



ANTI-CONCENTRATION IN MOST DIRECTIONS

AMIR YEHUDAYOFF

ABSTRACT. We prove anti-concentration for the inner product of two independent random vectors in the discrete cube. Our results imply Chakrabarti and Regev’s lower bound on the randomized communication complexity of the gap-hamming problem. They are also meaningful in the context of randomness extraction. The proof provides a framework for establishing anti-concentration in discrete domains. The argument has two different components. A local component that uses harmonic analysis, and a global (‘information theoretic’) component.

1. INTRODUCTION

Anti-concentration of a random processes means that the distribution of outcomes is not concentrated in a small region. No single outcome is obtained too often. It plays an important role in mathematics and computer science (see e.g. [11, 16, 1] and references within).

The standard example is a sum of i.i.d. random variables. If X in $\mathcal{X} = \{\pm 1\}^n$ is fixed, and B is uniformly distributed in \mathcal{X} , then the random integer $\langle B, X \rangle = \sum_i B_i X_i$ is anti-concentrated. The probability that $\langle B, X \rangle$ takes any specific value is at most $O(1/\sqrt{n})$. This was studied and generalized by Littlewood and Offord [11], Erdős [7], and many others. Higher dimensional analogs of this phenomenon were studied by Frankl and Füredi [8], Halász [9] and others.

It is interesting to understand the generality of this phenomenon (see also [15] and references within). Anti-concentration certainly fails when the entropy of B is not full. We can, for example, condition B on the pretty likely event that $\langle B, X \rangle = 0$.

Can we somehow recover anti-concentration? A natural suggestion is to allow X to be random as well. This indeed recovers anti-concentration, as the following theorem shows.

Theorem (Chakrabarti and Regev [4]). *There is a constant $c > 0$ so that the following holds. Let $\mathcal{A}, \mathcal{B} \subseteq \mathcal{X}$ be of sizes at least $2^{(1-c)n}$. If*

Partially supported by ISF grant 1162/15. This work was done while the author was visiting the Simons Institute for the Theory of Computing.

(A, B) is uniformly distributed in $\mathcal{A} \times \mathcal{B}$ then

$$\mathbb{P}[|\langle A, B \rangle| \leq c\sqrt{n}] \leq 1 - c.$$

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 translated it to the cube. Vidick [18] later simplified part of their argument, but stayed in the geometric setting. Sherstov [13] found a third proof that uses Talagrand's inequality from convex geometry [14] and ideas of Babai, Frankl and Simon from communication complexity [2].

We generalize the theorem above.

Theorem. *For every $\beta > 0$, there are $c, C > 0$ so that the following holds. Let $\mathcal{A}, \mathcal{B} \subseteq \mathcal{X}$ be so that $|\mathcal{B}| = 2^{\beta n}$ and $|\mathcal{A}| \geq 2^{(1-c)n}$. If (A, B) is uniformly distributed in $\mathcal{A} \times \mathcal{B}$ then for all $w \in \mathbb{Z}$,*

$$\mathbb{P}[\langle A, B \rangle = w] \leq \frac{C}{\sqrt{n}}.$$

This theorem is part of a more fundamental phenomenon. When studying anti-concentration, what we are ultimately interested in is proving point-wise estimates. Namely, we would like to control the *concentration probability* or the ℓ_∞ norm¹

$$\|\nu\|_\infty = \max_{\omega \in \Omega} \nu(\omega)$$

(see [16] and references within).

Our main result is the following sharp bound on the concentration probability. Let X be uniformly distributed in \mathcal{X} . Let \mathcal{B} be a family of vectors in \mathcal{X} of size $2^{\beta n}$. Let μ_X be the distribution of $\langle B, X \rangle$ with X fixed and B uniformly distributed in \mathcal{B} .

