ANTI-CONCENTRATION AND THE EXACT GAP-HAMMING PROBLEM

ANUP RAO AND AMIR YEHUDAYOFF

Abstract. We prove anti-concentration bounds for the inner product of two independent random vectors, and use these bounds to prove lower bounds in communication complexity. We show that if A, B are subsets of the cube $\{\pm 1\}^n$ with $|A| \cdot |B| \geq 2^{1.01n}$, and $X \in A$ and $Y \in B$ are sampled independently and uniformly, then the inner product $\langle X, Y \rangle$ takes on any fixed value with probability at most $O(1/\sqrt{n})$. In fact, we prove the following stronger "smoothness" statement:

$$\max_k \left| \Pr[\langle X, Y \rangle = k] - \Pr[\langle X, Y \rangle = k + 4] \right| \leq O(1/n).$$

We use these results to prove that the exact gap-hamming problem requires linear communication, resolving an open problem in communication complexity. We also conclude anti-concentration for structured distributions with low entropy. If $x \in \mathbb{Z}^n$ has no zero coordinates, and $B \subseteq \{\pm 1\}^n$ corresponds to a subspace of $\mathbb{F}_2^n$ of dimension $0.51n$, then $\max_k \Pr[\langle x, Y \rangle = k] \leq O(\sqrt{\ln(n)/n})$.

1. Introduction

Anti-concentration bounds establish that the distribution of outcomes of a random process is not concentrated in any small region. No single outcome is obtained too often. Anti-concentration plays an important role in mathematics and computer science. It is used in the study of roots of random polynomials [19], random matrix theory [15, 27], communication complexity [6, 29, 22], quantum computation [1], and more. In particular, as we discuss below, anti-concentration bounds are very useful to understand the communication complexity of the gap-hamming function.

A well-known context in which anti-concentration has been studied extensively is the sum of independent identically distributed random variables. If $Y \in \{\pm 1\}^n$ is uniformly distributed, then the probability that $\sum_{j=1}^n Y_j$ takes any specific value is at most $\binom{n}{\lceil n/2 \rceil}/2^n = O(\frac{1}{\sqrt{n}})$. This was studied and generalized by Littlewood and Offord [19], Erdős [7], and many others. The classical Littlewood-Offord problem is about understanding the anti-concentration of the inner product $\langle x, Y \rangle = \sum_{j=1}^n x_j Y_j$, for arbitrary $x \in \mathbb{R}^n$ and $Y \in \{\pm 1\}^n$ chosen uniformly. For example, Erdős proved that if x has no zero coordinates, then $\max_k \Pr[\langle x, Y \rangle = k] \leq \binom{n}{\lceil n/2 \rceil}/2^n = O(\frac{1}{\sqrt{n}})$.

It is interesting to understand the most general conditions under which such anti-concentration holds. Various generalizations were studied by Frankl and Füredi [10], Halász [11] and others (see [25] and references within). These results show that stronger bounds can be proved when the vector x satisfies stronger conditions. In this past work,

A.Y. is partially supported by ISF grant 1162/15. This work was done while the authors were visiting the Simons Institute for the Theory of Computing.

the vector Y is typically assumed to be uniformly distributed; indeed anti-concentration fails when the entries of Y are not independent. For example, if Y is sampled uniformly from the set of strings with exactly $\lceil n/2 \rceil$ entries that are 1, then for $x = 1^n$, we have that $\langle x, Y \rangle$ is always the same. Can we somehow recover anti-concentration when Y is not uniform?

We show that if extra structure holds then anti-concentration is still recovered although the entropy is small. For example, if we identify $\{\pm 1\}^n$ with the vector space $\mathbb{F}_2^n$, by associating -1 with 1 and 1 with 0, then our results imply:

Theorem 1. *There exists a constant $C > 0$ so that the following holds. If $x \in \mathbb{Z}^n$ has no zero coordinates, $B \subseteq \{\pm 1\}^n$ corresponds to a subspace of $\mathbb{F}_2^n$ of dimension $0.51n$, and $Y \in B$ is uniformly distributed, then*

$$\max_{k \in \mathbb{Z}} \Pr[\langle x, Y \rangle = k] \leq C \sqrt{\tfrac{\ln n}{n}}.$$

Theorem 1 is a direct consequence of Theorem 4 below.

Remark. *Theorem 1 and similar results can be used as a black box to prove the same bounds when x is a real-valued vector. To see this, think of the relevant real numbers as vectors in a finite dimensional vector space over the rationals. We omit the details here.*

Another natural setting is to consider the inner-product $\langle X, Y \rangle$ of two independent variables, neither of which may be uniform. Recent work has proved some interesting results under the assumption that X, Y have nice structure [28, 13], but what if the only assumption is that X, Y a uniformly distributed on large sets? The following theorem, proved by Chakrabarti and Regev [6] along the way to proving new lower bounds in communication complexity, shows that this does recover some anti-concentration:

Theorem (Chakrabarti and Regev [6])**.** *There is a constant $c > 0$ such that if $A, B \subseteq \{\pm 1\}^n$ are each of size at least $2^{(1-c)n}$ and $X \in A, Y \in B$ are sampled uniformly and independently, then*

$$\Pr[|\langle X, Y \rangle| \leq c\sqrt{n}] \leq 1 - c.$$

Alternate proofs of the same bound were subsequently given in [29, 22]. The theorem shows that $\langle X, Y \rangle$ does not land in an interval of length much smaller than $\sqrt{n}$ with high probability. The strongest anti-concentration bounds give point-wise estimates. We would like to control the *concentration probability*

$$\max_{k \in \mathbb{Z}} \Pr[\langle X, Y \rangle = k];$$

see [27] and references within.

In our work we prove a sharp bound on the point-wise concentration probability that holds for an overwhelming majority of directions x.

Theorem 2. *For every $\beta > 0$ and $\delta > 0$, there exists $C > 0$ such that the following holds. If $B \subseteq \{\pm 1\}^n$ is of size $2^{\beta n}$, and $Y \in B$ is uniformly distributed, then for all but $2^{n(1-\beta+\delta)}$ directions $x \in \{\pm 1\}^n$,*

$$\max_{k \in \mathbb{Z}} \Pr_Y[\langle x, Y \rangle = k] \leq \tfrac{C}{\sqrt{n}}.$$

In particular, if X is independent of Y and uniformly distributed in a set A of size $2^{n(1-\beta+2\delta)}$, then

$$\max_{k\in\mathbb{Z}} \Pr_Y[\langle X, Y\rangle = k] \leq \tfrac{C}{\sqrt{n}} + 2^{-\delta n}.$$

Our bound implies the result of Chakrabarti and Regev, but it is strictly stronger. It is also tight in the following senses. As mentioned above, the $O(\frac{1}{\sqrt{n}})$ bound is tight even when A and B are $\{\pm 1\}^n$. To see that the bound on the number of bad directions is sharp[1], observe that if $B \subset \{\pm 1\}^n$ is the set of y's with $\sum_{j=1}^n y_j = 0$, and $A \subset \{\pm 1\}^n$ is the set of x's with $\sum_{j=1}^n x_j = (1-2\epsilon)n$ for some small $\epsilon > 0$, then

$$|B| \approx \tfrac{1}{\sqrt{n}} 2^n \quad \& \quad |A| \approx 2^{h(\epsilon)n},$$

where $h(\epsilon)$ is the binary entropy function. Yet for every $x \in A$,

$$\Pr[|\langle x, Y\rangle| \leq 1] \geq \Omega(\tfrac{1}{\sqrt{\epsilon n}}).$$

The sets A, B do not satisfy the conclusions of Theorem 2, even though $|A| \cdot |B| \approx 2^{(1+h(\epsilon))n}$.

Our methods lead to even stronger conclusions about the distribution of $\langle x, Y\rangle$. We prove the following *smoothness* result:

Theorem 3. *For every $\beta, \epsilon > 0$, there is $C > 0$ so that the following holds. Suppose $B \subseteq \{\pm 1\}^n$ is a set with $|B| = 2^{\beta n}$, and $Y \in B$ is uniformly distributed. Then for all but $2^{(1-\beta+\epsilon)n}$ choices of $x \in \{\pm 1\}^n$, we have:*

$$\max_{k\in\mathbb{Z}} \left| \Pr[\langle x, Y\rangle = k] - \Pr[\langle x, Y\rangle = k+4] \right| \leq \frac{C}{n}.$$

In particular, if X is independent of Y and uniformly distributed in a set A of size $2^{(1-\beta+2\epsilon)n}$, then

$$\max_{k\in\mathbb{Z}} \left| \Pr[\langle X, Y\rangle = k] - \Pr[\langle X, Y\rangle = k+4] \right| \leq \frac{C}{n} + 2^{-\epsilon n}.$$

Theorem 2 is implied by Theorem 3. Indeed, if x is such that

$$\max_{k\in\mathbb{Z}} \left| \Pr[\langle x, Y\rangle = k] - \Pr[\langle x, Y\rangle = k+4] \right| \leq \frac{C}{n},$$

then for all k and $j \leq m$,

$$\Pr[\langle x, Y\rangle = k] \leq \Pr[\langle x, Y\rangle = k + 4j] + \frac{Cm}{n}.$$

But then we must have (for $m \approx \sqrt{n}$):

$$\Pr[\langle x, Y\rangle = k] \leq \frac{Cm}{n} + \frac{1}{m} \cdot \sum_{j=1}^{m} \Pr[\langle x, Y\rangle = k + 4j] \leq \frac{Cm}{n} + \frac{1}{m} \lesssim \frac{1}{\sqrt{n}}.$$

[1]This example is due to an anonymous reviewer.

Theorem 3 is proved in Section 4. It is sharp in the following two senses. First, even for the case $A = B = \{\pm 1\}^n$, there is a k so that[2]

$$|\Pr[\langle X, Y\rangle = k] - \Pr[\langle X, Y\rangle = k + 4| \geq \Omega(\tfrac{1}{n}).$$

So, $O(\frac{1}{n})$ is the best upper bound possible. Secondly, if $A = \{x \in \{\pm 1\}^n : n - \sum_{j=1}^n x_j = 0 \bmod 4\}$ and $B = \{y \in \{\pm 1\}^n : n - \sum_{j=1}^n y_j = 0 \bmod 4\}$, then because $n = \langle x, y\rangle \bmod 2$ for all $x, y \in \{\pm 1\}^n$, $\langle x, y\rangle \bmod 4$ is the same for every pair $x \in A, y \in B$. Thus, there are sets A, B with $|A| = |B| = 2^{n-1}$ so that for all $j \in \{1, 2, 3\}$,

$$|\Pr[\langle X, Y\rangle = 0] - \Pr[\langle X, Y\rangle = j]| = \Pr[\langle X, Y\rangle = 0] = \Omega(\tfrac{1}{\sqrt{n}})$$

So, 4 is the minimum gap for which an $O(\frac{1}{n})$ upper bound holds.

