Electronic Colloquium on Computational Complexity, Revision 1 of Report No. 70 (2019)

On Local Testability in the Non-Signaling Setting

Alessandro Chiesa
alexch@berkeley.edu
UC Berkeley

Peter Manohar
pmanohar@cs.cmu.edu
Carnegie Mellon University

Igor Shinkar
ishinkar@sfu.ca
Simon Fraser University

January 5, 2020

Abstract

Non-signaling strategies are a generalization of quantum strategies that have been studied in physics for decades, and have recently found applications in theoretical computer science. These applications motivate the study of local-to-global phenomena for *non-signaling functions*.

We prove that low-degree testing in the non-signaling setting is possible, assuming that the locality of the non-signaling function exceeds a threshold. We additionally show that if the locality is below the threshold then the test fails spectacularly, in that there exists a non-signaling function which passes the test with probability 1 and yet is maximally far from being low-degree.

Along the way, we present general results about the local testability of linear codes in the non-signaling setting. These include formulating natural definitions that capture the condition that a non-signaling function "belongs" to a given code, and characterizing the sets of local constraints that imply membership in the code. We prove these results by formulating a logical inference system for linear constraints on non-signaling functions that is complete and sound.

Keywords: non-signaling strategies; locally testable codes; low-degree testing; Fourier analysis

ISSN 1433-8092

Contents

1 Introduction

Locally testable codes (LTCs) are error correcting codes in which one can verify whether a given string belongs to the code by reading only a few (randomly chosen) bits from the string. Goldreich and Sudan [GS06] have described LTCs as the "combinatorial counterparts of the complexity theoretic notion of PCPs", motivating the standalone study of these objects.

In this work we study local testability for *non-signaling strategies*, which are a class of non-local strategies that generalize quantum strategies, capturing the maximum amount of "non-local correlation" that can occur under the assumption that spatially-isolated parties cannot communicate instantaneously. Non-signaling strategies have been studied in physics for decades [Ras85; KT92; PR94], in order to better understand quantum entanglement. Recently they have gained attention in computer science due to their applications to hardness of approximation [KRR16] and delegation of computation [KRR13; KRR14]. PCPs sound against non-signaling strategies (nsPCPs) underlie these applications, which motivates the study of local testability in the non-signaling setting.

Given an integer n, a field $\mathbb{F}$, and a locality parameter $k \leq n$, the object that we study is a *k-non-signaling function* $\mathcal{F} \colon [n] \to \mathbb{F}$, which extends the notion of a function $f \colon [n] \to \mathbb{F}$ as follows.[1]

Definition 1.1. *A k-**non-signaling function** $\mathcal{F} \colon [n] \to \mathbb{F}$ is a collection $\{\mathcal{F}_S\}_{S \subseteq [n] : |S| \leq k}$ where each $\mathcal{F}_S$ is a distribution over local functions $g \colon S \to \mathbb{F}$, and for any two subsets $R \subseteq S \subseteq [n]$ with $|S| \leq k$ it holds that the distribution $\mathcal{F}_R$ and the marginal distribution $\mathcal{F}_S|_R$ are equal.[2] (The set of all such $\mathcal{F}$ are the solutions to the k-relaxation in the Sherali–Adams hierarchy [SA90].)*

The evaluation of $\mathcal{F}$ on a set S is a single sample $g \colon S \to \mathbb{F}$ from the distribution $\mathcal{F}_S$. Intuitively, a k-non-signaling function is like a quantum function: evaluation is probabilistic and only happens once, just like quantum measurement; and $\mathcal{F}$ can only be evaluated on at most k points simultaneously, which is similar to the uncertainty principle. As k approaches n, $\mathcal{F}$ behaves more like a classical function and, when $k = n$, $\mathcal{F}$ is a distribution over functions $f \colon [n] \to \mathbb{F}$.

Local testability of non-signaling functions may sound like an oxymoron, because non-signaling functions, at least superficially, are collections of local distributions with no global structure that we can talk about. Yet prior work has shown that local-to-global phenomena *are possible*.

For example, [CMS18] shows that any non-signaling function passing the linearity test [BLR93] with high probability is well-approximated by a *quasi-distribution* supported on linear functions. This result was later used in [CMS19] to show that the exponential-length constant-query PCP of [Aro+98] is sound against non-signaling strategies.

The results obtained in [CMS18; CMS19] naturally raise the question of whether local testability in the non-signaling setting is possible for other codes, like those based on low-degree polynomials. After all, both linearity testing and low-degree testing do work in the quantum setting [NV18].

Recall that, in the classical setting, local testability plays a central role in PCP constructions, many of which can be described as having two main components.

- *Property testing:* check with few queries whether or not the given proof π belongs to a code $\mathbf{C}$.

[1] There are two distinct definitions of a non-signaling strategy, depending on whether the strategy is meant to represent isolated parties or a function. The former is used for MIPs [KRR13; KRR14], while the latter is used for PCPs and property testing [KRR13; KRR14; CMS18; CMS19]. We use the latter definition, although equivalent statements of all our results will hold when adopting the former definition (see the appendix in [CMS18]).

[2] A common relaxation of this condition requires that these two distributions are only statistically (or computationally) close. While we consider the standard definition, we note that this is without loss of generality as [CMS18] shows that every statistically (or computationally) non-signaling strategy is close to an (exact) non-signaling strategy.

- *Checking computation:* given that π is a codeword in $\mathbf{C}$ (or at least is close to a codeword), check with few queries whether or not π proves the desired statement.

This modular approach has enabled the study of local testability as a natural standalone goal, which in turn has led to improved PCP constructions.

Inspired by this state of affairs, we initiate the study of locally testable codes in general in the non-signaling setting, focusing specifically on the case of low-degree testing. We believe that, similarly to the classical setting, understanding local testability against non-signaling strategies will enable researchers to construct more efficient non-signaling PCPs.

1.1 Low-degree testing against non-signaling functions

We show that a simple low-degree test, the *evenly-spaced points test*, tests proximity to degree-d non-signaling functions when $k \geq O(d^2)$, and *fails* to test proximity when $k \leq O(d^2)$.

The evenly-spaced points test. Let $m, d \in \mathbb{N}$ and p be a prime with $p \geq d+2$. Given a function $f\colon \mathbb{F}_p^m \to \mathbb{F}_p$, the degree-$d$ evenly-spaced points test: (1) samples a random point $x \in \mathbb{F}_p^m$ and slope $h \in \mathbb{F}_p^m \setminus \{0^m\}$, (2) checks that $\sum_{i=0}^{d+1} c_i f(x+ih) = 0$, where $c_i = (-1)^i \binom{d+1}{i}$. It is well-known that if f passes the degree-d evenly-spaced points test with high probability, then f is close to (the evaluation of) an m-variate polynomial of total degree at most d [RS96]. Below we ask whether the test is also sound in the non-signaling setting.

Suppose that a k-non-signaling function $\mathcal{F}\colon \mathbb{F}_p^m \to \mathbb{F}_p$ passes the evenly-spaced points test with high probability. Can we deduce any global low-degree structure about $\mathcal{F}$?

In more detail, the probabilistic experiment that we consider is this: first we sample x and h according to the distribution of the evenly-spaced points test, and let the query set S be $\{x+ih : i \in \{0,\ldots,d+1\}\}$; then we sample a local function $g\colon S \to \mathbb{F}_p$ according to the distribution $\mathcal{F}_S$; and finally we check that $\sum_{i=0}^{d+1} c_i g(x+ih) = 0$.

The answer to the above question will, in general, depend on the locality parameter k of $\mathcal{F}$. At minimum, we need $k \geq d+2$ for otherwise we cannot even run the evenly-spaced points test (k is the maximum number of simultaneous queries to $\mathcal{F}$). At the other extreme, when k has the maximum value ($k = p^m$) then we are back to the classical case because $\mathcal{F}$ is now a distribution over functions $f\colon \mathbb{F}_p^m \to \mathbb{F}_p$; hence if $\mathcal{F}$ passes the test with high probability then (one can verify that) with high probability a function f sampled according $\mathcal{F}$ is close to low-degree. In fact, even when $k \geq O(d^m)$, we are in a trivial case, as one can query $\mathcal{F}$ on an interpolating set, a "cube of $(d+1)^m$ points".

We are thus interested in whether or not the test works for *non-trivial* values of k, namely when $O(d) \leq k < O(d^m)$, and thus we will assume that $m \geq 2$. In this regime, k is large enough to run the test, and yet is small enough so that one cannot query an interpolating set. Our first result shows that the test succeeds in the non-signaling setting when $k \geq O(d^2)$. This is a non-signaling analogue of the evenly-spaced points test, similar to how [CMS18] gives a non-signaling analogue of the linearity test of [BLR93].

Theorem 1 (informal)**.** *Let $\mathcal{F}\colon \mathbb{F}_p^m \to \mathbb{F}_p$ be a k-non-signaling function that passes the degree-d evenly-spaced points test with high probability, where $k \geq (d+2)^2$. Then $\mathcal{F}$ has a global individual degree-d structure.*

One drawback of the above theorem is that the conclusion only asserts that $\mathcal{F}$ has an individual degree-d structure, when one would like to conclude that $\mathcal{F}$ has a total degree-d structure. We discuss the difficulty of extending the result to total degree-d in Remark 2.3.

Our second result shows that the test fails when $k \leq O(d^2)$, and moreover it fails *even when the test passes with probability* 1, namely, it fails in the worst possible sense.

Theorem 2 (informal)**.** *For every k with $2d+2 \leq k < \frac{3}{16}(d+2)^2$, there exists a k-non-signaling function that passes the evenly-spaced points test with probability 1, and yet is $(1-\frac{1}{p})$-far from all degree-d k-non-signaling functions.*

Theorem 2 is surprising, as it reveals that in the non-signaling setting, there is a regime of k in which the low-degree test fails. This stands in sharp contrast to the fact that for linearity testing, there is *no regime* of k in which non-signaling linearity test fails. Our results thus suggest that low-degree testing is a qualitatively different task, as it has a regime of k where a natural test fails.

Theorem 2 also shows that in the case of bivariate testing when $m = 2$, there is (up to constants) no non-trivial value of k where the test succeeds. This is counterintuitive, as bivariate testing is a natural test for which we would expect some guarantee to hold (regardless of how weak), at the very least when the test passes with probability 1.

Other low-degree tests. We note that Theorem 2 generalizes to *any* low-degree test over an arbitrary finite field $\mathbb{F}$ that (1) works by checking constraints that lie along a line, and (2) has perfect completeness. Thus, Theorem 2 gives a strong negative result, as it proves that the requirement that $k \geq O(d^2)$ is necessary for a large class of natural tests.

Beyond low-degree testing. Our theorems on low-degree testing come from applying more general results that we prove about the structure of local characterizations for *any linear code*, in the non-signaling setting. We view our general results on local characterizations as a significant technical contribution within this paper, and we now discuss them.

1.2 Local characterizations and linear proofs

Local characterizations are fundamental to the study of locally testable codes [RS96]. They express membership in a given linear code via a set of low-weight constraints, and they naturally induce a canonical tester: sample a random low-weight constraint and check if the given word satisfies it. In order to prove the negative result in Theorem 2, we do not need to consider distributions on constraints, but instead we only need to study how constraints express code membership, via *exact* local characterizations [RS96]. Below we describe one of our main technical contributions, which informally consists of establishing necessary and sufficient conditions for when a constraint set is a local characterization for a code, in the non-signaling setting. We begin by recalling known facts about local characterizations in the classical setting, and then proceed to the non-signaling setting.

The classical setting. A *constraint set* $T \subseteq \mathbb{F}^n$ for a linear code $\mathbf{C} \subseteq \mathbb{F}^n$ is a subset of its dual code $\mathbf{C}^\perp$. A constraint set T is a *ℓ-local characterization* of $\mathbf{C}$ if every $\alpha \in T$ has at most ℓ non-zero entries, and the condition "$\langle \alpha, f \rangle = 0$ for every $\alpha \in T$" implies that $f \in \mathbf{C}$ (and conversely).

For example, the set $\{e_x + e_y - e_{x+y} : x, y \in \{0,1\}^n\}$ where e_x is the x-th standard basis vector in $\{0,1\}^{\{0,1\}^n}$ is a 3-local characterization of the Hadamard code, because $f(x)+f(y)-f(x+y) = 0$ for every $x, y \in \{0,1\}^n$ implies that f is a linear function, and conversely. As another example, the Reed–Muller code containing all polynomials $f \colon \mathbb{F}^m \to \mathbb{F}$ in m variables of total degree at most d

has a $(d+2)$-local characterization T, where T contains a constraint α for each subset S of $\mathbb{F}^m$ of size $d+2$ that is contained in a line.

There is a simple condition that is both necessary and sufficient for a constraint set T to be a local characterization for $\mathbf{C}$: the span of T equals $\mathbf{C}^\perp$. In this work it is useful to view this condition instead through the lens of mathematical logic, as follows. Given a constraint set T and $\alpha \in \mathbb{F}^n$, we define the notion of a linear proof.

Definition 1.2 (Linear proof)**.** *We write $T \vdash \alpha$ (T proves α) if there exists a sequence $(\alpha_0 := 0^n, \alpha_1, \ldots, \alpha_{r-1}, \alpha_r := \alpha)$ with each $\alpha_i \in \mathbb{F}^n$ such that, for every $i \in [r]$, one of the following holds:*
- *$\exists\, j < i$ and $b \in \mathbb{F}$ such that $\alpha_i = b\alpha_j$,*
- *$\exists\, j < i$ and $\gamma \in T$ such that $\alpha_i = \alpha_j + \gamma$,*
- *$\exists\, j_1, j_2 < i$ such that $\alpha_i = \alpha_{j_1} + \alpha_{j_2}$.*

As an example, suppose that $\alpha = \sum_{i=1}^{r} b_i\gamma_i$ with each $b_i \in \mathbb{F}$ and $\gamma_i \in T$. Then the sequence $(0^n, \gamma_1, b_1\gamma_1, \ldots, \gamma_r, b_r\gamma_r, \alpha_1, \ldots, \alpha_r)$, where each α_i is the partial sum $b_1\gamma_1 + \cdots + b_i\gamma_i$, gives a linear proof that $\alpha \in \mathrm{span}(T)$.

One can immediately see that $T \vdash \alpha$ if and only if $\alpha \in \mathrm{span}(T)$. In particular, we have the following lemma.

Lemma 1.3. *Linear proofs are (i) complete: if $\langle \gamma, f\rangle = 0$ for every $\gamma \in T$ implies $\langle \alpha, f\rangle = 0$, then $T \vdash \alpha$; and (ii) sound: if $\langle \gamma, f\rangle = 0$ for every $\gamma \in T$ and $T \vdash \alpha$, then $\langle \alpha, f\rangle = 0$. In particular, a constraint set T is a local characterization of a linear code $\mathbf{C}$ if and only if $T \vdash \mathbf{C}^\perp$.*

Our goal is to establish a non-signaling analogue of Lemma 1.3.

A motivating example. We illustrate via an example why a statement like Lemma 1.3 is non-trivial in the non-signaling setting. Let $n \in \mathbb{N}$ be even, and let $T = \{1^n - e_i : i \in [n]\} \subseteq \{0,1\}^n$, i.e. T contains every vector that is 1 in all but one of the coordinates, where it is 0. Classically, one can check that T is a $(n-1)$-local characterization of the code $\mathbf{C} = \{0^n\}$, as if $f \in \{0,1\}^n$ satisfies $\langle \alpha, f\rangle = 0$ for every $\alpha \in T$ (equivalently, $\sum_{\ell \neq i} f(\ell) = 0$ for every $i \in [n]$), then we must have $f = 0^n$, since n is even. This is because $T \vdash e_i$, and so f must satisfy $f(i) = \langle e_i, f\rangle = 0$.

However, there exist $(n-1)$-non-signaling functions that satisfy every constraint in T and yet are not identically 0; the non-signaling function which outputs uniformly random bits with parity 0 on every set of size exactly $n-1$ is one such example. In particular, T is not a $(n-1)$-local characterization of $\mathbf{C}$. To see why, let us examine where the classical argument that $f(i) = 0$ fails for $(n-1)$-non-signaling functions. Recall that an $(n-1)$-non-signaling function can only be evaluated simultaneously at $n-1$ points. Thus, while one can classically argue, for example, that $f(i) + f(j) = 0$ via the argument that $\sum_{\ell \neq i} f(\ell) = 0$ and $\sum_{\ell \neq j} f(\ell) = 0$ implies that $f(i) + f(j) = \sum_{\ell \neq i} f(\ell) + \sum_{\ell \neq j} f(\ell) = 0$, this reasoning is no longer valid in the non-signaling setting because it requires f to be simultaneously defined at every $i \in [n]$, which is $n > k = n-1$ points. In particular, any proof that $\langle e_i, f\rangle = 0$ from T requires f to be simultaneously defined on all n points, so this logical reasoning is not valid in the non-signaling setting.

An equivalence for non-signaling functions. We prove an analogous equivalence in the non-signaling setting, which informally states that a suitable notion of local characterization for any linear code is equivalent to being able to prove all low weight elements of $\mathbf{C}^\perp$ using *local* proofs. This equivalence is a strict generalization of Lemma 1.3. The example above can thus be viewed as a case where a low weight element of $\mathbf{C}^\perp$ (namely, e_i) has no local proof from a particular T.

We begin by formulating a notion of local characterization that works for constraint sets applied to non-signaling functions rather than (classical) functions. There are two main qualitative differences with the classical case. First, the definition depends on the locality parameter k because we need to specify the locality of the non-signaling functions that we consider. Second, the requirement that a non-signaling function "belongs" to a code $\mathbf{C}$ is expressed via a property that we call $\mathbf{C}$-*explainability*, on which we comment after the definition.

Definition 1.4 (informal)**.** *A constraint set $T \subseteq \mathbf{C}^\perp$ is a ℓ-**local characterization** for $(\mathbf{C}, k)$ if every $\alpha \in T$ has at most ℓ non-zero entries, and the set of k-non-signaling functions that satisfy every $\alpha \in T$ with probability 1 equals the set of k-non-signaling functions that are "$\mathbf{C}$-explainable".*

The term "$\mathbf{C}$-explainable" refers to the condition that the given non-signaling function is, with probability 1, consistent with the restriction of some codeword in $\mathbf{C}$. This condition is motivated by non-trivial properties of the Fourier spectrum of non-signaling functions that we discuss later on (see Section 2.5). For now, it suffices to say that if a non-signaling function $\mathcal{F}$ is $\mathbf{C}$-explainable then $\mathcal{F}$ satisfies natural *global* properties that extend code membership to the non-signaling setting.

We remark that Definition 1.4 reduces to the classical notion of local characterization when setting $k := n$. We now introduce the notion of local linear proofs that we use in our equivalence.

Definition 1.5 (k-local linear proof)**.** *Given a constraint set T and $\alpha \in \mathbb{F}^n$, we write $T \vdash_k \alpha$ if there exists a sequence $(\alpha_0 := 0^n, \alpha_1, \ldots, \alpha_{r-1}, \alpha_r := \alpha)$ with each $\alpha_i \in \mathbb{F}^n$ such that, for every $i \in [r]$, one of the following holds:*

- *$\exists\, j < i$ and $b \in \mathbb{F}$ such that $\alpha_i = b\alpha_j$*
- *$\exists\, j < i$ and $\gamma \in T$ such that $|\mathrm{supp}(\alpha_j) \cup \mathrm{supp}(\gamma)| \leq k$ and $\alpha_i = \alpha_j + \gamma$*
- *$\exists\, j_1, j_2 < i$ such that $|\mathrm{supp}(\alpha_{j_1}) \cup \mathrm{supp}(\alpha_{j_2})| \leq k$ and $\alpha_i = \alpha_{j_1} + \alpha_{j_2}$.*

Above, $\mathrm{supp}(\alpha)$ denotes the set of indices $i \in [n]$ where $\alpha_i \neq 0$, and $\mathrm{wt}(\alpha)$ is the size of $\mathrm{supp}(\alpha)$. Notice that Definition 1.5 is nearly identical to Definition 1.2: the only change is the addition of the restriction on the support size in the second and third bullets.

The motivation behind Definition 1.5 is the following fact: if $T \vdash_k \alpha$ then any k-non-signaling function that satisfies every constraint in T must satisfy α as well. Definition 1.5 thus captures a notion of constraint propagation for non-signaling functions. The restriction on the support size in the second and third bullets is there because querying a k-non-signaling function on more than k points simultaneously is undefined.

We now state our main technical contribution in this section, a non-signaling analogue of Lemma 1.3.

Theorem 3 (informal)**.** *k-local linear proofs are complete and sound for k-non-signaling functions. In particular, a constraint set T is a ℓ-local characterization for $(\mathbf{C}, k)$ if and only if $T \vdash_k \alpha$ for every $\alpha \in \mathbf{C}^\perp$ with $\mathrm{wt}(\alpha) \leq k$.*

Proving that k-local linear proofs are sound is straightforward; the interesting component of Theorem 3 is showing that k-local linear proofs are complete. We do this by showing that for every T there exists a k-non-signaling function that satisfies every α where $T \vdash_k \alpha$, and violates every α where $T \nvdash_k \alpha$ with probability $1 - \frac{1}{|\mathbb{F}|}$. This k-non-signaling function is very simple: the distribution $\mathcal{F}_S$ is the uniform distribution over all functions $f \colon S \to \mathbb{F}$ such that $\langle \alpha, f \rangle = 0$ for every α where $T \vdash_k \alpha$ and $\mathrm{supp}(\alpha) \subseteq S$.

When $k = n$ in Theorem 3 we recover the classical statement (Lemma 1.3). This is because when $k = n$, $T \vdash_k \alpha$ if and only if $T \vdash \alpha$. However, when $k < n$, the equivalence is qualitatively different from its classical analogue. While Lemma 1.3 essentially captures a simple linear algebraic statement (the constraints span the dual code), Theorem 3 is a non-trivial statement that *does not involve linear spaces*. This is because the requirement $T \vdash_k \alpha$ depends on k in a way that breaks linearity, as exhibited by our motivating example earlier.

Separating classical and non-signaling local characterizations. Theorem 2 is an application of Theorem 3 that shows that there is a gap between classical and non-signaling local characterizations. In Appendix A, we use Theorem 3 to prove the theorem below, showing a stronger gap between classical and non-signaling local characterizations.

Theorem 4 (Informal)**.** *Let $d \in \mathbb{N}$ with $d \geq 2$, and let $n \in \mathbb{N}$ such that $2n \equiv 0 \mod d$. There exists a code $\mathbf{C} \subseteq \{0,1\}^n$ and a constraint set T such that:*

- *For classical functions, T is a d-local characterization of $\mathbf{C}$, but*
- *For non-signaling functions, T is not a d-local characterization of $(\mathbf{C}, k)$, for all $k \leq O(n)$.*

Theorem 4 shows a strong separation between classical and non-signaling local characterizations, as T classically gives an $O(1)$-local characterization of $\mathbf{C}$, but is not an $O(1)$-local characterization of $\mathbf{C}$ for k-non-signaling functions, even when k is allowed be $\Omega(n)$, i.e. nearly maximally large.

On robust local characterizations. We have so far discussed *exact* local characterizations, which suffice for Theorem 2 presented in Section 1.1. Can we make general statements about *robust* local characterizations, which could be used to establish positive results such as Theorem 1? In Appendix B, we show that a suitable non-signaling analogue of robust local characterizations is related to the "proof length" of the k-local linear proof. An application of this result is that much of the analysis of the linearity test in [CMS18] is tight up to constants.

1.3 Roadmap

In Section 2 we provide an overview of the proofs of our results. Then, in Section 3 and Section 4 we formally define non-signaling functions, quasi-distributions, and discuss the relationship between them using Fourier analysis. In Section 5 we discuss what it means for a non-signaling function to "belong" to a given linear code. In Section 6 we prove that the non-signaling low-degree test works (Theorem 1). In Section 7 we prove an equivalence between local characterizations for non-signaling linear codes and local linear proofs (Theorem 3). We conclude in Section 8, by using Theorem 3 to show that the non-signaling low-degree test fails for small locality (Theorem 2).

2 Techniques

We outline the techniques used to prove our results. We begin by explaining the Fourier structure of non-signaling functions in Section 2.1. This structure is fundamental to the proofs of our results. We then outline the proof of Theorem 1 in Section 2.2. In Section 2.3 we outline our proof of the relationship between local characterizations and local linear proofs. In Section 2.4 we use the techniques and main theorem from Section 2.3 to show Theorem 2, that any low-degree lines test fails for non-signaling functions when $k \leq O(d^2)$. Finally, in Section 2.5 we justify our definition of "**C**-explainability".

Notation. A k-non-signaling function $\mathcal{F}$ is defined by local distributions $\mathcal{F}_S$ for each $S \subseteq [n]$ with $|S| \leq k$. Because of this, when studying non-signaling functions we naturally encounter situations where we only consider subsets of a domain containing at most k elements, or vectors in $\mathbb{F}^n$ of weight at most k. We introduce notation to make referring to these notions more convenient. For a subset $S \subseteq [n]$ we write $S \subseteq [n]_{\leq k}$ if $|S| \leq k$. For a vector $\alpha \in \mathbb{F}^n$, we let $\mathrm{supp}(\alpha) = \{i \in [n] : \alpha_i \neq 0\}$ and $\mathrm{wt}(\alpha) = |\mathrm{supp}(\alpha)|$. For a set of vectors $R \subseteq \mathbb{F}^n$, we let $R_{\leq k} \subseteq R$ denote the subset $\{\alpha \in R : \mathrm{wt}(\alpha) \leq k\}$. In particular, $\mathbb{F}^n_{\leq k}$ denotes the set $\{\alpha \in \mathbb{F}^n : \mathrm{wt}(\alpha) \leq k\}$. For a subset $S \subseteq [n]$, we use similar notation and let $R_{\subseteq S} = \{\alpha \in R : \mathrm{supp}(\alpha) \subseteq S\}$.

2.1 The Fourier structure of non-signaling functions

We make frequent use of Fourier analysis to state and establish properties of non-signaling functions. Below we recall basic facts about Fourier analysis, explain their application to quasi-distributions, and state an equivalence between non-signaling functions and quasi-distributions. This equivalence motivates a definition for the Fourier spectrum of a non-signaling function.

Refresher on Fourier analysis. Let $\mathbb{F}$ be the finite field of size q with characteristic p, and $\mathbb{F}_p$ the prime subfield of $\mathbb{F}$. The inner product of $F_1, F_2 \colon \mathbb{F}^n \to \mathbb{C}$ is $\langle F_1, F_2 \rangle := \frac{1}{q^n} \sum_{f \in \mathbb{F}^n} \overline{F_1(f)} F_2(f)$. The *character* corresponding to $\alpha \in \mathbb{F}^n$ is the function $\chi_\alpha \colon \mathbb{F}^n \to \mathbb{C}$ defined as $\chi_\alpha(f) := \omega^{\mathrm{Tr}(\langle \alpha, f \rangle)}$ where: $\mathrm{Tr} \colon \mathbb{F} \to \mathbb{F}_p$ is the trace map; $\langle \alpha, f \rangle$ is the inner product $\sum_{i=1}^n \alpha_i f_i$; $\omega = e^{2\pi i/p}$ is a primitive complex p-th root of unity; and ω^j is defined by thinking of $j \in \mathbb{F}_p$ as an integer in $\{0, 1, \ldots, p-1\}$. The characters $\{\chi_\alpha\}_{\alpha \in \mathbb{F}^n}$ form an orthonormal basis of the space of all functions $F \colon \mathbb{F}^n \to \mathbb{C}$, so every function $F \colon \mathbb{F}^n \to \mathbb{C}$ can be written as

$$F(\cdot) = \sum_{\alpha \in \mathbb{F}^n} \widehat{F}(\alpha) \chi_\alpha(\cdot) \ , \text{ where } \widehat{F}(\alpha) := \langle \chi_\alpha, F \rangle \ .$$

The values $\{\widehat{F}(\alpha)\}_{\alpha \in \mathbb{F}^n}$ are called the *Fourier coefficients* of F.