Theorem 1. *For every $\beta > 0$, there are $c, C > 0$ so that*

$$\mathbb{P}_X \left[\|\mu_X\|_\infty > \frac{C}{\sqrt{n}} \right] < C2^{-cn}.$$

To prove the theorem, we build a flexible framework for proving anti-concentration results in discrete domains. Think of the random variable $\langle B, X \rangle$ as built in n steps. It starts as 0, and $B_d X_d$ is added to $\langle B_{<d}, X_{<d} \rangle$ to generate $\langle B_{\leq d}, X_{\leq d} \rangle$. To analyze the behavior of this system, we first show that locally entropy often increasing (Section 2). This part of the argument uses harmonic analysis (even though our ultimate goal is not the ℓ_2 norm). The second part of the argument is macroscopic (Section 3). We identify a global event that guarantees that the small local increments in entropy yield substantial entropy

¹We consider only finite probability spaces in this text.

in the whole system. The last step is proving that the macroscopic event almost always holds. This is achieved by an encoding argument. Situations where the macroscopic entropy is not high can be described by a small number of bits.

There are several differences between our argument and the ones in [4, 18, 13]. The main difference is that the arguments from [4, 18, 13] are based, in one way or another, on the geometry of euclidean space. The arguments in [4, 18] prove a correlation inequality in gaussian space and translate it to the discrete world. It seems that such an argument can not yield effective bounds on the concentration probability in the discrete setting. A common ingredient to [4, 13] is showing that every set of large enough measure contains many almost orthogonal vectors (this is called ‘identifying the hard core’ in [13]). In [18] this part of the argument is replaced by a statement about a relevant matrix. Our argument does not contain any such step.

Related topics.

Communication complexity. Chakrabarti and Regev’s main motivation was understanding the randomized communication complexity of the gap-hamming problem. The gap-hamming problem was introduced by Indyk and Woodruff in the context of streaming [10]. Proving lower bounds on its communication complexity was a central open problem for almost ten years, until Chakrabarti and Regev solved it [4]. Vidick [18] and Sherstov [13] later simplified the proof.

Our results also imply the lower bound for the randomized communication complexity of the gap-hamming problem (see e.g. [13]). As opposed to [4, 18, 13], the proof presented here lies entirely in the discrete domain. The underlying ideas may therefore be of independent interest.

Pseudorandomness. Randomness is a computational resource [17]. There are many sources of randomness, and some of them are *weak* or imperfect. Randomness extractors allow to use weak sources of randomness as if they were perfect.

The study of randomness extractors is about constructing explicit maps that transform weak sources of randomness to almost uniform outputs. The main goal is generating a uniform output in the most general scenario possible. This often requires ingenious constructions.

The scenario described above fits nicely in the context of *two-source extractors*. A two-source extractor maps two independent random variables A and B with significant min-entropy to a single almost uniform output.

Chor and Goldreich [6] used Lindsey's lemma to show that inner product modulo two is a two-source extractor. The bit $\langle A, B \rangle \bmod 2$ is close to a uniform random bit as long as $|\mathcal{A}| \cdot |\mathcal{B}| \gg 2^n$. Bourgain [3], Raz [12] and Chattopadhyay and Zuckerman [5] constructed two-source extractors with much better parameters.

This work can be interpreted as studying a related but somewhat different question. The high-level suggestion is to investigate what other pseudorandom properties known extractors satisfy.

We already know that inner product is an excellent two-source extractor. Now we also know that over the integers inner product is anti-concentrated. This is not as good as being uniform, but inner product is not uniform over the integers (it is binomial).

2. MICROSCOPICALLY

Here we analyze the local behavior. The argument is spectral and uses harmonic analysis. We work over the abelian group $\mathbb{Z}_N = \mathbb{Z}/N\mathbb{Z}$ with $N = \lceil \sqrt{n} \rceil$. This choice of N allows to deduce the $O(\frac{1}{\sqrt{n}})$ estimate.

