. Suppose S is good, then

$$\Pr_{\boldsymbol{f} \sim \mathcal{D}_Y | \boldsymbol{S}=S}[\boldsymbol{f}(x^{(i)}) = r^{(i)}, \forall i \in [m]] = 2^{-m}.$$

In particular, $\mathcal{A}^{\boldsymbol{f}}$ follows $\mathcal{P}$ with probability exactly 2^{-m} over $\boldsymbol{f} \sim \mathcal{D}_Y$ conditioned on $\boldsymbol{S} = S$.

Proof. For every $i \in [m]$, we can write $x^{(i)} = (y^{(i)}, z^{(i)})$ where $y^{(i)} \in \{0,1\}^S, z^{(i)} \in \{0,1\}^{\overline{S}}$. Since S is good, for any pair of distinct indices i, j such that $y^{(i)} = y^{(j)}$, we must have $\Delta(z^{(i)}, z^{(j)}) < \delta \cdot \ell$. This implies either $z^{(i)} \notin C_{\overline{S}}$, or $z^{(j)} \notin C_{\overline{S}}$. Moreover, since the signs $(\boldsymbol{b}_y)_{y \in \{0,1\}^S}$ associated with the codewords are independent and uniform, we deduce that $\boldsymbol{f}(y^{(1)}, z^{(1)}), \ldots, \boldsymbol{f}(y^{(m)}, z^{(m)})$ are independent uniform bits, from which the desired statement follows. ◁

Next, we show $\boldsymbol{S}$ is good with high probability over $\boldsymbol{S} \sim \binom{[n]}{k}$.

▷ Claim 22. With probability $1 - m^2 \cdot n^{-\Omega(\delta \cdot \ell)}$ over $\boldsymbol{S} \sim \binom{[n]}{k}$, $\boldsymbol{S}$ is good.

Proof. For any fixed $1 \le i < j \le m$, let $T \coloneqq \{r \in [n] \mid x_r^{(i)} \neq x_r^{(j)}\}$. Observe that whenever $|T| \ge \delta \cdot \ell$,

$$\Pr_{\boldsymbol{S} \sim \binom{n}{k}}[\boldsymbol{S} \cap T = \emptyset] \le \binom{n - \delta \cdot \ell}{k} \Big/ \binom{n}{k} \le (\ell/n)^{\delta \cdot \ell} \le k^{0.01 \delta \cdot \ell},$$

where the second inequality follows from Fact 15, and the last inequality follows from the assumption that $\ell \le k^{0.99}$. Therefore,

$$\Pr_{\boldsymbol{S} \sim \binom{n}{k}}[x_{\boldsymbol{S}}^{(i)} = x_{\boldsymbol{S}}^{(j)} \wedge \Delta(x^{(i)}, x^{(j)}) \ge \delta \cdot \ell] \le n^{-\Omega(\delta \cdot \ell)}.$$

Finally, by union bound over all pairs (i, j),

$$\Pr_{\boldsymbol{S}}[\boldsymbol{S} \text{ is not good}] \le \sum_{1 \le i < j \le m} \Pr_{\boldsymbol{S}}[x_{\boldsymbol{S}}^{(i)} = x_{\boldsymbol{S}}^{(j)} \wedge \Delta(x^{(i)}, x^{(j)}) \ge \delta \cdot \ell] \le m^2 \cdot k^{-\Omega(\delta \cdot \ell)}.$$ ◁

With the above two claims in hand, we are ready to show (1). Let μ_Y, μ_N denote the distributions of the path which $\mathcal{A}^{\boldsymbol{f}}$ follows for $\boldsymbol{f} \sim \mathcal{D}_Y$ and $\boldsymbol{f} \sim \mathcal{D}_N$ respectively. Claims 21 and 22 together imply that for every path $\mathcal{P}$, $\mu_Y(\mathcal{P}) \ge (1 - m^2 \cdot k^{-\Omega(\delta \cdot \ell)}) \cdot 2^{-m}$. Hence, the total variation distance between μ_Y and μ_N is

$$d_{\mathrm{TV}}(\mu_Y, \mu_N) = \sum_{\mathcal{P}} \max(0, \mu_N(\mathcal{P}) - \mu_Y(\mathcal{P})) \le \sum_{\mathcal{P}} \mu_N(\mathcal{P}) \cdot (m^2 \cdot k^{-\Omega(\delta \cdot \ell)}) = m^2 \cdot k^{-\Omega(\delta \cdot \ell)}.$$

