

More efficient sifting for grid norms, and applications to multiparty communication complexity

Zander Kelley*

Xin Lyu†

Abstract

Building on the techniques behind the recent progress on the 3-term arithmetic progression problem [KM23], Kelley, Lovett, and Meka [KLM24] constructed the first explicit 3-player function $f : [N]^3 \rightarrow \{0, 1\}$ that demonstrates a strong separation between randomized and (non-)deterministic NOF communication complexity. Specifically, their hard function can be solved by a randomized protocol sending $O(1)$ bits, but requires $\Omega(\log^{1/3}(N))$ bits of communication with a deterministic (or non-deterministic) protocol.

We show a stronger $\Omega(\log^{1/2}(N))$ lower bound for their construction. To achieve this, the key technical advancement is an improvement to the sifting argument for grid norms of (somewhat dense) bipartite graphs. In addition to quantitative improvement, we qualitatively improve over [KLM24] by relaxing the hardness condition: while [KLM24] proved their lower bound for any function f that satisfies a strong two-sided pseudorandom condition, we show that a weak one-sided condition suffices. This is achieved by a new structural result for cylinder intersections (or, in graph-theoretic language, the set of triangles induced from a tripartite graph), showing that any small cylinder intersection can be efficiently covered by a sum of simple “slice” functions.

1 Introduction

In a recent work by Kelley and Meka [KM23], substantial quantitative progress was made on a well-known problem in additive combinatorics concerning the size of the largest set of integers $A \subseteq [N]$ which do not contain any three points x, y, z which are evenly spaced (i.e., which form a 3-AP). It was shown that such sets cannot have more than “quasipolynomial” density: $|A| \leq 2^{-\Omega(\log(N)^{1/12})} N$.¹ This comes reasonably close to the size of known constructions of large 3-AP free sets, which have size $|A| \geq 2^{-\Omega(\log(N)^{1/2})} N$.²

Since then, there have been several new applications in theoretical computer science and extremal combinatorics, which rely not on *result* of this work but on its *techniques*. Specifically, it has since been realized (first in a work of Kelley, Meka and Lovett [KLM24]) that the two main new ingredients from [KM23] (“sifting” [KM23, Section 4]) and “spectral positivity” [KM23, Section 5]) can be generalized to the graph theoretic setting, and are in fact best thought of as tools which are useful for the study of arbitrary bipartite graphs that have (at least) quasipolynomial edge-density. The main result of [KLM24] was an application of these new graph-theoretic tools to a problem in multiparty Number-On-Forehead (NOF) communication complexity. Subsequent

*Institute for Advanced Study (IAS). Supported by the Director’s Discretionary Fund. Email: awk@ias.edu.

†UC Berkeley. Supported by a Google Fellowship. Email: xinlyu@berkeley.edu

¹See [BS23b], [BS23a] for some subsequent improvements.

²[Beh46]. See also [Elk10], [GW10], [O’B11], [Hum24], [EHPS24] for some refinements.

additional applications (and new refinements) to these techniques were given in [AFK⁺24] and [FHHK24].

The specific applications in these works are on distant enough topics³ that it becomes somewhat difficult to give a succinct description of what these new techniques are *good for*, precisely. Nevertheless, we offer the following attempt at one. In any situation where one is faced with a graph-theoretic problem which can be solved with Szemerédi’s regularity lemma, on graphs with $\Omega(1)$ edge-density, then it is potentially reasonable to hope that this new machinery can lead to a solution for graphs with quasipolynomial edge-density. It is of course likely that for certain applications this hope is overly-optimistic, but we do not have a clear picture of what such cases should look like. The limitations of these new graph theoretic techniques are not yet well-understood.

This line of work (which develops graph-theoretic generalizations of the ideas introduced in [KM23]) is still relatively new, and it is unclear that any of the known results or underlying technical lemmas represent the “best version of themselves”, in a quantitative sense. In particular, each work mentioned so far proves some result about structures within a finite space of size N , which holds for all structures of quasipolynomial density: at least $2^{-\log(N)^c}$ for some constant $c > 0$. But, we are not aware of any application where it can be demonstrated that the result is *tight*, and we obtain the *right* quasipolynomial. In fact, it would be interesting to describe any application at all, along with an efficient-enough implementation of these new graph-theoretic techniques, where it can be shown that we obtain a tight answer, with the right quasipolynomial bound.

A core technical component needed in each of the works mentioned is an efficient “inverse theorem” for the grid norm (see (3) for its definition) of a bipartite graph: one needs to show (often by an argument involving sifting/dependent random choice) that for any dense bipartite graph with unusually large grid norm, it is possible to zoom in on some (reasonably sized) induced subgraph which has noticeably increased edge-density. Our main technical contribution is to give an improvement regarding this step.⁴ The specific result is given by Theorem 3.2.⁵ As an application, we give a quantitative improvement to the main result of [KLM24], which was the construction of an explicit function $\mathbb{1}_D : [N]^3 \rightarrow \{0, 1\}$ which is hard for deterministic 3-NOF protocols but easy for randomized 3-NOF protocols. In particular, our improved analysis shows that, in order to express the hard set D as a union of some R cylinder intersections, $R \geq 2^{\Omega(\log(N)^{1/2})}$ is required. This improves the prior result, $R \geq 2^{\Omega(\log(N)^{1/3})}$, from [KLM24].