2.1. Harmonic analysis. The group \mathbb{Z}_N acts on the vector space of functions from \mathbb{Z}_N to \mathbb{C} . This vector space is endowed with the standard inner product $\langle f, g \rangle = \sum_{z \in \mathbb{Z}_N} f(z) \overline{g(z)}$ where $\bar{\xi}$ is the complex conjugate of $\xi \in \mathbb{C}$. For $z \in \mathbb{Z}_N$, let S_z be the operator that shifts the function $f : \mathbb{Z}_N \rightarrow \mathbb{C}$ by z . That is, $S_z f(x) = f(x - z)$ for all x . The shifts are unitary and they commute. Let $\{e_z : z \in \mathbb{Z}_N\}$ be the set of the N normalized eigenvectors:

$$e_z(x) = \frac{\psi_a(x)}{\sqrt{N}},$$

where

$$\psi_a(x) = e^{2\pi i \frac{zx}{N}}.$$

The eigenvalue of e_z with respect to the shift S_1 is $\lambda_z = e^{-2\pi i \frac{z}{N}}$. The Fourier transform of f is $\widehat{f} : \mathbb{Z}_N \rightarrow \mathbb{C}$ defined by

$$\widehat{f}(z) = \langle f, e_z \rangle.$$

2.2. Local changes. The first observation is that the coefficients of the Fourier transform do not increase. Assume we have two distributions μ_1 and μ_{-1} on \mathbb{Z}_N . Think about them as distributions of two particles Z_1 and Z_{-1} . We use them to generate a new distribution

$$\nu = \gamma S_1 \mu_1 + (1 - \gamma) S_{-1} \mu_{-1}$$

where $\gamma \in [0, 1]$. Stated differently, we consider the particle

$$Z = \begin{cases} Z_1 + 1 & \text{with probability } \gamma, \\ Z_{-1} - 1 & \text{with probability } 1 - \gamma. \end{cases}$$

Observation 2. For all $z \in \mathbb{Z}_N$, since $|\lambda_z| = 1$ using convexity,

$$|\widehat{\nu}(z)|^2 \leq \gamma |\widehat{\mu}_1(z)|^2 + (1 - \gamma) |\widehat{\mu}_{-1}(z)|^2.$$

The main lemma is a quantitative estimate of the local change in the Fourier coefficients. For $x \in \{\pm 1\}$, let

$$\nu_x = \gamma S_x \mu_1 + (1 - \gamma) S_{-x} \mu_{-1}$$

Let $X \in \{\pm 1\}$ be distributed uniformly at random.

Lemma 3. Let $\gamma \in [\gamma_0, 1 - \gamma_0]$ with $\gamma_0 \in [0, \frac{1}{2}]$. For every $z \in \mathbb{Z}_N$,

$$\mathbb{E}_X |\widehat{\nu}_X(z)|^2 \leq (1 - \frac{\gamma_0}{2} + \frac{\gamma_0}{2} |\cos(4\pi \frac{z}{N})|) (\gamma |\widehat{\mu}_1(z)|^2 + (1 - \gamma) |\widehat{\mu}_{-1}(z)|^2).$$

Proof of Lemma 3. Start by bounding the geometric average:

$$\begin{aligned} & |\widehat{\nu}_1(z)| \cdot |\widehat{\nu}_{-1}(z)| \\ &= |\gamma \lambda_z \widehat{\mu}_1(z) + (1 - \gamma) \lambda_{-z} \widehat{\mu}_{-1}(z)| \cdot |\gamma \lambda_{-z} \widehat{\mu}_1(z) + (1 - \gamma) \lambda_z \widehat{\mu}_{-1}(z)| \\ &\leq \gamma^2 |\widehat{\mu}_1(z)|^2 + (1 - \gamma)^2 |\widehat{\mu}_{-1}(z)|^2 \\ &\quad + \gamma(1 - \gamma) |\widehat{\mu}_1(z) \widehat{\mu}_{-1}(z) (\lambda_{2z} + \lambda_{-2z})| \\ &\leq \gamma^2 |\widehat{\mu}_1(z)|^2 + (1 - \gamma)^2 |\widehat{\mu}_{-1}(z)|^2 \\ &\quad + \gamma(1 - \gamma) (|\widehat{\mu}_1(z)|^2 + |\widehat{\mu}_{-1}(z)|^2) |\cos(4\pi \frac{z}{N})| \\ &= \gamma |\widehat{\mu}_1(z)|^2 (\gamma + (1 - \gamma) |\cos(4\pi \frac{z}{N})|) \\ &\quad + (1 - \gamma) |\widehat{\mu}_{-1}(z)|^2 (1 - \gamma + \gamma |\cos(4\pi \frac{z}{N})|) \\ &\leq (1 - \gamma_0 + \gamma_0 |\cos(4\pi \frac{z}{N})|) (\gamma |\widehat{\mu}_1(z)|^2 + (1 - \gamma) |\widehat{\mu}_{-1}(z)|^2). \end{aligned}$$