Our proof builds a flexible framework for proving anti-concentration results in discrete domains. We use this framework to show that anti-concentration holds in a wide variety of settings. As we explain below, we show that bounds similar to those proved in [8, 21, 23, 11] apply even when the underlying distribution is not uniform. When Y is uniformly distributed, the additive structure of the entries in the direction vector x controls anti-concentration [10]. If x is unstructured, we get even stronger anti-concentration bounds for $\langle x, Y\rangle$. This idea is instrumental when analyzing random matrices [15, 27].

We choose the direction x from sets of the following form. We call a set $A \subset \mathbb{Z}^n$ a two-cube if $A = A_1 \times A_2 \times \cdots \times A_n$, where each $A_j = \{u_j, v_j\}$ consists of two distinct integers. The differences of A are the numbers $d_j = u_j - v_j$ for $j \in [n]$.

The following theorem describes three cases that yield different anti-concentration bounds. It shows that the additive structure of A is deeply related to the bounds we obtain. The less structured A is, the stronger the bounds are. The first bound in the theorem holds for arbitrary two-cubes. The second bound holds when all the differences $d_1, \ldots, d_n$ are distinct. The third bound applies in more general settings where the set of differences is unstructured. This is captured by the following definition. A set $S \subset \mathbb{N}$ of size n is called a Sidon set, or a Golomb ruler, if the number of solutions to the equation $s_1 + s_2 = s_3 + s_4$ for $s_1, s_2, s_3, s_4 \in S$ is $4 \cdot \binom{n}{2} + n$. In other words, every pair of integers has a distinct sum. Sidon sets were defined by Erdős and Turán [9] and have been studied by many others since. We say that $S \subset \mathbb{Z}$ is a weak Sidon set if the number of solutions to the equation $\epsilon_1 s_1 + \epsilon_2 s_2 = \epsilon_3 s_3 + \epsilon_4 s_4$ for $\epsilon_1, \ldots, \epsilon_4 \in \{\pm 1\}$ and $s_1, \ldots, s_4 \in S$ is at most $100n^2$. The number 100 can be replaced by any other constant, we use it here just to be concrete.

Theorem 4. *For every $\beta > 0$ and $\delta > 0$, there exists $C > 0$ such that the following holds. Let $A \subset \mathbb{Z}^n$ be a two-cube with differences $d_1, \ldots, d_n$. Let $B \subseteq \{\pm 1\}^n$ be of size $2^{\beta n}$ and Y be uniformly distributed in B.*

(1) For all but $2^{n(1-\beta+\delta)}$ directions $x \in A$,

$$\max_{k \in \mathbb{Z}} \Pr_Y [\langle x, Y\rangle = k] \leq C\sqrt{\tfrac{\ln(n)}{n}}.$$

[2]For an integer $k = \frac{n}{2} - \sqrt{n}$, we have $\binom{n}{k+1} - \binom{n}{k} = \binom{n}{k+1}\frac{n-2k-1}{n-k} \gtrsim \frac{2^n}{n}$.

(2) If $d_1, \ldots, d_n$ are distinct, then for all but $2^{n(1-\beta+\delta)}$ directions $x \in A$,

$$\max_{k \in \mathbb{Z}} \Pr_Y [\langle x, Y \rangle = k] \leq C\sqrt{\tfrac{\ln(n)}{n^3}}.$$

(3) If $\{d_1, \ldots, d_n\}$ is a weak Sidon set of size n, then for all but $2^{n(1-\beta+\delta)}$ directions $x \in A$,

$$\max_{k \in \mathbb{Z}} \Pr_Y [\langle x, Y \rangle = k] \leq C\sqrt{\tfrac{\ln(n)}{n^5}}.$$

To see why this is a generalization of past work, observe that if $Y \in \{\pm 1\}^n$ is uniformly distributed, then for any $x \in \mathbb{Z}^n$, the distribution of $\langle x, Y \rangle$ is identical to the distribution of $\langle X, Y \rangle$, where X is obtained by picking uniformly random signs for the coordinates of x. The number of directions in the support of X is 2^n, and the theorem above can be applied.

A similar idea proves Theorem 1. The key point is the assumption that B corresponds to a subspace of $\mathbb{F}_2^n$. Every element of B corresponds to a signing of x that gives the same distribution for $\langle x, Y \rangle$. We thus obtained a set A of distinct directions of size $|A| = |B|$. Because $|A| \cdot |B| \geq 2^{1.02n}$, we can apply Theorem 4 to prove Theorem 1.

The proof of Theorem 4 is given in Section 5. The first bound in Theorem 4 nearly implies Theorem 2. It is weaker by a factor of $\sqrt{\ln(n)}$. However, it holds for all two-cubes, not just the hypercube $\{\pm 1\}^n$. The second bound almost matches the sharp $O(1/n^{1.5})$ bound that holds when $(u_j, v_j) = (j, -j)$ for each j and Y is uniform in the hypercube [21, 23]. We believe that the $\sqrt{\ln(n)}$ factor is not needed, but were not able to eliminate it. The theorem is, in fact, part of a more general phenomenon. We postpone the full technical description to Section 5.

An application to Communication Complexity. These kinds of anti-concentration bounds are intimately connected to understanding the communication complexity of the *gap-hamming* function. The gap-hamming function $\mathsf{GH} = \mathsf{GH}_{n,k} : \{\pm 1\}^n \to \{0, 1, \star\}$ is defined by

$$\mathsf{GH}(x, y) = \begin{cases} 1 & \langle x, y \rangle \geq k, \\ 0 & \langle x, y \rangle \leq -k, \\ \star & \text{otherwise.} \end{cases}$$

Note that the Hamming distance between x and y is $\frac{n - \langle x, y \rangle}{2}$. This problem is well-studied in communication complexity; for background and definitions, see the books [18, 20]. Alice gets x, Bob gets y, and their goal is to compute $\mathsf{GH}(x, y)$. It is a promise problem; the protocol is allowed to compute any value when the input corresponds to a $\star$, and it needs to be correct only on the remaining inputs. The standard choice for k is $\lceil \sqrt{n} \rceil$, so we write GH_n to denote $\mathsf{GH}_{n, \lceil \sqrt{n} \rceil}$.

The gap-hamming problem was introduced by Indyk and Woodruff in the context of streaming algorithms [12], and was subsequently studied and used in many works and in various contexts (see [14, 30, 3, 4, 5] and references within). Proving a sharp $\Omega(n)$ lower bound on its randomized communication complexity was a central open problem for almost ten years, until Chakrabarti and Regev [6] solved it using the anti-concentration

bound mentioned above. Later, Vidick [29] and Sherstov [22] found simpler proofs. The difficulties in proving this lower bound are explained in [6, 22].

The exact gap-hamming function is defined by

$$\mathsf{EGH}_{n,k}(x,y) = \begin{cases} 1 & \langle x,y\rangle = k, \\ 0 & \langle x,y\rangle = -k, \\ \star & \text{otherwise.} \end{cases}$$

As before, we write EGH_n to denote $\mathsf{EGH}_{n,\lceil\sqrt{n}\rceil}$. The exact gap-hamming function is easier to compute than gap-hamming; the protocol only needs to worry about inputs whose inner product has magnitude *exactly* k. Proving a sharp lower bound on the randomized communication complexity of EGH was left as an open problem.

One of the difficulties in proving a lower bound for EGH is the following somewhat surprising property: *for infinitely many values of n, the deterministic communication complexity of EGH_n is* 2. The reason is that there is a simple deterministic protocol of length 2 that computes $\langle x,y\rangle \bmod 4$ for all n. This protocol corresponds to the sets A, B discussed with regards to Theorem 3 above. The players announce the parities of their inputs $\frac{n-\sum_{j=1}^n x_j}{2} \bmod 2$ and $\frac{n-\sum_{j=1}^n y_j}{2} \bmod 2$. These bits determine $\langle x,y\rangle \bmod 4$. Indeed, flipping a bit in x changes $\frac{n-\sum_{j=1}^n x_j}{2} \bmod 2$, and changes $\langle x,y\rangle$ by $+2 \bmod 4$. For example, this deterministic protocol computes EGH_n when $\sqrt{n}$ is an odd integer, because then we have $-\sqrt{n} \neq \sqrt{n} \bmod 4$.

We overcome this difficulty and show that EGH is extraordinary in that although it is a natural problem with communication complexity $O(1)$ for infinitely many values of n, the randomized communication complexity of EGH_n is at least $\Omega(n)$ for infinitely many values of n. Denote by $U_{n,k}$ the uniform distribution over the set of pairs $(x,y) \in \{\pm 1\}^n \times \{\pm 1\}^n$ so that $\langle x,y\rangle \in \{\pm k\}$.

Theorem 5. *There is universal constant $\alpha > 0$ such that for infinitely many values of n, any protocol that computes EGH_n over inputs from $U_{n,\lceil\sqrt{n}\rceil}$ with success probability $2/3$ must have communication complexity at least αn.*

There is a natural reduction between different parameters n, k, and from randomized protocols to distributional protocols. It turns out that the following theorem is stronger:

Theorem 6. *For every $\beta > 0$, there are constants $n_0 > 0$ and $\alpha > 0$ so that the following holds. Let n, k be positive even integers so that $n > n_0$ and $k < \alpha\sqrt{n}$. Any protocol that computes $\mathsf{EGH}_{n,k}$ over inputs from $U_{n,k}$ with success probability $2/3$ must have communication complexity at least $(1-\beta)n$.*

Theorems 5 and 6 are proved in Section 6. The results are sharp in the following two senses. First, if $k \neq n \bmod 2$ then $\mathsf{EGH}_{n,k}$ is trivial, and if k is odd then the deterministic communication complexity of $\mathsf{EGH}_{n,k}$ is 2. Secondly, for every $\alpha > 0$, there is $\beta > 0$ so that if $k > \alpha\sqrt{n}$ then the randomized communication complexity of $\mathsf{EGH}_{n,k}$ is at most $(1-\beta)n$. We sketch a randomized protocol for this here. In the randomized protocol, Alice gets x, Bob gets y and the public randomness is a sequence $I_1, I_2, \ldots, I_m$ of i.i.d. uniform elements in $[n]$ for $m \leq O(\frac{n}{\alpha^2})$. Although m is a constant factor larger than n, a standard coupon collector argument shows that the number of (distinct) elements

in the set $S = \{I_1, \ldots, I_m\}$ is at most $(1-\beta)n - 1$ with probability at least $\frac{5}{6}$. If $|S| > (1-\beta)n - 1$, the parties "abort", and otherwise Alice sends to Bob the value of x_s for all $s \in S$. Bob uses this data to compute $z = 1 + \mathsf{sign}\big(\sum_{j=1}^{m} x_{I_j} y_{I_j}\big)/2$. Bob sends the output of the protocol z to Alice. Chernoff's bound says that if $\mathsf{EGH}_{n,k}(x,y) \neq \star$ then $\Pr[z = \mathsf{EGH}_{n,k}(x,y)] \geq \frac{5}{6}$. The union bound implies that the overall success probability is at least $\frac{2}{3}$.