Cylinder Intersections. The basic combinatorial object of interest in the study of multiparty NOF communication complexity is the *cylinder intersection*. In the 3-party setting, a cylinder intersection S is a subset of a product space $X \times Y \times Z$ whose indicator function has the form

$$\mathbb{1}_S(x, y, z) = f(x, y)g(x, z)h(y, z).$$

Here, f, g , and h are indicator functions, each of which depends on only 2 out of 3 coordinates (so, S is the intersection of three “combinatorial cylinders”). For the purpose of obtaining complexity lower bounds for NOF communication protocols, it is important to understand the structural limitations of cylinder intersections (or at least, some particular limitation, which we can try to target).

³For example, [KLM24] is about lower bounds for communication complexity, while [AFK⁺24] is about upper bounds for graph algorithms.

⁴More specifically, our approach gives a more efficient inverse theorem for the $U(2, k)$ grid norm, which is the case most important for applications thus far. Our approach can be made to work also for general $U(\ell, k)$ grid norms, but for large ℓ (say, $\ell \approx k$) we do not seem to get any improvement over existing arguments.

⁵See also Eq. (15) for a related statement which is often more directly useful in applications.

Our main result is the following structural statement about small cylinder intersections (over $X \times Y \times Z$), which states that they can be somewhat efficiently covered by a small collection of simpler functions. To state it, we first describe this family of simple functions.

Definition 1.1 (Slice Function). *A slice function $s(x, y, z)$ (on $\Omega := X \times Y \times Z$) is an indicator function of the form*

$$s(x, y, z) = g(x, y)h(z) \quad \text{or} \quad g(x)h(y, z) \quad \text{or} \quad g(x, z)h(y).$$

That is, a slice function is just a combinatorial rectangle on $(X \times Y) \times Z$ or $X \times (Y \times Z)$ or $(X \times Z) \times Y$.

Theorem 1.2 (Slice Function “Removal Lemma”). *Let $f(x, y), g(x, z), h(y, z)$ be some indicator functions on $\Omega := X \times Y \times Z$, and consider the cylinder intersection*

$$F(x, y, z) := f(x, y)g(x, z)h(y, z).$$

Suppose that F is small, i.e.

$$\mathbb{E}_{(x, y, z) \in \Omega} [F(x, y, z)] \leq 2^{-d}.$$

Then, F has a reasonably small (pointwise) cover F' ,

$$F(x, y, z) \leq F'(x, y, z) := \sum_{i=1}^R s_i(x, y, z),$$

for some slice functions $s_i(x, y, z)$. Specifically,

- $\mathbb{E}_{(x, y, z) \in \Omega} [F'(x, y, z)] \leq 2^{-\Omega(d^{1/2})} \log |\Omega|$, and
- $\mathbb{E}[s_i] \geq 2^{-O(d)}$ for every $i \in [R]$ (consequently, $R \leq 2^{O(d)} \cdot \log(|\Omega|)$).⁶

Explicit Separations for Deterministic and Randomized NOF Complexity. As a corollary, we obtain the following improvement to the main result from [KLM24], which was the construction of an explicit function which is hard for deterministic 3-NOF protocols but easy for randomized 3-NOF protocols.

Theorem 1.3. *Let q be a prime power and k a large enough constant. Let $N = q^k$, and identify $[N] \cong \mathbb{F}_q^k$. Consider the following 3-player function over $[N] \times [N] \times [N]$.*

$$D(x, y, z) = \mathbb{1}[\langle x, y \rangle = \langle x, z \rangle = \langle y, z \rangle],$$

Then:

1. *The randomized NOF communication complexity of D (with public randomness) is $O(1)$.*
2. *The deterministic (or non-deterministic) NOF communication complexity of D is $\Omega((\log N)^{1/2})$.*

⁶It is possible to obtain a version of this statement with no dependence on the size of the ambient space Ω – that is, if one is willing to settle for a *fractional* cover by slice functions. See Theorem 4.1.

A pseudorandomness perspective. Quantitative improvements aside, we argue that the statement of Theorem 1.2 is interesting for some additional qualitative reasons. Firstly, it succinctly captures the exact weakness of cylinder intersections which was targeted in [KLM24] to obtain a communication lower bound.⁷ Additionally, it is interesting to compare the statement with some structurally related results in extremal combinatorics and unconditional pseudorandomness.