Thus,

$$\begin{aligned} & \min\{|\widehat{\nu}_1(z)|^2, |\widehat{\nu}_{-1}(z)|^2\} \\ & \leq (1 - \gamma_0 + \gamma_0 |\cos(4\pi \frac{z}{N})|) (\gamma |\widehat{\mu}_1(z)|^2 + (1 - \gamma) |\widehat{\mu}_{-1}(z)|^2). \end{aligned}$$

By Observation 2,

$$\begin{aligned} & \mathbb{E}_X |\widehat{\nu}_X(z)|^2 \\ & \leq (1 - \frac{\gamma_0}{2} + \frac{\gamma_0}{2} |\cos(4\pi \frac{z}{N})|) (\gamma |\widehat{\mu}_1(z)|^2 + (1 - \gamma) |\widehat{\mu}_{-1}(z)|^2). \quad \square \end{aligned}$$

3. MACROSCOPICALLY

We now analyze the global behavior. We use a *decision tree* to represent the system (Section 3.1). This representation allows to identify positions where the system ‘mixes’ (Section 3.2). We then show that all but a tiny fraction of positions are mixing (Section 3.3).

3.1. Representing the system. Think of the elements of \mathcal{X} as vectors (x_1, x_2, \dots, x_n) . Consider a full binary tree of depth n . The root v_0 has depth n and is labeled by the variable x_n . The two children of the root have depth $n - 1$ and are labelled by x_{n-1} . In general, all vertices of depth d are labelled by x_d . The depth of the leaves is 0. Every $x \in \mathcal{X}$ defines a walk from the root v_0 to a leaf in the tree. We identify between \mathcal{X} and the leaves in the tree.

Let \mathcal{B} be a collection of vectors in \mathcal{X} of size $|\mathcal{B}| = 2^{\beta n}$. Let P be the uniform distribution on \mathcal{B} . The elements of \mathcal{B} correspond to leaves in the tree. Every vertex v in the tree corresponds to the set $\mathcal{B}(v) \subseteq \mathcal{B}$ of all leaves in \mathcal{B} that are under v . Let P_v be the uniform distribution on $\mathcal{B}(v)$. Let γ_v be the distribution on $\{\pm 1\}$ that is inducted by P_v on the bit from v to its children. If the depth of v is $d = d(v)$ then γ_v is the marginal of P_v on the d 'th coordinate.

Fix a parameter $\gamma_0 \in (0, 1/2)$. Call a vertex v *unbiased* if

$$\gamma_v(1) \in [\gamma_0, 1 - \gamma_0].$$

Intuitively, unbiased vertices are positions where the entropy of the system can potentially grow. The unbiased count $\#\mathbf{ub}(v)$ of a vertex v is the number of unbiased vertices on the path from v to the root.

The following claim shows that if \mathcal{B} is large then there are many unbiased vertices.²

Claim 4. *For every $\alpha > 0$, the number of leaves v with $\#\mathbf{ub}(v) < \alpha n$ is at most $2^{n(\alpha + H(\alpha) + H(\beta/\log(1/\gamma_0)))}$.*

Proof. Encode a leaf v with $\#\mathbf{ub}(v) < \alpha n$ using the following data:

- (1) The depths at which the unbiased nodes appear. There are at most $2^{nH(\alpha)}$ such options.
- (2) The value of the path that reaches v at these depths. There are at most $2^{\alpha n}$ such options.
- (3) The depths at which the path that reaches v goes through the minority side of a vertex that is not unbiased. If there are δn such depths then $\gamma_0^{\delta n} \geq P(v) = 2^{-\beta n}$. There are, therefore, at most $2^{nH(\beta/\log(1/\gamma_0))}$ such options.