An application to Additive Combinatorics. Additive combinatorics studies the behavior of sets under algebraic operations [26]. It has many deep results, and connections to other areas of mathematics, as well as many applications in computer science. Our main result can be interpreted as showing that Hamming spheres are far from being sum-sets. Our results give quantitative bounds on the size of the intersection of any Hamming sphere with a sum-set.

Replace $\{\pm 1\}$ by the field $\mathbb{F}_2$ with two elements. The sum-set of $A \subseteq \mathbb{F}_2^n$ and $B \subseteq \mathbb{F}_2^n$ is

$$A + B = \{x + y : x \in A, y \in B\}.$$

If X and Y are sampled uniformly at random from A and B, then $X + Y$ is supported on $A + B$.

The cube $\mathbb{F}_2^n$ is endowed with a natural metric—the Hamming distance $\Delta(x,y)$. The sphere around 0 is the collection of all vectors with a fixed number of ones in them (a.k.a. a slice). The inner product $I = \sum_j (-1)^{X_j}(-1)^{Y_j}$ is similar to the inner product studied above (here $X_j, Y_j \in \{0,1\}$). The inner product is related to the Hamming distance by $I(X,Y) = n - 2\Delta(X,Y)$. We saw that if $|A| \cdot |B| > 2^{1.01n}$, then I is anti-concentrated. We can conclude that the distribution of the Hamming distance of $X+Y$ is anti-concentrated. The set $A+B$ is far from any slice. In particular, our results imply that for almost all choices of $a \in A$, we have that $|(a+B) \cap S| \leq O(|B|/\sqrt{n})$ for any slice S.

Techniques. Chakrabarti and Regev's proof uses the deep connection between the discrete cube and Gaussian space. They proved a geometric correlation inequality in Gaussian space, and then translated it to the cube. Vidick [29] later simplified part of their argument, but stayed in the geometric setting. Sherstov [22] found a third proof that uses Talagrand's inequality from convex geometry [24] and ideas of Babai, Frankl and Simon from communication complexity [2].

There are several differences between our argument and the ones in [6, 29, 22]. The main difference is that the arguments from [6, 29, 22] are based, in one way or another, on the geometry of Euclidean space. The arguments in [6, 29] prove a correlation inequality in Gaussian space and translate it to the discrete world. It seems that such an argument can not yield point-wise bounds on the concentration probability. A common ingredient in [6, 22] is a step showing that every set of large enough measure contains many almost orthogonal vectors (this is called 'identifying the hard core' in [22]). In [29] this part of the argument is replaced by a statement about a relevant matrix. Our argument does not contain such steps.

Let us briefly discuss our proof at a high level. The proof is based on harmonic analysis (Section 2). The argument consists of two parts. In the first part, we analyze

the Fourier behavior of $\langle x, Y\rangle$ for x fixed and Y random. We are able to identify a collection of good x's for which the Fourier spectrum of the distribution of $\langle x, Y\rangle$ decays rapidly. In the second part, we show that the number of bad x's is small by giving an explicit encoding of all of them.

Although the proofs of Theorem 3 and Theorem 4 follow similar strategies, we were not able to completely merge them.

2. Harmonic Analysis

We are interested in proving anti-concentration for integer-valued random variables. Harmonic analysis is a natural framework for studying such random variables [11]. Let Y be distributed in $\{\pm 1\}^n$. Let $x \in \mathbb{Z}^n$ be a direction. Let θ be uniformly distributed in $[0, 1]$, independently of Y. The idea is to use

$$\Pr_Y[\langle x, Y\rangle = k] = \mathop{\mathbb{E}}_Y \left[\mathop{\mathbb{E}}_\theta \left[\exp(2\pi i\theta \cdot (\langle x, Y\rangle - k)) \right] \right]$$

to bound

$$(\star) \qquad \max_{k\in\mathbb{Z}} \Pr_Y[\langle x, Y\rangle = k] \leq \mathop{\mathbb{E}}_\theta \left[\left| \mathop{\mathbb{E}}_Y \left[\exp(2\pi i\theta \cdot \langle x, Y\rangle) \right] \right| \right].$$

This inequality is useful for two reasons. First, the left hand side is a maximum over k, while the right hand side is not. So, there is one less quantifier to worry about. Secondly, the right hand side lives in the Fourier world, where it is easier to argue about the underlying operators. For example, when the coordinates of Y are independent, the expectation over Y breaks into a product of n simple terms.

3. The main technical Theorem

Our main technical bound is proved in this section. The following theorem controls the Fourier coefficients in most directions.

Theorem 7. *For every $\beta > 0$ and $\delta > 0$, there is $c > 0$ so that the following holds. Let $B \subseteq \{\pm 1\}^n$ be of size $2^{\beta n}$. For each $\theta \in [0, 1]$, for all but $2^{n(1-\beta+\delta)}$ directions $x \in \{\pm 1\}^n$,*

$$\left| \mathop{\mathbb{E}}_Y \left[\exp(2\pi i\theta \cdot \langle x, Y\rangle) \right] \right| < 2\exp(-cn\sin^2(4\pi\theta))$$

The rest of this section is devoted to proving the theorem.

3.1. A Single Direction.

In this section we analyze the behavior of $\langle x, Y\rangle$ for a single direction $x \in \mathbb{Z}^n$. We also focus on a single Fourier coefficient $\mathbb{E}_Y \left[\exp(i\eta \langle x, Y\rangle) \right]$ for a fixed angle $\eta \in [0, 2\pi]$.

We reveal the entropy of Y coordinate by coordinate. To keep track of this entropy, define the following functions $\gamma_1, \ldots, \gamma_n$ from $B = \mathrm{supp}(Y)$ to $\mathbb{R}$. For each $j \in [n]$, let

$$\gamma_j(y) = \gamma_j(y_{<j}) = \min_{\epsilon\in\{\pm 1\}} \Pr[Y_j = \epsilon | Y_{<j} = y_{<j}].$$

To understand the interaction between x and y, we use the following n measurements. For $j \in [n-1]$, define $\phi_j(x,y)$ to be half of the phase of the complex number

$$\mathop{\mathbb{E}}_{Y_{>j}|Y_j=1,Y_{<j}=y_{<j}} [\exp(i\eta \langle x_{>j}, Y_{>j}\rangle)] \cdot \overline{\mathop{\mathbb{E}}_{Y_{>j}|Y_j=-1,Y_{<j}=y_{<j}} [\exp(i\eta \langle x_{>j}, Y_{>j}\rangle)]}.$$

This quantity is not defined when $\gamma_j(y) = 0$. In this case, set $\phi_j(x,y)$ to be zero. Define $\phi_n(x,y)$ to be zero. The number $\phi_j(x,y)$ is determined by $y_{<j}$ and $x_{>j}$.

In the following we think of x as fixed, and of γ_j and ϕ_j as random variables that are determined by the random variable Y.

Lemma 8. *For each $x \in \mathbb{R}^n$, every random variable Y over $\{\pm 1\}^n$, and every angle $\eta \in \mathbb{R}$,*

$$\left| \mathop{\mathbb{E}}_{Y} [\exp(i\eta \langle x, Y\rangle)] \right|^2 \le \mathop{\mathbb{E}}_{Y} \left[\prod_{j\in[n]} (1 - \gamma_j \sin^2(\phi_j + x_j\eta)) \right].$$

Proof. The proof is by induction on n. We prove the base case of the induction and the inductive step simultaneously. Express

$$\begin{aligned} \left| \mathop{\mathbb{E}}_{Y} [\exp(i\eta \langle x, Y\rangle)] \right|^2 &= \left| \mathop{\mathbb{E}}_{Y_1} \left[\exp(i\eta x_1 Y_1) \cdot \mathop{\mathbb{E}}_{Y_{>1}|Y_1} [\exp(i\eta \langle x_{>1}, Y_{>1}\rangle)] \right] \right|^2 \\ &= \left| p_1 \exp(i\eta x_1) Z_1 + p_{-1} \exp(-i\eta x_1) Z_{-1} \right|^2, \end{aligned}$$

where for $\epsilon \in \{\pm 1\}$,

$$p_\epsilon = \Pr[Y_1 = \epsilon] \qquad \& \qquad Z_\epsilon = \mathop{\mathbb{E}}_{Y|Y_1=\epsilon} [\exp(i\eta \langle x_{>1}, Y_{>1}\rangle)].$$

When $n = 1$, we have $Z_1 = Z_{-1} = 1$. Rearranging,

$$\begin{aligned} &|p_1 \exp(i\eta x_1) Z_1 + p_{-1} \exp(-i\eta x_1) Z_{-1}|^2 \\ &= p_1^2 |Z_1|^2 + p_{-1}^2 |Z_{-1}|^2 + p_1 p_{-1} (Z_1 \overline{Z_{-1}} \exp(i2\eta x_1) + \overline{Z_1} Z_{-1} \exp(-i2\eta x_1)) \\ &= p_1^2 |Z_1|^2 + p_{-1}^2 |Z_{-1}|^2 + 2 p_1 p_{-1} |Z_1||Z_{-1}| \cos(2\phi_1 + 2x_1\eta). \end{aligned}$$

The last equality holds by the definition of ϕ_1.

There are two cases to consider. When $\cos(2\phi_1 + 2x_1\eta) < 0$, we continue to bound

$$\begin{aligned} &< p_1^2 |Z_1|^2 + p_{-1}^2 |Z_{-1}|^2 \\ &\le (p_1 |Z_1|^2 + p_{-1} |Z_{-1}|^2)(1 - \gamma_1) \\ &\le (p_1 |Z_1|^2 + p_{-1} |Z_{-1}|^2)(1 - \gamma_1 \sin^2(\phi_1 + x_1\eta)). \end{aligned}$$

Recall that γ_1 and ϕ_1 do not depend on Y. When $\cos(2\phi_1 + 2x_1\eta) \ge 0$, using the inequality $a^2 + b^2 \ge 2ab$, we bound

$$\begin{aligned} &\le p_1^2 |Z_1|^2 + p_{-1}^2 |Z_{-1}|^2 + p_1 p_{-1} (|Z_1|^2 + |Z_{-1}|^2) \cos(2\phi_1 + 2x_1\eta) \\ &= p_1 |Z_1|^2 (p_1 + p_{-1} \cos(2\phi + 2x_1\eta)) + p_{-1} |Z_{-1}|^2 (p_{-1} + p_1 \cos(2\phi_1 + 2x_1\eta)) \\ &\le (p_1 |Z_1|^2 + p_{-1} |Z_{-1}|^2)(1 - \gamma_1 + \gamma_1 \cos(2\phi_1 + 2x_1\eta)) \\ &= \mathop{\mathbb{E}}_{Y_1} \left[|Z_{Y_1}|^2 \right] (1 - 2\gamma_1 \sin^2(\phi_1 + x_1\eta)). \end{aligned}$$

When $n = 1$, we have proved the base case of the induction. When $n > 1$, apply induction on $|Z_\epsilon|^2$. □

3.2. **A Few Bad Directions.** Lemma 8 suggests proving that the expression

$$\sum_j \gamma_j \sin^2(\phi_j + x_j\eta)$$

is typically large. Namely, we aim to show that there are usually many coordinates j for which both γ_j and $\sin^2(\phi_j + x_j\eta)$ are bounded away from zero. Our approach is to explicitly encode the cases where this fails to hold.