For example, consider the collection of indicator functions F defined on the space $\Omega = \{0,1\}^n$ which have the form

$$F(x_1, x_2, \dots, x_n) = \prod_{S \in \binom{[n]}{w}} f_S(x_S).$$

It was shown by Bazzi⁸ that these functions (which are equivalently just the collection of width- w CNFs) have upper and lower pointwise approximators $F_\ell(x) \leq F(x) \leq F_u(x)$, which give a good approximation on average (for uniformly random points $x \in \{0,1\}^n$) in the sense that $\mathbb{E}_{x \in \{0,1\}^n} [F_u(x) - F_\ell(x)] \leq \varepsilon$. Furthermore, $F_\ell(x)$ and $F_u(x)$ are *simple functions* – specifically, polynomials of degree at most $k = \text{poly}(w, \log(n/\varepsilon))$. He uses this structural result to prove that k -wise independent distributions are pseudorandom for width- w CNFs, which follows because k -wise independent distributions are pseudorandom for degree k polynomials. Our structural result can be applied in much the same way: it implies that any set D which is evasive⁹ for small slice functions must also be reasonably evasive for small cylinder intersections – this fact is all that is needed for the proof of Theorem 1.3.

It was noted in [DETT10] that the existence of good degree- k sandwiching approximators $F_\ell(x) \leq F(x) \leq F_u(x)$, for some function F is, in fact, *equivalent* to the fact that every k -wise independent distribution is pseudorandom for F (with two-sided, additive error). The argument is based on linear programming duality, and is in fact quite generic: it extends easily to other settings involving a notion of pseudorandomness and a corresponding family of simple test functions.

An interesting consequence of this is that, if we have some proof of a pseudorandomness statement which is “abstract enough” (for example, it works for *all* k -wise independent distributions, rather than just a particular construction), then we can actually extract a *structural result* by linear programming duality. This is roughly how we obtain our main result, Theorem 1.2. First, we argue that an arbitrary distribution which is evasive for slice functions must be reasonably evasive for cylinder intersections. Then, by a different linear programming duality argument (which is needed for the one-sided and multiplicative quantification of error relevant here), we extract a good fractional cover, and then round it to obtain a cover.

Triangle Removal Lemma. Consider the following (unusual) formulation of Szemerédi’s triangle removal lemma. For the purpose of comparing with Theorem 1.2, we formulate it as a statement *about cylinder intersections*.

Theorem 1.4 (Triangle Removal Lemma). *Let $f(x, y), g(x, z), h(y, z)$ be some indicator functions on $\Omega := X \times Y \times Z$, and consider the cylinder intersection*

$$F(x, y, z) := f(x, y)g(x, z)h(y, z).$$

⁷Indeed, it can be checked that the key pseudorandomness statement, Theorem 2.4 from [KLM24], follows directly from our slice function removal lemma.

⁸[Baz09]. See also [Raz09], [DETT10], [Tal17].

⁹By “evasive”, we just mean a set D with a one-sided pseudorandomness property: no simple test function which captures a small fraction of Ω should be able to capture a large fraction of D .

Suppose that F is small, i.e.

$$\mathbb{E}_{(x,y,z) \in \Omega} [F(x,y,z)] \leq \varepsilon.$$

Then F has a reasonably small (pointwise) cover F' , which has the form

$$F(x,y,z) \leq F'(x,y,z) := f'(x,y) + g'(x,z) + h'(y,z).$$

Specifically,

$$\mathbb{E}_{(x,y,z) \in \Omega} [F'(x,y,z)] \leq \delta := \delta(\varepsilon)$$

for some $\delta(\varepsilon) \rightarrow 0$ as $\varepsilon \rightarrow 0$.

Recall that the standard formulation of the triangle removal lemma is stated in the language of graphs: it says that if an N -vertex graph has a small number of triangles (i.e., at most εN^3), then it is possible to remove a reasonably small number of edges (i.e., at most δN^2) in order to eliminate all triangles. One can observe that the two formulations differ only cosmetically.¹⁰

It is not known what should be the “right” quantitative dependence of δ on ε in the triangle removal lemma. Indeed, there is a large gap between the known upper bounds on $\delta(\varepsilon)$ (which are tower-type, see e.g. [Fox11]), and the best known lower bounds on plausible values for $\delta(\varepsilon)$, which follow from Behrend-type constructions of large sets of integers $A \subseteq [N]$ without 3-term arithmetic progressions. In light of these constructions, the best plausible dependence of δ on ε would be (roughly)

$$\delta(\varepsilon) \approx 2^{-\sqrt{\log(1/\varepsilon)}}.$$

Coincidentally, this corresponds to the same parameters we obtain in our slice function removal lemma. So, to conclude our comparison, we note the following. At present, the statements (Theorem 1.2 and Theorem 1.4) are incomparable: the former obtains substantially better parameters, but the latter provides a cover by a (strictly) simpler collection of functions – *cylinders* instead of *slice functions*. However, in the event that dramatically improved bounds for the triangle removal lemma were to be discovered, one could conceivably obtain a result strictly improving both statements.¹¹