Quasi-distributions. A *quasi-distribution* $\mathcal{Q}$ over functions $f \colon [n] \to \mathbb{F}$ is a distribution where the probability weights are complex numbers that "add up" to real probabilities. More formally, a quasi-distribution is a function $\mathcal{Q} \colon \mathbb{F}^n \to \mathbb{C}$ where $\sum_{f \in \mathbb{F}^n} \mathcal{Q}(f) = 1$. (We abuse notation and identify a function $f \colon [n] \to \mathbb{F}$ with the vector in $\mathbb{F}^n$ corresponding to its evaluation table.) We say that $\mathcal{Q}$ is *k-local* if the marginals $\mathcal{Q}|_S$ for each $S \subseteq [n]_{\leq k}$ are distributions, namely, if for each $S \subseteq [n]_{\leq k}$ and $g \colon S \to \mathbb{F}$ it holds that $\sum_{f \in \mathbb{F}^n : f|_S = g} \mathcal{Q}(f)$ is a non-negative real number. We can decompose a quasi-distribution $\mathcal{Q}$ according to the Fourier basis: we can write $\mathcal{Q}(f) = \sum_{\alpha \in \mathbb{F}^n} \widehat{\mathcal{Q}}(\alpha) \chi_\alpha(f)$, where $\{\chi_\alpha\}_{\alpha \in \mathbb{F}^n}$ are the characters and $\{\widehat{\mathcal{Q}}(\alpha)\}_{\alpha \in \mathbb{F}^n}$ are the Fourier coefficients of $\mathcal{Q}$.

Equivalence lemma. The following lemma shows that k-local quasi-distributions and k-non-signaling functions are equivalent, and exposes the Fourier structure of non-signaling functions.

Lemma 2.1. *A quasi-distribution $\mathcal{Q}$ is equivalent to a k-non-signaling function $\mathcal{F}$ if and only if for every $\alpha \in \mathbb{F}^n_{\leq k}$ it holds that $\widehat{\mathcal{Q}}(\alpha) = \frac{1}{q^n} \sum_{j \in \mathbb{F}_p} \omega^{-j} \Pr[\mathrm{Tr}(\langle \alpha, \mathcal{F} \rangle) = j]$, where the random variable $\langle \alpha, \mathcal{F} \rangle$ has the probability distribution given by*

$$\left\{ \Pr[\langle \alpha, \mathcal{F} \rangle = b] := \Pr_{f \leftarrow \mathcal{F}_{\mathrm{supp}(\alpha)}} \left[\sum_{i \in \mathrm{supp}(\alpha)} \alpha_i f(i) = b \right] \right\}_{b \in \mathbb{F}} .$$

The foregoing lemma motivates defining the Fourier coefficients of a k-non-signaling function $\mathcal{F}$ as follows: for every $\alpha \in \mathbb{F}^n$ with $\mathrm{wt}(\alpha) \leq k$ we define

$$\widehat{\mathcal{F}}(\alpha) := \frac{1}{q^n} \sum_{j \in \mathbb{F}_p} \omega^{-j} \Pr[\mathrm{Tr}(\langle \alpha, \mathcal{F} \rangle) = j] .$$

For more details on the above, including the proof of our Equivalence Lemma, see Section 4.

2.2 Low-degree testing

We outline the proof of Theorem 1. As a simple case, we first state and prove the theorem in the zero error case (when test passes with probability 1), and then we briefly explain how to extend the proof to the robust case (when the test passes with probability $1 - \varepsilon$).

The zero error case. Let $\mathbf{C}$ be the set of all m-variate polynomials of *individual* degree d. Formally, we first show the following.

Theorem 2.1 (formal version of Theorem 1, zero error case)**.** *Let $m, d \in \mathbb{N}$ and p be a prime with $p \geq d + 2$. Let $\mathcal{F} \colon \mathbb{F}_p^m \to \mathbb{F}_p$ be a k-non-signaling function, and suppose that $\mathcal{F}$ passes the degree-d evenly-spaced points test with probability 1. Then $\mathcal{F}$ (viewed as a $\lfloor k/(d+2) \rfloor$-non-signaling function) is $\mathbf{C}$-explainable.*

In the language of Section 1.2, Theorem 2.1 shows that T, the set of linear constraints checked by the degree-d evenly spaced points test, is a $(d+2)$-local characterization of $\mathbf{C}$.

Theorem 2.1 is a non-signaling analogue of the following classical fact: if $f \colon \mathbb{F}_p^m \to \mathbb{F}_p$ passes the evenly-spaced points test with probability 1, then f is a polynomial of total degree d. (Note that in Theorem 2.1 we only conclude that $\mathcal{F}$ has individual degree d. We remark on the difference after the proof.) Our proof of Theorem 2.1 can be interpreted as taking a local proof of the aforementioned classical fact, and lifting it to the non-signaling setting.

Concretely, let us consider the following simple classical statement.

Theorem 2.2 (folklore)**.** *Let $f \colon \mathbb{F}_p^m \to \mathbb{F}_p$ be a function such that, for every line L, f agrees with a univariate degree-d polynomial on L. Then for every $S \subseteq \mathbb{F}_p^m$, there exists a degree-$d$ function g such that $g|_S = f|_S$.*

There are multiple known proofs of Theorem 2.2. To demonstrate the challenges in the non-signaling setting, we first outline a standard classical proof of Theorem 2.2 that will not generalize to the non-signaling setting. The proof uses the following lemma.

Lemma 2.2. *Suppose that $f \colon \mathbb{F}_p^m \to \mathbb{F}_p$ is a polynomial where $\deg(f) = d < p$. Then there exists a line L in $\mathbb{F}_p^m$ such that $f|_L$ is a univariate polynomial of degree exactly d.*

The above lemma is shown by considering the function $f(a+tb)$ where $a, b \in \mathbb{F}_p^m$, and arguing that the coefficient of t^d in $f(a+tb)$ is a non-zero polynomial in the variables a, b, and hence does not vanish for all $a, b \in \mathbb{F}_p^m$. Thus, there exists a line $L(t) = a + tb$ for which the coefficient of t^d is non-zero, and therefore $f|_L$ has degree exactly d.

With the above lemma, one can prove Theorem 2.2 as follows. Suppose that f is a polynomial of degree $d' > d$. Then by the lemma there exists a line L such that $f|_L$ has degree $d' > d$, contradicting the fact that $f|_L$ has degree at most d.[3]

The above proof is a good example of a proof that will *not* extend to the non-signaling setting. This is because the proof of Lemma 2.2 is "global", in the sense that arguing about the polynomial coefficients of f requires "knowing" $f(a)$ for $\Omega(d^m)$ points $a \in \mathbb{F}_p^m$, as this is the minimum number of evaluations of f needed for all polynomial coefficients of f to be fixed. As indicated by Section 1.2, the types of classical proofs that will extend to the non-signaling setting are those with small locality, i.e. ones that require looking at $f(a)$ for only a small number of a at a time. This implies that the above proof will not work for k-non-signaling functions when $k \leq O(d^m)$, i.e., when k is a non-trivial value.

We instead present the following local proof of Theorem 2.2 for individual degree. Since this proof has small locality it will extend to the non-signaling setting, and thus imply Theorem 2.1.

Let $S \subseteq \mathbb{F}_p^m$. We wish to show that $f|_S = g|_S$ for some $g \in \mathbf{C}$. Let $S_0 = \emptyset$, and for each $i \in [m]$ define $S_i \subseteq \mathbb{F}_p^i$ to be the projection of S to the first i coordinates, so S_i is the set of all $(a_1, \ldots, a_i) \in \mathbb{F}_p^i$ such that $(a_1, \ldots, a_i, b_{i+1}, \ldots, b_m) \in S$ for some $(b_{i+1}, \ldots, b_m) \in \mathbb{F}_p^{m-i}$. Note that $S_m = S$.

We prove by induction that for every $i \in [m]$ and every $b_{i+1}, \ldots, b_m \in \mathbb{F}_p$ there exists an individual degree-d polynomial $g_i \colon \mathbb{F}_p^i \to \mathbb{F}_p$ such that $f|_{S_i \times \{(b_{i+1}, \ldots, b_m)\}} = g_i|_{S_i}$. This proves Theorem 2.2 for individual degree, as $S_m = S$. The base case ($i = 1$) holds since f looks degree-d on every line, so in particular f is degree-d on the line $\mathbb{F}_p \times \{(b_2, \ldots, b_m)\}$, which contains S_1.

We now argue the induction step. Suppose that the induction hypothesis holds for $i-1$ and every $b_i, \ldots, b_m \in \mathbb{F}_p$. The induction hypothesis implies that for each $j \in \{0, \ldots, d\}$, there exists an individual degree-d polynomial $g_{i-1}^{(j)} \colon \mathbb{F}_p^{i-1} \to \mathbb{F}_p$ such that $g_{i-1}^{(j)}|_{S_{i-1}} = f|_{S_{i-1} \times \{j\} \times \{(b_{i+1}, \ldots, b_m)\}}$. Let $g_i \colon \mathbb{F}_p^i \to \mathbb{F}_p$ be defined by interpolating the $g_{i-1}^{(j)}$'s along the i-th axis, i.e. $g_i(x_1, \ldots, x_i) := \sum_{j=0}^{d} \delta_j(x_i) \cdot g_{i-1}^{(j)}(x_1, \ldots, x_{i-1})$ where $\delta_j(y)$ is the unique degree-d univariate polynomial that is 1 if $y = j$ and 0 otherwise. We then argue that f agrees with g_i on $S_{i-1} \times \mathbb{F}_p$. This is because f agrees with g_i on $S_{i-1} \times \{0, \ldots, d\}$ (since here $g_i = g_{i-1}^{(j)} = f$ by the induction hypothesis), and therefore agrees with g_i on $S_{i-1} \times \mathbb{F}_p$ by polynomial interpolation, since f looks degree-d on any axis-parallel line along the i-th axis.

The above proof can be adapted to an $|S|\,(d+2)$-local proof (as stated above, it is $|S|\,p$-local). We thus conclude that if a k-non-signaling function $\mathcal{F}$ looks degree-d on every line L, then it also looks individual degree-d on every S where $|S|\,(d+2) \leq k$, i.e., $\mathcal{F}$ (viewed as a $\lfloor k/(d+2) \rfloor$-non-signaling function) is $\mathbf{C}$-explainable.

In the aforementioned argument, we have crucially required that $\mathcal{F}$ looks low-degree along *every* line, rather than merely on sets of evenly-spaced points, which are the only constraints checked by the test. Thus, we must show that if $\mathcal{F}$ passes the degree-d evenly-spaced points test with probability 1, then $\mathcal{F}$ looks degree-d on arbitrary subsets of any line. This last step can be viewed

[3]The argument as stated does not quite work, as the lemma only holds when $d' < p$. Here, we ignore this technicality to simplify the presentation of the argument.

as the following. Let T be the set of constraints checked by the evenly-spaced test, and let T' be the set of all low-weight line constraints (weight at most $k-d-2$) satisfied by degree-d polynomials. We show that $T \vdash_k T'$, so $\mathcal{F}$ (by Theorem 3) must also satisfy all constraints in T', and thus looks degree-d on arbitrary subsets of lines, which concludes the proof of Theorem 2.1.

Remark 2.3 (total degree vs. individual degree)**.** Theorem 1 only concludes that $\mathcal{F}$ has an individual degree-d structure, when one might expect to conclude that it has a *total* degree-d structure, as it passes the evenly-spaced points test along *random lines*. Indeed, this is the conclusion in the classical setting. The difficulty in establishing such a result comes from Lemma 2.2. The classical analysis of the low-degree test proceeds by induction, initially concluding that $f\colon \mathbb{F}^m \to \mathbb{F}$ is a polynomial that is degree-d in x_m and total degree-d in all the other variables, and hence is a total degree-$2d$ polynomial. Then, by using Lemma 2.2 one concludes that in fact f has total degree-d, not $2d$. A non-signaling analogue of Lemma 2.2 would allow us to conclude a total degree-d structure. However, as explained earlier the classical proof of Lemma 2.2 is not local, so it does not lift to a non-signaling one. Exploring whether or not the gap between total and individual degree is necessary in the non-signaling setting is thus an intriguing open question.

The robust case. We now explain how to adapt the above proof to the robust case. Our goal now is to show that $\mathcal{F}$ is close to a $\mathbf{C}$-explainable non-signaling function, where the distance between two k-non-signaling functions $\mathcal{F}$ and $\mathcal{G}$ is defined as

$$\Delta_k(\mathcal{F},\mathcal{G}) = \max_{S\subseteq[n],|S|\leq k} \Delta_{\mathrm{TV}}(\mathcal{F}_S,\mathcal{G}_S) \ ,$$

where Δ_{TV} is the total variation distance between distributions [CMS18]. As in the case of linearity testing in [CMS18], this is impossible, as the definition of distance requires that $\mathcal{F}$ be close to $\mathbf{C}$-explainable on all sets $S \subseteq \mathbb{F}_p^m$ with $|S| \leq k$. In particular, if $\mathcal{F}$ looks low-degree on all lines but one, then $\mathcal{F}$ will be very far from $\mathbf{C}$-explainable. Following [CMS18], we instead show that an appropriately defined self-correction of $\mathcal{F}$, denoted by $\hat{\mathcal{F}}$, is close to $\mathbf{C}$-explainable. Informally, $\hat{\mathcal{F}}(x)$ is defined by querying $\mathcal{F}$ on a random evenly-spaced line L passing through x, and then setting $\hat{\mathcal{F}}(x)$ to be the value at x obtained by locally decoding $\mathcal{F}$ along L. $\hat{\mathcal{F}}$ is a $\hat{k}$-non-signaling function, where $\hat{k} = k/(d+1)$.

We prove Theorem 1 via the following four steps:

1. Average to worst case reduction: we show that if $\mathcal{F}$ passes the evenly-spaced points test with high probability, then $\hat{\mathcal{F}}$ looks low-degree on *every* set of evenly-spaced set of points contained in a line L with high probability.
2. From evenly-spaced points to arbitrary subsets of a line: we show that if $\hat{\mathcal{F}}$ looks low-degree on every set of evenly-spaced set of points contained in a line, then $\hat{\mathcal{F}}$ looks low-degree on every subset of every line L.
3. Robust local characterization: we show that T, the set of constraints where the support of the constraint is contained in some line L, is a robust local characterization of $\mathbf{C}$, i.e. that if $\hat{\mathcal{F}}$ satisfies every $\alpha \in T$ with high probability, then $\hat{\mathcal{F}}$ satisfies every $\alpha \in \mathbf{C}^{\perp}_{\leq k'}$ with high probability, where $k' = \hat{k}/(d+2)$.
4. Finishing the proof: we show that if $\hat{\mathcal{F}}$ satisfies every $\alpha \in \mathbf{C}^{\perp}_{\leq k'}$ with high probability, then $\hat{\mathcal{F}}$ is close to a $\mathbf{C}$-explainable non-signaling function.

We have already discussed the proofs of the second and third steps in the zero error case. In the robust case, the main difference is that we now pay some small error in union bounds every time we use the fact that $\hat{\mathcal{F}}$ looks low-degree along an evenly-spaced line.

The first step follows from our non-trivial definition of $\hat{\mathcal{F}}$. Naively, one might define $\hat{\mathcal{F}}$ for each x by locally decoding its value from $\mathcal{F}$ along a random evenly-spaced line containing x. This does not work. Instead, we decode its value along the line $L_x(t) = x + iw_x$, where the slopes (the w_x's) are correlated so that $w_{x+y} = w_x + w_y - w_{0^n}$. These correlations, combined with the fact that $L(t)$ looks random for each x, allows us to show that $\hat{\mathcal{F}}$ looks low-degree on every evenly-spaced set of points.

The final step follows abstractly from the more general statements we show for all linear codes, and relies on our characterization of the Fourier spectrum of $\mathbf{C}$-explainable non-signaling functions.

2.3 Local characterizations and linear proofs

We outline the proof of Theorem 3; we assume familiarity with the notions introduced in Section 1.2. We begin by formally defining local characterizations.

Local characterizations. We say that a k-non-signaling function $\mathcal{F}$ is $\mathbf{C}$-*explainable* if for every $S \subseteq [n]_{\leq k}$, with probability 1 the function $f\colon S \to \mathbb{F}$ sampled from $\mathcal{F}_S$ is in $\mathbf{C}|_S$. (See Section 2.5 for a discussion of this definition.) Recall from Definition 1.4 that a subset $T \subseteq \mathbf{C}^\perp$ is an ℓ-local characterization of $(\mathbf{C}, k)$ if every $\alpha \in T$ has $\mathrm{wt}(\alpha) \leq \ell$ and the set of k-non-signaling functions $\mathcal{F}$ where $\Pr[\langle \alpha, \mathcal{F}\rangle = 0] = 1$ for every $\alpha \in T$ equals the set of $\mathbf{C}$-explainable k-non-signaling functions.

Outline of the proof. The proof of Theorem 3 has two directions: completeness and soundness. For soundness, we show that if $T \vdash_k \alpha$, then for any k-non-signaling function $\mathcal{F}$ where $\langle \gamma, \mathcal{F}\rangle = 0$ holds with probability 1 for every $\gamma \in T$, it also holds that $\langle \alpha, \mathcal{F}\rangle = 0$ with probability 1. Intuitively, this means that any k-non-signaling function satisfying every constraint in T must satisfy α as well, and therefore shows that our definition of "proof" makes sense. The proof of this direction is straightforward, and can be found in Section 7.1.

To show completeness, we explicitly construct a k-non-signaling function $\mathcal{F}$ that satisfies every constraint α where $T \vdash_k \alpha$ with probability 1, and satisfies every other constraint α with probability $\frac{1}{|\mathbb{F}|}$. Our construction of $\mathcal{F}$ makes crucial use of the notion of a *local subspace* that we introduce.

Definition 2.4. *A* k**-local subspace** $\mathcal{V}$ *is a subset of* $\mathbb{F}^n_{\leq k}$ *that looks like a subspace when restricted to local views of size at most* k*, i.e.,* $\mathcal{V}_{\subseteq S}$ *is a linear subspace in* $\mathbb{F}^n$ *for every* $S \subseteq [n]_{\leq k}$*.*

We show that for any k-local subspace $\mathcal{V}$ there is a k-non-signaling function $\mathcal{F}$ where $\Pr[\langle \alpha, \mathcal{F}\rangle = 0] = 1$ for every $\alpha \in \mathcal{V}$ and $\Pr[\langle \alpha, \mathcal{F}\rangle = 0] = \frac{1}{|\mathbb{F}|}$ otherwise. We then show that the set of α's provable from T, which is $\{\alpha \in \mathbb{F}^n_{\leq k} : T \vdash_k \alpha\}$, is a k-local subspace. This latter step is straightforward, and the proof is in Section 7.3. We now discuss the first step, which is non-trivial.

Non-signaling functions from local subspaces. Given a k-local subspace $\mathcal{V}$, we argue that there is a k-non-signaling function $\mathcal{F}$ where $\Pr[\langle \alpha, \mathcal{F}\rangle = 0] = 1$ for every $\alpha \in \mathcal{V}$, and $\Pr[\langle \alpha, \mathcal{F}\rangle = 0] = \frac{1}{|\mathbb{F}|}$ for every $\alpha \notin \mathcal{V}$. We construct $\mathcal{F} = \{\mathcal{F}_S\}_{S \subseteq [n] : |S| \leq k}$ by specifying its local distributions $\mathcal{F}_S$.

A distribution over functions $f\colon S \to \mathbb{F}$ is a function that maps each f to a non-negative real number such that the total sum is 1. With this viewpoint, we first define $\mathcal{F}_S$ as a *function* that maps each $f\colon S \to \mathbb{F}$ to a complex number. Then, we show that the total sum is 1 and that each f is mapped to a non-negative real number, so that the function $\mathcal{F}_S$ is indeed a distribution.

We define the function $\mathcal{F}_S\colon \mathbb{F}^S \to \mathbb{C}$ by specifying its Fourier coefficients:

$$\widehat{\mathcal{F}_S}(\alpha) := \begin{cases} \frac{1}{q^{|S|}} & \text{if } \alpha \in \mathcal{V} \\ 0 & \text{if } \alpha \notin \mathcal{V} \end{cases},$$

These "local" Fourier coefficients should *not* be confused with the Fourier coefficients of $\mathcal{F}$ that are defined in Section 2.1. In fact, at this point the non-signaling function $\mathcal{F}$ is not yet defined.

This completely specifies $\mathcal{F}_S$ as a function $\mathbb{F}^S \to \mathbb{C}$. We show that since $\mathcal{V}$ is a k-local subspace, $\mathcal{F}_S$ is in fact a distribution. First, $\sum_{f \in \mathbb{F}^S} \mathcal{F}_S(f) = 1$ because $\widehat{\mathcal{F}_S}(0^S) = 1/q^{|S|}$ since $\mathcal{V}$ is a k-local subspace, and thus must contain 0^n. Hence, it suffices to show that $\mathcal{F}_S(f) \in \mathbb{R}_{\geq 0}$ for each $f \in \mathbb{F}^S$. For each $f \in \mathbb{F}^S$ we have

$$\mathcal{F}_S(f) = \sum_{\alpha \in \mathbb{F}^S} \widehat{\mathcal{F}_S}(\alpha)\chi_\alpha(f) = \sum_{\alpha \in \mathcal{V}_{\subseteq S}} \widehat{\mathcal{F}_S}(\alpha)\chi_\alpha(f) \ ,$$

since we have defined $\mathcal{F}_S$ in this way using its Fourier coefficients. There are two cases: either $\langle \alpha, f \rangle = 0$ for every $\alpha \in \mathcal{V}_{\subseteq S}$, in which case the sum is $|\mathcal{V}_{\subseteq S}|/q^{|S|}$, or $\langle \alpha, f \rangle \neq 0$ for some $\alpha \in \mathcal{V}_{\subseteq S}$. In the latter case, we use the fact that $\mathcal{V}_{\subseteq S}$ is a linear subspace to show that the sum is 0. In either case, we conclude that $\mathcal{F}_S(f)$ is a non-negative real number, and therefore that $\mathcal{F}_S$ is a distribution. We note that in particular, $\mathcal{F}_S$ is the uniform distribution over all $f \colon S \to \mathbb{F}$ where $\langle \alpha, f \rangle = 0$ for all $\alpha \in \mathcal{V}_{\subseteq S}$.

Next, we argue that the collection of local distributions $\{\mathcal{F}_S\}_{S \subseteq [n]_{\leq k}}$ is indeed non-signaling. This follows from a lemma that we prove that shows that a collection of local distributions is non-signaling if and only if the Fourier coefficients of the local distributions (after removing the normalization factors) are the same. Thus the k-non-signaling function $\mathcal{F}$ is well-defined.

Finally, we show that $\mathcal{F}$ satisfies the desired properties. This follows from our definition of each $\mathcal{F}_S$, as the construction implies that the Fourier coefficients of $\mathcal{F}$ satisfy:

$$\widehat{\mathcal{F}}(\alpha) := \begin{cases} \frac{1}{q^n} & \text{if } \alpha \in \mathcal{V} \\ 0 & \text{if } \alpha \notin \mathcal{V} \end{cases} \ .$$

This corresponds to having $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = 1$ for every $\alpha \in \mathcal{V}$, and the random variable $\langle \alpha, \mathcal{F} \rangle$ having the uniform distribution when $\alpha \notin \mathcal{V}$, which completes the proof.

2.4 Low-degree testing fails for small locality

We discuss how to prove Theorem 2. We let $\mathbf{C}$ denote the linear code of polynomials $f \colon \mathbb{F}^m \to \mathbb{F}$ of total degree-d, and let T be the set of all $\alpha \in \mathbf{C}^\perp$ whose support is contained in a line in $\mathbb{F}^m$. Note that for any low-degree test that only checks line constraints and has perfect completeness, if we let T_0 be the set of constraints checked by this test, we will have $T_0 \subseteq T$.

The main combinatorial quantity that we use in our proof is the *rank* of an element $\alpha \in \mathbf{C}^\perp$, defined as

$$\mathsf{rank}_T(\alpha) := \min_{T' \subseteq T : \alpha \in \mathrm{span}(T')} \left|T'\right| \ .$$

Note that $\mathsf{rank}_T(\alpha)$ is a non-negative integer, as $\mathrm{span}(T) = \mathbf{C}^\perp$.

We now sketch the proof of Theorem 2 in three steps.

(1) Interval Cut Lemma. We show a generic lemma about the relationship between rank and provability from T. Informally, we show that in order for T to prove α of rank at least r, T must also prove some β of "intermediate" rank. Formally, we show that if there is an interval $[r/2, r)$ such that every β with rank in this interval is *not* provable from T, then every α of rank at least r is also not provable from T. We prove this *Interval Cut Lemma* via the fact that rank_T

is subadditive, that is, $\mathsf{rank}_T(\alpha+\beta) \leq \mathsf{rank}_T(\alpha) + \mathsf{rank}_T(\beta)$. Subadditivity implies that for every interval $[r/2, r)$, in order to prove a constraint of rank $\geq r$ from constraints of rank $< r/2$ there must be an intermediate constraint β with rank in $[r/2, r)$ bridging the gap.

(2) Two combinatorial facts. We prove two combinatorial facts about the dual code of $\mathbf{C}$.

- There exists $\alpha^* \in \mathbf{C}^\perp$ where $\mathrm{wt}(\alpha) = 2d+2$ and $\mathrm{supp}(\alpha) \subseteq \{(a, a^2, 0^{m-2}) : a \in \mathbb{F}\} \subseteq \mathbb{F}^m$, i.e., $\mathrm{supp}(\alpha)$ is contained along the curve $x_1^2 - x_2 = 0$ embedded on the plane $x_3 = x_4 = \cdots = x_m = 0$ of $\mathbb{F}^m$.

  Proof sketch. If f is an m-variate polynomial of total degree d then $f(t, t^2, 0, \ldots, 0)$ is a polynomial of degree $\leq 2d$ in t. Thus, there is an element $\alpha^* \in \mathbf{C}^\perp$ supported on this curve of weight $2d+2$ that checks some linear constraint. This shows the existence of the desired α^*.

- For every $\beta \in \mathbf{C}^\perp$ with $\mathsf{rank}_T(\beta) \in \{(d+2)/4, \ldots, (d+2)/2\}$ it holds that $\mathrm{wt}(\beta) \geq \frac{3}{16}(d+2)^2$.

  Proof sketch. Any β of rank r is the sum of *exactly* r line constraints, where each constraint is on a *distinct* line. Each new constraint adds at least $d+2$ weight to β, ignoring the weight that is removed by cancellation. The amount of cancellation is at most the number of intersection points, which is not too large when r is in $\{(d+2)/4, \ldots, (d+2)/2\}$, thus implying that $\mathrm{wt}(\beta) \geq \frac{3}{16}(d+2)^2$.

(3) Completing the proof. Theorem 2 follows from the Interval Cut Lemma, the two combinatorial facts, and Theorem 3. Any $\beta \in \mathbf{C}^\perp$ with rank in $[(d+2)/4, (d+2)/2)$ has weight $\geq \frac{3}{16}(d+2)^2$, and thus is *not* provable when $k < \frac{3}{16}(d+2)^2$. Since α^* has weight $2d+2$ and is supported only on the diagonal, it has rank $\geq d+1$, as each line constraint increases the number of points on the curve by at most 2, by Bézout's theorem. The Interval Cut Lemma implies that α^* is also not provable. The non-signaling function constructed in the proof of Theorem 3 thus passes the random lines test with probability 1 yet satisfies α^* with probability only $1/|\mathbb{F}|$. But, α^* must be satisfied with probability 1 by any non-signaling function that is "locally low-degree", which completes the proof.

2.5 Fourier spectrum of non-signaling linear codes

We have so far adopted the definition that a k-non-signaling function $\mathcal{F}$ is "in" a linear code $\mathbf{C} \subseteq \mathbb{F}^n$ if a function $f\colon S \to \mathbb{F}$ sampled from $\mathcal{F}_S$ is in $\mathbf{C}|_S$ with probability 1 for every $S \subseteq [n]_{\leq k}$. Indeed, we use this "$\mathbf{C}$-explainability" to define the notion of a *local characterization* (see Definition 1.4).

We now provide thorough justification for this choice. We view the definitions and results below as a conceptual contribution that sheds light on basic properties of non-signaling functions.

In the classical setting, a function $f\colon [n] \to \mathbb{F}$ "looks like" a codeword of $\mathbf{C}$ if, well, it equals some codeword in $\mathbf{C}$. The issue at hand is that, in the non-signaling setting, it is not immediately clear what it means for a non-signaling function $\mathcal{F}$ to be "in" $\mathbf{C}$ because $\mathcal{F}$ is a collection of local distributions. Below are two natural ways to capture this notion.