It follows that

$$\left| \Pr_{\boldsymbol{f} \sim \mathcal{D}_Y}[\mathcal{A}^{\boldsymbol{f}} = 1] - \Pr_{\boldsymbol{g} \sim \mathcal{D}_N}[\mathcal{A}^{\boldsymbol{g}} = 1] \right| \le d_{\mathrm{TV}}(\mu_Y, \mu_N) \le m^2 \cdot k^{-\Omega(\delta \cdot \ell)}.$$ ◀

3.3 Putting Everything Together

We are now ready to prove Theorem 16.

Proof of Theorem 16. Let $\mathcal{D}_Y$ and $\mathcal{D}_N$ be defined as in Definition 17. Let $\mathcal{D}'_Y$ denote the distribution of $\boldsymbol{f} \sim \mathcal{D}_Y$ conditioned on $\mathrm{corr}(\boldsymbol{f}, \mathcal{J}_k) \geq \tau_1$, and $\mathcal{D}'_N$ denote the distribution of $\boldsymbol{g} \sim \mathcal{D}'_N$ conditioned on $\mathrm{corr}(\boldsymbol{g}, \mathcal{J}_k) \leq \tau_2$. Note that $\mathcal{D}'_Y$ is fully supported on yes instances, and $\mathcal{D}'_N$ is fully supported on no instances. Moreover, Claim 19 implies that $d_{\mathrm{TV}}(\mathcal{D}_Y, \mathcal{D}'_Y) \leq \exp(-2^{k/2})$, and similarly, Claim 18 implies that $d_{\mathrm{TV}}(\mathcal{D}_N, \mathcal{D}'_N) \leq \exp(-2^{k/3})$. Combing with Lemma 20, for any m-query algorithm $\mathcal{A}^f$, it holds that

$$\begin{aligned}
&\left| \Pr_{\boldsymbol{f}\sim\mathcal{D}'_Y}[\mathcal{A}^{\boldsymbol{f}} = 1] - \Pr_{\boldsymbol{g}\sim\mathcal{D}'_N}[\mathcal{A}^{\boldsymbol{g}} = 1] \right| \\
\leq & d_{\mathrm{TV}}(\mathcal{D}'_Y, \mathcal{D}_Y) + \left| \Pr_{\boldsymbol{f}\sim\mathcal{D}_Y}[\mathcal{A}^{\boldsymbol{f}} = 1] - \Pr_{\boldsymbol{g}\sim\mathcal{D}_N}[\mathcal{A}^{\boldsymbol{g}} = 1] \right| + d_{\mathrm{TV}}(\mathcal{D}_N, \mathcal{D}'_N) \\
\leq & 2m^2 \cdot k^{-\Omega(\delta\cdot\ell)}.
\end{aligned}$$

Since

$$\delta \cdot \ell = H_2^{-1}(c/2 - 2/\ell) \cdot \ell \geq \Omega(\frac{c}{\log(2/c)} \log \tau_2) = \Omega(\frac{\log(1/\tau_1)}{\log(2/c)}).$$

By Yao's principle, we conclude that any algorithm must make $k^{\frac{\log(1/\tau_1)}{\log(2/c)}}$ queries. ◀

4 Quantum Tolerant Junta Tester

In this section, we introduce the quantum tolerant junta tester.