An Open Question for k -party NOF. It is interesting to understand to what extent the techniques introduced in [KLM24] can be extended to obtain explicit separations for k -party deterministic and randomized NOF protocols, for $k \geq 4$. Concretely, we ask: can our covering result Theorem 1.2, can be extended to apply to cylinder intersections $F(x_1, x_2, \dots, x_k)$ defined over larger product spaces $X_1 \times X_2 \times \dots \times X_k$? Can we also obtain a reasonably small cover F' of F by some slice functions? What about covering by “generalized rectangles” – functions of the form $g(x_S)h(x_T)$ for some partition $[k] = S \cup T$? It seems that such questions are related to the extent to which the new graph-theoretic machinery (mentioned in the introduction) can be extended to *hypergraphs*. If so, it seems likely this would be useful for applications beyond multiparty communication complexity, as well.

¹⁰Perhaps the only non-obvious step is to notice that the triangle removal lemma is equivalent (up to constant factors) to the special case of itself on tripartite graphs.

¹¹We point out that, as far as we are aware, it could be possible to give a strict *quantitative* improvement to just our slice function removal lemma itself. That is, we do not know of any sort of obstruction (Behrend-type or otherwise) to obtaining substantially improved parameters when we cover by slice functions. It would be interesting to find such an obstruction if there is one.

2 Preliminaries

We usually work with a finite universe Ω . We write $w \sim \Omega$ to denote that w is drawn from Ω uniformly at random. Alternatively, we also use $\mathcal{U}$ to denote the uniform distribution. A function $p : \Omega \rightarrow \mathbb{R}_{\geq 0}$ is called a probability density function if $\mathbb{E}_{w \sim \Omega}[p(w)] = \sum_w \frac{p(w)}{|\Omega|} = 1$. To sample from p , we draw each $w \in \Omega$ with probability $\frac{p(w)}{|\Omega|}$. For any set $S \subseteq \Omega$, $w \sim S$ denotes the uniform distribution over S . Finally, we will frequently use the following change-of-distribution identity:

$$\mathbb{E}_{w \sim p}[S(w)] = \frac{|S|}{|\Omega|} \cdot \mathbb{E}_{w \sim S}[p(w)]. \quad (1)$$

Let X be a finite domain. The ℓ_p -norm of a vector $u \in \mathbb{R}^X$ is defined as $\|u\|_p := \mathbb{E}_{x \sim X}[|u_x|^p]^{1/p}$. For two vectors $u, v \in \mathbb{R}^X$ of the same dimension, their inner product is $\langle u, v \rangle := \mathbb{E}_{x \sim X}[u_x \cdot v_x]$. For every $p \geq 1$, let $p^* := \frac{p}{p-1}$ be the Hölder conjugate of p . The Hölder inequality says that $\langle f, g \rangle \leq \|f\|_p \cdot \|g\|_{p^*}$. We collect some elementary equations regarding Hölder conjugates that we frequently use without notice:

$$\frac{1}{p-1} = p^* - 1 \quad , \quad p^* = \frac{p}{p-1} \quad , \quad \frac{1}{p} - 1 = -\frac{1}{p^*}. \quad (2)$$

For a matrix $M \in \mathbb{R}^{X \times Y}$ and vectors $u \in \mathbb{R}^X, v \in \mathbb{R}^Y$, we define the *normalized* matrix-vector product $Mg \in \mathbb{R}^X$ as $(Mg)(x) = \mathbb{E}_y[M(x, y) \cdot g(y)]$ and vector-matrix-vector product $f^\top Mg = \mathbb{E}_{x, y}[f(x)M(x, y)g(y)]$. Finally, given $\ell, k \geq 1$, the $U(\ell, k)$ -grid norm of M is defined as

$$\|M\|_{U(\ell, k)} := \left(\mathbb{E}_{x_{1:\ell}} \mathbb{E}_{y_{1:k}} \left[\prod_{i \in [\ell]} \prod_{j \in [k]} M(x_i, y_j) \right] \right)^{1/\ell k}. \quad (3)$$

2.1 Norms and Flat Norms of Vectors

For a vector $v \in \mathbb{R}_{\geq 0}^X$, we may define its ℓ_p -norm by

$$\|v\|_p := \mathbb{E}_{x \sim X}[(v_x)^p]^{1/p} = \max_{u \in \mathbb{R}^X} \left[\frac{\langle v, u \rangle}{\|u\|_{p^*}} \right]. \quad (4)$$

We consider a related notion of the *flat* p -norm¹², defined as

$$\|v\|_{\bar{p}} := \max_{S \subseteq X} \left\{ \|v|_S\|_1 \left(\frac{|S|}{|X|} \right)^{1/p} \right\}. \quad (5)$$

Thus, the flat p -norm is just the result one obtains by writing down the dual characterization of the p -norm, but allowing only *flat* witnesses: vectors u of the form $u = \mathbb{1}_S$ for some set S . Here, $v|_S \in \mathbb{R}^S$ denotes the restriction of v onto the subset S of coordinates (and $\|v|_S\|_1$ denotes the 1-norm normalized relative to S , i.e., $\|v|_S\|_1 = \mathbb{E}_{x \sim S}[v_x]$).