Definition 2.5 (informal)**.** *Let $\mathcal{F}\colon [n] \to \mathbb{F}$ be a k-non-signaling function.*

- *We say that $\mathcal{F}$ is* **C-supported** *if it is equivalent to a k-local quasi-distribution $\mathcal{Q}\colon \mathbb{F}^n \to \mathbb{C}$ that is supported on $\mathbf{C}$, namely, $\mathcal{Q}(f) = 0$ for all $f \notin \mathbf{C}$.*[4]

- *We say that $\mathcal{F}$ is* **C-explainable** *if, for all $S \subseteq [n]_{\leq k}$, the distribution $\mathcal{F}_S$ is supported on $\mathbf{C}|_S$. In other words, the output of $\mathcal{F}$ is always consistent with the restriction of some codeword in $\mathbf{C}$.*

[4] When $\mathbf{C}$ is the Hadamard code, this definition equals the notion of a *linear* non-signaling function from [CMS18].

The first definition is motivated by our Equivalence Lemma (Lemma 2.1), and imposes a "global" property on the non-signaling function. The second definition, implied by the first one, instead takes a "local" approach, imposing consistency with relevant restrictions of the code.

In the following lemma, we quantify the difference between the notions of "**C**-supported" and "**C**-explainable" by characterizing the Fourier spectrum in each case. For convenience, we denote by $\mathbf{C}^{\perp}_{\leq k}$ the set $\{\alpha \in \mathbf{C}^{\perp} : \mathrm{wt}(\alpha) \leq k\}$, which are the constraints with at most k non-zero entries.

Lemma 2.6 (informal)**.** *Let $\mathcal{F}\colon [n] \to \mathbb{F}$ be a k-non-signaling function.*

- *$\mathcal{F}$ is **C**-supported $\leftrightarrow$ the Fourier coefficients $\{\widehat{\mathcal{F}}(\alpha)\}_{\alpha \in \mathbb{F}^n_{\leq k}}$ are constant on each coset of $\mathbf{C}^{\perp}$.*
- *$\mathcal{F}$ is **C**-explainable $\leftrightarrow$ the Fourier coefficient $\widehat{\mathcal{F}}(\alpha)$ equals $\frac{1}{q^n}$ for every $\alpha \in \mathbf{C}^{\perp}_{\leq k}$.*

We additionally prove that the foregoing structure is robust to errors: $\mathcal{F}$ is close to being **C**-supported if and only if its Fourier coefficients are almost constant on every coset of $\mathbf{C}^{\perp}$; moreover $\mathcal{F}$ is close to being **C**-explainable if and only if $\widehat{\mathcal{F}}(\alpha)$ is close to $\frac{1}{q^n}$ for every $\alpha \in \mathbf{C}^{\perp}_{\leq k}$.

One may interpret Lemma 2.6 as "bad news" because it shows that the notions of "**C**-supported" and "**C**-explainable" are in fact *distinct*. Which one is the correct one to use? From the perspective of local testability, we may regard "**C**-supported" as more desirable, because it requires a global structure to hold. We prove that, fortunately, the two notions are equivalent up to a small change in parameters, reinforcing our belief that we have identified the right notions.

Lemma 2.7 (informal)**.** *Let $\mathcal{F}\colon [n] \to \mathbb{F}$ be a k-non-signaling function.*

- *If $\mathcal{F}$ is **C**-supported, then $\mathcal{F}$ is **C**-explainable.*
- *If $\mathcal{F}$ is **C**-explainable, then $\mathcal{F}$ (viewed as a $k/2$-non-signaling function) is **C**-supported.*

In light of the above, it suffices to study non-signaling functions that are **C**-explainable. We have used this notion in our results on local characterizations (see Definition 1.4), as it is more natural in this setting: the set of **C**-explainable k-non-signaling functions are precisely those that are consistent with the set of constraints $\mathbf{C}^{\perp}_{\leq k}$.

Detailed definitions and proofs can be found in Section 5. Below we provide proof sketches for Lemmas 2.6 and 2.7. The Fourier structure of non-signaling functions, discussed in Section 2.1, underlies all of these proofs.

2.5.1 Fourier spectrum of a C-supported function

We outline the proof of the first item of Lemma 2.6. A k-non-signaling function $\mathcal{F}$ that is **C**-supported is by definition equivalent to a quasi-distribution $\mathcal{Q}$ supported on **C**. We explain why all such non-signaling functions have Fourier coefficients that are constant on cosets of $\mathbf{C}^{\perp}$, that is, $\widehat{\mathcal{F}}(\alpha) = \widehat{\mathcal{F}}(\alpha')$ for every $\alpha, \alpha' \in \mathbb{F}^n_{\leq k}$ with $\alpha - \alpha' \in \mathbf{C}^{\perp}$. We compare the following two affine spaces:

$$V_1 = \left\{ \mathcal{Q}\colon \mathbb{F}^n \to \mathbb{C} \text{ s.t. } \sum_{f \in \mathbf{C}} \mathcal{Q}(f) = 1 \text{ and } \mathcal{Q}(f) = 0 \ \forall f \notin \mathbf{C} \right\} ,$$

$$V_2 = \left\{ \mathcal{Q}\colon \mathbb{F}^n \to \mathbb{C} \text{ s.t. } \widehat{\mathcal{Q}}(0^n) = \frac{1}{q^n} \text{ and } \widehat{\mathcal{Q}}(\alpha) = \widehat{\mathcal{Q}}(\alpha + \gamma) \ \forall \alpha \in \mathbb{F}^n, \gamma \in \mathbf{C}^{\perp} \right\} .$$

The affine space V_1 corresponds to quasi-distributions that are supported on $\mathbf{C}$, while V_2 corresponds to quasi-distributions whose Fourier coefficients satisfy the desired characterization. It suffices to prove that $V_1 = V_2$. First we show that $\dim(V_1) = \dim(V_2)$, and then that $V_1 \subseteq V_2$.

The dimension of V_1 is $|\mathbf{C}|-1$ because the $|\mathbf{C}|$ free terms are subject to a single linear constraint. The dimension of V_2 is $q^n/\left|\mathbf{C}^\perp\right|-1$ because the Fourier coefficients are constant on each coset of $\mathbf{C}^\perp$, and on each coset they may have an arbitrary value; the one exception is the coset $\mathbf{C}^\perp$, where the Fourier coefficients must be $\frac{1}{q^n}$. Recalling that $q^n = |\mathbf{C}| \cdot \left|\mathbf{C}^\perp\right|$, we deduce that $\dim(V_1) = \dim(V_2)$.

Next we show that $V_1 \subseteq V_2$. For any $\mathcal{Q} \in V_1$ and $\alpha \in \mathbb{F}^n$ we have by definition

$$\widehat{\mathcal{Q}}(\alpha) := \frac{1}{q^n} \cdot \sum_{f \in \mathbb{F}^n} \mathcal{Q}(f) \cdot \omega^{-\mathrm{Tr}(\langle \alpha, f \rangle)} \ .$$

Since $\mathcal{Q} \in V_1$, any function f in the support of $\mathcal{Q}$ must be in $\mathbf{C}$. Therefore, for any $\gamma \in \mathbf{C}^\perp$ have $\langle \gamma, f \rangle = 0$, so that $\omega^{\mathrm{Tr}(\langle \gamma, f \rangle)} = \omega^{\mathrm{Tr}(0)} = 1$. This implies that $\widehat{\mathcal{Q}}(\alpha) = \widehat{\mathcal{Q}}(\alpha + \gamma)$. Intuitively, when we shift α by γ the sum remains unchanged because each term in the sum is multiplied by $\omega^{-\mathrm{Tr}(\langle \gamma, f \rangle)} = 1$. Thus $V_1 \subseteq V_2$. Since $\dim(V_1) = \dim(V_2)$ and $V_1 \subseteq V_2$, we conclude that $V_1 = V_2$.

2.5.2 Fourier spectrum of a C-explainable function

We outline the proof of the second item of Lemma 2.6. The characterization of $\mathbf{C}$-explainable functions relies on the fact that the Fourier coefficient $\widehat{\mathcal{F}}(\alpha)$ is related to the distribution of the random variable $\langle \alpha, \mathcal{F} \rangle$, i.e., the distribution $(\Pr[\langle \alpha, \mathcal{F} \rangle = b])_{b \in \mathbb{F}}$. This intuition can be quantified via (a generalization of) the DFT matrix $M \in \mathbb{C}^{q \times q}$, which is the matrix defined as $M_{a,b} := \omega^{-\mathrm{Tr}(ab)}$ (entries are indexed by $\mathbb{F}$); M is invertible and $\frac{1}{\sqrt{q}} M$ is unitary.

Recall that the Fourier coefficients of $\mathcal{F}$ are defined as follows:

$$\forall\, \alpha \in \mathbb{F}^n_{\leq k} \quad \widehat{\mathcal{F}}(\alpha) := \frac{1}{q^n} \sum_{j \in \mathbb{F}_p} \omega^{-j} \Pr[\mathrm{Tr}(\langle \alpha, \mathcal{F} \rangle) = j] \ .$$

Letting $v := (\Pr[\langle \alpha, \mathcal{F} \rangle = b])_{b \in \mathbb{F}}$, expanding the definitions shows that $Mv = (q^n \widehat{\mathcal{F}}(a\alpha))_{a \in \mathbb{F}}$. The linear transformation M thus quantifies the relation between the distribution $(\Pr[\langle \alpha, \mathcal{F} \rangle = b])_{b \in \mathbb{F}}$ and the Fourier coefficients $(q^n \widehat{\mathcal{F}}(a\alpha))_{a \in \mathbb{F}}$.

Now, given a k-non-signaling function $\mathcal{F}$, we first show that $\mathcal{F}$ is $\mathbf{C}$-explainable if and only if $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = 1$ for every $\alpha \in \mathbf{C}^\perp_{\leq k}$. This follows from the fact that any local function $g \colon S \to \mathbb{F}$ that satisfies every $\alpha \in \mathbf{C}^\perp_{\subseteq S}$ can be extended into a codeword $f \in \mathbf{C}$. Using the matrix M, we can relate the condition that $\mathcal{F}$ satisfies every $\alpha \in \mathbf{C}^\perp_{\leq k}$ with probability 1 to its Fourier spectrum. Specifically, we have that $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = 1$ if and only if $(q^n \widehat{\mathcal{F}}(a\alpha))_{a \in \mathbb{F}} = M(1, 0, \ldots, 0)^\top$. Since $M(1, 0, \ldots, 0)^\top = (1, \ldots, 1)^\top$, we get that $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = 1$ if and only if $\widehat{\mathcal{F}}(a\alpha) = \frac{1}{q^n}$ for every $a \in \mathbb{F}$, completing the proof.

2.5.3 The relationship between C-supported and C-explainable

We outline the proof of Lemma 2.7. First note that Lemma 2.6 immediately implies that a $\mathbf{C}$-supported k-non-signaling function $\mathcal{F}$ is $\mathbf{C}$-explainable, because if $\mathcal{F}$ is $\mathbf{C}$-supported then $\widehat{\mathcal{F}}(\alpha) = \widehat{\mathcal{F}}(0^n) = \frac{1}{q^n}$ for every $\alpha \in \mathbf{C}^\perp_{\leq k}$, implying that $\mathcal{F}$ is $\mathbf{C}$-explainable.

Conversely, if $\mathcal{F}$ is $\mathbf{C}$-explainable, then for any $\alpha, \alpha' \in \mathbb{F}^n_{\leq k/2}$ with $\alpha - \alpha' \in \mathbf{C}^\perp$ we get that for any $b \in \mathbb{F}$,

$$\Pr[\langle \alpha, \mathcal{F} \rangle = b] = \Pr[\langle \alpha', \mathcal{F} \rangle + \langle \alpha - \alpha', \mathcal{F} \rangle = b] = \Pr[\langle \alpha', \mathcal{F} \rangle = b] \ ,$$

since $\Pr[\langle \alpha - \alpha', \mathcal{F} \rangle = 0] = 1$ as $\alpha - \alpha' \in \mathbf{C}^\perp$ and $\mathcal{F}$ is $\mathbf{C}$-explainable. This shows that the vectors $(\Pr[\langle \alpha, \mathcal{F} \rangle = b])_{b \in \mathbb{F}}$ and $(\Pr[\langle \alpha', \mathcal{F} \rangle = b])_{b \in \mathbb{F}}$ are equal, which implies that the Fourier coefficients $\widehat{\mathcal{F}}(\alpha)$ and $\widehat{\mathcal{F}}(\alpha')$ are equal. By Lemma 2.6, this completes the proof. Note that we crucially need $\mathrm{wt}(\alpha), \mathrm{wt}(\alpha') \leq k/2$ so that $\mathrm{wt}(\alpha - \alpha') \leq k$, as otherwise $\Pr[\langle \alpha - \alpha', \mathcal{F} \rangle = 0]$ is undefined.

3 Preliminaries

Throughout this paper we let $n \in \mathbb{N}$ be an arbitrary positive integer, and $k \in \mathbb{N}$ a positive integer that is at most n. We use $\mathbb{F}$ to denote the finite field of size q with characteristic p, and $\mathbb{F}_p$ to denote the prime subfield of $\mathbb{F}$. We often abuse notation and identify a function $f\colon [n] \to \mathbb{F}$ with its evaluation table in $\mathbb{F}^n$. For a vector $\alpha \in \mathbb{F}^n$ we let $\mathrm{supp}(\alpha) := \{i \in [n] : \alpha_i \neq 0\}$, and we let $\mathrm{wt}(\alpha) := |\mathrm{supp}(\alpha)|$. For a set of vectors $R \subseteq \mathbb{F}^n$, we let $R_{\leq \ell} \subseteq R$ denote the subset $\{\alpha \in R : \mathrm{wt}(\alpha) \leq \ell\}$. In particular, $\mathbb{F}^n_{\leq k}$ contains all vectors $\alpha \in \mathbb{F}^n$ of weight at most k. For a subset $S \subseteq [n]$, we let $R_{\subseteq S} = \{\alpha \in R : \mathrm{supp}(\alpha) \subseteq S\}$; we also write $S \subseteq [n]_{\leq \ell}$ if $|S| \leq \ell$.

3.1 Non-signaling functions

We define *non-signaling functions* and *quasi-distributions*, and introduce useful notation for them. The definitions are almost identical to those in [CMS18], but extended to any finite field.

Definition 3.1. *A k-non-signaling function $\mathcal{F}\colon [n] \to \mathbb{F}$ is a collection $\mathcal{F} = \{\mathcal{F}_S\}_{S \subseteq [n]_{\leq k}}$ where (i) each $\mathcal{F}_S$ is a distribution over functions $f\colon S \to \mathbb{F}$, and (ii) for every two subsets S and R each of size at most k, the restrictions of $\mathcal{F}_S$ and $\mathcal{F}_R$ to $S \cap R$ are equal as distributions.*

Note that any function $f\colon [n] \to \mathbb{F}$ induces a n-non-signaling function by setting $\mathcal{F}_S$ to be the distribution that outputs $f|_S$ with probability 1. More generally, any distribution $\mathcal{D}$ over functions $f\colon [n] \to \mathbb{F}$ induces a corresponding n-non-signaling function by defining $\mathcal{F}_S$ to be the distribution that samples $f \leftarrow \mathcal{D}$ and outputs $f|_S$.

Given a set $S \subseteq [n]_{\leq k}$ and function $g \in \mathbb{F}^S$, we define

$$\Pr\left[\, \mathcal{F}(S) = g \,\right] := \Pr[\, g \leftarrow \mathcal{F}_S \,] \ .$$

The non-signaling property in this notation is the following: for every two subsets $S, R \subseteq [n]_{\leq k}$ and every string $g \in \mathbb{F}^{S \cap T}$, $\Pr[\, \mathcal{F}(S)|_{S\cap T} = g \,] = \Pr[\, \mathcal{F}(T)|_{S\cap T} = g \,]$, where the probability is over the randomness of $\mathcal{F}$.

We extend the above notation to every $E \subseteq \mathbb{F}^S$ in the natural way by defining $\Pr\left[\, \mathcal{F}(S) \in E \,\right] := \Pr_{f \leftarrow \mathcal{F}_S}\left[\, f \in E \,\right]$. We highlight the case when E is an "inner product event", as we will encounter this case frequently.

Definition 3.2. *Let $\mathcal{F}\colon [n] \to \mathbb{F}$ be a k-non-signaling function. For $\alpha \in \mathbb{F}^n_{\leq k}$ and $b \in \mathbb{F}$, we define*

$$\Pr[\langle \alpha, \mathcal{F} \rangle = b] := \Pr_{f \leftarrow \mathcal{F}_{\mathrm{supp}(\alpha)}}\left[\sum_{i \in \mathrm{supp}(\alpha)} \alpha_i f(i) = b \right] \ .$$

Similarly, we define $\Pr[\mathrm{Tr}(\langle \alpha, \mathcal{F} \rangle) = j] := \sum_{b \in \mathbb{F}\colon \mathrm{Tr}(b) = j} \Pr[\langle \alpha, \mathcal{F} \rangle = b]$ for every $j \in \mathbb{F}_p$.

The probability above is well-defined since $\mathrm{wt}(\alpha) \leq k$, and so we query $\mathcal{F}$ on at most k points.

Since $\mathcal{F}$ is non-signaling, $\Pr[\langle \alpha, \mathcal{F} \rangle = b] = \Pr_{f \leftarrow \mathcal{F}_S}[\sum_{i \in S} \alpha_i f(i)]$ for any set $S \supseteq \mathrm{supp}(\alpha)$. The intuition behind the above definition is that the inner product $\langle \alpha, g \rangle$ for any $g\colon [n] \to \mathbb{F}$ can be computed only given $g|_{\mathrm{supp}(\alpha)}$, namely, given g restricted to a set of size at most k.

3.2 Quasi-distributions

A quasi-distribution extends the notion of a probability distribution by allowing probabilities to be complex, and is the main tool that we use to analyze non-signaling functions.

Definition 3.3.
- *A* **quasi-distribution** *is a function* $\mathcal{Q}\colon \mathbb{F}^n \to \mathbb{C}$ *where* $\sum_{f\in\mathbb{F}^n} \mathcal{Q}(f) = 1$.
- *For a set of functions* $R \subseteq \mathbb{F}^n$*, we say that* $\mathcal{Q}$ *is* **supported** *on* R *if* $\{f \in \mathbb{F}^n : \mathcal{Q}(f) \neq 0\} \subseteq R$.
- *For a positive integer* ℓ*, we say that* $\mathcal{Q}$ *is* ℓ**-local** *if the marginals* $\mathcal{Q}|_S$ *for each* $S \subseteq [n]_{\leq \ell}$ *are distributions (*$\sum_{f\in\mathbb{F}^n : f|_S = g} \mathcal{Q}(f)$ *is a non-negative real number for each* $S \subseteq [n]_{\leq \ell}$ *and* $g\colon S \to \mathbb{F}$*).*

If $\mathcal{Q}$ is ℓ-local, then for every subset $S \subseteq [n]_{\leq \ell}$, we may view $\mathcal{Q}|_S$ as a probability distribution over $\mathbb{F}^S$. If $\mathcal{Q}$ is ℓ-local then it is s-local for every $s \in \{0, 1, \ldots, \ell\}$.

Definition 3.4. *Given a quasi-distribution* $\mathcal{Q}$*, a subset* $S \subseteq [n]$*, and* $g \in \mathbb{F}^S$*, we define the* **quasi-probability** *of the event "*$\mathcal{Q}(S) = g$*" to be the following complex number*

$$\widetilde{\Pr}[\mathcal{Q}(S) = g\,] := \sum_{f\in\mathbb{F}^n : f|_S = g} \mathcal{Q}(f)\ .$$

(The tilde above Pr *denotes that quasi-probabilities are not necessarily non-negative real numbers.)*

Given a subset $E \subseteq \mathbb{F}^S$, we similarly define $\widetilde{\Pr}[\mathcal{Q}(S) \in E\,] := \sum_{f\in\mathbb{F}^n : f|_S \in E} \mathcal{Q}(f)$.

As for non-signaling functions, we highlight the case when E is an inner product event.

Definition 3.5. *Let* $\mathcal{Q}\colon \mathbb{F}^n \to \mathbb{C}$ *be a quasi-distribution. For* $\alpha \in \mathbb{F}^n$ *and* $b \in \mathbb{F}$*, we define*

$$\widetilde{\Pr}[\langle \alpha, \mathcal{Q}\rangle = b] := \sum_{f\in\mathbb{F}^n : \langle \alpha, f\rangle = b} \mathcal{Q}(f)\ .$$

Similarly, we define $\widetilde{\Pr}[\mathrm{Tr}(\langle \alpha, \mathcal{Q}\rangle) = j] := \sum_{b\in\mathbb{F}:\mathrm{Tr}(b)=j} \widetilde{\Pr}[\langle \alpha, \mathcal{Q}\rangle = b]$ *for every* $j \in \mathbb{F}_p$.

Definition 3.6 (statistical distance)**.** *Given a finite domain* $[n]$ *and an integer* $\ell \in \{1, \ldots, |D|\}$*, the* Δ_ℓ**-distance** *between two quasi-distributions* $\mathcal{Q}$ *and* $\mathcal{Q}'$ *is*

$$\Delta_\ell(\mathcal{Q}, \mathcal{Q}') := \max_{S\subseteq[n]_{\leq \ell}} \Delta(\mathcal{Q}|_S, \mathcal{Q}'|_S)\ ,$$

where $\Delta(\mathcal{Q}|_S, \mathcal{Q}'|_S) := \max_{E\subseteq\mathbb{F}^S} \left|\widetilde{\Pr}[\mathcal{Q}(S) \in E] - \widetilde{\Pr}[\mathcal{Q}'(S) \in E]\right|$.

We say that $\mathcal{Q}$ *and* $\mathcal{Q}'$ *are* ε**-close in the** Δ_ℓ**-distance** *if* $\Delta_\ell(\mathcal{Q}, \mathcal{Q}') \leq \varepsilon$*; else, they are* ε*-far.*

Remark 3.7 (distance for non-signaling functions)**.** The definition of Δ_ℓ-distance naturally extends to defining distances between k-non-signaling functions, as well as between quasi-distributions and k-non-signaling functions, provided that $\ell \leq k$.

The notion above generalizes the standard notion of statistical (total variation) distance: if $\mathcal{Q}$ and $\mathcal{Q}'$ are *distributions* then their Δ_n-distance equals their statistical distance. Also note that if $\mathcal{Q}$ and $\mathcal{Q}'$ are ℓ-local quasi-distributions then their Δ_ℓ-distance equals the maximum statistical distance, across all subsets $S \subseteq [n]$ with $|S| \leq \ell$, between the two *distributions* $\mathcal{Q}|_S$ and $\mathcal{Q}'|_S$ — in particular this means that any experiment that queries exactly one set of size at most ℓ cannot distinguish between the two quasi-distributions with probability greater than $\Delta_\ell(\mathcal{Q}, \mathcal{Q}')$.

We stress that $\Delta_\ell(\mathcal{Q}, \mathcal{Q}') = 0$ does *not* necessarily mean that $\mathcal{Q} = \mathcal{Q}'$! In fact, it is possible to have $\Delta_\ell(\mathcal{Q}, \mathcal{Q}') = 0$ while $\sum_{f\in U} |\mathcal{Q}(f) - \mathcal{Q}'(f)|$ is arbitrarily large. We also remark that the Δ_ℓ-distance is not necessarily upper bounded by 1, and is in general unbounded.

Definition 3.8 (approximate locality)**.** *Given a finite domain $[n]$, an integer $\ell \in \{1, \ldots, n\}$, and a real number $\varepsilon \geq 0$, a quasi-distribution $\mathcal{Q}$ over U_n is* (ℓ, ε)**-local** *if, for every subset $S \subseteq [n]_{\leq \ell}$ and every event $E \subseteq \{0,1\}^S$,*

$$\min_{x \in [0,1]} \{ \left| \widetilde{\Pr}[\mathcal{Q}(S) \in E] - x \right| \} \in [0, \varepsilon] \ .$$

Approximate locality generalizes the notion of (exact) locality as in Definition 3.3. Below, we state a lemma that if$\mathcal{Q}$ is (ℓ, ε)-local and is supported over a linear code $\mathbf{C}$, then there is an ℓ-local $\mathcal{Q}'$ over $\mathbf{C}$ that is close to $\mathcal{Q}$. The proof idea is similar to that of "smoothening" almost-feasible solutions to Sherali–Adams relaxations into feasible ones [RS09].

Lemma 3.9. *If $\mathcal{Q}$ is a (ℓ, ε)-local quasi-distribution over $\mathbf{C}$, then there is an ℓ-local quasi-distribution $\mathcal{Q}'$ over $\mathbf{C}$ such that $\Delta_\ell(\mathcal{Q}, \mathcal{Q}') < q^\ell \varepsilon$.*

We omit the proof of Lemma 3.9 as it is identical to the proof of lemma 7.8 in [CMS18], just replacing the field $\mathbb{F}_2 = \{0, 1\}$ with a general field $\mathbb{F}$.

4 Fourier analysis of non-signaling functions

We prove statements about the Fourier structure of non-signaling functions, and prove the Equivalence Lemma. In Section 4.1 we recall basic facts about Fourier analysis of functions over finite fields. In Section 4.2 we relate Fourier coefficients to probabilities and quasi-probabilities. In Section 4.3 we prove that non-signaling functions and quasi-distributions are equivalent notions.

4.1 Fourier analysis of functions over finite fields

We consider functions of the type $F\colon \mathbb{F}^n \to \mathbb{C}$. For two such functions F_1 and F_2, we define their inner product as $\langle F_1, F_2 \rangle := \frac{1}{q^n} \sum_{x \in \mathbb{F}^n} \overline{F_1(x)} F_2(x)$. For every $\alpha \in \mathbb{F}^n$, we define the *character* $\chi_\alpha\colon \mathbb{F}^n \to \mathbb{C}$ as $\chi_\alpha(x) := \omega^{\mathrm{Tr}(\langle \alpha, x \rangle)}$ where: (1) $\mathrm{Tr}\colon \mathbb{F} \to \mathbb{F}_p$ is the trace map; (2) $\langle \alpha, x \rangle$ is the inner product $\sum_{i=1}^n \alpha_i x_i$; (3) $\omega = e^{2\pi i/p}$ is a primitive complex p-th root of unity; and (4) ω^j is defined by thinking of $j \in \mathbb{F}_p$ as an integer in $\mathbb{Z}$. The functions $\{\chi_\alpha\}_{\alpha \in \mathbb{F}^n}$ form an orthonormal basis of the space of all functions $f\colon \mathbb{F}^n \to \mathbb{C}$, so every function $F\colon \mathbb{F}^n \to \mathbb{C}$ can be written as

$$F(\cdot) = \sum_{\alpha \in \mathbb{F}^n} \widehat{F}(\alpha) \chi_\alpha(\cdot) \ , \text{ where } \widehat{F}(\alpha) := \langle \chi_\alpha, F \rangle \ .$$

The values $\{\widehat{F}(\alpha)\}_{\alpha \in \mathbb{F}^n}$ are the *Fourier coefficients* of F. We recall and prove a few useful identifies.