² $H(\xi) = -\xi \log(\xi) - (1 - \xi) \log(1 - \xi)$ is the binary entropy function.

□

3.2. Mixing vertices. We analyze the behavior of the system for a fixed $x \in \mathcal{X}$. In the next section, we see what happens if x is random.

For a vertex v , define a distribution $\mu_v = \mu_{x,v}$ over the integers. If v has depth $d = d(v) > 0$, define μ_v to be the distribution of the inner product $\langle B_{\leq d}, x_{\leq d} \rangle$ where $B \sim P_v$. The distribution μ_v when v is a leaf gives mass 1 to the integer 0.

Call a vertex v with two children v_1 and v_{-1} z -mixing if it is unbiased and

$$(1) \quad \begin{aligned} & |\widehat{\mu}_v(z)| \\ & \leq (1 - \frac{\gamma_0}{2} + \frac{\gamma_0}{2} |\cos(4\pi \frac{z}{N})|) (\gamma |\widehat{\mu}_{v_1}(z)|^2 + (1 - \gamma) |\widehat{\mu}_{v_{-1}}(z)|^2). \end{aligned}$$

This definition makes sense with Lemma 3 in mind.

For a vertex v and a leaf $u \in \mathcal{B}(v)$, denote by $\#\text{mix}(v \rightarrow u)$ the number of z -mixing vertices (including v) on the path from v to u . Let $\alpha_0 > 0$ be a parameter. Recall that v_0 is the root of the tree. Define the set of ‘good’ leaves as

$$\mathcal{G} = \mathcal{G}_z = \{u \in \mathcal{B} : \#\text{mix}(v_0 \rightarrow u) \geq \frac{\alpha_0}{4} n\}.$$

Define

$$q(v) = q_z(v) = \min\{\#\text{mix}(v \rightarrow u) : u \in \mathcal{B}(v) \cap \mathcal{G}\};$$

when $\mathcal{B}(v) \cap \mathcal{G} = \emptyset$, define $q(v) = \min \emptyset = \infty$. A crucial observation is that if $\mathcal{G} \neq \emptyset$ then $q(v_0) \geq \frac{\alpha_0}{4} n$.

The measure $q(v)$ allows to control the Fourier coefficients at the vertex v .

Lemma 5. *For every $z \in \mathbb{Z}_N$ and every vertex v so that $\mathcal{B}(v) \neq \emptyset$,*

$$|\widehat{\mu}_v(z)|^2 \leq \frac{1}{N} (1 - \frac{\gamma_0}{2} + \frac{\gamma_0}{2} |\cos(4\pi \frac{z}{N})|)^{q(v)} + P_v(\neg \mathcal{G}).$$

The lemma is most interesting at the root v_0 :

$$(2) \quad |\widehat{\mu}_{v_0}(z)|^2 \leq \frac{1}{N} (1 - \frac{\gamma_0}{2} + \frac{\gamma_0}{2} |\cos(4\pi \frac{z}{N})|)^{\frac{\alpha_0 n}{4}} + P(\neg \mathcal{G}).$$

Proof. The proof is by induction. The induction base is when v is a leaf in \mathcal{B} . If v is not in \mathcal{G} then $P_v(\neg \mathcal{G}) = 1$ (in this case $q(v) = \infty$). If v is in \mathcal{G} then $q(v) = 0$. In both cases, the lemma holds since

$$|\widehat{\mu}_v(z)| = |\langle \mu_v, e_z \rangle| \leq \|\mu_v\|_1 \|e_z\|_\infty = \frac{1}{\sqrt{N}}.$$

For the induction step, denote by v_1 and v_{-1} the two children of v . Simplify notation:

$$q = q(v), \quad d = d(v) \quad \& \quad \gamma = \gamma_v(1).$$

Since

$$\langle B_{\leq d}, x_{\leq d} \rangle = \begin{cases} \langle B_{< d}, x_{< d} \rangle + x_d & B_d = 1 \\ \langle B_{< d}, x_{< d} \rangle - x_d & B_d = -1 \end{cases}$$

we can write

$$\mu_v = \gamma S_{x_d} \mu_{v_1} + (1 - \gamma) S_{-x_d} \mu_{v_{-1}}.$$

There are two cases to consider.