The following proposition states the flat p -norm and the usual ℓ_p -norm are closely related for all *smooth* vectors.

¹²This analytic quantity is closely related to the *weak* p -norm, and is well known, but seems to not go by any common name. See e.g. Exercise 1.1.12 in [G+08].

Proposition 2.1. *Let $k \geq 1$ be a parameter. Suppose $v \in \mathbb{R}_{\geq 0}^X$ is such that $\frac{\|v\|_\infty}{\|v\|_1} \leq 2^k$. Then, for every $\varepsilon \in (0, 1/10)$ and $p \geq \frac{10 \log(k)}{\varepsilon}$, it holds that*

$$\|v\|_p \geq \|v\|_{\bar{p}} \geq (1 - \varepsilon) \|v\|_p. \quad (6)$$

In fact, the first inequality holds for any $p \geq 1$ without any restriction on p .

Proof. To see the first inequality, simply note that for any $S \subseteq X$, we may consider the indicator vector $f \in \mathbb{R}^X$ defined by $f_x = \mathbf{1}_{\{x \in S\}}$. We have

$$\frac{\langle v, f \rangle}{\|f\|_{p^*}} = \frac{|S|}{|X|} \cdot \|v|_S\|_1 \cdot \left(\frac{|S|}{|X|} \right)^{-\frac{1}{p^*}} = \|v|_S\|_1 \left(\frac{|S|}{|X|} \right)^{\frac{1}{p}}.$$

This means any candidate maximizer $S \subseteq X$ for $\|v\|_{\bar{p}}$ gives a candidate maximizer for $\|v\|_p$. Optimizing over all possible S proves the inequality.

We turn to establish the second inequality. By properly scaling, we can assume W.L.O.G. $\|v\|_\infty = 1$. With this in mind, we consider a geometric decomposition

$$v = \sum_{j=0}^{-\infty} v^j$$

where v^j is the vector that contains entries of v with value $(2^{\varepsilon(-j-1)}, 2^{\varepsilon j}]$. Namely, $(v^j)_x = v_x$ if $v_x \in (2^{\varepsilon(-j-1)}, 2^{\varepsilon j}]$ and $v_x^j = 0$ otherwise. Since v^j 's have disjoint support, we have

$$\|v\|_p^p = \sum_{j=0}^{\infty} \|v^j\|_p^p.$$

For a suitable $d \geq \frac{\log(4 \frac{\|v\|_\infty}{\|v\|_p})}{\varepsilon} \approx \frac{k}{\varepsilon}$, we have

$$\left\| \sum_{j \geq d} v^j \right\|_p \leq \left\| \sum_{j \geq d} v^j \right\|_\infty \leq 2^{-\varepsilon d} \leq \frac{\|v\|_p}{4}.$$

Consequently,

$$\sum_{j=0}^{d-1} \|v^j\|_p^p = \|v\|_p^p - \left\| \sum_{j \geq d} v^j \right\|_p^p \geq \frac{1}{2} \|v\|_p^p.$$

By averaging principle, there exists $j^* < d$ such that $\|v^{j^*}\|_p^p \geq \frac{1}{2d} \|v\|_p^p$. Letting S be the support of v^{j^*} , we obtain

$$\begin{aligned} \|v|_S\|_1 \cdot \left(\frac{|S|}{|X|} \right)^{\frac{1}{p}} &\geq 2^{(-j^*-1)\varepsilon} \left(\frac{|S|}{|X|} \right)^{\frac{1}{p}} \\ &\geq 2^\varepsilon \|v^{j^*}\|_p \geq 2^\varepsilon \cdot \left(\frac{1}{2d} \right)^{1/p} \cdot \|v\|_p \\ &\geq (1 - O(\varepsilon)) \|v\|_p. \end{aligned} \quad (7)$$

To replace $(1 - O(\varepsilon))$ in (7) with a $(1 - \varepsilon)$, we can work with a smaller $\varepsilon' = \Theta(\varepsilon)$ in this argument. the proof will go through so long as we have, say, $p \geq \frac{2d}{\varepsilon'}$. $\square$

2.2 Norms and Flat Norms of Matrices

We can similarly develop a norm-vs-flat norm connection for matrices and operator norms. In particular, consider a matrix $M \in \mathbb{R}^{X \times Y}$ whose rows and columns are indexed by X and Y , respectively. We let $\|M\|_1 = \mathbb{E}_{x \sim X, y \sim Y} [M(x, y)]$ be the 1-norm of M .