Recall that Y is uniformly distributed in a set B of size $|B| = 2^{\beta n}$. Let $1 \geq \lambda > 1/n$ be a parameter. Set $0 < \kappa < \frac{1}{2}$ and $1 \geq \tau > 0$ to be parameters satisfying the conditions

$$H\left(\tfrac{1}{\log(1/\kappa)}\right) = \tau + H(\tau) = \lambda, \tag{1}$$

where H is the binary entropy function:

$$H(\xi) = \xi \log(1/\xi) + (1-\xi)\log(1/(1-\xi)).$$

The encoding is based on the following two sets:

$$J(y) = J_{B,\kappa}(y) = \{j \in [n] : \gamma_j(y) \geq \kappa\}$$

and

$$G(x,y) = G_{B,\kappa,\theta}(x,y) = \left\{j \in J(y) : \sin^2(\phi_j(x,y) + x_j\eta) \geq \tfrac{\sin^2(2\eta)}{4}\right\}.$$

We start by showing that there are few y's for which $|J(y)|$ is small.

Lemma 9. *The number of $y \in B$ with $|J(y)| \leq n(\beta - 3\lambda)$ is at most $2^{n(\beta-2\lambda)}$.*

Proof. If $3\lambda > \beta$, the statement is trivially true. So, in the rest of the proof, assume that $3\lambda \leq \beta$. Each $y \in B$ with $|J(y)| \leq n(\beta - 3\lambda)$ can be uniquely encoded by the following data:

- A vector $q \in \{\pm 1\}^t$ with $t = \lfloor n(\beta - 3\lambda)\rfloor$.
- A subset $S \subseteq [n]$ of size $|S| \leq \frac{n}{\log(1/\kappa)}$.

Let us describe the encoding. The vector q encodes the values taken by y in the coordinates $J(y)$. We do not encode $J(y)$ itself, only the values of y in the coordinates corresponding to $J(y)$. The set S includes $j \in [n]$ if and only if

$$\Pr[Y_j = y_j | Y_{<j} = y_{<j}] < \kappa.$$

Each string $y \in B$ has probability at least 2^{-n}. This implies that $\kappa^{|S|} \geq 2^{-n}$.

We can reconstruct y from q and S by iteratively computing y_1, then y_2, and so on, until we get to y_n. Whether or not $1 \in J(y)$ is determined even before we know y. If $1 \in J(y)$ then q tells us what y_1 is. If $1 \notin J(y)$ and $1 \in S$ then y_1 is the least likely value between ± 1. If $1 \notin J(y)$ and $1 \notin S$ then y_1 is the more likely value. Given the value of y_1, we can continue in the same way to compute the rest of y.

The number of choices for q is at most $2^{n(\beta-3\lambda)}$. The number of choices for S is at most $2^{nH(1/\log(1/\kappa))} = 2^{\lambda n}$. □

Next, we argue that there are few x's for which there are many y's with small $G(x,y)$.

Lemma 10. *The number of $x \in A$ for which*

$$\Pr_Y[|G(x,Y)| \leq \tau n] \geq 2^{-\lambda n}$$

is at most $2^{n(1-\beta+6\lambda)}$.

Proof. The lemma is proved by double-counting the edges in a bipartite graph. Let $\mathcal{X}$ be the set we are interested in:

$$\mathcal{X} = \left\{x : \Pr_Y[|G(x,Y)| \leq \tau n] \geq 2^{-\lambda n}\right\}.$$

The left side of the bipartite graph is $\mathcal{X}$ and the right side is B. Connect $x \in \mathcal{X}$ to $y \in B$ by an edge if and only if $G(x,y) \leq \tau n$. Let E denote the set of edges in this graph.

First, we bound the number of edges from below. The number of edges that touch each $x \in \mathcal{X}$ is at least $2^{-\lambda n}|B|$. It follows that

$$|E| \geq 2^{-\lambda n} \cdot |\mathcal{X}| \cdot |B|.$$

Next, we bound the number of edges from above. By Lemma 9, the number of $y \in B$ so that $|J(y)| \leq n(\beta - 3\lambda)$ is at most $2^{-2\lambda n}|B|$. We shall prove that the number of edges that touch each y with $|J(y)| > n(\beta - 3\lambda)$ is at most $2^{n(1-\beta+4\lambda)}$. It follows that

$$|E| \leq 2^{-2\lambda n} \cdot |\mathcal{X}| \cdot |B| + |B| \cdot 2^{n(1-\beta+4\lambda)}.$$

We can conclude that

$$\begin{aligned} 2^{-\lambda n} \cdot |\mathcal{X}| \cdot |B| &\leq 2^{-2\lambda n} \cdot |\mathcal{X}| \cdot |B| + |B| \cdot 2^{n(1-\beta+4\lambda)} \\ \Rightarrow |\mathcal{X}| &\leq 2^{n(1-\beta+6\lambda)}, \end{aligned}$$

since $\lambda n > 1$.

It remains to fix y so that $|J(y)| > n(\beta - 3\lambda)$ and bound its degree from above. This too is achieved by an encoding argument. Encode each x that is connected to y by an edge using the following data:

- A vector $q \in \{\pm 1\}^t$ with $t = \lfloor n(1 - \beta + 3\lambda) \rfloor$.
- The set $G(x,y)$.
- A vector $r \in \{\pm 1\}^s$ with $s = \lfloor \tau n \rfloor$.

Let us describe the encoding. The vector q specifies the values of x on coordinates not in $J(y)$. There are at most $n - n(\beta - 3\lambda) = n(1 - \beta + 3\lambda)$ such coordinates. The size of $G(x,y)$ is at most τn. The vector r specifies the values of x in the coordinates of $G(x,y)$, written in descending order.

The decoding of x from q, S and r is done as follows. Decode the coordinates of x in descending order from n to 1. If $n \notin J(y)$ then we read the value of x_n from q. If $n \in J(y)$ and $n \in G(x,y)$, we decode x_n by reading its value from r. If $n \in J(y)$ and $n \notin G(x,y)$, then

$$\sin^2(\phi_n(x,y) + x_n\eta) \leq \tfrac{\sin^2(2\eta)}{4}.$$

The number $\phi_n(x,y)$ does not depend on x. The following claim implies that there is at most one value of x_n that satisfies this property.

Claim 11. *For all $\varphi \in \mathbb{R}$ and $u, v \in \mathbb{Z}$,*

$$\max\{|\sin(\varphi + \eta u)|, |\sin(\varphi + \eta v)|\} \geq \tfrac{|\sin(\eta(u-v))|}{2}.$$

Proof. We wish to show that the two points on the unit circle of phase $\varphi + \eta u$ and $\varphi + \eta v$ cannot both be very close to the real line in general. Consider the map

$$\varphi \mapsto g(\varphi) = \max\{|\sin(\varphi + \eta u)|, |\sin(\varphi + \eta v)|\}.$$

Observe that the minimum of this map is attained when

$$|\sin(\varphi + \eta u)| = |\sin(\varphi + \eta v)|,$$

since if this is not the case, we can change φ by a little to reduce the larger of the two magnitudes. Now, $|\sin(\alpha)| = |\sin(\beta)|$ when $\alpha + \beta$ is an integer multiple of $\pi/2$. Thus, the two magnitudes are equal exactly when $\varphi = -\frac{\eta(u+v)}{2} + t\pi/2$, for some integer t. By symmetry, it is enough to consider $t \in \{0, 1\}$, so we obtain that

$$\begin{aligned} g(\varphi) &\geq \min\{g(-\eta(u+v)/2), g(-\eta(u+v)/2 + \pi/2)\} \\ &\geq |g(-\eta(u+v)/2) \cdot g(-\eta(u+v)/2 + \pi/2)| \\ &\geq |\sin(\eta(u-v)/2) \cdot \cos(\eta(u-v)/2)| \\ &= \frac{|\sin(\eta(u-v))|}{2}. \end{aligned}$$

□

The claim implies that we can indeed reconstruct x_n. Given x_n, we can similarly reconstruct x_{n-1}, since ϕ_{n-1} depends only on y and x_n. Continuing in this way, we can reconstruct $x_{n-2}, \ldots, x_1$. The total number of choices for q, S, r is at most $2^{n(1-\beta+3\lambda)+nH(\tau)+\tau n} = 2^{n(1-\beta+4\lambda)}$. □

Proof of Theorem 7. Set $\lambda = \frac{\delta}{6}$. By Lemma 8,

$$\left|\mathop{\mathbb{E}}_{Y}\left[\exp(2\pi i\theta \langle x, Y\rangle)\right]\right| \leq \sqrt{\mathop{\mathbb{E}}_{Y}\left[\exp\left(-\sum_{j=1}^{n}\gamma_j \sin^2(\phi_j + 2\pi\theta x_j)\right)\right]}.$$

Whenever x is such that

$$\Pr_{Y}[G(x, Y) \leq \tau n] < 2^{-\lambda n}, \tag{2}$$

we can bound

$$\mathop{\mathbb{E}}_{Y}\left[\exp\left(-\sum_{j=1}^{n}\gamma_j \sin^2(\phi_j + 2\pi\theta x_j)\right)\right] \leq \exp(-\tfrac{\kappa}{4} n\tau \sin^2(4\pi\theta)) + 2^{-\lambda n}.$$

Since $\sqrt{a+b} \leq \sqrt{a} + \sqrt{b}$ for $a, b \geq 0$, for such an x we can bound

$$\begin{aligned} \left|\mathop{\mathbb{E}}_{Y}\left[\exp(2\pi i\theta \langle x, Y\rangle)\right]\right| &\leq \exp(-\tfrac{\kappa}{8} n\tau \sin^2(4\pi\theta)) + 2^{-\lambda n/2} \\ &\leq 2\exp(-cn \sin^2(4\pi\theta)). \end{aligned}$$

Lemma 10 promises that there are at most $2^{n(1-\beta+\delta)}$ choices for x that does not satisfy (2).

□

4. Smoothness

To prove smoothness, we use Theorem 7. The constant 4π on the r.h.s. of the bound in the theorem corresponds to a step size of 4.

Proof of Theorem 3. Theorem 7 with $\delta = \frac{\epsilon}{2}$ promises that for each $\theta \in [0,1]$, the size of

$$A_\theta = \left\{ x \in \{\pm 1\}^n : \left| \mathop{\mathbb{E}}_Y \left[\exp(2\pi i \theta \langle x, Y \rangle) \right] \right| > 2 \exp(-cn \sin^2(4\pi\theta)) \right\}$$

is at most $2^{n(1-\beta+\delta)}$. For each x, define $S_x = \{\theta \in [0,1] : x \in A_\theta\}$.