Parseval's identity. For every two functions $F, G\colon \mathbb{F}^n \to \mathbb{C}$,

$$\langle F, G \rangle = \frac{1}{q^n} \sum_{x \in \mathbb{F}^n} \overline{F(x)} G(x) = \sum_{\alpha \in \mathbb{F}^n} \overline{\widehat{F}(\alpha)} \widehat{G}(\alpha) \ .$$

Proof.

$$\begin{aligned} \frac{1}{q^n} \sum_{x \in \mathbb{F}^n} \overline{F(x)} G(x) &= \frac{1}{q^n} \sum_{x \in \mathbb{F}^n} \overline{\left(\sum_{\alpha \in \mathbb{F}^n} \widehat{F}(\alpha) \chi_\alpha(x) \right)} \left(\sum_{\beta \in \mathbb{F}^n} \widehat{G}(\beta) \chi_\beta(x) \right) \\ &= \sum_{\alpha \in \mathbb{F}^n} \sum_{\beta \in \mathbb{F}^n} \overline{\widehat{F}(\alpha)} \widehat{G}(\beta) \langle \chi_\alpha, \chi_\beta \rangle = \sum_{\alpha \in \mathbb{F}^n} \overline{\widehat{F}(\alpha)} \widehat{G}(\alpha) \ , \end{aligned}$$

since $\{\chi_\alpha\}_{\alpha \in \mathbb{F}^n}$ are orthonormal. □

Plancherel's identity. As a corollary of the above,

$$\frac{1}{q^n} \sum_{x \in \mathbb{F}^n} |F(x)|^2 = \sum_{\alpha \in \mathbb{F}^n} |\widehat{F}(\alpha)|^2 \ .$$

The case of indicator functions. When analyzing non-signaling functions and quasi-distributions we will apply the above identities in the case where F is an indicator function $\mathbf{1}_E$ for a set $E \subseteq \mathbb{F}^n$. In this case, by Plancherel's identity we have that $|E|/q^n = \sum_{\alpha \in \mathbb{F}^n} |\widehat{\mathbf{1}_E}(\alpha)|^2$. In particular, by the Cauchy–Schwarz inequality, this implies that

$$\|\widehat{\mathbf{1}_E}\|_1 = \sum_{\alpha \in \mathbb{F}^n} |\widehat{\mathbf{1}_E}(\alpha)| \leq \sqrt{\sum_{\alpha \in \mathbb{F}^n} |\widehat{\mathbf{1}_E}(\alpha)|^2} \cdot \sqrt{\sum_{\alpha \in \mathbb{F}^n} 1} \leq \sqrt{|E|/q^n} \cdot q^{n/2} = \sqrt{|E|} \ .$$

If we let $F(x) = \mathbf{1}_E(x)$, then Parseval's identity becomes the following lemma.

Lemma 4.1. *Let $G\colon \mathbb{F}^n \to \mathbb{C}$ be a function and $E \subseteq \mathbb{F}^n$. Then*

$$\frac{1}{q^n}\sum_{x\in E} G(x) = \frac{1}{q^n}\sum_{\alpha\in\mathbb{F}^n} \overline{\widehat{\mathbf{1}_E}(\alpha)} \sum_{x\in\mathbb{F}^n} G(x)\omega^{-\mathrm{Tr}(\langle\alpha,x\rangle)} = \sum_{\alpha\in\mathbb{F}^n} \overline{\widehat{\mathbf{1}_E}(\alpha)}\widehat{G}(\alpha) \ .$$

4.2 Relating the Fourier spectrum to the probabilities of events

A quasi-distribution $\mathcal{Q}$ is a function $\mathcal{Q}\colon \mathbb{F}^n \to \mathbb{C}$ that maps a function $f\colon [n] \to \mathbb{F}$ (identified with the corresponding vector $\mathbb{F}^n$) to $\mathcal{Q}(f)$. We can write $\mathcal{Q}(\cdot) = \sum_{\alpha\in\mathbb{F}^n} \widehat{\mathcal{Q}}(\alpha)\chi_\alpha(\cdot)$, where $\{\chi_\alpha\}_{\alpha\in\mathbb{F}^n}$ are the characters and $\{\widehat{\mathcal{Q}}(\alpha)\}_{\alpha\in\mathbb{F}^n}$ are $\mathcal{Q}$'s Fourier coefficients. For $S \subseteq [n]$ and $\alpha \in \mathbb{F}^S$, we abuse notation and use $\widehat{\mathcal{Q}}(\alpha)$ to refer to $\widehat{\mathcal{Q}}(\beta)$ where $\beta \in \mathbb{F}^n$ has $\beta_i = \alpha_i$ for all $i \in S$ and 0 otherwise.

The lemma below relates the inner product quasi-probabilities defined in Definition 3.5 to the Fourier coefficients of $\mathcal{Q}$.

Lemma 4.2. *Let $\mathcal{Q}\colon \mathbb{F}^n \to \mathbb{C}$ be a quasi-distribution. For every $\alpha \in \mathbb{F}^n$,*

$$\widehat{\mathcal{Q}}(\alpha) = \frac{1}{q^n}\sum_{j\in\mathbb{F}_p} \omega^{-j}\widetilde{\Pr}[\mathrm{Tr}(\langle\alpha,\mathcal{Q}\rangle) = j] \ .$$

Proof of Lemma 4.2. By definition,

$$\begin{aligned}\widehat{\mathcal{Q}}(\alpha) = \langle\chi_\alpha,\mathcal{Q}(\cdot)\rangle &= \frac{1}{q^n}\sum_f \overline{\chi_\alpha(f)}\mathcal{Q}(f) = \frac{1}{q^n}\sum_{j\in\mathbb{F}_p}\omega^{-j}\sum_{f:\chi_\alpha(f)=\omega^j}\mathcal{Q}(f)\\ &= \frac{1}{q^n}\sum_{j\in\mathbb{F}_p}\omega^{-j}\sum_{f:\mathrm{Tr}(\langle\alpha,f\rangle)=j}\mathcal{Q}(f) = \frac{1}{q^n}\sum_{j\in\mathbb{F}_p}\omega^{-j}\widetilde{\Pr}[\mathrm{Tr}(\langle\alpha,\mathcal{Q}\rangle) = j] \ . \qquad \square\end{aligned}$$

The above lemma implies that the Fourier coefficients $(\widehat{\mathcal{Q}}(a\alpha))_{a\in\mathbb{F}}$ are determined by the quasi-probabilities $(\Pr[\langle\alpha,\mathcal{Q}\rangle = b])_{b\in\mathbb{F}}$, as the quasi-probabilities $(\Pr[\langle\alpha,\mathcal{Q}\rangle = b])_{b\in\mathbb{F}}$ determine the quasi-probabilities $(\Pr[\langle a\alpha,\mathcal{Q}\rangle = b])_{b\in\mathbb{F}}$ for every $a \in \mathbb{F}$. In fact, there is a linear transformation M that maps $(\Pr[\langle\alpha,\mathcal{Q}\rangle = b])_{b\in\mathbb{F}}$ to $(q^n\widehat{\mathcal{Q}}(a\alpha))_{a\in\mathbb{F}}$. Below, we state a well-known lemma about M.

Lemma 4.3. *Let $M \in \mathbb{C}^{q\times q}$ be the matrix defined as $M_{a,b} := \omega^{-\mathrm{Tr}(ab)}$ (entries are indexed by $\mathbb{F}$). Then M is invertible and $\frac{1}{\sqrt{q}}M$ is unitary (namely, $M^\dagger \cdot M = qI$). In particular, for every vector $(v_b)_{b\in\mathbb{F}}$ with values in $\mathbb{C}$, the map $(v_b)_{b\in\mathbb{F}} \mapsto (\sum_{b\in\mathbb{F}}\omega^{-\mathrm{Tr}(ab)}v_b)_{a\in\mathbb{F}}$ is a bijection.*

We additionally prove the following lemma, which relates the Fourier spectrum of the quasi-distribution $\mathcal{Q}|_S$ to the Fourier spectrum of $\mathcal{Q}$.

Lemma 4.4. *Let $\mathcal{Q}\colon \mathbb{F}^n \to \mathbb{C}$ be a quasi-distribution. Let $S \subseteq [n]$, and let $\mathcal{Q}|_S$ denote the restriction of $\mathcal{Q}$ to S, namely, $\mathcal{Q}|_S$ is the quasi-distribution from $\mathbb{F}^S$ to $\mathbb{C}$ where $\mathcal{Q}|_S(g) := \sum_{f:f|_S=g}\mathcal{Q}(f)$. Then for every $\alpha \in \mathbb{F}^S$ it holds that $q^{|S|}\widehat{\mathcal{Q}|_S}(\alpha) = q^n\widehat{\mathcal{Q}}(\alpha)$.*[5]

Proof of Lemma 4.4.

$$q^{|S|}\widehat{\mathcal{Q}|_S}(\alpha) = \sum_{g\in\mathbb{F}^S}\mathcal{Q}|_S(g)\omega^{-\mathrm{Tr}(\langle\alpha,g\rangle)} = \sum_{f\in\mathbb{F}^n}\mathcal{Q}(f)\omega^{-\mathrm{Tr}(\langle\alpha,f\rangle)} = q^n\widehat{\mathcal{Q}}(\alpha) \ . \qquad \square$$

[5]The vector α in $\widehat{\mathcal{Q}}(\alpha)$ is treated as a element in $\mathbb{F}^n$ with $\alpha_j = 0$ for all $j \notin S$

If $\mathcal{F}\colon [n] \to \mathbb{F}$ is a k-non-signaling function, then for any $\alpha \in \mathbb{F}^n_{\le k}$ and $b \in \mathbb{F}$ we have defined $\Pr[\langle \alpha, \mathcal{F}\rangle = b]$ in Definition 3.2 to be $\Pr_{f \leftarrow \mathcal{F}_{\mathrm{supp}(\alpha)}}[\langle \alpha, f\rangle = b]$. Note that the probability is well-defined since $\mathrm{wt}(\alpha) \le k$ (so we query $\mathcal{F}$ on at most k points). Also note that Lemma 4.2 implies that, for every $\alpha \in \mathbb{F}^n_{\le k}$, we can define the Fourier coefficient $\widehat{\mathcal{F}}(\alpha)$ of $\mathcal{F}$ as

$$\widehat{\mathcal{F}}(\alpha) := \frac{1}{q^n} \sum_{j \in \mathbb{F}_p} \omega^{-j} \Pr[\mathrm{Tr}(\langle \alpha, \mathcal{F}\rangle) = j] \ .$$

With the above definitions, we can prove the following two corollaries of Lemma 4.1. The first is for non-signaling functions, and the second is for quasi-distributions.

Corollary 4.5. *For any k-non-signaling function $\mathcal{F}\colon [n] \to \mathbb{F}$, set $S \subseteq [n]$, and event $E \subseteq \mathbb{F}^S$,*

$$\Pr[\mathcal{F}(S) \in E] = \sum_{\alpha \in \mathbb{F}^S} \overline{\widehat{\mathbf{1}_E}(\alpha)} \sum_{j \in \mathbb{F}_p} \omega^{-j} \Pr[\mathrm{Tr}(\langle \alpha, \mathcal{F}\rangle) = j] = q^n \sum_{\alpha \in \mathbb{F}^S} \overline{\widehat{\mathbf{1}_E}(\alpha)} \widehat{\mathcal{F}}(\alpha) \ .$$

Proof. Apply Lemma 4.1 with $G\colon \mathbb{F}^S \to \mathbb{C}$ defined as $G(x) := \Pr[\mathcal{F}_S(i) = x_i \ \forall i \in S]$. □

Corollary 4.6. *For any quasi-distribution $\mathcal{Q}\colon \mathbb{F}^n \to \mathbb{C}$, set $S \subseteq [n]$, and event $E \subseteq \mathbb{F}^S$,*

$$\widetilde{\Pr}[\mathcal{Q}(S) \in E] = \sum_{f\colon f(S) \in E} \mathcal{Q}(f) = \sum_{\alpha \in \mathbb{F}^S} \overline{\widehat{\mathbf{1}_E}(\alpha)} \sum_{j \in \mathbb{F}_p} \omega^{-j} \widetilde{\Pr}[\mathrm{Tr}(\langle \alpha, \mathcal{Q}\rangle) = j] = q^n \sum_{\alpha \in \mathbb{F}^S} \overline{\widehat{\mathbf{1}_E}(\alpha)} \widehat{\mathcal{Q}}(\alpha) \ .$$

Proof. Apply Lemma 4.1 to the function $G\colon \mathbb{F}^S \to \mathbb{C}$ that is the quasi-distribution $\mathcal{Q}|_S$. Then observe that for every $\alpha \in \mathbb{F}^S$, $q^{|S|}\widehat{\mathcal{Q}|_S}(\alpha) = q^n \widehat{\mathcal{Q}}(\alpha)$ by Lemma 4.4. □

The above two lemmas allow us to bound the distance between a k-non-signaling function $\mathcal{F}$ and a quasi-distribution $\mathcal{Q}$ in terms of their Fourier spectra.

Lemma 4.7. *Let $\mathcal{F}\colon [n] \to \mathbb{F}$ be a k-non-signaling function and $\mathcal{Q}\colon \mathbb{F}^n \to \mathbb{C}$ a quasi-distribution. For any set $S \subseteq [n]_{\le k}$ and event $E \subseteq \mathbb{F}^S$,*

$$\left|\Pr[\mathcal{F}(S) \in E] - \widetilde{\Pr}[\mathcal{Q}(S) \in E]\right| \le q^n \sum_{\alpha \in \mathbb{F}^S} \left|\widehat{\mathbf{1}_E}(\alpha)\right| \left|\widehat{\mathcal{F}}(\alpha) - \widehat{\mathcal{Q}}(\alpha)\right| \ .$$

In particular, $\Delta_k(\mathcal{Q}, \mathcal{F}) \le q^{n+k/2} \max_{\alpha \in \mathbb{F}^n_{\le k}} |\widehat{\mathcal{F}}(\alpha) - \widehat{\mathcal{Q}}(\alpha)|$.

Corollary 4.8. *Let $\mathcal{F}\colon [n] \to \mathbb{F}$ be a k-non-signaling function and $\mathcal{Q}\colon \mathbb{F}^n \to \mathbb{C}$ a quasi-distribution. Then $\Delta_k(\mathcal{Q}, \mathcal{F}) = 0$ if and only if $\widehat{\mathcal{F}}(\alpha) = \widehat{\mathcal{Q}}(\alpha)$ for every $\alpha \in \mathbb{F}^n_{\le k}$.*

Proof of Lemma 4.7. The first equation follows immediately from Corollary 4.5 and Corollary 4.6. For the second part of the lemma,

$$\begin{aligned}
\Delta_k(\mathcal{Q}, \mathcal{F}) &\le \max_{S \subseteq [n]_{\le k}} \max_{E \subseteq \mathbb{F}^S} q^n \sum_{\alpha \in \mathbb{F}^S} \left|\overline{\widehat{\mathbf{1}_E}(\alpha)}\right| \left|\widehat{\mathcal{F}}(\alpha) - \widehat{\mathcal{Q}}(\alpha)\right| \\
&\le q^n \max_{S \subseteq [n]_{\le k}} \left(\left(\max_{E \subseteq \mathbb{F}^S} \sum_{\alpha \in \mathbb{F}^S} \left|\widehat{\mathbf{1}_E}(\alpha)\right| \right) \max_{\alpha \in \mathbb{F}^S} \left|\widehat{\mathcal{F}}(\alpha) - \widehat{\mathcal{Q}}(\alpha)\right| \right) \\
&\le q^n \left(\max_{S \subseteq [n]_{\le k}} q^{|S|/2} \right) \max_{\alpha \in \mathbb{F}^n_{\le k}} \left|\widehat{\mathcal{F}}(\alpha) - \widehat{\mathcal{Q}}(\alpha)\right| \\
&\le q^{n+k/2} \max_{\alpha \in \mathbb{F}^n_{\le k}} \left|\widehat{\mathcal{F}}(\alpha) - \widehat{\mathcal{Q}}(\alpha)\right| \ .
\end{aligned}$$

□

Proof of Corollary 4.8. If $\widehat{\mathcal{F}}(\alpha) = \widehat{\mathcal{Q}}(\alpha)$ for every $\alpha \in \mathbb{F}^n_{\leq k}$, then by Lemma 4.7 it follows that $\Delta_k(\mathcal{Q}, \mathcal{F}) = 0$. Conversely, if $\Delta_k(\mathcal{Q}, \mathcal{F}) = 0$, then for every $\alpha \in \mathbb{F}^n_{\leq k}$ and $j \in \mathbb{F}_p$ it holds that $\widetilde{\Pr}[\mathrm{Tr}(\langle \alpha, \mathcal{Q} \rangle) = j] = \Pr[\mathrm{Tr}(\langle \alpha, \mathcal{F} \rangle) = j]$, as these are both events. This implies that $q^n \widehat{\mathcal{Q}}(\alpha) = \sum_{j \in \mathbb{F}_p} \omega^{-j} \widetilde{\Pr}[\mathrm{Tr}(\langle \alpha, \mathcal{Q} \rangle) = j] = \sum_{j \in \mathbb{F}_p} \omega^{-j} \Pr[\mathrm{Tr}(\langle \alpha, \mathcal{F} \rangle) = j] = q^n \widehat{\mathcal{F}}(\alpha)$. □

Suppose that we are given a collection of local distributions $(\mathcal{F}_S)_{S \subseteq [n]_{\leq k}}$, namely, $\mathcal{F}_S$ is a distribution over functions $f \colon S \to \mathbb{F}$. We can think of each local distribution $\mathcal{F}_S$ as a function $\mathcal{F}_S \colon \mathbb{F}^S \to \mathbb{C}$, and in this way define for each local distribution $\mathcal{F}_S$ the Fourier coefficients $\widehat{\mathcal{F}_S}(\alpha)$ for each $\alpha \in \mathbb{F}^n_{\subseteq S}$. In the following lemma, we characterize when $(\mathcal{F}_S)_{S \subseteq [n]_{\leq k}}$ is k-non-signaling in terms of the Fourier spectra of the local distributions.

Lemma 4.9. *Let $(\mathcal{F}_S)_{S \subseteq [n]_{\leq k}}$ be a collection of local distributions. Then $(\mathcal{F}_S)_{S \subseteq [n]_{\leq k}}$ is a k-non-signaling function if and only if $q^{|S|} \widehat{\mathcal{F}_S}(\alpha) = q^{|R|} \widehat{\mathcal{F}_R}(\alpha)$ for every $S \subseteq [n]_{\leq k}$, $R \subseteq S$, and $\alpha \in \mathbb{F}^n_{\subseteq R}$.*

Proof. Suppose $(\mathcal{F}_S)_{S \subseteq [n]_{\leq k}}$ is a k-non-signaling function. Fix $S \subseteq [n]_{\leq k}$, $R \subseteq S$, and $\alpha \in \mathbb{F}^n_{\subseteq R}$. Since the collection of local distributions is k-non-signaling we have that $\mathcal{F}_S|_R = \mathcal{F}_R$. Therefore by Lemma 4.4 we have that $q^{|S|} \widehat{\mathcal{F}_S}(\alpha) = q^{|R|} \widehat{\mathcal{F}_R}(\alpha)$.

Now, fix $S \subseteq [n]_{\leq k}$ and $R \subseteq S$. Applying Corollary 4.6 to the distributions $\mathcal{F}_S$ and $\mathcal{F}_R$, we see that if $q^{|S|} \widehat{\mathcal{F}_S}(\alpha) = q^{|R|} \widehat{\mathcal{F}_R}(\alpha)$ for every $\alpha \in \mathbb{F}^n_{\subseteq R}$, then $\mathcal{F}_S|_R \equiv \mathcal{F}_R$. Hence, $(\mathcal{F}_S)_{S \subseteq [n]_{\leq k}}$ is k-non-signaling. □

4.3 Equivalence between non-signaling functions and quasi-distributions

We show that k-non-signaling functions and k-local quasi-distributions are equivalent. Every k-local quasi-distribution $\mathcal{Q}$ induces a k-non-signaling function $\mathcal{F}$ (Proposition 4.10). Conversely, every k-non-signaling function $\mathcal{F}$ can be described by a k-local quasi-distribution $\mathcal{Q}$ (Proposition 4.11). In fact, the set of such quasi-distributions is an affine subspace of co-dimension $\sum_{i=0}^{k} \binom{n}{i} \cdot (q-1)^i$ in $\mathbb{C}^{q^n}$. The first direction of the equivalence is elementary; the other direction is the interesting one.

The aforementioned result is a special case of a result of Abramsky and Brandenburger [AB11] that establishes an equivalence between *non-signaling empirical models* (a general notion of non-signaling experiments in the language of sheaf theory) and quasi-distributions over *global sections*. Our result strengthens this equivalence by giving an explicit characterization of the affine subspace of quasi-distributions describing a non-signaling function, by leveraging Fourier-analytic tools. This also extends to any finite field $\mathbb{F}$ the equivalence lemma for $\mathbb{F}_2$ presented in [CMS18].[6]

Proposition 4.10. *For every k-local quasi-distribution $\mathcal{Q}$ over functions $f \colon [n] \to \mathbb{F}$ there exists a k-non-signaling function $\mathcal{F} \colon [n] \to \mathbb{F}$ such that $\Delta_k(\mathcal{Q}, \mathcal{F}) = 0$.*

Proof. For every subset $S \subseteq [n]_{\leq k}$, define $\mathcal{F}_S$ to be the distribution over functions $f \colon S \to \mathbb{F}$ where $\Pr[\mathcal{F}_S \text{ outputs } f] := \widetilde{\Pr}[\mathcal{Q}(S) = f(S)]$, namely, such that $\mathcal{F}_S \equiv \mathcal{Q}|_S$. Note that $\mathcal{F}_S$ is a distribution because $\mathcal{Q}$ is k-local, so the relevant probabilities are in $[0, 1]$ and sum to 1. The definition immediately implies that $\Pr[\mathcal{F}(S) = g] = \widetilde{\Pr}[\mathcal{Q}(S) = g]$ for every string $g \in \mathbb{F}^S$, and so $\Delta_k(\mathcal{Q}, \mathcal{F}) = 0$. We are left to argue that $\mathcal{F} = \{\mathcal{F}_S\}_{S \subseteq [n]_{\leq k}}$ is k-non-signaling. Let $S \subseteq [n]_{\leq k}$,

[6] The characterization further extends to functions taking values in any finite alphabet Σ (not necessarily a field) by adding an abelian group structure to Σ (for example, by identifying Σ with $\mathbb{Z}/\,|\Sigma|\,\mathbb{Z}$), and then using analogous tools from Fourier analysis over finite abelian groups.

and let $R \subseteq S$. By definition of $\mathcal{F}$ and Lemma 4.4 we have that for every $\alpha \in \mathbb{F}^R$, $q^{|S|}\widehat{\mathcal{F}_S}(\alpha) = q^{|S|}\widehat{\mathcal{Q}|_S}(\alpha) = q^{|R|}\widehat{\mathcal{Q}|_R}(\alpha) = q^{|R|}\widehat{\mathcal{F}_R}(\alpha)$. By Lemma 4.9, it follows that $\mathcal{F}$ is k-non-signaling. □

Proposition 4.11. *For every k-non-signaling function $\mathcal{F}\colon [n] \to \mathbb{F}$, there exists a k-local quasi-distribution $\mathcal{Q}$ over functions $f\colon [n] \to \mathbb{F}$ such that $\Delta_k(\mathcal{F}, \mathcal{Q}) = 0$. Moreover, the set of such $\mathcal{Q}$'s (viewed as vectors in $\mathbb{C}^{q^n}$) is the affine subspace of co-dimension $\sum_{i=0}^{k} \binom{n}{i} \cdot (q-1)^i$ in $\mathbb{C}^{q^n}$ given by $\mathcal{Q}_0 + \operatorname{span}\{\chi_\alpha : \alpha \in \mathbb{F}^n, \operatorname{wt}(\alpha) > k\}$, where $\mathcal{Q}_0$ is any solution.*

Proof. Let $\mathcal{Q}$ be a quasi-distribution over functions $f\colon [n] \to \mathbb{F}$. By Corollary 4.8, it holds that $\Delta_k(\mathcal{Q}, \mathcal{F}) = 0$ if and only if $\widehat{\mathcal{F}}(\alpha) = \widehat{\mathcal{Q}}(\alpha)$ for all $\alpha \in \mathbb{F}^n_{\leq k}$.

Let $\mathcal{Q}_0$ be the quasi-distribution with Fourier coefficients $\widehat{\mathcal{Q}}(\alpha) := \widehat{\mathcal{F}}(\alpha)$ for all α of weight at most k and $\widehat{\mathcal{Q}}(\alpha) := 0$ otherwise. Consider the affine subspace $\mathcal{Q}_0 + \operatorname{span}\{\chi_\alpha : \alpha \in \mathbb{F}^n, \operatorname{wt}(\alpha) > k\}$. By Corollary 4.8, every quasi-distribution $\mathcal{Q}$ in the affine subspace satisfies $\Delta_k(\mathcal{Q}, \mathcal{F}) = 0$. We note that this affine subspace has dimension $\sum_{i=0}^{k} \binom{n}{i} \cdot (q-1)^i$.

Conversely, suppose that $\mathcal{Q}$ satisfies $\Delta_k(\mathcal{Q}, \mathcal{F}) = 0$. Then by Corollary 4.8 it holds that $\widehat{\mathcal{Q}}(\alpha) = \widehat{\mathcal{F}}(\alpha)$ for all $\alpha \in \mathbb{F}^n_{\leq k}$, which implies that $\mathcal{Q}$ is in the aforementioned affine subspace. Hence, the affine subspace contains all $\mathcal{Q}$ such that $\Delta_k(\mathcal{Q}, \mathcal{F}) = 0$. □

5 Non-signaling linear codes

We wish to define what it means for a non-signaling function $\mathcal{F}\colon [n] \to \mathbb{F}$ to be "in" a linear code $\mathbf{C} \subseteq \mathbb{F}^n$. We introduce two natural definitions for the above goal. The first definition is motivated by the equivalence between non-signaling functions and quasi-distributions established in Section 4.3. The second definition is motivated by a notion of local consistency.

For each of the two definitions, we *characterize* the Fourier spectrum of non-signaling strategies that satisfy the definition, in the exact and in the robust case. Also, we prove a strong relationship between the two definitions, showing that they are *equivalent* (up to a small loss in parameters). The compelling structure that we uncover supports our choice of definitions.

For this section, we remind the reader that a linear code $\mathbf{C}$ over $\mathbb{F}$ with block length n is a linear subspace of $\mathbb{F}^n$. We equivalently also view $\mathbf{C}$ as a linear subspace of the set of all functions $f\colon [n] \to \mathbb{F}$. The dual code of $\mathbf{C}$ is the linear subspace $\mathbf{C}^\perp := \{\alpha : \langle \alpha, f \rangle = 0 \ \forall f \in \mathbf{C}\} \subseteq \mathbb{F}^n$.

5.1 Quasi-distributions supported on linear codes

The equivalence between non-signaling functions and quasi-distributions in Section 4.3 suggests a natural way to capture when a non-signaling function is "in" a given linear code.

Definition 5.1. *Given a k-non-signaling strategy $\mathcal{F}\colon [n] \to \mathbb{F}$, code $\mathbf{C} \subseteq \mathbb{F}^n$ and parameter $k' \leq k$, we say that $\mathcal{F}$ is* $(\mathbf{C}, k')$**-supported** *if there exists a k'-local quasi-distribution $\mathcal{Q}\colon \mathbb{F}^n \to \mathbb{C}$ supported on $\mathbf{C}$ such that $\Delta_{k'}(\mathcal{Q}, \mathcal{F}) = 0$.*

In light of the characterization of the Fourier spectra of quasi-distributions equivalent to a given non-signaling function in Section 4.3, it is natural to ask if the Fourier spectrum of a quasi-distribution supported on $\mathbf{C}$ has a special structure. In the following lemma, we characterize the Fourier spectrum of quasi-distributions supported on a given linear code $\mathbf{C}$. Informally, we show that the condition "Fourier coefficients are constants on cosets of $\mathbf{C}^\perp$" is necessary and sufficient.

Lemma 5.2. *Let $\mathbf{C} \subseteq \mathbb{F}^n$ be a linear code. A quasi-distribution $\mathcal{Q}\colon \mathbb{F}^n \to \mathbb{C}$ is supported on $\mathbf{C}$ if and only if $\widehat{\mathcal{Q}}(\alpha) = \widehat{\mathcal{Q}}(\alpha')$ for all $\alpha, \alpha' \in \mathbb{F}^n$ such that $\alpha - \alpha' \in \mathbf{C}^\perp$.*

The foregoing statement immediately gives us a corollary about non-signaling functions.

Corollary 5.3. *A k-non-signaling strategy $\mathcal{F}\colon [n] \to \mathbb{F}$ is $(\mathbf{C}, k')$-supported if and only if for all $\alpha, \alpha' \in \mathbb{F}^n_{\leq k}$ such that $\alpha - \alpha' \in \mathbf{C}^\perp$ it holds that $\widehat{\mathcal{F}}(\alpha) = \widehat{\mathcal{F}}(\alpha')$.*

Next, we wish to study the Fourier spectrum of a quasi-distribution $\mathcal{Q}$ that is merely *close* to being supported on $\mathbf{C}$. For this case, we give the following "robust" version of Lemma 5.2.