▶ **Theorem 23.** *Let $k \in \mathbb{N}$ and $0 < \tau_2 < \tau_1^2 < 1$ be parameters. Then there exists a non-adaptive quantum algorithm with query complexity $\mathrm{poly}(k, 1/\tau)$ which*

- *accepts f with probability at least $2/3$ if f has correlation at least τ_1 to some k-junta;*
- *rejects f with probability at least $2/3$ if f has correlation at most τ_2 to every k-junta.*

The quantum tolerant junta tester described in Theorem 23 is implemented by Algorithm 1. It is essentially an adaptation of the tester from [3], tailored to the following refined trade-off between advantage and influence:

▶ **Lemma 24.** *Let $f : \{0,1\}^n \to \{\pm 1\}$ be a Boolean function and $T \subseteq [n]$ be an arbitrary set. Then*

$$1 - \mathrm{Inf}_{\overline{T}}(f) \leq \mathrm{corr}(f, \mathcal{J}_T) \leq \sqrt{1 - \mathrm{Inf}_{\overline{T}}(f)}.$$

The upper bound was shown in [3] as an intermediate step, and the lower bound was shown in [22].

We reprove this lemma here for completeness.

Proof. Let $f_{\mathrm{avg},T}$ be as given in Claim 7. It holds that $\mathrm{corr}(f, \mathcal{J}_T) = \mathbb{E}_{\boldsymbol{x}}[|f_{\mathrm{avg},T}(\boldsymbol{x})|]$. Moreover,

$$f_{\mathrm{avg},T}(x) = \sum_{S\subseteq[n]} \widehat{f}_S \mathbb{E}_{\boldsymbol{y}\sim\{0,1\}^n}[\chi_S(\boldsymbol{y}) \mid \boldsymbol{y}_S = x_S] = \sum_{S\subseteq T} \widehat{f}_S \chi_S(x).$$

Using Parseval's identity, we have $\mathbb{E}_{\boldsymbol{x}}[f_{\mathrm{avg},T}(\boldsymbol{x})^2] = \sum_{S\subseteq T} \widehat{f}_S^2 = 1 - \mathrm{Inf}_{\overline{T}}(f)$. Finally, since $f_{\mathrm{avg},T}(x) \in [-1, 1]$ for all $x \in \{0,1\}^n$, we get

$$\mathbb{E}_{\boldsymbol{x}}[f_{\mathrm{avg},T}(\boldsymbol{x})^2] \leq \mathbb{E}_{\boldsymbol{x}}[|f_{\mathrm{avg},T}(\boldsymbol{x})|] \leq \sqrt{\mathbb{E}_{\boldsymbol{x}}[f_{\mathrm{avg},T}(\boldsymbol{x})^2]}.$$

This completes the proof. ◀

Algorithm 1 Quantum tolerant junta tester.

Input: Quantum oracle access to f, an integer k, and parameters $\varepsilon_1, \varepsilon_2$ with $(1-2\varepsilon_1)^2 > 1-2\varepsilon_2$

Output: "Yes" or "No"

1: $\varepsilon \leftarrow ((1-2\varepsilon_1)^2 - (1-2\varepsilon_2))/2$
2: $\tau \leftarrow \varepsilon/3$
3: $M \leftarrow \frac{100k\log k \cdot \log(1/\tau)}{\tau^2}$
4: Run FourierSample^f M times to obtain $\boldsymbol{S}^{(1)}, \ldots, \boldsymbol{S}^{(M)} \sim \mathcal{S}_f$.
5: $\boldsymbol{p}_j \leftarrow 0$ for all $j \in [n]$
6: **for** $i \in [M]$ **do**
7: **if** $|\boldsymbol{S}^{(i)}| \le k$ **then**
8: $\boldsymbol{p}_j \leftarrow \boldsymbol{p}_j + 1/M$ for all $j \in \boldsymbol{S}^{(i)}$
9: **end if**
10: **end for**
11: $\widetilde{\boldsymbol{C}} \leftarrow \{j \in [n] : \boldsymbol{p}_j \ge \frac{\tau^2}{2k}\}$
12: $\boldsymbol{q}_R = 0$ for all $R \subseteq \widetilde{\boldsymbol{C}}$ with $|R| \le k$
13: **for** $i \in [M]$ **do**
14: $\boldsymbol{q}_R \leftarrow \boldsymbol{q}_R + 1/M$ for all $\boldsymbol{S}^{(i)} \subseteq R \subseteq \widetilde{\boldsymbol{C}}$ with $|R| \le k$
15: **end for**
16: $\widetilde{\mathbf{corr}} \leftarrow \max_{R \subseteq \widetilde{\boldsymbol{C}} : |R| \le k} \boldsymbol{q}_R$
17: **if** $\widetilde{\mathbf{corr}} > ((1-2\varepsilon_1)^2 + (1-2\varepsilon_2))/2$ **then**
18: **return** "Yes"
19: **else**
20: **return** "No"
21: **end if**