Non-mixing. If v is not mixing then $q = \min\{q(v_1), q(v_{-1})\}$. Since $P_v(-\mathcal{G}) = \gamma P_{v_1}(-\mathcal{G}) + (1 - \gamma) P_{v_{-1}}(-\mathcal{G})$, Observation 2 and induction imply

$$|\widehat{\mu}_v(z)|^2 \leq \frac{1}{N} (1 - \frac{\gamma_0}{2} + \frac{\gamma_0}{2} |\cos(4\pi \frac{z}{N})|)^q + P_v(-\mathcal{G}).$$

Mixing. If v is mixing then $q = 1 + \min\{q(v_1), q(v_{-1})\}$. In this case, by definition and induction,

$$\begin{aligned} |\widehat{\mu}_v(z)| &\leq (1 - \frac{\gamma_0}{2} + \frac{\gamma_0}{2} |\cos(4\pi \frac{z}{N})|) (\gamma |\widehat{\mu}_{v_1}(z)|^2 + (1 - \gamma) |\widehat{\mu}_{v_{-1}}(z)|^2) \\ &\leq \frac{1}{N} (1 - \frac{\gamma_0}{2} + \frac{\gamma_0}{2} |\cos(4\pi \frac{z}{N})|)^q + P_v(-\mathcal{G}). \end{aligned}$$

□

3.3. Many good leaves. The previous section (Lemma 5) highlights the role of the good leaves \mathcal{G} . We need to show that \mathcal{G} is typically almost full.

Lemma 6. *There is a constant $c = c(\alpha_0) > 0$ so that for each $z \in \mathbb{Z}_N$,*

$$\mathbb{E}_X[P(-\mathcal{G}_z)] < 2^{-cn} + \frac{2^{n(\alpha_0 + H(\alpha_0) + H(\beta/\log(1/\gamma_0)))}}{|\mathcal{B}|}.$$

Proof.

$$\mathbb{E}_X[P(-\mathcal{G}_z)] = \frac{1}{|\mathcal{B}|} \sum_{v \in \mathcal{B}} \mathbb{P}_X[\#\text{mix}(v) < \frac{\alpha_0}{4} n].$$

By Claim 4, the number of leaves v with $\#\text{ub}(v) < \alpha_0 n$ is at most $2^{n(\alpha_0 + H(\alpha_0) + H(\beta/\log(1/\gamma_0)))}$. We can thus focus on the rest of the leaves. Let v be a leaf in \mathcal{B} with $K = \#\text{ub}(v) \geq \alpha_0 n$. Let u_1, \dots, u_K be the unbiased vertices on the path from the root to v . Denote by E_k the indicator random variable for the event that u_k is z -mixing (namely, (1) holds for u_k).

Claim 7. *For each $k > 1$, we have $\mathbb{E}[E_k | E_1, \dots, E_{k-1}] \geq \frac{1}{2}$.*

Proof. Fix $u = u_k$ of depth d . Denote its two children by u_1 and u_{-1} . Fix $X_{< d}$ so that $\mu_X(u_1)$ and $\mu_X(u_{-1})$ are fixed as well. Let X_d be uniform in $\{\pm 1\}$. Lemma 3 implies that for at least one choice of X_d the vertex u is mixing. □

By the claim, the sequence of random variables S_0 and $S_k = E_1 + E_2 + \dots + E_k - \frac{k}{2}$ is a submartingale. Azuma's inequality implies that

$$\mathbb{P}[\#\text{mix}(v) < \frac{\alpha_0}{4}n] \leq \mathbb{P}[S_K - S_0 < -\frac{K}{4}] \leq e^{-\frac{K}{32}}.$$