Given $\ell, r \geq 1$ and a *non-negative* matrix M , define the *flat* (ℓ, r) -norm of M as

$$\|M\|_{\bar{\ell}, \bar{r}} = \max_{S \subseteq X, T \subseteq Y} \left\{ \|M|_{S \times T}\|_1 \cdot \left(\frac{|S|}{|X|} \right)^{\frac{1}{\bar{\ell}}} \left(\frac{|T|}{|Y|} \right)^{\frac{1}{\bar{r}}} \right\}. \quad (8)$$

We can then compare the flat norms with the usual operator norm. In particular, we define the (ℓ, r) -norm of a non-negative matrix M as

$$\|M\|_{\ell, r} = \max_{\substack{f: X \rightarrow \mathbb{R}_{\geq 0}, \\ g: Y \rightarrow \mathbb{R}_{\geq 0}}} \left\{ \frac{f^\top M g}{\|f\|_{\ell^*} \|g\|_{r^*}} \right\} := \max_{\|g\|_{r^*} \leq 1} \|Mg\|_{\ell}. \quad (9)$$

From an operator norm perspective, one might find it easier to understand $\|M\|_{\ell, r}$ as $\|M\|_{\ell \leftarrow r^*}$ (i.e., how M maps a r^* -norm bounded vector to a vector with bounded ℓ -norm).

The following lemma connects flat norms with operator norms for smooth matrices.

Lemma 2.2. *Suppose $M \in \mathbb{R}_{\geq 0}^{X \times Y}$ is such that $\|M\|_{\infty} \leq \|M\|_1 \cdot 2^d$. Then, for any $\varepsilon \in (0, 1)$, provided that $\min(\ell, r) \geq \frac{20 \log d}{\varepsilon}$, we have*

$$\|M\|_{\ell, r} \geq \|M\|_{\bar{\ell}, \bar{r}} \geq (1 - \varepsilon) \cdot \|M\|_{\ell, r}. \quad (10)$$

Proof. To see the first inequality, simply note that for any $S \subseteq X, T \subseteq Y$, we may consider the indicator functions $f := \mathbf{1}_{x \in S}$ and $g := \mathbf{1}_{y \in T}$, and derive that

$$\begin{aligned} & \frac{f^\top M g}{\|f\|_{\ell^*} \|g\|_{r^*}} \\ &= \|M|_{S \times T}\|_1 \cdot \left(\frac{|S|}{|X|} \right)^{1 - \frac{1}{\ell^*}} \left(\frac{|T|}{|Y|} \right)^{1 - \frac{1}{r^*}} \\ &= \|M|_{S \times T}\|_1 \cdot \left(\frac{|S|}{|X|} \right)^{\frac{1}{\bar{\ell}}} \left(\frac{|T|}{|Y|} \right)^{\frac{1}{\bar{r}}} \end{aligned}$$

This means any candidate subsets S, T for $\|M\|_{\bar{\ell}, \bar{r}}$ give a candidate maximizer for $\|M\|_{\ell, r}$ as well. Optimizing over all possible S, T proves the first inequality.

We now prove the second inequality. Assume without loss of generality that $\|M\|_1 = 1$. So we have $\|M\|_{\infty} \leq 2^d$ and consequently $\|M\|_{\infty \leftarrow p} \leq 2^d$ for any $p \geq 1$. Say $\tau := \|M\|_{\ell, r}$ is achieved by witnesses $f: X \rightarrow \mathbb{R}_{\geq 0}$ and $g: Y \rightarrow \mathbb{R}_{\geq 0}$ with $\|f\|_{\ell^*} = \|g\|_{r^*} = 1$. Consider the following two optimization problems

$$\max_{h: X \rightarrow \mathbb{R}_{\geq 0}} \left\{ \frac{h^\top M g}{\|h\|_{\ell^*}} \right\} \quad \text{vs.} \quad \max_{h: X \rightarrow \{0, 1\}} \left\{ \frac{h^\top M g}{\|h\|_{\ell^*}} \right\}$$

The optimum of the left problem is realized by taking $h = f$ and, by Proposition 2.1, there exists $\bar{f} : X \rightarrow \{0, 1\}$ such that

$$\frac{\bar{f}^\top M g}{\|\bar{f}\|_{\ell^\star}} \geq (1 - \varepsilon) \frac{f^\top M g}{\|f\|_{\ell^\star}},$$

provided that $\ell \geq \frac{10 \log d}{\varepsilon}$. Next, one can similarly find $\bar{g} : Y \rightarrow \{0, 1\}$ (provided that $r \geq \frac{10 \log d}{\varepsilon}$) such that

$$\frac{\bar{f}^\top M \bar{g}}{\|\bar{f}\|_{\ell^\star} \|\bar{g}\|_{r^\star}} \geq (1 - \varepsilon) \frac{\bar{f}^\top M g}{\|\bar{f}\|_{\ell^\star} \|g\|_{r^\star}} \geq \frac{(1 - O(\varepsilon)) \cdot \tau}{\|\bar{g}\|_{r^\star} \|\bar{f}\|_{\ell^\star}}.$$

The sets indicated by $\bar{f}, \bar{g}$ would then witness that $\|M\|_{\bar{\ell}, \bar{r}} \geq (1 - O(\varepsilon)) \|M\|_{\ell, r}$, as desired. $\square$

3 Spread Matrices and Their Products

This section refines, simplifies, and improves the technique from [KLM24, Section 4].