Regarding query complexity, the tester only queries f during the execution of FourierSample^f at line 4. Consequently, it makes $M = \frac{100k\log k \cdot \log(1/\tau)}{\tau^2}$ quantum queries in total, as desired. The remainder of this section is dedicated to proving the correctness of the algorithm.

Before presenting the formal proof, we provide a brief overview of how the algorithm proceeds. First, the algorithm draws M random subsets $\boldsymbol{S}^{(1)}, \ldots, \boldsymbol{S}^{(M)} \sim \mathcal{S}_f$. Then from line 5 to 11, the tester utilizes these samples to compute $\boldsymbol{p}_j$ as an estimate $\mathrm{Inf}_j^{\le k}[f]$, and identifies a set $\widetilde{\boldsymbol{C}}$ of coordinates with high estimated level-k influence.

Next, from line 12 to 15, the tester uses the sampled subsets to compute $\boldsymbol{q}_R$ as an estimate $1 - \mathrm{Inf}_R[f]$ for all $R \subseteq C$ with size $|R| = k$. Finally, the algorithm computes $\widetilde{\mathbf{corr}}$ to approximate $\mathrm{corr}(f, \mathcal{J}_k)$ and compares this value with a certain threshold to determine whether to accept or reject f.

▶ **Lemma 25.** *Fix any $k, \varepsilon_1, \varepsilon_2$ with $(1-2\varepsilon_1)^2 > 1-2\varepsilon_2$. Let $\mathcal{A}^f$ denote Algorithm 1. Then:*
- *for every f with $\mathrm{corr}(f, \mathcal{J}_k) > 1-2\varepsilon_1$, with high probability $\mathcal{A}^f$ outputs "YES"; and*
- *for every f with $\mathrm{corr}(f, \mathcal{J}_k) < 1-2\varepsilon_2$, with high probability $\mathcal{A}^f$ outputs "NO".*

Proof. Fix any Boolean function $f : \{0,1\}^n \to \{\pm 1\}$. Recall that Algorithm 1 computes $\widetilde{\mathbf{corr}}$ as an estimation of $\mathrm{corr}(f, \mathcal{J}_k)$ at line 16. It suffices to show that for any Boolean function f, with high probability $\mathrm{corr}^2(f, \mathcal{J}_k) - \varepsilon < \widetilde{\mathbf{corr}} < \mathrm{corr}(f, \mathcal{J}_k) + \varepsilon$ for $\varepsilon := ((1-2\varepsilon_1)^2 - (1-2\varepsilon_2))/2$.

Let $C := \{j \in [n] \mid \mathrm{Inf}_j^{\le k}[f] \ge \tau^2/k\}$ denote the set of coordinates with level-k influence at least τ^2/k. We show that Algorithm 1 computes a superset $\widetilde{\boldsymbol{C}} \supseteq C$ with high probability at line 11.

▷ Claim 26. It always holds that $|\widetilde{\boldsymbol{C}}| \leq \frac{2k^2}{\tau^2}$. Moreover, with probability at least $1 - k^{-2}$, $C \subseteq \widetilde{\boldsymbol{C}}$.

Proof. The first statement holds because $\sum_{j\in[n]} \boldsymbol{p}_j \leq k$.