Fix x such that $|S_x| \le 2^{-\delta n}$. Bound

$$\begin{aligned}
&\left| \Pr_Y[\langle x, Y \rangle = k] - \Pr_Y[\langle x, Y \rangle = k+4] \right| \\
&= \left| \mathop{\mathbb{E}}_Y \left[\int_0^1 \exp(2\pi i \theta(\langle x, Y \rangle - k)) - \exp(2\pi i \theta(\langle x, Y \rangle - k - 4)) \, \mathrm{d}\theta \right] \right| \\
&\le \int_0^1 \left| \exp(4\pi i \theta) - \exp(-4\pi i \theta) \right| \cdot \left| \mathop{\mathbb{E}}_Y \left[\exp(2\pi i \theta \langle x, Y \rangle) \right] \right| \mathrm{d}\theta \\
&\le 2 \int_0^1 \left| \sin(4\pi\theta) \right| \cdot \left| \mathop{\mathbb{E}}_Y \left[\exp(2\pi i \theta \langle x, Y \rangle) \right] \right| \mathrm{d}\theta.
\end{aligned}$$

Continue to bound

$$\begin{aligned}
&\int_0^1 \left| \sin(4\pi\theta) \right| \cdot \left| \mathop{\mathbb{E}}_Y \left[\exp(2\pi i \theta \langle x, Y \rangle) \right] \right| \mathrm{d}\theta \\
&\le 2^{-\delta n} + 2 \int_0^1 \left| \sin(4\pi\theta) \right| \cdot \exp(-cn \sin^2(4\pi\theta)) \, \mathrm{d}\theta.
\end{aligned}$$

The integral goes around the circle twice, and it is identical in each quadrant. So,

$$\begin{aligned}
&\int_0^1 \left| \sin(4\pi\theta) \right| \cdot \exp(-cn \sin^2(4\pi\theta)) \, \mathrm{d}\theta \\
&= 8 \int_0^{1/8} \sin(4\pi\theta) \cdot \exp(-cn \sin^2(4\pi\theta)) \, \mathrm{d}\theta \\
&\le 32\pi \int_0^\infty \theta \cdot \exp(-16cn\theta^2) \, \mathrm{d}\theta \\
&\le \tfrac{c_1}{n} \int_0^\infty \phi \cdot \exp(-\phi^2) \, \mathrm{d}\phi \le \tfrac{C}{n},
\end{aligned}$$

where $c_1, C > 0$ depend on ϵ, and we used $\frac{\eta}{\pi} \le \sin(\eta) \le \eta$ for $0 \le \eta \le \frac{\pi}{2}$.

Finally, because

$$\mathop{\mathbb{E}}_x |S_x| = \mathop{\mathbb{E}}_\theta \tfrac{|A_\theta|}{2^n} \le 2^{n(-\beta+\delta)},$$

by Markov's inequality, the number of $x \in \{\pm 1\}^n$ for which $|S_x| > 2^{-\delta n}$ is at most $2^{(1-\beta+2\delta)n} = 2^{(1-\beta+\epsilon)n}$. ∎

5. Anti-concentration in General Two-Cubes

Now we move to the setting where the direction x is chosen from an arbitrary two-cube $A \subset \mathbb{Z}^n$ with differences $d_1, \ldots, d_n$; our goal is to prove Theorem 4. The way we measure the structure of A follows ideas of Halász [11]. For an integer $\ell > 0$, define $r_\ell(A)$ to be the number of elements $(\epsilon, j) \in \{\pm 1\}^{2\ell} \times [n]^{2\ell}$ that satisfy

$$\epsilon_1 \cdot d_{j_1} + \cdots + \epsilon_{2\ell} \cdot d_{j_{2\ell}} = 0.$$

The smaller $r_\ell(A)$ is, the less structured A is.

The theorem below shows that $r_\ell(A)$ allows us to control the concentration probability. More concretely, for $C > 0$ and $\ell > 0$, define

$$R_{C,\ell}(A) = \frac{C^\ell r_\ell(A)}{n^{2\ell+1/2}} + \exp(-\tfrac{n}{C}).$$

Define

$$R_C(A) = \inf\{R_{C,\ell}(A) : \ell \in \mathbb{N}\}.$$

This is essentially the bound on the concentration probability that Halász obtained in [11] when Y is uniform in $\{\pm 1\}^n$. Our upper bounds are slightly weaker. Let

$$\mu_C(A) = \inf\left\{\mu \in [0,1] : \exists \nu \in (0,1] \;\; \mu^{(1+\nu)^2} \geq 3\exp(-\tfrac{\nu n}{C}) + \tfrac{R_C(A)}{50\sqrt{\nu}}\right\},$$

where we adopt the convention that the infimum of the empty set is 1. Before stating the theorem, let us go over the three examples from Theorem 4:

(1) For arbitrary A, since $r_1(A) \leq O(n^2)$, we get[3] $\mu_C(A) \leq O(\frac{\sqrt{\ln n}}{\sqrt{n}})$ with $\nu = \frac{1}{\ln(1/R_{C,1}(A))}$.
(2) When all the differences are distinct, since $r_1(A) \leq O(n)$, we get $\mu_C(A) \leq O(n^{-1.5}\sqrt{\ln n})$ with $\nu = \frac{1}{\ln(1/R_{C,1}(A))}$.
(3) When $\{\pm d_1, \ldots, \pm d_n\}$ is a Sidon set, since $r_2(A) \leq O(n^2)$, we get $\mu_C(A) \leq O(n^{-2.5}\sqrt{\ln n})$ with $\nu = \frac{1}{\ln(1/R_{C,2}(A))}$.

More generally, when $R_C(A)$ is bound from below by some polynomial in $\frac{1}{n}$ then $\mu_C(A)$ is at most $O(R_C(A)\sqrt{\log(4/R_C(A))})$.

Theorem 12. *For every $\beta > 0$ and $\delta > 0$, there is $C > 0$ so that the following holds. Let $B \subseteq \{\pm 1\}^n$ be of size $2^{\beta n}$. Let Y be uniformly distributed in B. Let $A \subset \mathbb{Z}^n$ be a two-cube. Then, for all but $2^{n(1-\beta+\delta)}$ directions $x \in A$,*

$$\mathop{\mathbb{E}}_{\theta}\left[\left|\mathop{\mathbb{E}}_{Y}\left[\exp(2\pi i\theta \cdot \langle x, Y\rangle)\right]\right|\right] \leq \mu_C(A).$$

Before moving on, we discuss a fourth extreme example. When $A_j = \{2^j, -2^j\}$ for each $j \in [n]$, we have $r_\ell(A) \leq (2\ell n)^\ell$. In this case, setting $\ell = \Omega(n)$ gives exponentially small anti-concentration with $\nu = 1$. This result is trivial, but it illustrates that the mechanism underlying the proof yields strong bounds in many settings.

By $(\star)$ from Section 2 and the explanation above, we see that Theorem 12 implies Theorem 4. The rest of this section is devoted to the proof of Theorem 12. The

[3]Here and below the big O notation hides a constant that may depend on C.

high-level structure of the proof is similar to that of Theorem 7. However, there are several new technical challenges that we need to overcome.

The main technical challenge that needs to be overcome has to do with the definition of the set G. The G defined in the previous section depends on the angle θ. This is problematic for the proof in the generality we are working with now. So, we need to find a different set of *good* coordinates, one that depends only on x and y. Our solution is based on the following claim, which quantifies the strict convexity of the map $\zeta \mapsto \zeta^{1+\nu}$ for $\nu > 0$. We defer the proof to Appendix A.

Claim 13. *For every $\kappa > 0$, there is a constant $c_1 > 0$ so that the following holds. For every random variable $W \in \{\pm 1\}$ such that*

$$\min\left\{ \Pr[W = 1], \Pr[W = -1] \right\} \geq \kappa,$$

every $\alpha_1 \geq 2\alpha_{-1} \geq 0$ and every $0 < \nu \leq 1$,

$$\mathbb{E}\left[\alpha_W\right]^{1+\nu} \leq (1 - c_1\nu)\,\mathbb{E}\left[\alpha_W^{1+\nu}\right].$$

5.1. **A Single Direction.** The following lemma generalizes Lemma 8. Recall the definition of γ_j, ϕ_j and $J(y)$ from Sections 3.1 and 3.2.

Lemma 14. *For every $\kappa > 0$, there is a constant $c_0 > 0$ so that the following holds. For every $0 < \nu \leq 1$, every angle $\eta \in \mathbb{R}$, every direction $x \in \mathbb{Z}^n$, and every random variable Y over $\{\pm 1\}^n$,*

$$\left| \mathop{\mathbb{E}}_{Y} \left[\exp(i\eta \left\langle x, Y \right\rangle) \right] \right|^{1+\nu} \leq \mathop{\mathbb{E}}_{Y} \left[\prod_{j \in J} (1 - c_0 \nu \sin^2(\phi_j + x_j \eta) \right].$$

Proof. The proof is by induction on n. If $1 \notin J$, the proof holds by induction. The base case of $n = 1$ is trivial. So assume that $1 \in J$. Express

$$\mathop{\mathbb{E}}_{Y} \left[\exp(i\eta \left\langle x, Y \right\rangle) \right] = p_1 \exp(i\eta x_1) Z_1 + p_{-1} \exp(-i\eta x_1) Z_{-1},$$

where for $\epsilon \in \{\pm 1\}$,

$$p_\epsilon = \Pr[Y_1 = \epsilon] \qquad \& \qquad Z_\epsilon = \mathop{\mathbb{E}}_{Y | Y_1 = \epsilon} \left[\exp(i\eta \left\langle x_{>1}, Y_{>1} \right\rangle) \right].$$

When $n = 1$, we have $Z_1 = Z_{-1} = 1$. Using the definition of ϕ_1,

$$\begin{aligned}
&|p_1 \exp(i\eta x_1) Z_1 + p_{-1} \exp(-i\eta x_1) Z_{-1}|^2 \\
&= p_1^2 |Z_1|^2 + p_{-1}^2 |Z_{-1}|^2 + p_1 p_{-1} (Z_1 \overline{Z_{-1}} \exp(i2\eta x_1) + \overline{Z_1} Z_{-1} \exp(-i2\eta x_1)) \\
&= p_1^2 |Z_1|^2 + p_{-1}^2 |Z_{-1}|^2 + 2 p_1 p_{-1} |Z_1||Z_{-1}| \cos(2\phi_1 + 2x_1\eta) \\
&= p_1^2 |Z_1|^2 + p_{-1}^2 |Z_{-1}|^2 + 2 p_1 p_{-1} |Z_1||Z_{-1}| \\
&\qquad - 2 p_1 p_{-1} |Z_1||Z_{-1}| (1 - \cos(2\phi_1 + 2x_1\eta)) \\
&= \mathbb{E}\left[|Z_{Y_1}| \right]^2 - 4 p_1 p_{-1} |Z_1||Z_{-1}| \sin^2(\phi_1 + x_1\eta),
\end{aligned}$$

Without loss of generality, assume that $|Z_1| \geq |Z_{-1}|$. There are two cases to consider. The first case is that Z_1 and Z_{-1} are comparable in magnitude: $|Z_1| \leq 2|Z_{-1}|$. In this

case, we can continue the bound by

$$\begin{aligned}&\leq \mathbb{E}\left[|Z_{Y_1}|\right]^2 - 2p_1p_{-1}|Z_1|^2\sin^2(\phi_1 + x_1\eta)\\&\leq \mathbb{E}\left[|Z_{Y_1}|\right]^2(1 - 2\kappa(1-\kappa)\sin^2(\phi_1 + x_1\eta)),\end{aligned}$$

since $1 \in J$. This gives

$$\begin{aligned}&|p_1\exp(i\eta x_1)Z_1 + p_{-1}\exp(-i\eta x_1)Z_{-1}|^{1+\nu}\\&\leq \mathbb{E}\left[|Z_{Y_1}|\right]^{1+\nu}(1 - 2\kappa(1-\kappa)\sin^2(\phi_1 + x_1\eta))^{(1+\nu)/2}\\&\leq \mathbb{E}\left[|Z_{Y_1}|^{1+\nu}\right](1 - \kappa(1-\kappa)\sin^2(\phi_1 + x_1\eta)),\end{aligned}$$

since the map $\zeta \mapsto \zeta^{1+\nu}$ is convex.