Lemma 5.4. *Let $\mathbf{C} \subseteq \mathbb{F}^n$ be a linear code, and let $\mathcal{Q}$ be a quasi-distribution.*

- *Suppose that there exists a quasi-distribution $\mathcal{Q}'$ supported on $\mathbf{C}$ such that $\Delta_k(\mathcal{Q}, \mathcal{Q}') \leq \delta$. Then for all $\alpha, \alpha' \in \mathbb{F}^n_{\leq k}$ and $\alpha - \alpha' \in \mathbf{C}^\perp$ it holds that $\left|\widehat{\mathcal{Q}}(\alpha) - \widehat{\mathcal{Q}}(\alpha')\right| \leq \frac{2\delta}{q^n}$.*
- *Conversely, suppose that for all $\alpha, \alpha' \in \mathbb{F}^n_{\leq k}$ and $\alpha - \alpha' \in \mathbf{C}^\perp$ it holds that $\left|\widehat{\mathcal{Q}}(\alpha) - \widehat{\mathcal{Q}}(\alpha')\right| \leq \frac{2\delta}{q^n}$. Then there exists a quasi-distribution $\mathcal{Q}'$ supported on $\mathbf{C}$ such that $\Delta_k(\mathcal{Q}, \mathcal{Q}') \leq q^{k/2} \cdot 2\delta$.*

We note that in Lemma 5.4, neither quasi-distribution is required to be local.

5.1.1 Proof of Lemma 5.2

Define the affine spaces

$$V_1 = \left\{ \mathcal{Q}\colon \mathbb{F}^n \to \mathbb{C} \text{ s.t. } \sum_{f \in \mathbf{C}} \mathcal{Q}(f) = 1 \text{ and } \mathcal{Q}(f) = 0 \ \forall f \notin \mathbf{C} \right\} \ ,$$
$$V_2 = \left\{ \mathcal{Q}\colon \mathbb{F}^n \to \mathbb{C} \text{ s.t. } \widehat{\mathcal{Q}}(0^n) = \frac{1}{q^n} \text{ and } \widehat{\mathcal{Q}}(\alpha) = \widehat{\mathcal{Q}}(\alpha+\gamma) \ \forall \alpha \in \mathbb{F}^n, \gamma \in \mathbf{C}^\perp \right\} \ .$$

It suffices to prove that $V_1 = V_2$. First we show that $\dim(V_1) = \dim(V_2)$. The dimension of V_1 is $|\mathbf{C}| - 1$ because the $|\mathbf{C}|$ free terms are subject to a single linear constraint. The dimension of V_2 is $q^n/\left|\mathbf{C}^\perp\right| - 1$ because the Fourier coefficients are constant on each coset of $\mathbf{C}^\perp$, and on each coset they can take on an arbitrary value; the one exception is the coset $\mathbf{C}^\perp$, on which the Fourier coefficients must be $\frac{1}{q^n}$. Recalling that $q^n = |\mathbf{C}| \cdot |\mathbf{C}^\perp|$, we deduce that $\dim(V_1) = \dim(V_2)$.

Next we show that $V_1 \subseteq V_2$. Fix $\mathcal{Q} \in V_1$. Since $\sum_f \mathcal{Q}(f) = 1$, we have $\widehat{\mathcal{Q}}(0^n) = \frac{1}{q^n} \sum_f \mathcal{Q}(f) \cdot \omega^0 = \frac{1}{q^n}$. Moreover, for any $\alpha \in \mathbb{F}^n$ and $\gamma \in \mathbf{C}^\perp$,

$$\widehat{\mathcal{Q}}(\alpha+\gamma) = \frac{1}{q^n} \cdot \sum_f \mathcal{Q}(f) \cdot \omega^{-\mathrm{Tr}(\langle \alpha+\gamma f \rangle)} = \frac{1}{q^n} \cdot \sum_f \mathcal{Q}(f) \cdot \omega^{-\mathrm{Tr}(\langle \alpha, f \rangle)} \cdot \omega^{-\mathrm{Tr}(\langle \gamma, f \rangle)} \ .$$

Since $\mathcal{Q} \in V_1$, if $\mathcal{Q}(f) \neq 0$ then $f \in \mathbf{C}$ and hence $\omega^{\mathrm{Tr}(\langle \gamma, f \rangle)} = \omega^{\mathrm{Tr}(0)} = 1$. Therefore,

$$\widehat{\mathcal{Q}}(\alpha+\gamma) = \frac{1}{q^n} \cdot \sum_f \mathcal{Q}(f) \cdot \omega^{-\mathrm{Tr}(\langle \alpha, f \rangle)} = \widehat{\mathcal{Q}}(\alpha) \ .$$

Thus $V_1 \subseteq V_2$. Since $\dim(V_1) = \dim(V_2)$ and $V_1 \subseteq V_2$, we conclude that $V_1 = V_2$.

5.1.2 Proof of Lemma 5.4

Suppose $\mathcal{Q}\colon \mathbb{F}^n \to \mathbb{C}$ is a quasi-distribution such that there exists a quasi-distribution $\mathcal{Q}'$ supported on $\mathbf{C}$ with $\Delta_k(\mathcal{Q}, \mathcal{Q}') \leq \delta$. Fix $\alpha \in \mathbb{F}^n_{\leq k}$, so that $S = \mathrm{supp}(\alpha)$ has $|S| \leq k$. Since $\Delta_k(\mathcal{Q}, \mathcal{Q}') \leq \delta$, we have that $\sum_{g \in \mathbb{F}^S} \left| \widetilde{\Pr}[\mathcal{Q}(S) = g] - \widetilde{\Pr}[\mathcal{Q}'(S) = g] \right| \leq \delta$. Therefore,

$$\begin{aligned}
\left| \widehat{\mathcal{Q}}(\alpha) - \widehat{\mathcal{Q}}'(\alpha) \right| &\leq \frac{1}{q^n} \sum_{j \in \mathbb{F}_p} |\omega^{-j}| \left| \widetilde{\Pr}[\mathrm{Tr}(\langle \alpha, \mathcal{Q} \rangle) = j] - \widetilde{\Pr}[\mathrm{Tr}(\langle \alpha, \mathcal{Q}' \rangle) = j] \right| \\
&= \frac{1}{q^n} \sum_{j \in \mathbb{F}_p} \sum_{g \in \mathbb{F}^S : \mathrm{Tr}(\langle \alpha, g \rangle) = j} \left| \widetilde{\Pr}[\mathcal{Q}(S) = g] - \widetilde{\Pr}[\mathcal{Q}'(S) = g] \right| \\
&\leq \frac{1}{q^n} \sum_{g \in \mathbb{F}^S} \left| \widetilde{\Pr}[\mathcal{Q}(S) = g] - \widetilde{\Pr}[\mathcal{Q}'(S) = g] \right| \leq \frac{\delta}{q^n} \ .
\end{aligned}$$

By Lemma 5.2, we know that for every $\alpha, \alpha' \in \mathbb{F}^n$ such that $\alpha - \alpha' \in \mathbf{C}^\perp$ it holds that $\left| \widehat{\mathcal{Q}}'(\alpha) - \widehat{\mathcal{Q}}'(\alpha') \right| = 0$. Hence, for every $\alpha, \alpha' \in \mathbb{F}^n_{\leq k}$ such that $\alpha - \alpha' \in \mathbf{C}^\perp$ it holds that

$$\left| \widehat{\mathcal{Q}}(\alpha) - \widehat{\mathcal{Q}}(\alpha') \right| \leq \left| \widehat{\mathcal{Q}}(\alpha) - \widehat{\mathcal{Q}}'(\alpha) \right| + \left| \widehat{\mathcal{Q}}'(\alpha) - \widehat{\mathcal{Q}}'(\alpha') \right| + \left| \widehat{\mathcal{Q}}'(\alpha') - \widehat{\mathcal{Q}}(\alpha') \right|$$

$$\leq \frac{\delta}{q^n} + 0 + \frac{\delta}{q^n} = \frac{2\delta}{q^n} \ .$$

Now, suppose that $\mathcal{Q}$ is a quasi-distribution such that $\left|\widehat{\mathcal{Q}}(\alpha) - \widehat{\mathcal{Q}}(\alpha')\right| \leq \frac{2\delta}{q^n}$ for all $\alpha, \alpha' \in \mathbb{F}^n_{\leq k}$ such that $\alpha - \alpha' \in \mathbf{C}^\perp$. For each $\alpha \in \mathbb{F}^n$, let γ_α be an element of the coset $\alpha + \mathbf{C}^\perp$ of minimal weight (ties are broken arbitrarily). Define $\mathcal{Q}'$ to be the quasi-distribution where $\widehat{\mathcal{Q}'}(\alpha) := \widehat{\mathcal{Q}}(\gamma_\alpha)$ if $\mathrm{wt}(\gamma_\alpha) \leq k$ and 0 otherwise. By construction, for any $\alpha, \alpha' \in \mathbb{F}^n$ such that $\alpha - \alpha' \in \mathbf{C}^\perp$ it holds that $\widehat{\mathcal{Q}'}(\alpha) = \widehat{\mathcal{Q}'}(\alpha')$, so $\mathcal{Q}'$ is supported on $\mathbf{C}$ by Lemma 5.2. Let $\alpha \in \mathbb{F}^n_{\leq k}$. By construction, we know that $\left|\widehat{\mathcal{Q}}(\alpha) - \widehat{\mathcal{Q}'}(\alpha)\right| \leq \left|\widehat{\mathcal{Q}}(\alpha) - \widehat{\mathcal{Q}}(\gamma_\alpha)\right| + \left|\widehat{\mathcal{Q}}(\gamma_\alpha) - \widehat{\mathcal{Q}'}(\alpha)\right| \leq \frac{2\delta}{q^n} + 0 = \frac{2\delta}{q^n}$, since $\alpha - \gamma_\alpha \in \mathbf{C}^\perp$ and $\mathrm{wt}(\gamma_\alpha) \leq \mathrm{wt}(\alpha) \leq k$. Therefore, by Lemma 4.7 we have that $\Delta_k(\mathcal{Q}, \mathcal{Q}') \leq q^{k/2} \cdot 2\delta$.

5.2 Locally-explainable non-signaling functions

We introduce another natural definition that captures when a non-signaling function $\mathcal{F}$ is "in" a given linear code $\mathbf{C} \subseteq \mathbb{F}^n$. This time we take the perspective of local consistency, namely, we shall require that the output of $\mathcal{F}$ is always consistent with a codeword in $\mathbf{C}$.

Definition 5.5. *Given a k-non-signaling strategy $\mathcal{F}\colon [n] \to \mathbb{F}$, code $\mathbf{C} \subseteq \mathbb{F}^n$, and parameter $k' \leq k$, we say that $\mathcal{F}$ is* $(\mathbf{C}, k')$**-explainable** *if for every set $S \subseteq [n]_{\leq k'}$ it holds that $\Pr[\mathcal{F}(S) \in \mathbf{C}|_S] = 1$.*

Note that $\mathcal{F}$ is $(\mathbf{C}, k')$-explainable if and only if $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = 1$ for every $\alpha \in \mathbf{C}^\perp_{\leq k'}$. The non-trivial direction of the equivalence is implied by the following lemma.

Lemma 5.6. *Let $\mathbf{C} \subseteq \mathbb{F}^n$ be a linear code, $S \subseteq [n]_{\leq k}$, and $g\colon S \to \mathbb{F}$. If $\langle \alpha, g \rangle = 0$ for every $\alpha \in \mathbf{C}^\perp_{\subseteq S}$, then there is a codeword $f \in \mathbf{C}$ such that $f|_S = g$.*

Proof. Since $\mathbf{C} \subseteq \mathbb{F}^n$ is a linear code, there is a *pivotal set* $P \subseteq [n]$ of size $|P| = \dim(\mathbf{C})$ such that for all $y\colon P \to \mathbb{F}$ there is a unique codeword $f \in \mathbf{C}$ satisfying $f|_P = y$. Such P need not be unique.

Let $P^* \subseteq [n]$ be a pivotal set such that $|P^* \cap S|$ is maximal, and let $P_S := P^* \cap S$. Define $f'\colon P^* \to \mathbb{F}$ by letting $f'(i) = g(i)$ for all $i \in P_S$, and letting $f'(j)$ be arbitrary for all $j \in P^* \setminus P_S$. Since P^* is a pivotal set, there exists a unique $f \in \mathbf{C}$ such that $f|_{P^*} = f'$.

It remains to show that $f|_S = g$. Let $i \in S$. If $i \in P_S$, then $f(i) = f'(i) = g(i)$, as required. Suppose that $i \notin P_S$. Since P^* is maximal, there exists $\alpha \in \mathbf{C}^\perp$ such that $\alpha_i = 1$ and $\mathrm{supp}(\alpha) \subseteq P_S \cup \{i\} \subseteq S$. Indeed, if no such α exists then for any codeword $h \in \mathbf{C}$, $h(i)$ is not determined by $\{h(j) : j \in P_S\}$. Hence, the set $P_S \cup \{i\}$ can be extended into a pivotal set for $\mathbf{C}$, which contradicts the maximality of P^*. Therefore, such an α exists. Since $\langle \alpha, f \rangle = 0$ and $\langle \alpha, g \rangle = 0$, we get that $0 = \langle \alpha, f \rangle - \langle \alpha, g \rangle = f(i) + \sum_{j \in P_S} \alpha_j f(j) - g(i) - \sum_{j \in P_S} \alpha_j g(j) = f(i) + \sum_{j \in P_S} \alpha_j g(j) - g(i) - \sum_{j \in P_S} \alpha_j g(j) = f(i) - g(i)$, and therefore $f(i) = g(i)$. We conclude that $f|_S = g$, as required. □

We provide a characterization of the Fourier spectrum of $\mathbf{C}$-explainable non-signaling functions, both in the exact and in the robust cases, as captured by the respective lemmas below. Both lemmas make crucial use of Lemma 4.3.

Lemma 5.7. *Let $\mathcal{F}\colon [n] \to \mathbb{F}$ be a k-non-signaling function. Then $\mathcal{F}$ is $(\mathbf{C}, k')$-explainable if and only if $\widehat{\mathcal{F}}(\alpha) = \frac{1}{q^n}$ for every $\alpha \in \mathbf{C}^\perp_{\leq k'}$.*

Proof. We know that $\mathcal{F}$ is $(\mathbf{C}, k')$-explainable if and only if for every $\alpha \in \mathbf{C}^{\perp}_{\leq k'}$ it holds that $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = 1$. By Lemma 4.3, we know that $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = 1$ if and only if $\widehat{\mathcal{F}}(a\alpha) = \frac{1}{q^n}$ for every $a \in \mathbb{F}$, as M is invertible and maps the distribution $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = 1$ and $\Pr[\langle \alpha, \mathcal{F} \rangle = b] = 0$ for all other b to the vector 1^q. We conclude the proof by noting that if $\alpha \in \mathbf{C}^{\perp}_{\leq k'}$ then $a\alpha \in \mathbf{C}^{\perp}_{\leq k'}$ for any $a \in \mathbb{F}$. □

Lemma 5.8. *Let $\mathcal{F}\colon [n] \to \mathbb{F}$ be a k-non-signaling function, and let $\alpha \in \mathbb{F}^n_{\leq k}$.*

- *If $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] \geq 1 - \varepsilon$, then $|\widehat{\mathcal{F}}(a\alpha) - \frac{1}{q^n}| \leq \frac{2\varepsilon}{q^n}$ for every $a \in \mathbb{F}$.*
- *If $|\widehat{\mathcal{F}}(a\alpha) - \frac{1}{q^n}| \leq \frac{\varepsilon}{q^n}$ for every $a \in \mathbb{F}$, then $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] \geq 1 - \varepsilon$.*

Proof. Suppose that $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] \geq 1 - \varepsilon$. This immediately implies that, for every $a \in \mathbb{F}$, $\Pr[\langle a\alpha, \mathcal{F} \rangle = 0] \geq 1 - \varepsilon$. Therefore,

$$\begin{aligned}
\left| q^n \widehat{\mathcal{F}}(a\alpha) - 1 \right| &= \left| -1 + \sum_{b \in \mathbb{F}} \omega^{-\mathrm{Tr}(ab)} \Pr[\langle a\alpha, \mathcal{F} \rangle = b] \right| \\
&\leq |-1 + \Pr[\langle a\alpha, \mathcal{F} \rangle = 0]| + \sum_{b \neq 0} |\omega^{-\mathrm{Tr}(ab)}| \, |\Pr[\langle a\alpha, \mathcal{F} \rangle = b]| \\
&\leq \varepsilon + \sum_{b \neq 0} \Pr[\langle a\alpha, \mathcal{F} \rangle = b] \\
&= \varepsilon + (1 - \Pr[\langle a\alpha, \mathcal{F} \rangle = 0]) \leq 2\varepsilon \ .
\end{aligned}$$

This proves the first direction.

For the second direction, let $v \in \mathbb{C}^q$ be the vector where $v_b = \Pr[\langle \alpha, \mathcal{F} \rangle = b]$ and let $w \in \mathbb{C}^q$ be the vector where $w_a = q^n \widehat{\mathcal{F}}(a\alpha)$. Note that $Mv = w$, where M is the matrix from Lemma 4.3. Suppose that $|\widehat{\mathcal{F}}(a\alpha) - \frac{1}{q^n}| \leq \frac{\varepsilon}{q^n}$ for every $a \in \mathbb{F}$, so that $|w_a - 1| \leq \varepsilon$ for every $a \in \mathbb{F}$. Then, we have that $\|w - 1^q\|^2_{\ell_2} \leq q\varepsilon^2$, so that $\|w - 1^q\|_{\ell_2} \leq \varepsilon\sqrt{q}$. Let $u \in \mathbb{C}^q$ be the vector where $u_0 = 1$ and $u_b = 0$ for all other $b \in \mathbb{F}$. Observe that $Mu = 1^q$. Since $\frac{1}{\sqrt{q}} M$ is unitary, we have that $\|\frac{1}{\sqrt{q}} M(v - u)\|_{\ell_2} = \|\frac{1}{\sqrt{q}}(w - 1^q)\|_{\ell_2} \leq \frac{1}{\sqrt{q}} \cdot \varepsilon\sqrt{q} = \varepsilon$. Therefore, $|v_b - u_b| \leq \varepsilon$ for all $b \in \mathbb{F}$. In particular, $|v_0 - 1| \leq \varepsilon$, so that $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] \geq 1 - \varepsilon$. □

5.3 The relationship between the two definitions

We have given two natural definitions of what it means for a non-signaling function to be in a linear code. Which of the two definitions is more "correct"? Lemma 5.2 and Lemma 5.7 show that Definition 5.1 implies Definition 5.5, in the sense that if $\mathcal{F}$ is $(\mathbf{C}, k')$-supported then $\mathcal{F}$ is $(\mathbf{C}, k')$-explainable. We prove that, conversely, Definition 5.5 implies Definition 5.1 up to a factor of 2 in the locality k'. We conclude that the two definitions are essentially equivalent.

Lemma 5.9. *Let $\mathbf{C} \subseteq \mathbb{F}^n$ be a linear code, and let $\mathcal{F}\colon [n] \to \mathbb{F}$ be a k-non-signaling function.*

- *If $\mathcal{F}$ is $(\mathbf{C}, k')$-supported then $\mathcal{F}$ is $(\mathbf{C}, k')$-explainable.*
- *If $\mathcal{F}$ is $(\mathbf{C}, k')$-explainable then $\mathcal{F}$ is $(\mathbf{C}, k'/2)$-supported.*

Remark 5.10. For specific choices of $\mathbf{C}$ one can achieve stronger versions of the above lemma. For example, when $\mathbf{C}$ is the Hadamard code (all linear functions), one can prove the lemma with $k' - 1$ in place of $k'/2$. Also, *some* gap in locality is necessary: taking again $\mathbf{C}$ to be the Hadamard

code, there exists a non-signaling function $\mathcal{F}$ that is $(\mathbf{C}, k)$-explainable and $(\mathbf{C}, k-1)$-supported but *not* $(\mathbf{C}, k)$-supported. (The foregoing statements are shown implicitly in [CMS18].)

Proof. Lemma 5.2 and Lemma 5.7 imply the first direction, as any $(\mathbf{C}, k')$-supported k-non-signaling function $\mathcal{F}$ satisfies $\widehat{\mathcal{F}}(\alpha) = \widehat{\mathcal{F}}(0^n) = \frac{1}{q^n}$ for every $\alpha \in \mathbf{C}^{\perp}_{\leq k'}$, implying that $\mathcal{F}$ is $(\mathbf{C}, k')$-explainable.

We now prove the second direction. Fix $\alpha \in \mathbf{C}^{\perp}_{\leq k'}$, and let $S := \{i \in [n] : \alpha_i \neq 0\}$. Note that $|S| \leq k'$ since $|S| = \mathrm{wt}(\alpha)$. We first show that $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = 1$. Indeed, since $\mathcal{F}$ is $(\mathbf{C}, k')$-explainable, we have that

$$\begin{aligned}\Pr[\langle \alpha, \mathcal{F} \rangle = 0] &\geq \Pr[\langle \alpha, \mathcal{F} \rangle = 0 \wedge \exists\, f \in \mathbf{C} \text{ s.t. } \mathcal{F}(S) = f|_S] \\ &= \Pr[\langle \alpha, f \rangle = 0 \wedge \exists\, f \in \mathbf{C} \text{ s.t. } \mathcal{F}(S) = f|_S] \\ &= \Pr[\exists\, f \in \mathbf{C} \text{ s.t. } \mathcal{F}(S) = f|_S] = 1 \ ,\end{aligned}$$

and so $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = 1$.

Now, for any $\alpha, \alpha' \in \mathbb{F}^n_{\leq k'/2}$ with $\alpha - \alpha' \in \mathbf{C}^{\perp}$ we get that for any $b \in \mathbb{F}$,

$$\Pr[\langle \alpha, \mathcal{F} \rangle = b] = \Pr[\langle \alpha', \mathcal{F} \rangle + \langle \alpha - \alpha', \mathcal{F} \rangle = b] = \Pr[\langle \alpha', \mathcal{F} \rangle = b] \ ,$$

since $\Pr[\langle \alpha - \alpha', \mathcal{F} \rangle = 0] = 1$ as $\alpha - \alpha' \in \mathbf{C}^{\perp}$ with $\mathrm{wt}(\alpha - \alpha') \leq k'$. This shows that the vectors $(\Pr[\langle \alpha, \mathcal{F} \rangle = b])_{b \in \mathbb{F}}$ and $(\Pr[\langle \alpha', \mathcal{F} \rangle = b])_{b \in \mathbb{F}}$ are the same. Thus, $\widehat{\mathcal{F}}(\alpha) = \widehat{\mathcal{F}}(\alpha')$, by the definition of $\mathcal{F}$'s Fourier coefficients. By Lemma 5.2, it follows that $\mathcal{F}$ is $(\mathbf{C}, k'/2)$-supported. □

6 Low-degree testing

In this section, we prove Theorem 1. Throughout this section, we let $m, d, k \in \mathbb{N}$ be parameters, and p be a prime where $p \geq d+2$. Let $\mathsf{RS}^{\otimes m}[\mathbb{F}_p, d]$ denote the code of polynomials from $\mathbb{F}_p^m$ to $\mathbb{F}_p$ of individual degree at most d in each variable. We let $\mathcal{F}\colon \mathbb{F}_p^m \to \mathbb{F}_p$ be a k-non-signaling function. For a subset $S \subseteq \mathbb{F}_p$ and $a \in S$, we let $\delta_{a,S}(x) = \frac{\prod_{a' \in S \setminus \{a\}}(x-a')}{\prod_{a' \in S \setminus \{a\}}(a-a')}$ be the degree $|S|-1$ polynomial satisfying $\delta_a(x) = \begin{cases} 1 & \text{if } x = a \ , \\ 0 & \text{if } x \in S \setminus \{a\} \ . \end{cases}$

We begin by recalling the degree-d evenly-spaced points test.

Definition 6.1 (Evenly-spaced points test)**.** *Given a function $f\colon \mathbb{F}_p^m \to \mathbb{F}_p$, the degree-$d$ evenly-spaced points test:*

1. *samples a random point $x \in \mathbb{F}_p^m$ and slope $h \in \mathbb{F}_p^m \setminus \{0^m\}$,*
2. *checks that $\sum_{i=0}^{d+1} c_i f(x+ih) = 0$, where $c_i = (-1)^i \binom{d+1}{i}$.*

Definition 6.2 (Evenly-spaced self-correction)**.** *The evenly-spaced self-correction of $\mathcal{F}$, denoted $\hat{\mathcal{F}}$, is a $\lfloor k/(d+1) \rfloor$-non-signaling function defined as follows. Let $w_0, w_1, \ldots, w_m \in \mathbb{F}_p^m$ be independent uniformly random vectors in $\mathbb{F}_p^m$, and for each $x \in \mathbb{F}_p^m$ let $w_x = w_0 + \sum_{i=1}^m x_i w_i$. For an input set S and $g\colon S \to \mathbb{F}_p$, the distribution of $\hat{\mathcal{F}}(S)$ correction is defined as*

$$\Pr[\hat{\mathcal{F}}(S) = g] = \Pr_{w_0,\ldots,w_m,\mathcal{F}}[-\sum_{j=1}^{d+1} c_j \mathcal{F}(x+jw_x) = g(x) \ \forall x \in S] \ .$$

Our main theorem is the following.

Theorem 6.1 (Formal version of Theorem 1)**.** *Let p be a prime, and let $m, d \in \mathbb{N}$ with $p \geq d+2$. Let $\mathcal{F}\colon \mathbb{F}_p^m \to \mathbb{F}_p$ be a k-non-signaling function. Suppose that $\mathcal{F}$ passes the degree-d evenly-spaced points test with probability $1-\varepsilon$. Then there exists a k'-non-signaling function $\mathcal{G}\colon \mathbb{F}_p^m \to \mathbb{F}_p$ that is $(\mathsf{RS}^{\otimes m}[\mathbb{F}_p, d], k')$-supported such that $\Delta_{k'}(\hat{\mathcal{F}}, \mathcal{G}) \leq O(p^{3k'/2}(d+1)^m \varepsilon)$, where $k' = \lfloor \frac{k}{2(d+1)(d+2)} \rfloor - 3$.*

When $\varepsilon = 0$, the theorem can be simplified considerably to the statement below.

Theorem 6.2 (Formal version of Theorem 1, zero error case)**.** *Let p be a prime, and let $m, d \in \mathbb{N}$ with $p \geq d+2$. Let $\mathcal{F}\colon \mathbb{F}_p^m \to \mathbb{F}_p$ be a k-non-signaling function. Suppose that $\mathcal{F}$ passes the degree-d evenly-spaced points test with probability 1. Then $\mathcal{F}$ is $(\mathsf{RS}^{\otimes m}[\mathbb{F}_p, d], \lfloor k/(d+2) \rfloor - 1)$-supported.*

We prove Theorem 6.1 via the following statements, each proved in one of the following sections.