3.1 Proof Overview

Before developing the results, let us comment on how we obtain a better sifting argument than prior works [KM23, KLM24].

Review of [KLM24]. At the core of all the “sifting” arguments is a “density increment” argument. To review the [KLM24] proof: we start with a non-negative matrix $M \in \mathbb{R}^{N \times N}$ that is (1) bounded: $\|M\|_\infty = 1$ and (2) relatively dense: $\|M\|_1 = \mathbb{E}_{(x,y)}[M(x,y)] \geq 2^{-d}$. We inspect how “pseudorandom” M is, as measured by its $U(2, k)$ -grid norm $\|M\|_{U(2,k)}$. There are two cases: If M is “pseudorandom” enough in the sense that $\|M\|_{U(2,k)} \leq (1 + \varepsilon) \|M\|_1$, we can proceed to argue other nice properties of M . Otherwise, M is not pseudorandom and we have $\|M\|_{U(2,k)} \geq (1 + \varepsilon) \|M\|_1$. In this case, one tries to find a sub-matrix $X' \times Y' \subseteq [N] \times [N]$, such that the average of M increases by a multiplicative factor of $(1 + \varepsilon)$ where ε is a small constant, and the sub-matrix is of size at least $\frac{1}{2^{O(kd)}}$. On the sub-matrix, one can iterate this argument. Since each iteration increases the density of M by at least $(1 + \varepsilon)$. The density increment step only happens for at most d/ε times, and we are guaranteed a final submatrix of density at least $2^{-O(d^2 k)}$. In the most relevant parameter regime, one first fixes $d = \log(\frac{1}{\text{density of the object}})$ and chooses $k \approx d$. This leads to a $2^{O(d^3)}$ loss in the size of the final sub-matrix, and is ultimately the reason that [KLM24] ended up with a $\Omega((\log N)^{1/3})$ lower bound for the 3-party NOF communication model.

Our improvement. We show that, with a new and simpler way to sift, one gets a trade-off between how much larger the mean of M increases, and how much smaller the sub-matrix becomes. In particular, we show the following: in each iteration of the density increment argument, if the average of M only increases by a single $(1 + \varepsilon)$ factor, then the sub-matrix $X' \times Y'$ only gets smaller by a factor of $2^{-O(k)}$. On the other hand, if we can only increase onto a sub-matrix $X' \times Y'$ of size much smaller than $2^{-\Omega(k)}$, then the average of M will increase by a significantly larger amount. Roughly speaking, there always exists a parameter $1 \leq \gamma \leq d$ and a sub-matrix $X' \times Y'$ of size $2^{-\gamma \cdot k}$ such that $\mathbb{E}_{(x,y) \sim X' \times Y'}[M(x,y)]$ increases by a factor of $(1 + \varepsilon)^{(d-\gamma)k}$. Iterating based on this claim allows us to find a sub-matrix M with the desired pseudorandom property, yet the size of the final sub-matrix is at least $2^{-O(kd)} N^2$.

One might find it easier to understand the improvement by drawing an analogy to vector ℓ_k -norms. Say we have a vector $v \in [0, 1]^N$ that we know $\|v\|_1 \geq 2^{-d}$ and $\|v\|_k := \mathbb{E}[v_i^k]^{1/k} > (1 + \varepsilon) \mathbb{E}[v_i]$. This can happen for (at least) two possible reasons: (1) there is a subset $S \subseteq [N]$ such that $|S| \geq 2^{-k}N$ and $v_i > (1 + \frac{1}{2})\|v\|_1$ for $i \in S$, or (2) there is a much smaller subset $S \subseteq [N]$ of size only $|S| > 2^{-dk}N$, but $v_i > 2^d \cdot \|v\|_1$ for $i \in S$. Both cases lead to a significantly larger ℓ_k -norm of v . Still, to run an efficient “density increment” argument on v , one must account for the trade-off between $|S|$ and $\min_{i \in S} \{v_i\}$.

Currently, our improvement only applies to the $U(2, k)$ -grid norms¹³, mostly due to the nice interpretation of $U(2, k)$ norm as “the k -th moment of the pairwise inner products between rows of M ”, as well as the flat-vs-operator norm connection established in Section 2. It remains to be seen whether a similar improvement can be made for the case of 3-APs [KM23].