The second case is when $|Z_1| > 2|Z_{-1}|$. Recall that we have already shown

$$\begin{aligned}&|p_1\exp(i\eta x_1)Z_1 + p_{-1}\exp(-i\eta x_1)Z_{-1}|^2\\&= \mathbb{E}\left[|Z_{Y_1}|\right]^2 - 4p_1p_{-1}|Z_1||Z_{-1}|\sin^2(\phi_1 + x_1\eta)\\&\leq \mathbb{E}\left[|Z_{Y_1}|\right]^2.\end{aligned}$$

Claim 13 implies that

$$\begin{aligned}&|p_1\exp(i\eta x_1)Z_1 + p_{-1}\exp(-i\eta x_1)Z_{-1}|^{1+\nu}\\&\leq \mathbb{E}\left[|Z_{Y_1}|\right]^{1+\nu}\\&\leq (1 - c_1\nu)\cdot\mathbb{E}\left[|Z_{Y_1}|^{1+\nu}\right]\\&\leq (1 - c_1\nu\sin^2(\phi_j + x_j\eta))\cdot\mathbb{E}\left[|Z_{Y_1}|^{1+\nu}\right].\end{aligned}$$

Finally, setting $c_0 = \min\{c_1, \kappa(1-\kappa)\}$, we get a bound that applies in both cases:

$$\left|\mathop{\mathbb{E}}_{Y}\left[\exp(i\eta\left\langle x, Y\right\rangle)\right]\right|^{1+\nu} \leq (1 - c_0\nu\sin^2(\phi_j + x_j\eta))\cdot\mathbb{E}\left[|Z_{Y_1}|^{1+\nu}\right].$$

This proves the base case of the induction and also allows to perform the inductive step. □

5.2. **An Average Direction.** In this section we analyze the bound from the previous section for an average direction X in a two-cube $A \subset \mathbb{Z}^n$. This step has no analogy in the proof of Theorem 7. To compute the expectation over an average direction, we reveal the entropy of X coordinate by coordinate in reverse order (from the n'th coordinate to the first one).

In analogy with $\gamma_1, \ldots, \gamma_n$, define the following functions $\mu_1, \ldots, \mu_n$. For each $j \in [n]$, let

$$\mu_j(x) = \mu_j(x_{>j}) = \min_{\epsilon \in A_j}\Pr[X_j = \epsilon | X_{>j} = x_{>j}];$$

this is well-defined for x in $A = \mathrm{supp}(X)$. In analogy with the definition of $J(y)$, let

$$J'(x) = \{j \in [n] : \mu_j(x) \geq \kappa\}.$$

In this section, we define the set G differently, but use the same notation. Let

$$G(x,y) = G_{A,B,\kappa}(x,y) = J'(x) \cap J(y).$$

Recall that γ_j, ϕ_j and $J(\cdot)$ depend on the set B, on $y \in B$ and on $x \in \mathbb{Z}^n$. In the following lemma, we fix an arbitrary $y \in B$, and take the expectation over a random $X \in A$. We allow G to be a random set that depends on X, and ϕ_j to be a random variable that depends on $X_{>j}$.

Lemma 15. *For every $\kappa > 0$ and $0 < c_0 \le 1$, there is a constant $c > 0$ so that the following holds. For every $0 < \nu \le 1$, every angle $\eta \in \mathbb{R}$, every $B \subseteq \{\pm 1\}^n$, every $y \in B$, every random variable X taking values in a two-cube $A \subseteq \mathbb{Z}^n$ with differences $d_j = u_j - v_j$,*

$$\mathop{\mathbb{E}}_{X}\left[\prod_{j\in J}(1 - c_0\nu \sin^2(\phi_j + X_j\eta))\right]^{1+\nu} \le \mathop{\mathbb{E}}_{X}\left[\exp\Big(-c\nu\sum_{j\in G}\sin^2(d_j\eta)\Big)\right].$$

Proof. The proof is by induction on n. Recall that ϕ_j and μ_j is determined by $x_{>j}$. In particular, whether or not $n \in G(x,y)$ does not depend on x. If $n \notin G(x,y)$, the proof holds by induction, or is trivially true for $n = 1$. So assume that $n \in G(x)$. Start with

$$\begin{aligned}&\mathop{\mathbb{E}}_{X}\left[\prod_{j\in J}(1 - c_0\zeta \sin^2(\phi_j + X_j\eta))\right]\\ &= \mathop{\mathbb{E}}_{X_n}\left[(1 - c_0\zeta \sin^2(\phi_n + X_n\eta))Z_{X_n}\right],\end{aligned}$$

where for $a \in A_n := \{u_n, v_n\}$,

$$Z_a = \mathop{\mathbb{E}}_{X|X_n=a}\left[\prod_{j\in J: j<n}(1 - c_0 \sin^2(\phi_j + X_j\eta))\right].$$

If $n = 1$, then $Z_u = Z_v = 1$. Assume without loss of generality that $Z_u \ge Z_v$. There are two cases to consider. The first case is that $Z_u > 2Z_v$. In this case, Claim 13 implies

$$\begin{aligned}\mathop{\mathbb{E}}_{X}\left[\prod_{j\in J}(1 - c_0\nu \sin^2(\phi_j + X_j\eta))\right]^{1+\nu} &\le \mathbb{E}\left[Z_{X_n}\right]^{1+\nu}\\ &\le (1 - c_1\nu)\,\mathbb{E}\left[Z_{X_n}^{1+\nu}\right]\\ &\le \exp(-c_1\nu)\,\mathbb{E}\left[Z_{X_n}^{1+\nu}\right].\end{aligned}$$

The second case is when $Z_u \le 2Z_v$. By Claim 11,

$$\max\left\{|\sin(\phi_n + u\eta)|, |\sin(\phi_n + v\eta)|\right\} \ge \tfrac{\sin(d_n\eta)}{2}.$$

Since $\mu_n(x) \ge \kappa$,

$$\begin{aligned}&\mathop{\mathbb{E}}_{X_n}\left[(1 - c_0\nu \sin^2(\phi_n + X_n\eta))Z_{X_n}\right]^{1+\nu}\\ &\le (\mathop{\mathbb{E}}_{X_n}\left[Z_{X_n}\right] - \kappa c_0\nu\tfrac{\sin^2(d_n\eta)}{4}\tfrac{Z_u}{2})^{1+\nu}\\ &\le (\mathop{\mathbb{E}}_{X_n}\left[Z_{X_n}\right](1 - \tfrac{\kappa c_0\nu}{8}\sin^2(d_n\eta)))^{1+\nu}\\ &\le \mathop{\mathbb{E}}_{X_n}\left[Z_{X_n}^{1+\nu}\right]\exp(-\tfrac{c_0\kappa\nu}{8}\sin^2(d_n\eta)).\end{aligned}$$

In both cases,

$$\mathop{\mathbb{E}}_{X}\left[\prod_{j\in J}(1-c_0\sin^2(\phi_j+X_j\eta))\right]^{1+\nu}\leq \exp(-c\nu\sin^2(d_n\eta))\mathop{\mathbb{E}}_{X_n}\left[Z_{X_n}^{1+\nu}\right],$$

for some constant $c(\kappa,c_0)>0$. This proves the base case of the induction and also allows to perform the inductive step.

□

5.3. **Putting It Together.**

Proof of Theorem 12. Let $\mu>0$ and $0<\nu\leq 1$ be so that

$$\mu^{(1+\nu)^2}\geq 3\exp(-\tfrac{\nu n}{C})+\tfrac{R_C(A)}{50\sqrt{\nu}};$$

if no such μ,ν exist then the theorem is trivially true. Let

$$A_0=\left\{x\in A:\mathop{\mathbb{E}}_{\theta}\left[\left|\mathop{\mathbb{E}}_{Y}\left[\exp(2\pi i\theta\cdot\langle x,Y\rangle)\right]\right|\right]\geq\mu\right\}.$$

Denote the size of A_0 by $2^{\alpha n}$. Assume towards a contradiction that $\alpha+\beta\geq 1+\delta$. Let X be uniformly distributed in A_0, independently of Y and θ. Let $\lambda=\frac{\delta}{7}$, and let κ be as in (1). By Lemma 14,

$$\begin{aligned}
&\mathop{\mathbb{E}}_{X,\theta}\left[\left|\mathop{\mathbb{E}}_{Y}\left[\exp(i2\pi\theta\left\langle x,Y\right\rangle)\right]\right|\right]^{(1+\nu)^2}\\
&\leq\mathop{\mathbb{E}}_{X,\theta}\left[\left|\mathop{\mathbb{E}}_{Y}\left[\exp(i2\pi\theta\left\langle x,Y\right\rangle)\right]\right|^{1+\nu}\right]^{1+\nu}\\
&\leq\mathop{\mathbb{E}}_{X,\theta}\left[\mathop{\mathbb{E}}_{Y}\left[\prod_{j\in J}(1-c_0\nu\sin^2(\phi_j+x_j2\pi\theta)\right]\right]^{1+\nu}.
\end{aligned}$$