For the second statement, observe that for each $j \in [n]$, $\boldsymbol{p}_j$ is the average of M independent random bits with mean $\Pr_{\boldsymbol{S}\sim\mathcal{S}_f}[j \in \boldsymbol{S} \wedge |\boldsymbol{S}| \leq k] = \mathrm{Inf}_j^{\leq k}[f]$. By Chernoff bound, with probability $1 - \exp(\log k \cdot \log \tau)$, we have $\boldsymbol{p}_j \geq \frac{\tau^2}{2k}$. The desired statement then follows from a union bound over all $j \in C$. ◁

Next, we observe that for each $R \subseteq \widetilde{\boldsymbol{C}}$ with size $|R| \leq k$, $\boldsymbol{q}_R$ is the average of M independent random bits with mean $\Pr_{\boldsymbol{S}\sim\mathcal{S}_f}[\boldsymbol{S} \subseteq R] = 1 - \mathrm{Inf}_{\overline{R}}[f]$. By Hoeffding's inequality, with probability at least $1 - \exp(100k \log k \cdot \log \tau)$ we have $|\boldsymbol{q}_R - (1 - \mathrm{Inf}_{\overline{R}}[f])| \leq \tau$. In which case, we can further apply Lemma 24 to obtain

$$\mathrm{corr}^2(f, \mathcal{J}_R) - \tau < \boldsymbol{q}_R < \mathrm{corr}(f, \mathcal{J}_R) + \tau.$$

Then by applying union bound over all $R \subseteq \widetilde{\boldsymbol{C}}$ with size $|R| = k$, with probability at least $1 - \exp(-\Omega(k \log k))$,

$$\max_{R\subseteq\widetilde{\boldsymbol{C}}:|R|\leq k} \mathrm{corr}^2(f, \mathcal{J}_R) - \tau < \widetilde{\mathbf{corr}} < \max_{R\subseteq\widetilde{\boldsymbol{C}}:|R|\leq k} \mathrm{corr}(f, \mathcal{J}_R) + \tau.$$

When $C \subseteq \widetilde{\boldsymbol{C}}$, by Lemma 8, we further have

$$\mathrm{corr}^2(f, \mathcal{J}_k) - \varepsilon = \mathrm{corr}^2(f, \mathcal{J}_k) - 2\tau - \tau < \widetilde{\mathbf{corr}} < \mathrm{corr}(f, \mathcal{J}_k) + \varepsilon.$$ ◀

References

1. Andris Ambainis, Aleksandrs Belovs, Oded Regev, and Ronald de Wolf. Efficient quantum algorithms for (gapped) group testing and junta testing. In *Proceedings of the Twenty-Seventh Annual ACM-SIAM Symposium on Discrete Algorithms*, SODA '16, pages 903–922, USA, 2016. Society for Industrial and Applied Mathematics. doi:10.1137/1.9781611974331.CH65.
2. Alp Atıcı and Rocco A Servedio. Quantum algorithms for learning and testing juntas. *Quantum Information Processing*, 6(5):323–348, 2007. doi:10.1007/s11128-007-0061-6.
3. Zongbo Bao, Yuxuan Liu, Penghui Yao, Zekun Ye, and Jialin Zhang. Efficient non-adaptive quantum algorithms for tolerant junta testing. In *2026 SIAM Symposium on Simplicity in Algorithms (SOSA)*, pages 100–126. Society for Industrial and Applied Mathematics, 2026. doi:10.1137/1.9781611978964.9.
4. Aleksandrs Belovs. Quantum algorithm for distribution-free junta testing. In René van Bevern and Gregory Kucherov, editors, *Computer Science - Theory and Applications - 14th International Computer Science Symposium in Russia, CSR 2019, Novosibirsk, Russia, July 1-5, 2019, Proceedings*, volume 11532 of *Lecture Notes in Computer Science*, pages 50–59. Springer, 2019. doi:10.1007/978-3-030-19955-5_5.
5. Ethan Bernstein and Umesh Vazirani. Quantum complexity theory. *SIAM Journal on Computing*, 26(5):1411–1473, 1997. doi:10.1137/S0097539796300921.
6. Eric Blais. Improved bounds for testing juntas. In Ashish Goel, Klaus Jansen, José D. P. Rolim, and Ronitt Rubinfeld, editors, *Approximation, Randomization and Combinatorial Optimization. Algorithms and Techniques*, pages 317–330, Berlin, Heidelberg, 2008. Springer Berlin Heidelberg. doi:10.1007/978-3-540-85363-3_26.
7. Eric Blais. Testing juntas nearly optimally. In *Proceedings of the Forty-First Annual ACM Symposium on Theory of Computing*, STOC '09, pages 151–158, New York, NY, USA, 2009. Association for Computing Machinery. doi:10.1145/1536414.1536437.