Lemma 6.3 (Average to worst case)**.** *Suppose that $\mathcal{F}$ passes the degree-d evenly-spaced points test with probability $1-\varepsilon$, i.e. that $\Pr_{h \neq 0^m, x, \mathcal{F}}[\sum_{i=0}^{d+1} c_i \mathcal{F}(x+ih) = 0] \geq 1-\varepsilon$. Then for every $x, h \in \mathbb{F}_p^m$ with $h \neq 0^m$ it holds that $\Pr_{\hat{\mathcal{F}}}[\sum_{i=0}^{d+1} c_i \hat{\mathcal{F}}(x+ih) = 0] \geq 1-(d+1)\varepsilon$.*

Lemma 6.4 (Evenly-spaced points to axis-parallel lines)**.** *Let $\hat{\mathcal{F}}\colon \mathbb{F}_p^m \to \mathbb{F}_p$ be a $\hat{k}$-non-signaling function, with $\hat{k} \geq 2d+3$. Suppose that for every $x, h \in \mathbb{F}_p^m$ with $h \neq 0^m$ it holds that $\Pr_{\hat{\mathcal{F}}}[\sum_{i=0}^{d+1} c_i \hat{\mathcal{F}}(x+ih) = 0] \geq 1-\varepsilon$. Then for every $b_1, \ldots, b_m \in \mathbb{F}_p$, $i \in [m]$, and $S \subseteq \mathbb{F}_p \setminus \{b_i\}$ of size $|S| = d+1 \leq \hat{k} - d - 2$ it holds that*

$$\Pr_{\hat{\mathcal{F}}}[\hat{\mathcal{F}}(b_1, \ldots, b_m) = \sum_{a \in S} \delta_{a,S}(b_i) \cdot \hat{\mathcal{F}}(b_1, \ldots, b_{i-1}, a, b_{i+1}, \ldots, b_m)] \geq 1 - p\varepsilon \ .$$

Lemma 6.5 (Robust local characterization)**.** *Let $\hat{\mathcal{F}}\colon \mathbb{F}_p^m \to \mathbb{F}_p$ be a $\hat{k}$-non-signaling function, and suppose that for every $i \in [m]$, $b_1, \ldots, b_m \in \mathbb{F}_p$, and $S' \subseteq \mathbb{F}_p \setminus \{b_i\}$ of size $|S'| = d+1$ it holds that*

$$\Pr_{\hat{\mathcal{F}}}[\hat{\mathcal{F}}(b_1, \ldots, b_m) = \sum_{a \in S'} \delta_{a,S'}(b_i) \cdot \hat{\mathcal{F}}(b_1, \ldots, b_{i-1}, a, b_{i+1}, \ldots, b_m)] \geq 1 - \varepsilon \ .$$

Then for every $S \subseteq \mathbb{F}_p^m$ with $|S| \leq \lfloor \hat{k}/(d+2) \rfloor$ it holds that $\Pr[\hat{\mathcal{F}}(S) \in \mathsf{RS}^{\otimes m}[\mathbb{F}_p, d]\big|_S] \geq 1 - 2\,|S| \cdot (d+1)^{m-1}\varepsilon$.

Lemma 6.6. *Let $\hat{\mathcal{F}}\colon \mathbb{F}_p^m \to \mathbb{F}_p$ be a $\hat{k}$-non-signaling function, and let $k' \leq \hat{k}$. Suppose that for every $S \subseteq \mathbb{F}_p^m$ with $|S| \leq k'$ it holds that $\Pr[\hat{\mathcal{F}}(S) \in \mathsf{RS}^{\otimes m}[\mathbb{F}_p, d]\big|_S] \geq 1 - \varepsilon$. Then there exists a $k'/2$-non-signaling function $\mathcal{G}$ that is $(\mathsf{RS}^{\otimes m}[\mathbb{F}_p, d], k'/2)$-supported and $\Delta_{k'/2}(\mathcal{F}, \mathcal{G}) \leq (p^{k'/2}+1) \cdot p^{k'/4} \cdot 2\varepsilon$.*

6.1 Step 1: Average to worst case reduction

We prove Lemma 6.3. Fix $x, h \in \mathbb{F}_p^m$ with $h \neq 0^m$. Observe that $w_{x+ih} = w_x + i(w_h - w_0)$. Therefore

$$\begin{aligned}
&\Pr_{\hat{\mathcal{F}}}[\sum_{i=0}^{d+1} c_i \hat{\mathcal{F}}(x+ih) = 0] = \Pr_{w_0,\ldots,w_m,\mathcal{F}}[\sum_{i=0}^{d+1} c_i \sum_{j=1}^{d+1} -c_j \mathcal{F}(x+ih+jw_{x+ih}) = 0] \\
&\Pr_{w_0,\ldots,w_m,\mathcal{F}}[\sum_{j=1}^{d+1} -c_j \sum_{i=0}^{d+1} c_i \mathcal{F}(x+ih+jw_x+ij(w_h-w_0)) = 0] \\
&\geq \Pr_{w_0,\ldots,w_m,\mathcal{F}}[\sum_{i=0}^{d+1} c_i \mathcal{F}(x+jw_x+i(h+j(w_h-w_0))) = 0 \ \forall j \in [d+1]] \\
&\geq 1 - (d+1)\varepsilon \ ,
\end{aligned}$$

where last inequality is by union bound, using the fact that for $j \neq 0$ the vectors $x + jw_x$ and $h + j(w_h - w_0)$ are independent and uniformly random vectors in $\mathbb{F}_p^m$. Indeed, for $h \neq 0^m$ the vector $h + j(w_h - w_0)$ is equal to $h + \sum_{i=1}^m jh_i w_i$, and hence is uniformly random, and $x + jw_x$ is equal to $x + jw_0 + \sum_{i=1}^m jx_i w_i$, which is also uniformly random and independent of $h + j(w_h - w_0)$ because the term w_0 is independent of all other w_i's.

6.2 Step 2: From evenly-spaced points to axis-parallel lines

We prove Lemma 6.4. Note that by the assumption for every $x \in \mathbb{F}_p^m$ and $h = e_i$ it holds that $\Pr_{\hat{\mathcal{F}}}[\sum_{\ell=0}^{d+1} c_\ell \hat{\mathcal{F}}(x + \ell e_i) = 0] \geq 1 - \varepsilon$. Fix $i \in [m]$, $b_1, \ldots, b_{i-1}, b_{i+1}, \ldots b_m \in \mathbb{F}_p$, and $S \subseteq \mathbb{F}_p$ of size $|S| = d+2$. We claim that

$$\Pr[\exists\, g \in \mathsf{RS}[\mathbb{F}_p, d] \text{ s.t. } \hat{\mathcal{F}}(b_1, \ldots, b_{i-1}, s, b_{i+1}, \ldots b_m) \equiv g(s) \ \forall s \in S] \geq 1 - p\varepsilon \ . \tag{1}$$

This clearly implies Lemma 6.4.

In order to prove Eq. (1), let us order the elements of S as $s_1 \leq s_2 \leq \cdots \leq s_{d+2}$ by treating s_i's as integers in $\{0, 1, 2, \ldots, p-1\} \subseteq \mathbb{N}$. For each $j \in \mathbb{F}_p$ define the interval $I_j = \{j, j+1, j+2, \ldots, j+d+1\}$, and denote $Q_j = \{s \in S : s \leq j + d + 2\} \cup I_j$. We claim that with high probability $\hat{\mathcal{F}}$ on the points corresponding to Q_j agrees with some univariate polynomial. Specifically, we prove the following claim.

Claim 6.7. *For each $j = 0, 1 \ldots, p-1$ it holds that*

$$\Pr[\exists\, g \in \mathsf{RS}[\mathbb{F}_p, d] \text{ s.t. } \hat{\mathcal{F}}(b_1, \ldots, b_{i-1}, \ell, b_{i+1}, \ldots b_m) \equiv g(\ell)\ \forall \ell \in Q_j] \geq 1 - (j+1)\varepsilon \ .$$

It is clear that the statement of Claim 6.7 for $j = p-1$ implies Eq. (1), and hence proves Lemma 6.4.

Proof of Claim 6.7. The proof is by induction in j. For the base case of $j = 0$ we query $\hat{\mathcal{F}}$ on the set I_0. By the assumption of the lemma we have

$$\Pr_{\hat{\mathcal{F}}}[\exists\, g \in \mathsf{RS}[\mathbb{F}_p, d] \text{ s.t. } \hat{\mathcal{F}}(b_1, \ldots, b_{i-1}, \ell, b_{i+1}, \ldots b_m) = \ell\ \forall \ell \in I_0] \geq 1 - \varepsilon \ ,$$

and the claim hold since for $j = 0$ we have $Q_0 = I_0$.

For the induction step suppose that the statement of the claim holds for $j-1$, i.e., with probability $1 - j\varepsilon$ there exists a univariate polynomial g of degree d that agrees with $\hat{\mathcal{F}}$ on Q_{j-1}. Note that the set $Q'_{j-1} = Q_{j-1} \setminus \{j-1\}$ contains the $d+1$ consecutive points $\{j, j+1, \ldots, j+d\}$, and these points uniquely define the polynomial g of degree d. Therefore

$$\Pr[\exists\, g \in \mathsf{RS}[\mathbb{F}_p, d] \text{ s.t. } \hat{\mathcal{F}}(b_1, \ldots, b_{i-1}, \ell, b_{i+1}, \ldots b_m) \equiv g(\ell)\ \forall \ell \in Q'_{j-1}] \geq 1 - j\varepsilon \ .$$

On the other hand, by the assumption that $\hat{\mathcal{F}}$ satisfies evenly-spaced constraints with probability $\geq 1 - \varepsilon$ we have

$$\Pr[\exists\, g' \in \mathsf{RS}[\mathbb{F}_p, d] \text{ s.t. } \hat{\mathcal{F}}(b_1, \ldots, b_{i-1}, \ell, b_{i+1}, \ldots b_m) \equiv g'(\ell)\ \forall \ell \in I_j] \geq 1 - \varepsilon \ .$$

Note that the above statement requires that $p \geq d+2$, as passing the evenly-spaced points test along a line corresponds to being a univariate polynomial only when $p \geq d+2$.

Query $\hat{\mathcal{F}}$ on the set $Q'_{j-1} \cup I_j$. By union bound, the above events both hold with probability $\geq 1-(j+1)\varepsilon$, so that there exists $g, g' \in \mathsf{RS}[\mathbb{F}_p, d]$ such that $\hat{\mathcal{F}}(b_1, \ldots, b_{i-1}, \ell, b_{i+1}, \ldots b_m) = g(\ell)$ for all $\ell \in Q'_{j-1}$ and $\hat{\mathcal{F}}(b_1, \ldots, b_{i-1}, \ell, b_{i+1}, \ldots b_m) = g'(\ell)$ for all $\ell \in I_j$. However, since Q'_{j-1} intersects I_j on $d+1$ points it must be the case that $g = g'$. Hence, $\hat{\mathcal{F}}(b_1, \ldots, b_{i-1}, \ell, b_{i+1}, \ldots b_m) = g(\ell)$ for all $\ell \in Q'_{j-1} \cup I_j$. Since $Q_j \subseteq Q'_{j-1} \cup I_j$, this proves the induction step. Since $\left|Q'_{j-1} \cup I_j\right| \leq |S| + d + 2$ and $|Q_j| \leq |S| + d + 2$, it follows that this is valid for all S with $|S| \leq \hat{k} - d - 2$, which completes the proof of Claim 6.7. □

6.3 Step 3: A robust local characterization of low-degree polynomials

We prove Lemma 6.5. We begin by introducing some notation. For each $i \in [m]$ define $S_i \subseteq \mathbb{F}_p^i$ to be the projection of S to the first i coordinates, i.e. that

$$S_i := \{(a_1, \ldots, a_i) \in \mathbb{F}_p^i : \exists\, b_{i+1}, \ldots b_m \in \mathbb{F}_p \text{ s.t. } (a_1, \ldots, a_i, b_{i+1}, \ldots, b_m) \in S\} \ .$$

Note that $S_m = S$. For any set $R \subseteq \mathbb{F}_p^i$ and $b_{i+1}, \ldots, b_m \in \mathbb{F}_p$, we define the extension of R as

$$R^{(b_{i+1}, \ldots, b_m)} := R \times \{(b_{i+1}, \ldots, b_m)\} \ .$$

We prove by induction that for every $i \in [m]$ and every $b_{i+1}, \ldots, b_m \in \mathbb{F}_p$ it holds that

$$\Pr[\exists\, g_i \in \mathsf{RS}^{\otimes i}[\mathbb{F}_p, d] \text{ s.t. } \hat{\mathcal{F}}(a_1, \ldots, a_i, b_{i+1}, \ldots, b_m) = g_i(a_1, \ldots, a_i)\ \forall (a_1, \ldots, a_i) \in S_i] \geq 1 - \varepsilon_i \ , \tag{2}$$

where $\varepsilon_i = |S| \sum_{j=1}^{i} (d+1)^{j-1}\varepsilon$. Note that the above probability is well-defined since we query $\hat{\mathcal{F}}$ on the set $S_i^{(b_{i+1},\ldots,b_m)}$, which contains at most $|S| \leq k' \leq k$ points. Eq. (2) proves Lemma 6.5, as $S_m = S$ and $\varepsilon_m \leq |S| \cdot (\sum_{j=1}^{m}(d+1)^{j-1}) \cdot \varepsilon \leq \frac{(d+1)^m - 1}{d} \cdot |S|\,\varepsilon \leq 2(d+1)^{m-1}\,|S|\,\varepsilon$.

We first show the base case: $i = 1$. Let $R \subseteq S_1$ be a subset of size $\min(d+1, |S_1|)$. Query $\hat{\mathcal{F}}$ on $S_1^{(b_2,\ldots,b_m)}$, and let $g_1 \in \mathsf{RS}^{\otimes 1}[\mathbb{F}_p, d]$ be the univariate polynomial $g_1(x) = \sum_{a \in R} \delta_{a,R}(x) \cdot \hat{\mathcal{F}}(a, b_2, \ldots, b_m)$. Then,

$$\Pr[\exists\, g_1 \in \mathsf{RS}^{\otimes 1}[\mathbb{F}_p, d] \text{ s.t. } \hat{\mathcal{F}}(a_1, b_2, \ldots, b_m) = g_1(a_1) \text{ for all } a_1 \in S_1] \geq 1 - |S_1|\,\varepsilon = 1 - \varepsilon_1 \ .$$

This is because if $a_1 \in R$, then $\hat{\mathcal{F}}(a, b_2, \ldots, b_m) = g_1(a)$ is trivially true by definition of g_1, and if $a_1 \in S_1 \setminus R$ then this is true because

$$\hat{\mathcal{F}}(a_1, b_2, \ldots, b_m) = \sum_{a \in R} \delta_{a,R}(a_1) \cdot \hat{\mathcal{F}}(a, b_2, \ldots, b_m) = g_1(a_1) \ ,$$

with probability $1 - \varepsilon$, by the assumption on $\hat{\mathcal{F}}$.

We now show the induction step. Suppose that Eq. (2) holds for $i-1$ and every $b_i, \ldots, b_m \in \mathbb{F}_p$. Let $R_i = S_i \cup (S_{i-1} \times \{0, \ldots, d\})$. Fix $b_{i+1}, \ldots, b_m \in \mathbb{F}_p$, and query $\hat{\mathcal{F}}$ on $R_i^{(b_{i+1},\ldots,b_m)}$. Note that this is well-defined since $\left|R_i^{(b_{i+1},\ldots,b_m)}\right| \leq |S_i| + (d+1)\,|S_{i-1}| \leq (d+2)\,|S| \leq (d+2)k' \leq k$.

We have that $S_{i-1} \times \{0, \ldots, d\} \subseteq R_i$, and so for every $j \in \{0, \ldots, d\}$, the induction hypothesis implies (by setting $b_i = j$) that with probability at least $1 - \varepsilon_{i-1}$ there exists $g_{i-1}^{(j)} \in \mathsf{RS}^{\otimes i-1}[\mathbb{F}_p, d]$ such that $\hat{\mathcal{F}}(a_1, \ldots, a_{i-1}, j, b_{i+1}, \ldots, b_m) = g_{i-1}^{(j)}(a_1, \ldots, a_{i-1})$ for every $(a_1, \ldots, a_{i-1}) \in S_{i-1}$.

For $j \in \{0, \ldots, d\}$, let $\delta_j(x) := \delta_{j,\{0,\ldots,d\}}(x)$. Let $g_i \in \mathsf{RS}^{\otimes i}[\mathbb{F}_p, d]$ be defined as $g_i(x_1, \ldots, x_i) = \sum_{j=0}^{d} \delta_j(x_i) \cdot g_{i-1}^{(j)}(x_1, \ldots, x_{i-1})$. We show that

$$\Pr[\hat{\mathcal{F}}(a_1, \ldots, a_i, b_{i+1}, \ldots, b_m) = g_i(a_1, \ldots, a_i) \text{ for all } (a_1, \ldots, a_i) \in S_i] \geq 1 - \varepsilon_i \ .$$

Indeed, note that for any $(a_1, \ldots, a_i) \in S_i$ we have $(a_1, \ldots, a_{i-1}) \in S_{i-1}$, and so with probability $1 - (d+1)\varepsilon_{i-1}$ all the $g_{i-1}^{(j)}$'s exist, and so we have that for all $(a_1, \ldots, a_i) \in S_i$ it holds that

$$g_i(a_1, \ldots, a_i) = \sum_{j=0}^{d} \delta_j(a_i) \cdot g_{i-1}^{(j)}(a_1, \ldots, a_{i-1}) = \sum_{j=0}^{d} \delta_j(a_i) \cdot \hat{\mathcal{F}}(a_1, \ldots, a_{i-1}, j, b_{i+1}, \ldots, b_m)$$

By the assumption on $\hat{\mathcal{F}}$, we have

$$\Pr[\sum_{j=0}^{d} \delta_j(a_i) \cdot \hat{\mathcal{F}}(a_1, \ldots, a_{i-1}, j, b_{i+1}, \ldots, b_m) = \hat{\mathcal{F}}(a_1, \ldots, a_i, b_{i+1}, \ldots, b_m)\ \forall (a_1, \ldots, a_i) \in S_i] \geq 1 - |S_i|\,\varepsilon \ .$$

Combining the two equations shows that $g_i(a_1, \ldots, a_i) = \hat{\mathcal{F}}(a_1, \ldots, a_i, b_{i+1}, \ldots, b_m)$ with probability $1 - (d+1)\varepsilon_{i-1} - |S_i|\,\varepsilon \geq 1 - (d+1)\varepsilon_{i-1} - |S|\,\varepsilon = 1 - \varepsilon_i$, as required, completing the proof.

6.4 Step 4: Completing the proof

The following generic lemma immediately implies Lemma 6.6.

Lemma 6.8. *Let $\mathbf{C} \subseteq \mathbb{F}^n$ be a linear code, and let $\mathcal{F}$ be a k-non-signaling function. Suppose that $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] \geq 1 - \varepsilon$ for every $\alpha \in \mathbf{C}^{\perp}_{\leq k}$. Then there exists a $k/2$-non-signaling function $\mathcal{G}$ that is $(\mathbf{C}, k/2)$-supported such that $\Delta_{k/2}(\mathcal{F}, \mathcal{G}) \leq (q^{k/2} + 1) \cdot q^{k/4} \cdot 2\varepsilon$.*

Proof. Let $\alpha, \alpha' \in \mathbb{F}^n_{\leq k/2}$ such that $\alpha - \alpha' \in \mathbf{C}^{\perp}$. Then

$$\begin{aligned}
\left|\widehat{\mathcal{F}}(\alpha) - \widehat{\mathcal{F}}(\alpha')\right| &= \left|\frac{1}{q^n} \sum_{j \in \mathbb{F}_p} \omega^{-j} (\Pr[\mathrm{Tr}(\langle \alpha, \mathcal{F} \rangle) = j] - \Pr[\mathrm{Tr}(\langle \alpha', \mathcal{F} \rangle) = j])\right| \\
&\leq \frac{1}{q^n} \sum_{j \in \mathbb{F}_p} \left|\omega^{-j}\right| \left|\Pr[\mathrm{Tr}(\langle \alpha, \mathcal{F} \rangle) = j] - \Pr[\mathrm{Tr}(\langle \alpha', \mathcal{F} \rangle) = j]\right| \\
&\leq \frac{1}{q^n} \sum_{j \in \mathbb{F}_p} \left|\Pr[\mathrm{Tr}(\langle \alpha, \mathcal{F} \rangle) = j \wedge \langle \alpha - \alpha', \mathcal{F} \rangle \neq 0] - \Pr[\mathrm{Tr}(\langle \alpha', \mathcal{F} \rangle) = j \wedge \langle \alpha - \alpha', \mathcal{F} \rangle \neq 0]\right| \\
&\leq \frac{1}{q^n} \sum_{j \in \mathbb{F}_p} \left(\left|\Pr[\mathrm{Tr}(\langle \alpha, \mathcal{F} \rangle) = j \wedge \langle \alpha - \alpha', \mathcal{F} \rangle \neq 0]\right| + \left|\Pr[\mathrm{Tr}(\langle \alpha', \mathcal{F} \rangle) = j \wedge \langle \alpha - \alpha', \mathcal{F} \rangle \neq 0]\right|\right) \\
&\leq \frac{2}{q^n} \cdot \Pr[\langle \alpha - \alpha', \mathcal{F} \rangle \neq 0] \leq \frac{2\varepsilon}{q^n} \ ,
\end{aligned}$$

since $\alpha - \alpha' \in \mathbf{C}^{\perp}$ and $\mathrm{wt}(\alpha - \alpha') \leq k$. By Lemma 5.4, there exists a quasi-distribution $\mathcal{Q}$ supported on $\mathbf{C}$ such that $\Delta_{k/2}(\mathcal{F}, \mathcal{Q}) \leq q^{k/4} \cdot 2\varepsilon$. By Lemma 3.9, there exists a quasi-distribution $\mathcal{Q}'$ supported on $\mathbf{C}$ such that $\Delta_{k/2}(\mathcal{Q}, \mathcal{Q}') \leq q^{k/2} \cdot q^{k/4} \cdot 2\varepsilon$. Letting $\mathcal{G}$ be the $k/2$-non-signaling function corresponding to $\mathcal{Q}'$ completes the proof. □

7 Local characterizations and linear proofs

We prove Theorem 3 in this section. For this section, we let $k' \leq k$ be an integer. We let $\mathbf{C} \subseteq \mathbb{F}^n$ be a linear code, and $T \subseteq \mathbb{F}^n$ be a set of constraints. Given a k-non-signaling function $\mathcal{F}$, we say that $\mathcal{F}$ satisfies a constraint $\alpha \in \mathbb{F}^n_{\leq k}$ if $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = 1$.

Definition 7.1. *We let* $\mathsf{Consistent}(T, k)$ *denote the set of k-non-signaling functions $\mathcal{F}$ where* $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = 1$ *for every $\alpha \in T$. That is,* $\mathsf{Consistent}(T, k)$ *is the set of k-non-signaling functions that are consistent with T.*

We note that by Lemma 5.7, $\mathsf{Consistent}(\mathbf{C}^{\perp}_{\leq k'}, k)$ is the set of k-non-signaling functions that are $(\mathbf{C}, k')$-explainable.

With the above definition, the definition of local characterization can be rephrased as follows.

Definition 7.2. *For $\ell \leq k' \leq k$, a set of constraints $T \subseteq \mathbf{C}^{\perp}_{\leq \ell}$ is a ℓ-local characterization of $(\mathbf{C}, k', k)$ if* $\mathsf{Consistent}(T, k)$ *equals the set of k-non-signaling functions that are $(\mathbf{C}, k')$-explainable, i.e. that* $\mathsf{Consistent}(T, k) = \mathsf{Consistent}(\mathbf{C}^{\perp}_{\leq k'}, k)$.

In this language, [CMS18] shows that $T = \{e_x + e_y - e_{x+y} : x, y \in \{0,1\}^n\}$ is a 3-local characterization of $(\mathbf{C}, k-1, k)$, where $\mathbf{C}$ is the Hadamard code.

We briefly recall the definition of a k-local linear proof introduced in Section 1.2.

Definition 7.3 (k-local linear proof)**.** *Given a constraint set T and $\alpha \in \mathbb{F}^n$, we write $T \vdash_k \alpha$ if there exists a sequence $(\alpha_0 := 0^n, \alpha_1, \ldots, \alpha_{r-1}, \alpha_r := \alpha)$ with each $\alpha_i \in \mathbb{F}^n$ such that, for every $i \in [r]$, one of the following holds:*

- *$\exists\, j < i$ and $b \in \mathbb{F}$ such that $\alpha_i = b\alpha_j$*
- *$\exists\, j < i$ and $\gamma \in T$ such that $|\mathrm{supp}(\alpha_j) \cup \mathrm{supp}(\gamma)| \leq k$ and $\alpha_i = \alpha_j + \gamma$*
- *$\exists\, j_1, j_2 < i$ such that $|\mathrm{supp}(\alpha_{j_1}) \cup \mathrm{supp}(\alpha_{j_2})| \leq k$ and $\alpha_i = \alpha_{j_1} + \alpha_{j_2}$.*

Theorem 3 is stated formally as the theorem below.

Theorem 7.1 (Formal version of Theorem 3)**.** *k-local linear proofs are complete and sound for k-non-signaling functions. In particular, for $\ell \leq k' \leq k$, a set of constraints $T \subseteq \mathbf{C}^{\perp}_{\leq \ell}$ is a ℓ-local characterization of $(\mathbf{C}, k', k)$ if and only if $T \vdash_k \mathbf{C}^{\perp}_{\leq k'}$.*

The proof of Theorem 7.1 relies on the notion of a k-local subspace, which we define below.

Definition 7.4. *A k-local subspace $\mathcal{V}$ is a subset of $\mathbb{F}^n_{\leq k}$ where $\mathcal{V}_{\subseteq S} \subseteq \mathbb{F}^n$ is a linear subspace for every $S \subseteq [n]_{\leq k}$.*

We prove Theorem 7.1 by showing the following three lemmas.

Lemma 7.5 (Soundness)**.** *If $T \vdash_k \alpha$, then* $\mathsf{Consistent}(T, k) = \mathsf{Consistent}(T \cup \{\alpha\}, k)$.

Lemma 7.6. *For every k-local subspace $\mathcal{V} \subseteq \mathbb{F}^n_{\leq k}$, there exists a k-non-signaling function $\mathcal{F}$ such that $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = 1$ for every $\alpha \in \mathcal{V}$, and $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = \frac{1}{|\mathbb{F}|}$ otherwise.*

Lemma 7.7. *$\{\alpha \in \mathbb{F}^n : T \vdash_k \alpha\} \subseteq \mathbb{F}^n_{\leq k}$ is a k-local subspace.*

The following corollary follows immediately from Lemma 7.6 and Lemma 7.7.

Corollary 7.8 (Strong completeness)**.** *There exists a k-non-signaling function such that* $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = 1$ *for every α where $T \vdash_k \alpha$, and* $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = \frac{1}{|\mathbb{F}|}$ *otherwise.*

Proof of Theorem 7.1. Completeness and soundness are shown in Corollary 7.8 and Lemma 7.5. It remains to show the equivalence for local characterizations.

Suppose that $T \vdash_k \mathbf{C}^{\perp}_{\leq k'}$. Then by Lemma 7.5 we have that $\mathsf{Consistent}(T, k) = \mathsf{Consistent}(T \cup \mathbf{C}^{\perp}_{\leq k'}, k)$. Since $T \subseteq \mathbf{C}^{\perp}_{\leq \ell}$ and $\ell \leq k'$, we get that $T \subseteq \mathbf{C}^{\perp}_{\leq k'}$. Hence, $\mathsf{Consistent}(T, k) = \mathsf{Consistent}(T \cup \mathbf{C}^{\perp}_{\leq k'}, k) = \mathsf{Consistent}(\mathbf{C}^{\perp}_{\leq k'}, k)$, as required.

Conversely, suppose that T is an ℓ-local characterization of $(\mathbf{C}, k', k)$. By Lemma 7.6 and Lemma 7.7, there exists a k-non-signaling function $\mathcal{F}$ such that $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = 1$ for every $\alpha \in \mathbb{F}^n_{\leq k}$ such that $T \vdash_k \alpha$, and $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = \frac{1}{|\mathbb{F}|}$ otherwise. Since $T \vdash_k \alpha$ for every $\alpha \in T$, it follows that $\mathcal{F} \in \mathsf{Consistent}(T, k)$, which implies that $\mathcal{F} \in \mathsf{Consistent}(\mathbf{C}^{\perp}_{\leq k'}, k)$ as T is an ℓ-local characterization of $(\mathbf{C}, k', k)$. This implies that $T \vdash_k \alpha$ for all $\alpha \in \mathbf{C}^{\perp}_{\leq k'}$, since for all such α it holds that $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = 1$, and thus $T \vdash_k \alpha$. Hence, $T \vdash_k \mathbf{C}^{\perp}_{\leq k'}$, as required. □

7.1 Proof of Lemma 7.5

It is clear that from the definition that $\mathsf{Consistent}(T, k) \supseteq \mathsf{Consistent}(T \cup \{\alpha\}, k)$ for all $\alpha \in \mathbb{F}^n$. Below we prove the containment in the other direction. Suppose that $T \vdash_k \alpha$, and let $(\alpha_0 = 0^n, \alpha_1, \ldots, \alpha_r = \alpha)$ be a k-local proof of α from T. Let $\mathcal{F} \in \mathsf{Consistent}(T, k)$, that is, $\mathcal{F}$ is a k-non-signaling function such that $\forall \gamma \in T$, $\Pr[\langle \gamma, \mathcal{F} \rangle = 0] = 1$. We prove by induction that for $i \in [r]$ it holds that $\Pr[\langle \alpha_i, \mathcal{F} \rangle = 0] = 1$.