By Lemma 15, we can continue

$$\begin{aligned}
&=\mathop{\mathbb{E}}_{Y,\theta}\left[\mathop{\mathbb{E}}_{X}\left[\prod_{j\in J}(1-c_0\nu\sin^2(\phi_j+x_j2\pi\theta)\right]\right]^{1+\nu}\\
&\leq\mathop{\mathbb{E}}_{Y,\theta}\left[\mathop{\mathbb{E}}_{X}\left[\prod_{j\in J}(1-c_0\nu\sin^2(\phi_j+x_j2\pi\theta)\right]^{1+\nu}\right]\\
&\leq\mathop{\mathbb{E}}_{X,Y,\theta}\left[\exp(-c\nu D(\theta))\right],
\end{aligned}$$

where

$$D(\theta)=D_{x,y}(\theta)=\sum_{j\in G(x,y)}\sin^2(2\pi\theta d_j).$$

By Lemma 9, $|J(y)|>n(\beta-3\lambda)$ for all but $2^{n(\beta-2\lambda)}$ choices for y. Similarly, $|J'(x)|>n(\alpha-3\lambda)$ for all but $2^{n(\alpha-2\lambda)}$ choices of x. By assumption, $\beta-3\lambda+\alpha-3\lambda\geq\lambda$. Since

$$|G(x,y)| \geq |J(y)| + |J'(x)| - n,$$

$$\begin{aligned}
&\Pr[|G(X,Y)| \leq \lambda n] \\
&\leq \Pr[|J(Y)| \leq n(\beta - 3\lambda)] + \Pr[|J'(X)| \leq n(\alpha - 3\lambda)] \\
&\leq 2^{-2\lambda n} + 2^{-2\lambda n}.
\end{aligned}$$

Next, we need a claim that is essentially identical to one used in the standard proof of Halász inequality [26]:

Claim. *Let x, y be so that $G(x,y) \geq \lambda n$. For every $0 \leq \rho \leq \frac{\lambda n}{4}$ and integer $\ell > 0$,*

$$\Pr_\theta[D(\theta) < \rho] \leq \frac{4r_\ell(A)}{(\lambda n)^{2\ell+1/2}}\sqrt{\rho}.$$

Given the claim, for every x, y so that $G(x,y) \geq \lambda n$ and $\ell > 0$,

$$\begin{aligned}
&\mathop{\mathbb{E}}_\theta[\exp(-c\nu D(\theta))] \\
&= \int_0^1 \Pr_\theta[\exp(-c\nu D(\theta)) > t]\,\mathrm{d}t \\
&\leq \exp(-\tfrac{c\nu\lambda n}{4}) + \int_{\exp(-c\nu\lambda n/4)}^1 \Pr_\theta[D(\theta) < -\tfrac{\ln t}{c\nu}]\,\mathrm{d}t \\
&\leq \exp(-\tfrac{c\nu\lambda n}{4}) + \frac{4r_\ell(A)}{(\lambda n)^{2\ell+1/2}} \int_0^1 \sqrt{-\tfrac{\ln t}{c\nu}}\,\mathrm{d}t.
\end{aligned}$$

The integral $\int_0^1 \sqrt{-\ln t}\,\mathrm{d}t \leq 1$ converges to a constant. For an appropriate $C = C(\beta,\delta) > 0$ and $\ell > 0$, we get the desired contradiction.

$$\begin{aligned}
\mu^{(1+\nu)^2} &\leq 2 \cdot 2^{-2\lambda n} + \exp(-\tfrac{c\nu\lambda n}{4}) + \frac{4r_\ell(A)}{\sqrt{c\nu}(\lambda n)^{2\ell+1/2}} \\
&< 3\exp(-\tfrac{\nu n}{C}) + \tfrac{R_C(A)}{50\sqrt{\nu}}.
\end{aligned}$$

Proof of Claim. Let $G = G(x,y)$. Observe that

$$\begin{aligned}
&\mathop{\mathbb{E}}_\theta\left[(|G| - 2D(\theta))^{2\ell}\right] \\
&= \mathop{\mathbb{E}}_\theta\left[\Big(\sum_{j\in G}\cos(4\pi d_j\theta)\Big)^{2\ell}\right] \\
&= 2^{-2\ell}\mathop{\mathbb{E}}_\theta\left[\Big(\sum_{j\in G}\exp(4\pi i d_j\theta) + \exp(-4\pi i d_j\theta)\Big)^{2\ell}\right] \\
&\leq 2^{-2\ell} r_\ell(A);
\end{aligned}$$

the last equality follows from the fact that of the $\leq (2|G|)^{2\ell}$ terms in the expansion, the only ones that survive are the ones with phase 0. There are at most $r_\ell(A)$ such terms, and each contributes 1.

By Markov's inequality, since $|G| \geq \lambda n$,

$$\Pr_{\theta}[D(\theta) \leq \tfrac{\lambda n}{4}] \leq \Pr_{\theta}\left[(|G| - 2D(\theta))^{2\ell} \geq (\tfrac{\lambda n}{2})^{2\ell}\right] \leq \frac{2^{-2\ell} r_{\ell}(A)}{(\lambda n/2)^{2\ell}} = \frac{r_{\ell}(A)}{(\lambda n)^{2\ell}}.$$

This proves the claim for $\rho = \frac{\lambda n}{4}$.

It remains to prove the claim for $\rho < \frac{\lambda n}{4}$. This part uses Kemperman's theorem [16] from group theory (in fact Kneser's theorem [17] for abelian groups suffices). Kemperman's theorem says that if a group is endowed with Haar measure μ, then for any compact subsets A, B of the group, $\mu(AB) \geq \min\{\mu(A) + \mu(B), 1\}$.

Think of $[0, 1)$ as the group $\mathbb{R}/\mathbb{Z}$. Let

$$S_{\rho} = \{\theta \in \mathbb{R}/\mathbb{Z} : D(\theta) \leq \rho\}.$$

We claim that the m-fold sum $S_{\rho} + S_{\rho} + \cdots + S_{\rho} \subseteq \mathbb{R}/\mathbb{Z}$ is contained in $S_{\rho m^2}$. Indeed,

$$\begin{aligned} |\sin(\eta_1 + \eta_2)| &= |\sin(\eta_1)\cos(\eta_2) + \sin(\eta_2)\cos(\eta_1)| \\ &\leq |\sin(\eta_1)| + |\sin(\eta_2)|, \end{aligned}$$

and so

$$\begin{aligned} \sin^2(\eta_1 + \cdots + \eta_m) &\leq (|\sin(\eta_1)| + \cdots + |\sin(\eta_m)|)^2 \\ &\leq m(\sin^2(\eta_1) + \cdots + \sin^2(\eta_m)). \end{aligned}$$

It follows that

$$\begin{aligned} D(\theta_1 + \theta_2 + \cdots + \theta_m) &\leq m(D(\theta_1) + D(\theta_2) + \cdots + D(\theta_m)) \\ &\leq m^2 \max\{D(\theta_1), D(\theta_2), \ldots, D(\theta_m)\}. \end{aligned}$$

Kemperman's theorem thus implies that

$$|S_{\rho m^2}| \geq |S_{\rho} + \cdots + S_{\rho}| \geq m|S_{\rho}|,$$

as long as $S_{\rho m^2}$ is not all of $\mathbb{R}/\mathbb{Z}$. Since

$$\mathop{\mathbb{E}}_{\theta}[D(\theta)] = \sum_{j \in G} \mathop{\mathbb{E}}_{\theta}\left[\sin^2(2\pi\theta d_j)\right] = \frac{|G|}{2},$$

we can deduce that $|S_{\lambda n/4}| = \Pr_{\theta}[D(\theta) \leq \frac{\lambda n}{4}]$ is strictly less than one. Hence, $S_{\lambda n/4}$ is not the full group $\mathbb{R}/\mathbb{Z}$. Setting m to be the largest integer so that $m^2\rho \leq \frac{\lambda n}{4}$, we can conclude

$$\Pr_{\theta}[D(\theta) \leq \rho] \leq \tfrac{1}{m} \Pr_{\theta}[D(\theta) \leq \rho m^2] \leq \tfrac{1}{m} \Pr_{\theta}[D(\theta) \leq \tfrac{\lambda n}{4}]. \qquad \square$$

$\square$

6. The lower bound for EGH

First, we show how to use Theorem 6 to prove Theorem 5.

Proof of Theorem 5. The main observation is that for every integer t, from a protocol that solves $\mathsf{EGH}_{tn,tk}$ over the distribution $U_{tn,tk}$, we get a randomized protocol that solves $\mathsf{EGH}_{n,k}$. The reduction is constructed as follows. Given inputs $x, y \in \{\pm 1\}^n$, first they repeat each input bit t times to obtain $x', y' \in \{\pm 1\}^{tn}$. Then they sample a uniformly random $z \in \{\pm 1\}^{tn}$ using shared randomness, and compute $x'', y'' \in \{\pm 1\}^{tn}$ by setting $x''_j = x'_j z_j$ and $y''_j = y'_j z_j$ for all $j \in [n]$. Finally, they randomly permute the coordinates of x'', y'' to obtain x''', y'''. The result is that x''', y''' are uniformly distributed among all inputs with inner product that is equal to t times the inner product of x, y. The pair (x''', y''') was generated with no communication. Finally, they run the protocol for $\mathsf{EGH}_{tn,tk}$ on x''', y'''.

Now, let α, n_0 be the constants from Theorem 6. Let $t > 0$ and $n > n_0$ be integers so that both n/t and $k = \sqrt{n}/t$ are even and $k \le \alpha\sqrt{n/t}$. By Theorem 6, any protocol for $\mathsf{EGH}_{n/t,k}$ over $U_{n/t,k}$ requires $\Omega(n/t)$ communication. By the reduction above, any protocol for $\mathsf{EGH}_n = \mathsf{EGH}_{n,\sqrt{n}}$ yields a protocol for $\mathsf{EGH}_{n/t,k}$. □

Proof of Theorem 6. Suppose the assertion of the theorem is false. By a standard argument in communication complexity, the space of inputs can be partitioned into rectangles $R_1, \ldots, R_L$ with $L \le 2^{(1-\beta)n}$, where the output of the protocol on each R_ℓ is fixed.