8 Eric Blais, Clément L. Canonne, Talya Eden, Amit Levi, and Dana Ron. Tolerant junta testing and the connection to submodular optimization and function isomorphism. *ACM Trans. Comput. Theory*, 11(4):24:1–24:33, 2019. doi:10.1145/3337789.

9 Guy Blanc, Alexandre Hayderi, Caleb Koch, and Li-Yang Tan. The sample complexity of smooth boosting and the tightness of the hardcore theorem. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1431–1450, 2024. doi:10.1109/FOCS61266.2024.00092.

10 Nader H. Bshouty. Almost Optimal Distribution-Free Junta Testing. In Amir Shpilka, editor, *34th Computational Complexity Conference (CCC 2019)*, volume 137 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 2:1–2:13, Dagstuhl, Germany, 2019. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.CCC.2019.2.

11 Mark Bun, Robin Kothari, and Justin Thaler. The polynomial method strikes back: Tight quantum query bounds via dual polynomials. *Theory of Computing*, 16(10):1–71, 2020. doi:10.4086/toc.2020.v016a010.

12 Chris Calabro. *The Exponential Complexity of Satisfiability Problems*. PhD thesis, University of California at San Diego, 2009.

13 Thomas Chen, Shivam Nadimpalli, and Henry Yuen. Testing and learning quantum juntas nearly optimally. In Nikhil Bansal and Viswanath Nagarajan, editors, *Proceedings of the 2023 ACM-SIAM Symposium on Discrete Algorithms, SODA 2023, Florence, Italy, January 22-25, 2023*, pages 1163–1185. SIAM, 2023. doi:10.1137/1.9781611977554.CH43.

14 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. Society for Industrial and Applied Mathematics, 2024. doi:10.1137/1.9781611977912.151.

15 Xi Chen and Shyamal Patel. New lower bounds for adaptive tolerant junta testing. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1778–1786, 2023. doi:10.1109/FOCS57990.2023.00108.

16 Xi Chen, Shyamal Patel, and Rocco A. Servedio. A mysterious connection between tolerant junta testing and agnostically learning conjunctions. *arXiv preprint arXiv:2504.16065*, 2025. doi:10.48550/arXiv.2504.16065.

17 Xi Chen, William Pires, Toniann Pitassi, and R. A. Servedio. Testing juntas and junta subclasses with relative error. In Nika Haghtalab and Ankur Moitra, editors, *Proceedings of Thirty Eighth Conference on Learning Theory*, volume 291 of *Proceedings of Machine Learning Research*, pages 5214–5245. PMLR, 30 June–04 July 2025. URL: https://proceedings.mlr.press/v291/chen25h.html.

18 Xi Chen, Rocco A. Servedio, Li-Yang Tan, Erik Waingarten, and Jinyu Xie. Settling the query complexity of non-adaptive junta testing. *J. ACM*, 65(6), 2018. doi:10.1145/3213772.

19 Hana Chockler and Dan Gutfreund. A lower bound for testing juntas. *Inf. Process. Lett.*, 90(6):301–305, 2004. doi:10.1016/J.IPL.2004.01.023.

20 Anindya De, Elchanan Mossel, and Joe Neeman. Junta Correlation is Testable . In *2019 IEEE 60th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1549–1563, Los Alamitos, CA, USA, November 2019. IEEE Computer Society. doi:10.1109/FOCS.2019.00090.

21 Vitaly Feldman, Parikshit Gopalan, Subhash Khot, and Ashok Kumar Ponnuswami. New results for learning noisy parities and halfspaces. In *2006 47th Annual IEEE Symposium on Foundations of Computer Science (FOCS'06)*, pages 563–574, 2006. doi:10.1109/FOCS.2006.51.

22 Eldar Fischer, Guy Kindler, Dana Ron, Shmuel Safra, and Alex Samorodnitsky. Testing juntas. *Journal of Computer and System Sciences*, 68(4):753–787, 2004. Special Issue on FOCS 2002. doi:10.1016/j.jcss.2003.11.004.