□

3.4. Putting it together.

Proof of Theorem 1. Let X be uniformly distributed in \mathcal{X} . Let \mathcal{B} be a family of vectors in \mathcal{X} of size $|\mathcal{B}| = 2^{\beta n}$. Let $\alpha_0, \gamma_0 > 0$ be so that $4(\alpha_0 + H(\alpha_0)) \leq \beta$ and $4H(\beta/\log(1/\gamma_0)) \leq \beta$. By Lemma 6, there is a constant $c' = c'(\beta) > 0$ so that for each $z \in \mathbb{Z}_N$,

$$\mathbb{E}_X[P(-\mathcal{G}_z)] < 2^{-c'n} + \frac{2^{n(\alpha_0 + H(\alpha_0) + H(\beta/\log(1/\gamma_0)))}}{|\mathcal{B}|}.$$

By Markov's inequality and the union bound,

$$\mathbb{P}_X[\exists z \in \mathbb{Z}_N \ P(-\mathcal{G}_z) > 2^{-cn}] < 2N2^{-2cn} \leq C2^{-cn}$$

for some $c, C > 0$ that depend on β .

Fix x so that $P(-\mathcal{G}_z) \leq 2^{-cn}$ for all $z \in \mathbb{Z}_N$. Let $\mu = \mu_x(v_0)$. By (2), for all $z \in \mathbb{Z}_N$

$$|\hat{\mu}(z)| > \frac{1}{\sqrt{N}} \sqrt{(1 - \frac{\gamma_0}{2} + \frac{\gamma_0}{2} |\cos(4\pi \frac{z}{N})|)^{\frac{\alpha_0 n}{4}} + 2^{-cn}}.$$

Since $N \approx \sqrt{n}$ and $\cos(\xi) \approx 1 - \xi^2$ near zero,

$$\|\hat{\mu}\|_1 \leq \frac{C}{\sqrt{N}}.$$

Thus, for every $w \in \mathbb{Z}_N$,

$$\mu(w) = \sum_z \hat{\mu}(z) e_z(w) \leq \|\hat{\mu}\|_1 \|e_z\|_\infty \leq \frac{C}{N}.$$

□

Acknowledgement. I wish to thank Oded Regev, Avishay Tal and David Woodruff for helpful conversations. Anup Rao contributed to the getting the sharp bound on the concentration probability (the previous version of the text had a weaker bound).

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. Bourgain. More on the sum-product phenomenon in prime fields and its applications. *International Journal of Number Theory*, 1(01):1–32, 2005.

- [4] 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.
- [5] E. Chattopadhyay and D. Zuckerman. Explicit two-source extractors and resilient functions. In *STOC*, pages 670–683, 2016.
- [6] B. Chor and O. Goldreich. Unbiased bits from sources of weak randomness and probabilistic communication complexity. *SIAM Journal on Computing*, 17(2):230–261, 1988.
- [7] P. Erdős. On a lemma of littlewood and offord. *Bulletin of the American Mathematical Society*, 51(12):898–902, 1945.
- [8] P. Frankl and Z. Füredi. Solution of the littlewood-offord problem in high dimensions. *Annals of Mathematics*, pages 259–270, 1988.
- [9] G. Halász. Estimates for the concentration function of combinatorial number theory and probability. *Periodica Mathematica Hungarica*, 8(3-4):197–211, 1977.
- [10] P. Indyk and D. Woodruff. Tight lower bounds for the distinct elements problem. In *FOCS*, pages 283–288, 2003.
- [11] 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.
- [12] R. Raz. Extractors with weak random seeds. In *STOC*, pages 11–20, 2005.
- [13] A. A. Sherstov. The communication complexity of gap hamming distance. *Theory of Computing*, 8(1):197–208, 2012.
- [14] 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.
- [15] T. Tao and V. Vu. A sharp inverse littlewood-offord theorem. *Random Structures & Algorithms*, 37(4):525–539, 2010.
- [16] 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.
- [17] L. Trevisan and S. Vadhan. Extracting randomness from samplable distributions. In *FOCS*, pages 32–42, 2000.
- [18] 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.

DEPARTMENT OF MATHEMATICS, TECHNION-IIT
 Email address: amir.yehudayoff@gmail.com