Let X, Y be i.i.d. uniformly at random in $\{\pm 1\}^n$. Let E denote the event that $|\langle X, Y\rangle| = k$. Define the collection of "typical" rectangles as

$$\mathbb{T} = \Big\{ \ell \in [L] \ : \ \Pr_{X,Y}[E|R_\ell] \ge \tfrac{\Pr_{X,Y}[E]}{10} \quad \& \quad \Pr_{X,Y}[R_\ell] \ge 2^{-\left(1-\frac{\beta}{2}\right)n} \Big\}.$$

For $\alpha \le 2$, because $k = n \bmod 2$, we have $\Pr_{X,Y}[E] \ge \frac{p}{\sqrt{n}}$ for some universal constant $p > 0$. The contribution of non-typical rectangles is small:

$$\begin{aligned}\sum_{\ell \notin \mathbb{T}} \Pr_{X,Y}[R_\ell|E] &= \tfrac{1}{\Pr_{X,Y}[E]} \sum_{\ell \notin \mathbb{T}} \Pr_{X,Y}[R_\ell] \Pr_{X,Y}[E|R_\ell] \\ &< \tfrac{1}{\Pr_{X,Y}[E]} \Big(L 2^{-\left(1-\frac{\beta}{2}\right)n} + \tfrac{\Pr_{X,Y}[E]}{10} \Big) < \tfrac{1}{5},\end{aligned}$$

for n large enough. Because $k = -k \bmod 4$ and $|k| < \alpha\sqrt{n}$, for each $\ell \in \mathbb{T}$, Theorem 3 with $\epsilon \ge \frac{\beta}{2}$ implies that

$$\begin{aligned}&|\Pr_{X,Y}[\langle X, Y\rangle = k|R_\ell \wedge E] - \Pr_{X,Y}[\langle X, Y\rangle = -k|R_\ell \wedge E]| \\ &= |\Pr_{X,Y}[\langle X, Y\rangle = k|R_j] - \Pr_{X,Y}[\langle X, Y\rangle = -k|R_j]| \cdot \tfrac{1}{\Pr_{X,Y}[E|R_j]} \\ &\le \alpha\sqrt{n}\tfrac{c_0}{n} \cdot \tfrac{10\sqrt{n}}{p} < \tfrac{1}{6},\end{aligned}$$

for α small enough. So, the probability of error conditioned on R_ℓ for $\ell \in \mathbb{T}$ is at least $\frac{5}{12}$. The total probability of error is at least

$$\sum_{\ell \in \mathbb{T}} \Pr_{X,Y}[R_\ell | E] \cdot \tfrac{5}{12} > \tfrac{4}{5} \cdot \tfrac{5}{12} = \tfrac{1}{3}.$$

This contradicts the correctness of the protocol. □

Acknowledgements. We wish to thank James Lee, Oded Regev, Avishay Tal and David Woodruff for helpful conversations. We also wish to thank the two anonymous reviewers for many valuable comments.

References

[1] S. Aaronson and A. Arkhipov. The computational complexity of linear optics. In *STOC*, pages 333–342, 2011.
[2] L. Babai, P. Frankl, and J. Simon. Complexity classes in communication complexity theory. In *FOCS*, pages 337–347, 1986.
[3] J. Brody and A. Chakrabarti. A multi-round communication lower bound for gap hamming and some consequences. In *CCC*, pages 358–368, 2009.
[4] J. Brody, A. Chakrabarti, O. Regev, T. Vidick, and R. De Wolf. Better gap-hamming lower bounds via better round elimination. In *Approximation, Randomization, and Combinatorial Optimization. Algorithms and Techniques*, pages 476–489. 2010.
[5] A. Chakrabarti, G. Cormode, and A. McGregor. A near-optimal algorithm for estimating the entropy of a stream. *ACM Transactions on Algorithms*, 6(3):51, 2010.
[6] A. Chakrabarti and O. Regev. An optimal lower bound on the communication complexity of gap-hamming-distance. *SIAM Journal on Computing*, 41(5):1299–1317, 2012.
[7] P. Erdös. On a lemma of littlewood and offord. *Bulletin of the American Mathematical Society*, 51(12):898–902, 1945.
[8] P. Erdos. Extremal problems in number theory. *Proc. Symp. Pure Math.*, 8, 1965.
[9] P. Erdos and P. Turan. On a problem of sidon in additive number theory and on some related problems. *J. London Math. Soc.*, 1941.
[10] P. Frankl and Z. Furedi. Solution of the littlewood-offord problem in high dimensions. *Annals of Mathematics*, pages 259–270, 1988.
[11] G. Halász. Estimates for the concentration function of combinatorial number theory and probability. *Periodica Mathematica Hungarica*, 8(3-4):197–211, 1977.
[12] P. Indyk and D. Woodruff. Tight lower bounds for the distinct elements problem. In *FOCS*, pages 283–288, 2003.
[13] V. Jain, A. Sah, and M. Sawhney. Singularity of discrete random matrices, 2021.
[14] T. S. Jayram, R. Kumar, and D. Sivakumar. The one-way communication complexity of hamming distance. *Theory of Computing*, 4(1):129–135, 2008.
[15] Kahn, Komlos, and Szemeredi. On the probability that a random +/-1 matrix is singular. *JAMS: Journal of the American Mathematical Society*, 8, 1995.
[16] J. H. B. Kemperman. On products of sets in a locally compact group. *Fund. Math.*, 56:5168, 1964.
[17] M. Kneser. Summenmengen in lokalkompakten abelschen gruppen. *Math. Z.*, 66:88110, 1956.
[18] E. Kushilevitz and N. Nisan. *Communication Complexity*. Cambridge University Press, 2006.
[19] J. E. Littlewood and A. C. Offord. On the number of real roots of a random algebraic equation (iii). *Rec. Math. (Mat. Sbornik) N.S*, 12(3):277–286, 1943.
[20] A. Rao and A. Yehudayoff. *Communication complexity and applications*. Cambridge University Press, 2020.
[21] A. Sarkozy, , and E. Szemeredi. Uber ein problem von erdos und moser. *Acta Arithmetica*, 1965.

[22] A. A. Sherstov. The communication complexity of gap hamming distance. *Theory of Computing*, 8(1):197–208, 2012.
[23] R. Stanley. Weyl groups, the hard lefschetz theorem, and the sperner property. *SIAM J. Algebraic Discrete Methods*, 1980.
[24] M. Talagrand. Concentration of measure and isoperimetric inequalities in product spaces. *Publications Mathématiques de l'Institut des Hautes Etudes Scientifiques*, 81(1):73–205, 1995.
[25] T. Tao and V. Vu. A sharp inverse littlewood-offord theorem. *Random Structures & Algorithms*, 37(4):525–539, 2010.
[26] T. Tao and V. H. Vu. Additive combinatorics. *Cambridge University Press*.
[27] T. Tao and V. H. Vu. Inverse littlewood-offord theorems and the condition number of random discrete matrices. *Annals of Mathematics*, pages 595–632, 2009.
[28] T. Tran. The smallest singular value of random combinatorial matrices, 2020.
[29] T. Vidick. A concentration inequality for the overlap of a vector on a large set. *Chicago Journal of Theoretical Computer Science*, 1:1–12, 2012.
[30] D. P. Woodruff. The average-case complexity of counting distinct elements. In *ICDT*, pages 284–295, 2009.

Appendix A. Strict Convexity

Proof of Claim 13. If $\alpha_1 = 0$, then the claim is trivially true. So, assume that $\alpha_1 > 0$. Without loss of generality, we may also assume that $\kappa > 0$ is small enough so that $4^\kappa > \exp(\kappa + \kappa^2)$.

Let $p = \Pr[W = 1] \in [\kappa, 1-\kappa]$ and $\xi = \frac{\alpha_{-1}}{\alpha_1} \in [0, \frac{1}{2}]$. So,

$$\frac{\mathbb{E}\left[\alpha_W\right]^{1+\nu}}{\mathbb{E}\left[\alpha_W^{1+\nu}\right]} = \frac{(p+(1-p)\xi)^{1+\nu}}{p+(1-p)\xi^{1+\nu}}.$$

We need to upper bound this ratio by $1 - c_1\nu$, for some constant c_1 that depends only on κ. Let

$$\Phi(\xi, p, \nu) = (p+(1-p)\xi^{1+\nu}) - (p+(1-p)\xi)^{1+\nu}.$$

We shall argue that there is a constant $c_1 = c_1(\kappa) > 0$ such that $\Phi(\xi, p, \nu) \geq c_1\nu$. This completes the proof, since

$$\frac{(p+(1-p)\xi)^{1+\nu}}{(p+(1-p)\xi^{1+\nu})} = 1 - \frac{\Phi(\xi,p,\nu)}{(p+(1-p)\xi^{1+\nu})} < 1 - c_1\nu.$$

First, we show that for every ν and ξ, the function $\Phi(\xi, p, \nu)$ is minimized when $p = \kappa$. Consider

$$\begin{aligned}\frac{\partial\Phi}{\partial p} &= 1 - \xi^{1+\nu} - (1+\nu)(p+(1-p)\xi)^\nu(1-\xi) \\ &\geq 1 - \xi^{1+\nu} - (1+\nu)(1-\xi) \\ &\geq \xi(1+\nu-\xi^\nu) > 0,\end{aligned}$$

since $\xi^\nu < 1$. So, the minimum is achieved when $p = \kappa$.

Second, we claim that for every ν and p, the function $\Phi(\xi, p, \nu)$ is minimized when $\xi = \frac{1}{2}$. Consider

$$\begin{aligned}\frac{\partial \Phi}{\partial \xi} &= (1-p)(1+\nu)\xi^{\nu} - (1+\nu)(p+(1-p)\xi)^{\nu}(1-p) \\ &= (1-p)(1+\nu)(\xi^{\nu} - (p+(1-p)\xi)^{\nu}) < 0,\end{aligned}$$

since $p + (1-p)\xi > \xi$. So, the minimum is achieved when $\xi = 1/2$.

Third, we control the derivative with respect to ν for $\xi = \frac{1}{2}$ and $p = \kappa$. Consider

$$\begin{aligned}\frac{\partial \Phi}{\partial \nu}(\tfrac{1}{2}, \kappa, \nu) &= (1-\kappa)\ln(\tfrac{1}{2})(\tfrac{1}{2})^{1+\nu} - \ln(\tfrac{1+\kappa}{2})(\tfrac{1+\kappa}{2})^{1+\nu} \\ &\geq (\tfrac{1}{2})^2((1-\kappa)\ln(\tfrac{1}{2}) - \ln(\tfrac{1+\kappa}{2})(1+\kappa)^{1+\nu}),\end{aligned}$$

since $\nu \leq 1$. The expression

$$(1-\kappa)\ln(\tfrac{1}{2}) - \ln(\tfrac{1+\kappa}{2})(1+\kappa)^{1+\nu}$$

only increases with ν. When $\nu = 0$, this expression is

$$\ln(\tfrac{2^{2\kappa}}{(1+\kappa)^{1+\kappa}}) \geq \ln(\tfrac{4^{\kappa}}{\exp(\kappa(1+\kappa))}) > 0,$$

since $4^{\kappa} > \exp(\kappa + \kappa^2)$. This proves that $\frac{\partial \Phi}{\partial \nu}(\frac{1}{2}, \kappa, \nu) > c_1$ for some constant $c_1(\kappa) > 0$.

Finally,

$$\Phi(\xi, p, \nu) \geq \Phi(\tfrac{1}{2}, \kappa, \nu) = \int_0^{\nu} \frac{\partial \Phi}{\partial \nu}(\tfrac{1}{2}, \kappa, \zeta)\, \mathrm{d}\zeta \geq \int_0^{\nu} c_1\, \mathrm{d}\zeta = c_1 \nu. \quad \square$$

School of Computer Science, University of Washington
E-mail address: anuprao@cs.washington.edu

Department of Mathematics, Technion-IIT
E-mail address: amir.yehudayoff@gmail.com