For the base case of $i = 0$ it must be the case that $\alpha_0 = 0^n$. Therefore, $\Pr[\langle \alpha_0, \mathcal{F} \rangle = 0] = 1$. For the induction step let $i \geq 1$, and consider the following three cases.

1. There exists $j < i$ and $b \in \mathbb{F} \setminus \{0^n\}$ such that $\alpha_i = b\alpha_j$. Then,

   $$\Pr[\langle \alpha_i, \mathcal{F} \rangle = 0] = \Pr[b\langle \alpha_j, \mathcal{F} \rangle = 0] = \Pr[\langle \alpha_j, \mathcal{F} \rangle = 0] = 1 \ ,$$

   where the last equality uses the induction hypothesis.

2. There exist $j < i$ and $\gamma \in T$ such that $\alpha_i = \alpha_j + \gamma$ with $|\mathrm{supp}(\alpha_j) \cup \mathrm{supp}(\gamma)| \leq k$. Since $\mathcal{F} \in \mathsf{Consistent}(T, k)$ we have that $\Pr[\langle \gamma, \mathcal{F} \rangle = 0] = 1$, as $\gamma \in T$. Therefore,

   $$\Pr[\langle \alpha_i, \mathcal{F} \rangle = 0] = \Pr[\langle \alpha_j, \mathcal{F} \rangle + \langle \gamma, \mathcal{F} \rangle = 0] \geq \Pr[\langle \alpha_j, \mathcal{F} \rangle = 0 \wedge \langle \gamma, \mathcal{F} \rangle = 0] = 1 \ ,$$

   as required. Note that $\Pr[\langle \alpha_j, \mathcal{F} \rangle = 0 \wedge \langle \gamma, \mathcal{F} \rangle = 0]$ is well-defined since $|\mathrm{supp}(\alpha_{j_1}) \cup \mathrm{supp}(\gamma)| \leq k$.

3. There exist $j_1, j_2 < i$ such that $\alpha_i = \alpha_{j_1} + \alpha_{j_2}$ and $|\mathrm{supp}(\alpha_{j_1}) \cup \mathrm{supp}(\alpha_{j_2})| \leq k$. By the induction hypothesis we know that $\Pr[\langle \alpha_{j_1}, \mathcal{F} \rangle = 0] = 1$ and $\Pr[\langle \alpha_{j_2}, \mathcal{F} \rangle = 0] = 1$. Thus,

   $$\Pr[\langle \alpha_i, \mathcal{F} \rangle = 0] = \Pr[\langle \alpha_{j_1}, \mathcal{F} \rangle + \langle \alpha_{j_2}, \mathcal{F} \rangle = 0] \geq \Pr[\langle \alpha_{j_1}, \mathcal{F} \rangle = 0 \wedge \langle \alpha_{j_2}, \mathcal{F} \rangle = 0] = 1 \ ,$$

   and therefore $\Pr[\langle \alpha_i, \mathcal{F} \rangle = 0] = 1$. Again, we require $|\mathrm{supp}(\alpha_{j_1}) \cup \mathrm{supp}(\alpha_{j_2})| \leq k$ in order for the last probability to be well-defined.

In particular, this implies that $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = \Pr[\langle \alpha_r, \mathcal{F} \rangle = 0] = 1$, and hence $\mathcal{F} \in \mathsf{Consistent}(T \cup \{\alpha\}, k)$. Therefore $\mathsf{Consistent}(T, k) \subseteq \mathsf{Consistent}(T \cup \{\alpha\}, k)$, which completes the proof of Lemma 7.5.

7.2 Proof of Lemma 7.6

We define $\mathcal{F}$ specifying its local distributions $\mathcal{F}_S$ for each $S \subseteq [n]_{\leq k}$. We define the function $\mathcal{F}_S \colon \mathbb{F}^S \to \mathbb{C}$ by specifying its (local) Fourier coefficients as follows. We set the Fourier coefficient $\widehat{\mathcal{F}_S}(\alpha)$ to be $\frac{1}{q^{|S|}}$ if $\alpha \in \mathcal{V}$, and 0 otherwise.

We now show that each $\mathcal{F}_S$ is a distribution. For any $f \colon S \to \mathbb{F}$ we have

$$\mathcal{F}_S(f) = \sum_{\alpha \in \mathbb{F}^S} \widehat{\mathcal{F}_S}(\alpha)\chi_\alpha(f) = \sum_{\alpha \in \mathcal{V}_{\subseteq S}} \frac{1}{q^{|S|}}\chi_\alpha(f) = \frac{1}{q^{|S|}} \sum_{\alpha \in \mathcal{V}_{\subseteq S}} \omega^{\mathrm{Tr}(\langle \alpha, f \rangle)} \ .$$

For each $b \in \mathbb{F}$, let $\mathcal{V}_b \subseteq \mathcal{V}_{\subseteq S}$ be the set of $\alpha \in \mathcal{V}_{\subseteq S}$ where $\langle \alpha, f \rangle = b$. Let $\pi \colon \mathcal{V}_{\subseteq S} \to \mathbb{F}$ be the map where $\pi(\alpha) = \langle \alpha, f \rangle$. Since $\mathcal{V}_{\subseteq S}$ is a subspace, π is a homomorphism. It follows that either $\mathcal{V}_0 = \mathcal{V}_{\subseteq S}$ or $|\mathcal{V}_b| = |\mathcal{V}_0|$ for every $b \in \mathbb{F}$. In the first case, $\sum_{\alpha \in \mathcal{V}_{\subseteq S}} \omega^{\mathrm{Tr}(\langle \alpha, f \rangle)} = |\mathcal{V}_{\subseteq S}| \geq 0$. In the second case,

$$\sum_{\alpha \in \mathcal{V}_{\subseteq S}} \omega^{\mathrm{Tr}(\langle \alpha, f \rangle)} = \sum_{b \in \mathbb{F}} \sum_{\alpha \in \mathcal{V}_b} \omega^{\mathrm{Tr}(b)} = \sum_{b \in \mathbb{F}} |\mathcal{V}_b| \, \omega^{\mathrm{Tr}(b)} = |\mathcal{V}_0| \sum_{b \in \mathbb{F}} \omega^{\mathrm{Tr}(b)} = 0 \ .$$

This implies that in either case, $\mathcal{F}_S(f) \geq 0$, and so $\mathcal{F}_S$ is a distribution.

We now show that the collection of local distributions $\{\mathcal{F}_S\}_{S \subseteq [n]_{\leq k}}$ is indeed non-signaling. This follows from Lemma 4.9. If $\alpha \in \mathcal{V}$ then we have that $q^{|S|}\widehat{\mathcal{F}_S}(\alpha) = 1 = q^{|R|}\widehat{\mathcal{F}_R}(\alpha)$ for every $S, R \in [n]_{\leq k}$ such that $\mathrm{supp}(\alpha) \subseteq S \cap R$, and otherwise we have $q^{|S|}\widehat{\mathcal{F}_S}(\alpha) = 0 = q^{|R|}\widehat{\mathcal{F}_R}(\alpha)$. Thus, the collection of local distributions is a k-non-signaling function $\mathcal{F}$.

It remains to show that $\mathcal{F}$ satisfies the desired property. Observe that for every α, $q^n\widehat{\mathcal{F}}(\alpha) = q^{|\mathrm{supp}(\alpha)|}\widehat{\mathcal{F}}_{\mathrm{supp}(\alpha)}(\alpha) = 1$ if $\alpha \in \mathcal{V}$, and otherwise $\widehat{\mathcal{F}}(\alpha) = 0$. By Lemma 4.3 it follows that $\mathcal{F}$ has the desired properties.

7.3 Proof of Lemma 7.7

Let $\mathcal{V} = \{\alpha \in \mathbb{F}^n : T \vdash_k \alpha\}$. We show that $\mathcal{V}$ is a k-local subspace. Let $S \subseteq [n]_{\leq k}$. We need to show that $\mathcal{V}_{\subseteq S}$ is a linear subspace of $\mathbb{F}^n$. We first observe that 0^n is always in the set, as $T \vdash_k 0^n$ always.

Let $\alpha \in \mathcal{V}_{\subseteq S}$ and let $b \in \mathbb{F} \setminus \{0\}$. Then we have that $T \vdash_k \alpha$ which implies that $T \vdash_k b\alpha$. Since $\mathrm{supp}(b\alpha) = \mathrm{supp}(\alpha) \subseteq S$, it follows that $b\alpha \in \mathcal{V}_{\subseteq S}$.

Let $\alpha, \beta \in \mathcal{V}_{\subseteq S}$. Then, since $|\mathrm{supp}(\alpha) \cup \mathrm{supp}(\beta)| \leq |S| \leq k$ we have that $(\alpha, \beta, \alpha + \beta)$ is a hyperedge in Γ_k. Thus, since $T \vdash_k \{\alpha, \beta\}$ it follows that $T \vdash_k \alpha + \beta$. Since $\mathrm{supp}(\alpha + \beta) \subseteq \mathrm{supp}(\alpha) \cup \mathrm{supp}(\beta) \subseteq S$, it follows that $\alpha + \beta \in \mathcal{V}_{\subseteq S}$.

We have thus shown that $\mathcal{V}_{\subseteq S}$ is a linear subspace of $\mathbb{F}^n$, which completes the proof.

8 Low-degree testing fails for small locality

In this section, we prove Theorem 2. The proof relies heavily on Theorem 3.

We let $\mathbf{C}$ be the linear code of m-variate polynomials $P\colon \mathbb{F}^m \to \mathbb{F}$ of total degree at most d, with $m \geq 2$, and let T be the set of α's in $\mathbf{C}^\perp$ where the $\mathrm{supp}(\alpha)$ is contained in exactly one line.

We define the *rank* of an element in $\mathbf{C}^\perp$ to be

$$\mathsf{rank}_T(\alpha) := \min_{T' \subseteq T : \alpha \in \mathrm{span}(T')} \left|T'\right| \ .$$

Note that since $\mathrm{span}(T) = \mathbf{C}^\perp$, the rank of α is well-defined for all $\alpha \in \mathbf{C}^\perp$.

We let T_0 denote the subset of T that only contains elements whose support is *evenly-spaced* along a line and has weight $d+2$. With this notation, the non-signaling evenly-spaced test (i) samples $\alpha \leftarrow T_0$ uniformly at random, and (ii) checks that $\langle \alpha, \mathcal{F} \rangle = 0$.

The main theorem we prove is stated below, and is the formal statement of Theorem 2.

Theorem 8.1 (Formal version of Theorem 2)**.** *For every k with $2d+2 \leq k < \frac{3}{16}(d+2)^2$, there exists a k-non-signaling function such that $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = 1$ for every $\alpha \in T_{\leq k}$, and yet $\Delta_{2d+2}(\mathcal{F}, \mathcal{F}') \geq (1 - \frac{1}{|\mathbb{F}|})$ for every $(2d+2)$-non-signaling function $\mathcal{F}'$ that is $(\mathbf{C}, 2d+2)$-explainable.*

We begin the proof of Theorem 8.1 by showing the following lemma. This lemma follows from earlier statements, and outlines a sufficient condition to prove Theorem 8.1

Lemma 8.1. *Suppose that there exists $\alpha^* \in \mathbf{C}^\perp$ with $\mathrm{wt}(\alpha^*) = 2d+2$ such that for every $k < \frac{3}{16}(d+2)^2$ it holds that $T \not\vdash_k \alpha^*$. Then for every k with $2d+2 \leq k < \frac{3}{16}(d+2)^2$ there exists a k-non-signaling function $\mathcal{F}$ such that $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = 1$ for every $\alpha \in T_{\leq k}$, and yet $\Delta_{2d+2}(\mathcal{F}, \mathcal{F}') \geq 1 - \frac{1}{|\mathbb{F}|}$ for every $(2d+2)$-non-signaling function $\mathcal{F}'$ that is $(\mathbf{C}, 2d+2)$-explainable.*

Proof. Applying Corollary 7.8, for every k with $2d+2 \leq k < \frac{3}{16}(d+2)^2$, we get that there exists a k-non-signaling function $\mathcal{F}$ such that $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = 1$ for every $\alpha \in T_{\leq k}$ and $\Pr[\langle \alpha^*, \mathcal{F} \rangle = 0] = \frac{1}{|\mathbb{F}|}$. Let $\mathcal{F}'$ be a $(2d+2)$-non-signaling function that is $(\mathbf{C}, 2d+2)$-explainable. Since for every $S \subseteq \mathbb{F}^m$ with $|S| \leq 2d+2$ we have that $\Pr[\mathcal{F}'(S) \in \mathbf{C}|_S] = 1$ and $\alpha^* \in \mathbf{C}^\perp$ has $\mathrm{wt}(\alpha^*) = 2d+2$, it follows that $\Pr[\langle \alpha^*, \mathcal{F}' \rangle = 0] = 1$. Therefore, $\Delta_{2d+2}(\mathcal{F}, \mathcal{F}') \geq |\Pr[\langle \alpha^*, \mathcal{F} \rangle = 0] - \Pr[\langle \alpha^*, \mathcal{F}' \rangle = 0]| = 1 - \frac{1}{|\mathbb{F}|}$. □

By Lemma 8.1 it suffices to find such an α^*. We let $\alpha^* \in \mathbf{C}^\perp$ be any constraint where $\mathrm{supp}(\alpha)$ has size $2d+2$ and is contained on the curve $x_1^2 - x_2 = 0$ embedded on the plane $x_3 = x_4 = \cdots = x_m = 0$ in $\mathbb{F}^m$. We note that α^* is one of the constraints that checks that $P(t, t^2, 0, \ldots, 0)$ is a univariate polynomial of degree at most $2d$ in t.

We show that α^* satisfies the desired properties in two main lemmas. We first show the following generic lemma, which gives us a way to prove that $T \not\vdash_k \alpha^*$.

Lemma 8.2 (Interval cut Lemma)**.** *Fix $\alpha \in \mathbf{C}^\perp$. Suppose that there exists $r \in \mathbb{R}$ with $2 \leq r \leq \mathsf{rank}_T(\alpha)$ such that for every $\beta \in \mathbf{C}^\perp$ with $\mathsf{rank}_T(\beta) \in [r/2, r)$ it holds that $T \not\vdash_k \beta$. Then $T \not\vdash_k \alpha$.*

We then show that every $\beta \in \mathbf{C}^\perp$ of rank in $[(d+2)/4, (d+2)/2)$ must have large weight, implying that they are not provable from T when k is small.

Lemma 8.3. *For every $\beta \in \mathbf{C}^\perp$ with $\mathsf{rank}_T(\beta) \in [(d+2)/4, (d+2)/2)$ it holds that $\mathrm{wt}(\beta) \geq \frac{3}{16}(d+2)^2$. In particular, if $k < \frac{3}{16}(d+2)^2$ then $T \not\vdash_k \beta$.*

With the above two lemmas, we now finish the proof of Theorem 8.1.

Proof of Theorem 8.1. Let $k < \frac{3}{16}(d+2)^2$. We first show that $\mathsf{rank}_T(\alpha^*) \geq d+1$. Since the curve $x_1^2 - x_2 = 0$ is irreducible in $\mathbb{F}[x_1, x_2, \ldots, x_m]$, any line L intersects the curve on at most 2 distinct points. It follows that $\mathsf{rank}_T(\alpha^*) \geq (2d+2)/2 = d+1$, as any constraint $\beta \in T$ can only have at most 2 points on the curve $x_1^2 - x_2 = 0$.

Since $k < \frac{3}{16}(d+2)^2$, Lemma 8.3 implies that $T \nvdash_k \beta$ for every β with $\mathsf{rank}_T(\beta) \in [(d+2)/4, (d+2)/2)$. Thus, by Lemma 8.2 it follows that $T \nvdash_k \alpha^*$. Hence, α^* satisfies the assumptions of Lemma 8.1, and so applying Lemma 8.1 completes the proof of Theorem 8.1. □

Next we turn to the proofs of Lemma 8.2 and Lemma 8.3.

Proof of Lemma 8.2. First, observe that by definition of rank, $\mathsf{rank}_T(\alpha_1 + \alpha_2) \leq \mathsf{rank}_T(\alpha_1) + \mathsf{rank}_T(\alpha_2)$. By the assumption of the lemma, there exists $r \in \mathbb{R}$ with $2 \leq r \leq \mathsf{rank}_T(\alpha)$ such that for every $\beta \in \mathbf{C}^\perp$ with $\mathsf{rank}_T(\beta) \in [r/2, r)$ it holds that $T \nvdash_k \beta$. We need to show that $T \nvdash_k \alpha$.

Suppose toward a contradiction that $T \vdash_k \alpha$. Then there exists a path $(\alpha_1, \ldots, \alpha_t = \alpha)$ in $\Gamma_k(\mathbf{C}^\perp, T)$ from 0^n to α. Let S_1 be the set of α_i's such that $\mathsf{rank}_T(\alpha_i) < r/2$, and let S_2 be the set of α_i's such that $\mathsf{rank}_T(\alpha_i) \geq r$. Note that $S_1 \cup S_2 = \{\alpha_1, \ldots, \alpha_t\}$, as otherwise there would exist some i such that α_i has rank in $[r/2, r)$, which would contradict the assumption that $T \vdash_k \alpha_i$ for all $i \in [t]$.

Since $\mathsf{rank}_T(\alpha) \geq r$ it follows that $\alpha \in S_2$, and hence $S_2 \neq \emptyset$. Let ℓ be the smallest index such that $\alpha_\ell \in S_2$. We have that $\alpha_\ell \neq 0^n$ since $\alpha_\ell \in S_2$, and there does not exist $i < \ell$ and $b \in \mathbb{F} \setminus \{0\}$ such that $\alpha_\ell = b\alpha_i$, as then $\mathsf{rank}_T(\alpha_i) = \mathsf{rank}_T(\alpha_\ell) \geq r$, thus contradicting the minimality of ℓ. Suppose that there exists $i < \ell$ and $\gamma \in T$ such that $\alpha_\ell = \alpha_i + \gamma$. By the minimality of ℓ, we must have that $\alpha_i \in S_1$, and hence $r \leq \mathsf{rank}_T(\alpha_\ell) \leq \mathsf{rank}_T(\alpha_i) + \mathsf{rank}_T(\gamma) < r/2 + 1 \leq r/2 + r/2 = r$, which is also a contradiction. Therefore, there must either exist $j_1, j_2 < \ell$ such that $\alpha_\ell = \alpha_{j_1} + \alpha_{j_2}$. By the minimality of ℓ, we must have that $\alpha_{j_1}, \alpha_{j_2} \in S_1$, and hence $r \leq \mathsf{rank}_T(\alpha_\ell) \leq \mathsf{rank}_T(\alpha_{j_1}) + \mathsf{rank}_T(\alpha_{j_1}) < r/2 + r/2 = r$, which is, again, a contradiction. In all cases we have reached a contradiction to the assumption that $T \vdash_k \alpha$, which completes the proof of Lemma 8.2. □

Remark 8.4. We note that in the foregoing proof we only required that rank_T is subadditive, i.e., that $\mathsf{rank}_T(\alpha_1 + \alpha_2) \leq \mathsf{rank}_T(\alpha_1) + \mathsf{rank}_T(\alpha_2)$, $\mathsf{rank}_T(\alpha) = 1$ for every $\alpha \in T$, and $\mathsf{rank}_T(0^n) = 0$. Thus, the Interval Cut Lemma holds for any such subadditive function.

Proof of Lemma 8.3. Let $\beta \in \mathbf{C}^\perp$ be such that $\mathsf{rank}_T(\beta) = r \in [(d+2)/4, (d+2)/2)$. Then there exist lines $L_1, \ldots, L_r$ such that $\beta = \sum_{i=1}^r \beta_i$ where $\mathrm{supp}(\beta_i) \subseteq L_i$. The L_i's must be distinct, as otherwise we could add two constraints contained in the same line and we would then get $\mathsf{rank}_T(\beta) < r$. We have that $\mathrm{wt}(\beta_i) \geq d+2$ for each i. Hence, $\mathrm{wt}(\beta) \geq r(d+2) - 2\binom{r}{2}$, since each β_i contributes at least $d+2$ to the weight, and there are at most $\binom{r}{2}$ intersection points as each of the r lines is distinct. The function $f(r) = r(d+2) - r^2$ for $r \in [\frac{d+2}{4}, \frac{d+2}{2})$ is minimized when $r = \frac{d+2}{4}$, and hence $\mathrm{wt}(\beta) \geq r(d+2) - r^2 \geq \frac{3}{16}(d+2)^2$, which completes the proof. □

A Separating classical and non-signaling local characterizations

In this section we prove Theorem 4.

Theorem A.1 (Formal version of Theorem 4)**.** *Fix $d \in \mathbb{N}$ with $d \geq 2$, and let $n \in \mathbb{N}$ such that $2n \equiv 0 \mod d$. There exists a code $\mathbf{C} \subseteq \mathbb{F}_2^n$ and a constraint set T such that*

- *For classical functions, T is a d-local characterization of $\mathbf{C}$,*
- *For non-signaling functions, T is not a d-local characterization of $(\mathbf{C}, 2d, k)$ for every $k \leq \frac{n-d}{6}$.*

Theorem A.1 follows from the lemma below.

Lemma A.1. *Fix $d \in \mathbb{N}$ with $d \geq 2$, and let $n \in \mathbb{N}$ such that $2n \equiv 0 \mod d$. There exists a collection $T \subseteq \mathbb{F}_2^n$ of $2n/d - 2$ linearly independent vectors such that:*

1. $\mathrm{wt}(v) = d$ *for all* $v \in T$.
2. $\mathrm{wt}(\sum_{v \in T} v) = 2d$.
3. *For all $S \subseteq T$ of size $|S| \leq n/2$ it holds that* $\mathrm{wt}(\sum_{v \in S} v) \geq \frac{d|S|}{3}$.

A lemma of similar flavor is used in [BHR05] to prove that a random LDPC code is not locally testable, as well as in the proof of lower bounds for SoS algorithms used to solve the MAX-3-XOR problem [Gri01; Sch08]. We now prove Theorem A.1 assuming Lemma A.1.

Proof of Theorem A.1. Let $T \subseteq \mathbb{F}_2^n$ be the set of vectors from Lemma A.1, and let $\mathbf{C} = \mathrm{span}(T)^\perp$. By construction, T is a classical d-local characterization of $\mathbf{C}$.

Fix $k < \frac{n-d}{6}$, and let $\alpha = \sum_{v \in T} v$. Note that $\alpha \in \mathrm{span}(T)$ and that by Lemma A.1, $\mathrm{wt}(\alpha) = 2d$. By Theorem 7.1, to show that T is not a d-local characterization of $(\mathbf{C}, 2d, k)$, it suffices to show that $T \nvdash_k \alpha$.

We have that $\mathsf{rank}_T(\alpha) = |T| = 2n/d - 2$. Suppose that $T \vdash_k \alpha$. Then by Lemma 8.2, there exists $\beta \in \mathrm{span}(T)$ such that $\mathsf{rank}_T(\beta) \in [(n/d-1)/2, n/d-1)$ and $T \vdash_k \beta$. Since T is a linearly independent set of vectors, we must have that $\beta = \sum_{v \in S} v$ for some $S \subseteq T$ and that $\mathsf{rank}_T(\beta) = |S|$. Therefore, $|S| \in [(n/d-1)/2, n/d-1)$, and in particular $|S| \leq n/d - 1 \leq n/2$. By Item 3 of Lemma A.1, we get that $\mathrm{wt}(\beta) \geq \frac{d|S|}{3} \geq \frac{n-d}{6}$. This implies that $T \nvdash_k \beta$, since $k < \frac{n-d}{6} \leq \mathrm{wt}(\beta)$, a contradiction. Thus, $T \nvdash_k \alpha$, which completes the proof. □

Proof of Lemma A.1. Let $G = (V, E)$ be a d-regular $(|V|/2, d/3)$-edge expander with $|V| = 2n/d$ vertices and $|E| = n$ edges, and denote the edges by $e_1, \ldots, e_n$. For each vertex $u \in V$ let $v_u \in \mathbb{F}_2^n$ be defined as

$$(v_u)_i = \begin{cases} 1, & \text{if } e_i \text{ is adjacent to } v \\ 0, & \text{otherwise} \quad . \end{cases}$$

Let $T' = \{v_u : u \in V\}$ be the collection of all vectors corresponding to the vertices of V.

We claim that for all $S \subseteq T'$ of size $|S| \leq n/d$ it holds that $\mathrm{wt}(\sum_{v \in S} v) \geq \frac{|S| \cdot d}{3}$. Indeed, let $V_S \subseteq V$ be the subset of vertices corresponding to the vectors in S. Observe that $\mathrm{wt}(\sum_{v \in S} v) = |E(V_S, V \setminus V_S)|$, and hence, since G is an $(n/d, d/3)$-expander graph, it follows that $\mathrm{wt}(\sum_{v \in S} v) = |E(V_S, V \setminus V_S)| \geq \frac{d|V_S|}{3} = \frac{d|S|}{3}$. Also, observe that T' is *not* linearly independent. However, the following claim shows that removing any vector from T' produces a linearly independent set of vectors.

Claim A.2. *Suppose that* $\sum_{i=1}^{n} a_i v_i = 0$ *for some* $a_i \in \mathbb{F}_2$. *Then either* $a_i = 0$ *for all* $i \in [n]$ *or* $a_i = 1$ *for all* $i \in [n]$.

Proof. Let $V_S \subseteq V$ be the set of vertices corresponding to a_i's, i.e, $S = \{i : a_i = 1\}$. Then $\sum_{i=1}^{n} a_i v_i = |E(V_S, V \setminus V_S)|$. By edge expansion the graph G is connected, and hence V_S is either $\emptyset$ or V, and the claim follows. □

Finally, we define the set T by removing two arbitrary vectors with disjoint support from T'. By Claim A.2 the vectors in T are linearly independent. Since $\mathrm{wt}(\sum_{v \in T'} v) = 0$ and the two vectors removed have disjoint support, we see that $\mathrm{wt}(\sum_{v \in T} v) = 2d$. We also have that $\mathrm{wt}(\sum_{v \in S} v) \geq \frac{|S| \cdot d}{3}$ for all $S \subseteq T$ of size at most $|S| \leq n/d$, as this property held for T', and $T \subseteq T'$. □

B On robust local characterizations

In this section, we prove a robust analogue of Theorem 3. We consider the case where a non-signaling function $\mathcal{F}$ satisfies *every* constraint α in T with high probability (as opposed to probability 1, as in the exact case). This is different from the classical case where we assume that a function f satisfies a random constraint α in T (sampled from a distribution over T) with high probability. In the non-signaling setting, the assumption that $\mathcal{F}$ satisfies every constraint with high probability is typical, as for natural codes (e.g., Hadamard and Reed–Muller codes), $\mathcal{F}$ satisfying a *random* constraint α with high probability implies that its self-correction satisfies *every* constraint $\alpha \in T$ with high probability.

Informally, we let $\mathsf{nsrank}_T(\alpha)$ denote the length of the shortest k-local linear proof of α from T. We then show the following robust analogue of Theorem 3.

Theorem 5. *Let $T \subseteq \mathbf{C}^\perp$ be set of constraints each of weight at most k.*

1. *Suppose that a k-non-signaling function $\mathcal{F}$ satisfies every $\alpha \in T$ with probability at least $1-\varepsilon$. Then $\mathcal{F}$ satisfies every α where $T \vdash_k \alpha$ with probability at least $1 - \mathsf{nsrank}_T(\alpha)\varepsilon$.*

2. *Conversely, there exists a k-non-signaling function $\mathcal{F}$ that satisfies every α where $T \vdash_k \alpha$ with probability exactly $1 - \mathrm{wt}(\alpha)\varepsilon$, and every other α with probability $\frac{1}{|\mathbb{F}|}$.*

We additionally show that $\mathsf{nsrank}_T(\alpha) \geq \mathrm{wt}(\alpha)/\mathrm{wt}(T)$ where $\mathrm{wt}(T) = \max_{\gamma \in T} \mathrm{wt}(\gamma)$. We then show that if $\mathbf{C} = \{(b, \ldots, b) : b \in \mathbb{F}_2\}$ is the repetition code and $T = \{e_i + e_j : i, j \in [n]\}$ is the natural 2-local test, then $\mathsf{nsrank}_T(\alpha) = \mathrm{wt}(\alpha)/2$, showing that Theorem 5 is tight for some choice of $\mathbf{C}$ (by replacing ε with $\varepsilon/2$ in the second bullet). Finally, if $\mathbf{C}$ is the Hadamard code, then $\mathrm{wt}(\alpha)/3 \leq \mathsf{nsrank}_T(\alpha) \leq \mathrm{wt}(\alpha)$, implying that Theorem 5 is tight up to a constant factor in this case.