23 E. N. Gilbert. A comparison of signalling alphabets. *The Bell System Technical Journal*, 31(3):504–522, 1952. doi:10.1002/j.1538-7305.1952.tb01393.x.

24 Uffe Haagerup. The best constants in the khintchine inequality. *Studia Mathematica*, 70(3):231–283, 1981. doi:10.4064/sm-70-3-231-283.

25 Shirley Halevy and Eyal Kushilevitz. Distribution-free property-testing. *SIAM J. Comput.*, 37(4):1107–1138, 2007. doi:10.1137/050645804.

26 Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *Journal of the American Statistical Association*, 58(301):13–30, 1963. doi:10.2307/2282952.

27 Vishnu Iyer, Avishay Tal, and Michael Whitmeyer. Junta Distance Approximation with Sub-Exponential Queries. In Valentine Kabanets, editor, *36th Computational Complexity Conference (CCC 2021)*, volume 200 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 24:1–24:38, Dagstuhl, Germany, 2021. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.CCC.2021.24.

28 Vinayak M. Kumar. Most juntas saturate the hardcore lemma. In *2026 SIAM Symposium on Simplicity in Algorithms (SOSA)*, pages 142–150. Society for Industrial and Applied Mathematics, 2026. doi:10.1137/1.9781611978964.11.

29 Amit Levi and Erik Waingarten. Lower bounds for tolerant junta and unateness testing via rejection sampling of graphs. In Avrim Blum, editor, *10th Innovations in Theoretical Computer Science Conference, ITCS 2019, San Diego, California, USA, January 10-12, 2019*, volume 124 of *LIPIcs*, pages 52:1–52:20. Schloss Dagstuhl – Leibniz-Zentrum für Informatik, 2019. doi:10.4230/LIPIcs.ITCS.2019.52.

30 Zhengyang Liu, Xi Chen, Rocco A. Servedio, Ying Sheng, and Jinyu Xie. Distribution-free junta testing. *ACM Trans. Algorithms*, 15(1), 2018. doi:10.1145/3264434.

31 Elchanan Mossel, Ryan O'Donnell, and Rocco P. Servedio. Learning juntas. In *Proceedings of the Thirty-Fifth Annual ACM Symposium on Theory of Computing*, STOC '03, pages 206–212, New York, NY, USA, 2003. Association for Computing Machinery. doi:10.1145/780542.780574.

32 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*, STOC 2024, pages 1039–1050, New York, NY, USA, 2024. Association for Computing Machinery. doi:10.1145/3618260.3649687.

33 Ramesh Krishnan S. Pallavoor, Sofya Raskhodnikova, and Erik Waingarten. Approximating the distance to monotonicity of boolean functions. *Random Struct. Algorithms*, 60(2):233–260, 2022. doi:10.1002/RSA.21029.

34 Michal Parnas, Dana Ron, and Ronitt Rubinfeld. Tolerant property testing and distance approximation. *Journal of Computer and System Sciences*, 72(6):1012–1042, 2006. doi:10.1016/j.jcss.2006.03.002.

35 Mert Sağlam. Near log-convexity of measured heat in (discrete) time and consequences. In *2018 IEEE 59th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 967–978, 2018. doi:10.1109/FOCS.2018.00095.

36 Rocco A. Servedio, Li-Yang Tan, and John Wright. Adaptivity Helps for Testing Juntas. In David Zuckerman, editor, *30th Conference on Computational Complexity (CCC 2015)*, volume 33 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 264–279, Dagstuhl, Germany, 2015. Schloss Dagstuhl – Leibniz-Zentrum für Informatik. doi:10.4230/LIPIcs.CCC.2015.264.

37 Gregory Valiant. Finding correlations in subquadratic time, with applications to learning parities and the closest pair problem. *J. ACM*, 62(2), 2015. doi:10.1145/2728167.

38 Revaz R. Varshamov. Estimate of the number of signals in error correcting codes. *Doklady Akademii Nauk SSSR*, 117(5):739–741, 1957.