In Appendix B.1 we prove part 1 of Theorem 5, and in Appendix B.2 we prove part 2. Finally, in Appendix B.3 we show that Theorem 5 is tight for the repetition code and is tight up to a constant factor for the Hadamard code.

B.1 Part 1 of Theorem 5

We prove part 1 of Theorem 5. In order to do this, we must first formally define nsrank_T. We define

$$\mathsf{nsrank}_T(\alpha) := \min_P \mathsf{cost}_P(\alpha) \ ,$$

where the minimum is taken over all k-local linear proofs P of α from T, and $\mathsf{cost}_P(\alpha)$ is defined according to the following definition:

Definition B.1. *Let $P = (\alpha_0, \ldots, \alpha_r)$ be a proof of α from T as in Definition 1.5. For each $i \in \{0, \ldots, r\}$, we define $\mathsf{cost}_P(\alpha_i)$ recursively as follows.*

1. *(Base case) $\mathsf{cost}_P(\alpha_0) = \mathsf{cost}_P(0^n) = 0$.*

2. *(Case 1) if there exists $j < i$ and $b \in \mathbb{F}$ such that $\alpha_i = b\alpha_j$, then $\mathsf{cost}_P(\alpha_i) = \mathsf{cost}_P(\alpha_j)$.*

3. *(Case 2) if there exists $j < i$ and $\gamma \in T$ such that $|\mathrm{supp}(\alpha_j) \cup \mathrm{supp}(\gamma)| \leq k$ and $\alpha_i = \alpha_j + \gamma$, then $\mathsf{cost}_P(\alpha_i) = \mathsf{cost}_P(\alpha_j) + 1$.*

4. *(Case 3) if there exists $j_1, j_2 < i$ such that $|\mathrm{supp}(\alpha_{j_1}) \cup \mathrm{supp}(\alpha_{j_2})| \leq k$ and $\alpha_i = \alpha_{j_1} + \alpha_{j_2}$, then* $\mathsf{cost}_P(\alpha_i) = \mathsf{cost}_P(\alpha_{j_1}) + \mathsf{cost}_P(\alpha_{j_2})$.

If more than one of the above cases hold for a particular α_i, then $\mathsf{cost}_P(\alpha_i)$ is defined to be the minimum over all possible cases.

We note that $\mathsf{nsrank}_T(\alpha)$ implicitly depends on k. In fact, when $k = n$ we have that $\mathsf{nsrank}_T(\alpha) = \mathsf{rank}_T(\alpha)$, which motivates nsrank as a non-signaling analogue of rank.

Using the definition above, we prove part 1 of Theorem 5. Suppose that $\mathcal{F}$ is a k-non-signaling function such that $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] \geq 1 - \varepsilon$ for every $\alpha \in T$. Let $\alpha \in \mathbb{F}^n_{\leq k}$ be such that $T \vdash_k \alpha$. We show that $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] \geq 1 - \mathsf{nsrank}_T(\alpha)\varepsilon$.

Let $P = (\alpha_0, \ldots, \alpha_r)$ be a proof of α from T such that $\mathsf{cost}_P(\alpha)$ is minimal, i.e., such that $\mathsf{nsrank}_T(\alpha) = \min_P \mathsf{cost}_P(\alpha)$. Let $\mathsf{cost}_P(\alpha_i)$ be the non-negative integers assigned to each $\alpha_i \in P$. We prove that for all $i \in \{0, \ldots, r\}$ it holds that $\Pr[\langle \alpha_i, \mathcal{F} \rangle = 0] \geq 1 - \mathsf{cost}_P(\alpha_i)\varepsilon$. The proof is by induction on i.

For the base case of $i = 0$ indeed holds $\Pr[\langle \alpha_0, \mathcal{F} \rangle = 0] = 1 = 1 - \mathsf{cost}_P(\alpha_0)\varepsilon$. For the induction step let $i \geq 1$, and consider the following three cases.

1. Case 1: Suppose there exists $j < i$ and $b \in \mathbb{F}$ such that $\alpha_i = b\alpha_j$. By the induction hypothesis $\Pr[\langle \alpha_j, \mathcal{F} \rangle = 0] \geq 1 - \mathsf{cost}_P(\alpha_j)\varepsilon$, and hence

$$\Pr[\langle \alpha_i, \mathcal{F} \rangle = 0] = \Pr[\langle b\alpha_j, \mathcal{F} \rangle = 0] = \Pr[\langle \alpha_j, \mathcal{F} \rangle = 0] \geq 1 - \mathsf{cost}_P(\alpha_j)\varepsilon = 1 - \mathsf{cost}_P(\alpha_i)\varepsilon \ .$$

2. Case 2: Suppose there exists $j < i$ and $\gamma \in T$ such that $|\mathrm{supp}(\alpha_j) \cup \mathrm{supp}(\gamma)| \leq k$ and $\alpha_i = \alpha_j + \gamma$. Then by the induction hypothesis $\Pr[\langle \alpha_j, \mathcal{F} \rangle = 0] \geq 1 - \mathsf{cost}_P(\alpha_j)\varepsilon$, and hence

$$\Pr[\langle \alpha_i, \mathcal{F} \rangle = 0] = \Pr[\langle \alpha_j + \gamma, \mathcal{F} \rangle = 0] \geq \Pr[\langle \alpha_j, \mathcal{F} \rangle = 0 \wedge \langle \gamma, \mathcal{F} \rangle = 0] \geq 1 - \mathsf{cost}_P(\alpha_j)\varepsilon - \varepsilon = 1 - \mathsf{cost}_P(\alpha_i)\varepsilon \ ,$$

   by union bound. Therefore, also in this case $\Pr[\langle \alpha_i, \mathcal{F} \rangle = 0] \geq 1 - \mathsf{cost}_P(\alpha_i)\varepsilon$.

3. Case 3: Otherwise, there exists $j_1, j_2 < i$ such that $|\mathrm{supp}(\alpha_{j_1}) \cup \mathrm{supp}(\alpha_{j_2})| \leq k$ and $\alpha_i = \alpha_{j_1} + \alpha_{j_2}$. Then,

$$\Pr[\langle \alpha_i, \mathcal{F} \rangle = 0] = \Pr[\langle \alpha_{j_1} + \alpha_{j_2}, \mathcal{F} \rangle = 0] \geq \Pr[\langle \alpha_{j_1}, \mathcal{F} \rangle = 0 \wedge \langle \alpha_{j_2}, \mathcal{F} \rangle = 0] \geq 1 - \mathsf{cost}_P(\alpha_{j_1})\varepsilon - \mathsf{cost}_P(\alpha_{j_2})\varepsilon \ ,$$

   by the induction hypothesis and union bound. Since $\mathsf{cost}_P(\alpha_i) = \mathsf{cost}_P(\alpha_{j_1}) + \mathsf{cost}_P(\alpha_{j_2})$, it follows that $\Pr[\langle \alpha_i, \mathcal{F} \rangle = 0] \geq 1 - \mathsf{cost}_P(\alpha_i)\varepsilon$, as required.

By induction, we conclude that $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] \geq 1 - \mathsf{cost}_P(\alpha_r)\varepsilon = 1 - \mathsf{nsrank}_T(\alpha)\varepsilon$, which completes the proof.

B.2 Part 2 of Theorem 5

We prove part 2 of Theorem 5 by showing the following lemma.

Lemma B.2. *Let $\mathsf{cost}\colon \mathbb{F}^n \to \mathbb{Z}_{\geq 0}$ be a function such that for every $\alpha \in \mathbb{F}^n$, if $\alpha = \sum_{i=1}^n \alpha_i e_i$, then $\mathsf{cost}(\alpha) = \sum_{i:\alpha_i \neq 0} \mathsf{cost}(e_i)$. Let $M\colon \mathbb{F}^n \to \mathbb{F}^n$ be a linear transformation. Let $\mathcal{W}$ be a k-local subspace, and let $\varepsilon \geq 0$ be such that $1 - \frac{q}{q-1}\mathsf{cost}(M\alpha)\varepsilon \geq 0$ for every $\alpha \in \mathcal{W}$. Then there exists a k-non-signaling function $\mathcal{F}$ such that $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = 1 - \mathsf{cost}(M\alpha)\varepsilon$ for every $\alpha \in \mathcal{W}$, and $\Pr[\langle \alpha, \mathcal{F} \rangle = 0] = \frac{1}{|\mathbb{F}|}$ otherwise.*

Part 2 of Theorem 5 follows from Lemma B.2 by setting $\mathsf{cost}(e_i) = 1$ for each $i \in [n]$, and letting M be the identity matrix.

Note that the assumptions on cost in Lemma B.2 imply that $\mathsf{cost}(0^n) = 0$, and for every $\alpha \in \mathbb{F}^n$ and $b \in \mathbb{F} \setminus \{0\}$, we have that $\mathsf{cost}(\alpha) = \mathsf{cost}(b\alpha)$. In addition, if we let $\pi_i \colon \mathbb{F}^n \to \mathbb{F}$ be the projection map $\alpha \mapsto \alpha_i$, and let $h_i \colon \mathbb{F}^n \to \mathbb{Z}$ be the map which sends $\alpha \mapsto 1$ if $\alpha_i \neq 0$ and $\alpha \mapsto 0$ otherwise, then $\mathsf{cost}(\alpha) = \sum_{i=1}^{n} h_i(\alpha)\mathsf{cost}(e_i)$.

The following lemma will be used in the proof. We delay the proof of the lemma until after the proof of Lemma B.2.

Lemma B.3. *Let $\mathcal{V} \subseteq \mathbb{F}^n$ be a linear subspace. Then for every subspace $\mathcal{V}_0 \subseteq \mathcal{V}$ of co-dimension 1 it holds that $\frac{1}{q-1}\sum_{\alpha \in \mathcal{V} \setminus \mathcal{V}_0} \mathsf{cost}(\alpha) - \sum_{\alpha \in \mathcal{V}_0} \mathsf{cost}(\alpha) \geq 0$.*

Proof of Lemma B.2. As in the proof of Lemma 7.6, we define each $\mathcal{F}_S$ first as a function $\mathbb{F}^S \to \mathbb{C}$ by specifying its Fourier coefficients. In particular, we set $\widehat{\mathcal{F}_S} = \frac{1}{q^{|S|}} \cdot (1 - \frac{q}{q-1}\mathsf{cost}(M\alpha)\varepsilon)$ if $\alpha \in \mathcal{W}$, and 0 otherwise.

We now finish the proof assuming that each $\mathcal{F}_S$ is in fact a distribution. By Lemma 4.9, it follows that the collection of local distributions $\mathcal{F} = \{\mathcal{F}_S\}_{S \subseteq [n]_{\leq k}}$ is k-non-signaling, and by Lemma 4.3 it follows that $\mathcal{F}$ has the desired properties, completing the proof.

It remains to show that each $\mathcal{F}_S$ is a distribution. Since $\mathsf{cost}(0^n) = 0$, we have that $\widehat{\mathcal{F}_S}(0^S) = \frac{1}{q^{|S|}}$, and hence $\sum_{f \in \mathbb{F}^S} \mathcal{F}_S(f) = 1$. So, it remains to show that $\mathcal{F}_S(f) \geq 0$ for each $f \in \mathbb{F}^S$.

Let $\mathcal{V} = \mathcal{W}_{\subseteq S}$. Note that by definition of $\mathcal{F}_S$ we have that

$$\mathcal{F}_S(f) = \sum_{\alpha \in \mathbb{F}^S} \widehat{\mathcal{F}_S}(\alpha)\chi_\alpha(f) = \sum_{\alpha \in \mathcal{V}} \left(1 - \frac{q}{q-1}\mathsf{cost}(M\alpha)\varepsilon\right) \cdot \frac{1}{q^{|S|}} \omega^{-\mathrm{Tr}(\langle \alpha, f \rangle)} \ ,$$

since $\widehat{\mathcal{F}_S}(\alpha) = 0$ when $\alpha \notin \mathcal{V}$. For any $\alpha \in \mathcal{V}$, if $\langle \alpha, f \rangle = 0$ then $\sum_{b \in \mathbb{F} \setminus \{0\}} \omega^{-\mathrm{Tr}(\langle b\alpha, f \rangle)} = q - 1$. Otherwise, $\sum_{b \in \mathbb{F} \setminus \{0\}} \omega^{-\mathrm{Tr}(\langle b\alpha, f \rangle)} = -1$.

Let $\mathcal{V}_0 \subseteq \mathcal{V}$ be the subspace containing all $\alpha \in \mathcal{V}$ such that $\langle \alpha, f \rangle = 0$. Since $\mathsf{cost}(M\alpha) = \mathsf{cost}(M(b\alpha))$ for all $b \in \mathbb{F} \setminus \{0\}$, the above computation shows that

$$q^{|S|}\mathcal{F}_S(f) = \sum_{\alpha \in \mathcal{V}_0} \left(1 - \frac{q}{q-1}\mathsf{cost}(M\alpha)\varepsilon\right) + \frac{-1}{q-1} \cdot \sum_{\alpha \in \mathcal{V} \setminus \mathcal{V}_0} \left(1 - \frac{q}{q-1}\mathsf{cost}(M\alpha)\varepsilon\right) \ .$$

There are two cases. If $\mathcal{V}_0 = \mathcal{V}$, then $q^{|S|}\mathcal{F}_S(f) = \sum_{\alpha \in \mathcal{V}_0} \left(1 - \frac{q}{q-1}\mathsf{cost}(M\alpha)\varepsilon\right) \geq 0$ by assumption. If $\mathcal{V}_0 \subsetneq \mathcal{V}$, then $\mathcal{V}_0$ is a subspace of co-dimension 1, as it is specified by one linear constraint. Let $\gamma \in \mathcal{V} \setminus \mathcal{V}_0$. Then

$$\begin{aligned} q^{|S|}\mathcal{F}_S(f) &= \sum_{\alpha \in \mathcal{V}_0} \left(1 - \frac{q}{q-1}\mathsf{cost}(M\alpha)\varepsilon + \frac{-1}{q-1} \cdot \sum_{b \in \mathbb{F} \setminus \{0\}} 1 - \frac{q}{q-1}\mathsf{cost}(M(\alpha + b\gamma))\varepsilon\right) \\ &= \frac{q}{q-1}\varepsilon \cdot \sum_{\alpha \in \mathcal{V}_0} \left(-\mathsf{cost}(M\alpha) + \frac{1}{q-1} \cdot \sum_{b \in \mathbb{F} \setminus \{0\}} \mathsf{cost}(M(\alpha + b\gamma))\right) \ . \end{aligned}$$

If $M\gamma = 0^n$, then we have that $\mathsf{cost}(M\alpha) = \mathsf{cost}(M(\alpha + b\gamma))$ for every $b \in \mathbb{F}$, which implies that the above sum is 0. Hence, $\mathcal{F}_S(f) \geq 0$ in this case. If $M\gamma \neq 0^n$, then $M\mathcal{V}_0 \subsetneq M\mathcal{V}$ is a subspace of co-dimension 1. The remainder of the proof follows from Lemma B.3 applied to the subspaces $M\mathcal{V}_0 \subseteq M\mathcal{V}$. □

We now prove Lemma B.3

Proof of Lemma B.3. Let $\mathcal{V}_0 \subseteq \mathcal{V}$ be a subspace of co-dimension 1. Since $\mathcal{V}_0 \neq \mathcal{V}$, there exists an element $\gamma \in \mathcal{V} \setminus \mathcal{V}_0$. We have that

$$\frac{1}{q-1}\sum_{\alpha\in\mathcal{V}\setminus\mathcal{V}_0}\mathsf{cost}(\alpha) - \sum_{\alpha\in\mathcal{V}_0}\mathsf{cost}(\alpha) = \sum_{\alpha\in\mathcal{V}_0}\left(-\mathsf{cost}(\alpha) + \frac{1}{q-1}\sum_{b\neq 0}\mathsf{cost}(\alpha+b\gamma)\right)$$
$$= \sum_{i=1}^{n}\mathsf{cost}(e_i)\sum_{\alpha\in\mathcal{V}_0}\left(-h_i(\alpha) + \frac{1}{q-1}\sum_{b\neq 0}h_i(\alpha+b\gamma)\right) \ .$$

Let $i \in [n]$. Observe that if $\gamma_i = 0$, then $-h_i(\alpha) + \frac{1}{q-1}\sum_{b\neq 0} h_i(\alpha+b\gamma) = 0$ for every $\alpha \in \mathcal{V}_0$. Let $i \in [n]$ such that $\gamma_i \neq 0$. Observe that if $h_i(\alpha) = 0$, then $-h_i(\alpha) + \frac{1}{q-1}\sum_{b\neq 0} h_i(\alpha+b\gamma) = 1$, as $h_i(\alpha+b\gamma) = 1$ for every $b \in \mathbb{F}\setminus\{0\}$ as $\gamma_i \neq 0$, and $h_i(\alpha) = 0$. If $h_i(\alpha) = 1$, then $-h_i(\alpha) + \frac{1}{q-1}\sum_{b\neq 0} h_i(\alpha+b\gamma) = -\frac{1}{q-1}$, as then there exists a unique $b^* \in \mathbb{F}\setminus\{0\}$ such that $h_i(\alpha+b^*\gamma) = 0$ and $h_i(\alpha+b\gamma) = 1$ for all other b.

Now, either $h_i(\alpha) = 0$ for every $\alpha \in \mathcal{V}_0$, or $h_i(\alpha) = 1$ for some $\alpha \in \mathcal{V}_0$. In the first case, we have that $\sum_{\alpha\in\mathcal{V}_0}\left(-h_i(\alpha) + \frac{1}{q-1}\sum_{b\neq 0} h_i(\alpha+b\gamma)\right) = |\mathcal{V}_0| \geq 0$, as each term in the sum is 1. The second case is more complicated. If $h_i(\alpha) = 1$ for some $\alpha \in \mathcal{V}_0$, then we have that $\sum_{\alpha\in\mathcal{V}_0}\left(-h_i(\alpha) + \frac{1}{q-1}\sum_{b\neq 0} h_i(\alpha+b\gamma)\right) = |\{\alpha\in\mathcal{V}_0 : h_i(\alpha) = 0\}| - \frac{1}{q-1}|\{\alpha\in\mathcal{V}_0 : h_i(\alpha) = 1\}|$. In this case, the linear homomorphism $\pi_i\colon \mathcal{V}_0 \to \mathbb{F}$ has $\pi_i(\alpha) \neq 0$ for some $\alpha \in \mathcal{V}_0$, which implies that $|\{\alpha\in\mathcal{V}_0 : \pi(\alpha) = 0\}| = |\{\alpha\in\mathcal{V}_0 : \pi(\alpha) = b\}|$ for every $b \in \mathbb{F}$. In particular, $|\{\alpha\in\mathcal{V}_0 : h_i(\alpha) = 0\}| = \frac{1}{q-1}|\{\alpha\in\mathcal{V}_0 : h_i(\alpha) = 1\}|$. This implies that $\sum_{\alpha\in\mathcal{V}_0}\left(-h_i(\alpha) + \frac{1}{q-1}\sum_{b\neq 0} h_i(\alpha+b\gamma)\right) = 0$. Hence,

$$\frac{1}{q-1}\sum_{\alpha\in\mathcal{V}\setminus\mathcal{V}_0}\mathsf{cost}(\alpha) - \sum_{\alpha\in\mathcal{V}_0}\mathsf{cost}(\alpha) = \sum_{i=1}^{n}\mathsf{cost}(e_i)\sum_{\alpha\in\mathcal{V}_0}\left(-h_i(\alpha) + \frac{1}{q-1}\sum_{b\neq 0}h_i(\alpha+b\gamma)\right) \geq 0 \ ,$$

as $\sum_{\alpha\in\mathcal{V}_0}\left(-h_i(\alpha) + \frac{1}{q-1}\sum_{b\neq 0} h_i(\alpha+b\gamma)\right) \geq 0$ for each $i \in n$. □

B.3 On the tightness of Theorem 5

We now show that Theorem 5 is tight when $\mathbf{C}$ is the repetition code and is tight up to a factor of 3 when $\mathbf{C}$ is the Hadamard code. We begin by stating the following proposition.

Proposition B.4. *Let T be a set of local constraints, and let $k \geq 0$. Then,*

- $\mathsf{nsrank}_T(\alpha) \geq \mathsf{rank}_T(\alpha)$.
- $\mathsf{nsrank}_T \geq \mathrm{wt}(\alpha)/\ell$, *where* $\ell = \max_{\alpha\in T}\mathrm{wt}(\alpha)$.

The first statement follows immediately from the fact that any k-local proof of length r can be mapped to a proof of length $\leq r$ as in Definition 1.2. The second statement follows immediately from the first one and the fact that if $\mathsf{rank}_T(\alpha) = r$ then $\mathrm{wt}(\alpha) \leq r\ell$.

Let $\mathbf{C}$ be the Hadamard code and $T = \{e_x + e_y - e_{x+y} : x, y \in \mathbb{F}^n\}$. In [CMS18] it is shown implicitly that $\mathsf{nsrank}_T(\alpha) \leq \mathrm{wt}(\alpha) - 2$. The above shows that $\mathsf{nsrank}_T(\alpha) \geq \mathrm{wt}(\alpha)/3$. This implies that for the Hadamard code, Theorem 5 is tight up to a factor of 3.

We now show the following lemma, which implies that Theorem 5 is tight for the repetition code.

Lemma B.5. *If $\mathbf{C} = \{0^n, 1^n\} \subseteq \{0,1\}^n$ is the repetition code and $T = \{e_i + e_j : i, j \in [n]\}$ is the canonical test, then $\mathsf{nsrank}_T(\alpha) = \mathrm{wt}(\alpha)/2$.*

Proof. Observe that $\mathbf{C}^\perp = \{\alpha \in \{0,1\}^n : \sum_{i=1}^n \alpha_i = 0\}$. Note that in particular, $\mathrm{wt}(\alpha)$ is even for every $\alpha \in \mathbf{C}^\perp$. Let $\alpha \in \mathbf{C}^\perp$, and let $i_1, \ldots, i_\ell$ be the set of indices in $[n]$ such that $\alpha_{i_j} \neq 0$. Then $\alpha = (e_{i_1} + e_{i_2}) + (e_{i_3} + e_{i_4}) + \cdots + (e_{i_{\ell-1}} + e_{i_\ell}) = \alpha_1 + \cdots + \alpha_{\ell/2}$. Observe that if $T \vdash_k \alpha$, then $\mathrm{wt}(\alpha) \leq k$. Hence, the above gives a k-local proof of α of length $\ell/2$, and so $\mathsf{nsrank}_T(\alpha) \leq \ell/2$. The earlier proposition implies that $\mathsf{nsrank}_T(\alpha) = \ell/2$, completing the proof. □

Acknowledgements

We are grateful to Thomas Vidick for suggesting using irreducible curves to extend our initial non-testability result for axis-parallel lines to the case of general lines.

This research was supported in part by: the NSF Graduate Research Fellowship Program (under Grant No. DGE1745016); the ARCS Foundation; and donations from the Ethereum Foundation and the Interchain Foundation. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the author(s) and do not necessarily reflect the views of the National Science Foundation.

References

[AB11] Samson Abramsky and Adam Brandenburger. "The sheaf-theoretic structure of non-locality and contextuality". In: *New Journal of Physics* 13.11 (2011), p. 113036.

[Aro+98] Sanjeev Arora, Carsten Lund, Rajeev Motwani, Madhu Sudan, and Mario Szegedy. "Proof verification and the hardness of approximation problems". In: *Journal of the ACM* 45.3 (1998). Preliminary version in FOCS '92., pp. 501–555.

[BHR05] Eli Ben-Sasson, Prahladh Harsha, and Sofya Raskhodnikova. "Some 3CNF Properties Are Hard to Test". In: *SIAM Journal on Computing* 35.1 (2005), pp. 1–21.

[BLR93] Manuel Blum, Michael Luby, and Ronitt Rubinfeld. "Self-Testing/Correcting with Applications to Numerical Problems". In: *Journal of Computer and System Sciences* 47.3 (1993), pp. 549–595.

[CMS18] Alessandro Chiesa, Peter Manohar, and Igor Shinkar. "Testing Linearity against Non-Signaling Strategies". In: *Proceedings of the 33rd Annual Conference on Computational Complexity*. CCC '18. 2018, 17:1–17:37.

[CMS19] Alessandro Chiesa, Peter Manohar, and Igor Shinkar. "Probabilistic Checking Against Non-Signaling Strategies from Linearity Testing". In: *Proceedings of the 10th Innovations in Theoretical Computer Science Conference*. ITCS '19. 2019, 25:1–25:17.

[GS06] Oded Goldreich and Madhu Sudan. "Locally testable codes and PCPs of almost-linear length". In: *Journal of the ACM* 53 (4 2006). Preliminary version in STOC '02., pp. 558–655.

[Gri01] Dima Grigoriev. "Linear lower bound on degrees of Positivstellensatz calculus proofs for the parity". In: *Theoretical Computer Science* 259.1 (2001), pp. 613–622.

[KRR13] Yael Kalai, Ran Raz, and Ron Rothblum. "Delegation for Bounded Space". In: *Proceedings of the 45th ACM Symposium on the Theory of Computing*. STOC '13. 2013, pp. 565–574.

[KRR14] Yael Tauman Kalai, Ran Raz, and Ron D. Rothblum. "How to delegate computations: the power of no-signaling proofs". In: *Proceedings of the 46th ACM Symposium on Theory of Computing*. STOC '14. Full version available at `https://eccc.weizmann.ac.il/report/2013/183/`. 2014, pp. 485–494.

[KRR16] Yael Tauman Kalai, Ran Raz, and Oded Regev. "On the Space Complexity of Linear Programming with Preprocessing". In: *Proceedings of the 7th Innovations in Theoretical Computer Science Conference*. ITCS '16. 2016, pp. 293–300.

[KT92] Leonid A Khalfin and Boris S Tsirelson. "Quantum/classical correspondence in the light of Bell's inequalities". In: *Foundations of physics* 22.7 (1992), pp. 879–948.

[NV18] Anand Natarajan and Thomas Vidick. "Low-Degree Testing for Quantum States, and a Quantum Entangled Games PCP for QMA". In: *Proceedings of the 59th IEEE Symposium on Foundations of Computer Science*. FOCS '18. 2018, pp. 731–742.

[PR94] Sandu Popescu and Daniel Rohrlich. "Quantum nonlocality as an axiom". In: *Foundations of Physics* 24.3 (1994), pp. 379–385.

[RS09] Prasad Raghavendra and David Steurer. "Integrality Gaps for Strong SDP Relaxations of UNIQUE GAMES". In: *Proceedings of the 50th IEEE Symposium on Foundations of Computer Science*. FOCS '09. Full version at `http://people.eecs.berkeley.edu/~prasad/Files/cspgaps.pdf`. 2009, pp. 575–585.

[RS96] Ronitt Rubinfeld and Madhu Sudan. "Robust Characterizations of Polynomials with Applications to Program Testing". In: *SIAM Journal on Computing* 25.2 (1996), pp. 252–271.

[Ras85] Peter Rastall. "Locality, Bell's theorem, and quantum mechanics". In: *Foundations of Physics* 15.9 (1985), pp. 963–972.

[SA90] Hanif D. Sherali and Warren P. Adams. "A Hierarchy of Relaxations between the Continuous and Convex Hull Representations for Zero-One Programming Problems". In: *SIAM Journal on Discrete Mathematics* 3.3 (1990), pp. 411–430.

[Sch08] Grant Schoenebeck. "Linear Level Lasserre Lower Bounds for Certain k-CSPs". In: *Proceedings of the 49th IEEE Symposium on Foundations of Computer Science*. FOCS '08. 2008, pp. 593–602.

ECCC ISSN 1433-8092

https://eccc.weizmann.ac.il