Organization of the section. With the setup from Section 2, the density increment argument is compactly presented in Section 3.2. We also refine the “spectral positivity” part of the [KLM24] argument and present a simplified proof in Section 3.3. Finally, Theorem 3.5 combines the two ingredients and concludes this section.

3.2 Spreadness of Matrices

We define a notion of spreadness for matrices. Roughly, it says that a spread matrix has no large sub-matrix with significantly larger entries on average.

Definition 3.1. A matrix $M \in \mathbb{R}_{\geq 0}^{X \times Y}$ is (t, ε) -spread, if $\|M\|_{\bar{t}, \bar{t}} \leq \|M\|_1 \cdot (1 + \varepsilon)$.

We compare Definition 3.1 with the definition of spreadness from [KLM24]. The spreadness of [KLM24] is parameterized by two parameters $k\mathbb{N}$ and $\varepsilon > 0$. A matrix M is spread in the sense of [KLM24], if for any submatrix $M' \in \mathbb{R}^{X' \times Y'}$ of size $|X' \times Y'| \geq 2^{-k}|X \times Y|$, one has that $\|M'\|_1 \leq \|M\|_1 \cdot (1 + \varepsilon)$. Namely, the density of every “reasonably large” subrectangle can only be larger than the density of M be at most a $(1 + \varepsilon)$ factor. Our Definition 3.1, on the other hand, should be (roughly) read as for every submatrix M' of size $(1 - \varepsilon)^{k \cdot \tau}$ for some $\tau \geq 1$, one has $\|M'\|_1 \leq \|M\|_1 \cdot (1 + \varepsilon)^\tau$. So, our definition is stronger in that we will consider submatrices of size significantly smaller than 2^{-k} , and we impose proportionally more relaxed density upper bounds for smaller submatrices.

The key technical lemma of [KLM24] asserts that a spread matrix has bounded $U(2, k)$ -norm. We state and prove our version of the lemma in the following.

Lemma 3.2. Let $k, d \geq 1, \varepsilon \in (0, 1/5)$ be parameters such that $k \geq 20d/\varepsilon$. Suppose $M \in [0, 1]^{X \times Y}$ is a matrix that is (k, ε) -spread and $\|M\|_1 \geq 2^{-d}$. Then, $\|M\|_{U(2, k)} \leq (1 + O(\varepsilon))\|M\|_1$.

Proof. The condition of M implies that $\frac{\|M\|_\infty}{\|M\|_1} \leq 2^d$. As such, Definition 3.1 and Lemma 2.2 imply that $\|M\|_{k, k} \leq (1 + O(\varepsilon))\|M\|_{\bar{k}, \bar{k}} \leq (1 + O(\varepsilon))\|M\|_1$. For each $x \in X$, let M_x be the x -th row

¹³We can extend this to $U(\ell, k)$ -norms by a tensoring trick. Still, the improvement is only significant in the regime that $\ell \ll k$.

vector of M . For a proper choice of $p \approx \frac{10d}{\varepsilon}$, we have

$$\begin{aligned}
& \|M\|_{U(2,k)}^{2k} \\
&= \mathbb{E}_{x, x' \sim X} [\langle Mx, Mx' \rangle^k] \\
&= \mathbb{E}_{x \sim X} [\|M \cdot M_x^\top\|_k^k] \\
&\leq \mathbb{E}_{x \sim X} [(\|M\|_{k,p} \cdot \|M_x\|_{p^*})^k] \\
&\leq \|M\|_{k,k}^k \cdot \mathbb{E}_x \left[\left(\|M_x\|_1^{1/p^*} \|M_x\|_\infty^{\frac{p^*-1}{p^*}} \right)^k \right] \\
&\leq \|M\|_{k,k}^k \cdot \|M\|_{\frac{k}{p^*}, 1}^{\frac{k}{p^*}} \cdot 2^{\frac{dk}{p}} \\
&\leq \|M\|_{k,k}^{2k} \|M\|_{k,k}^{\frac{k}{p^*} - k} \cdot 2^{\frac{dk}{p}}.
\end{aligned} \tag{11}$$

Here, the fourth-line is using the definition of operator norm, and the last line is due to $\|M\|_{k,k} \geq \|M\|_{\frac{k}{p^*}, 1}$. Raising both sides to the power of $\frac{1}{2k}$ and using the fact $\|M\|_{k,k} \geq \|M\|_1 \geq 2^{-d}$, we obtain

$$\begin{aligned}
\|M\|_{U(2,k)} &\leq \|M\|_{k,k} \cdot \|M\|_{k,k}^{-\frac{1}{p}} \cdot 2^{\frac{d}{2p}} \\
&\leq \|M\|_{k,k} \cdot 2^{2d/p} \\
&\leq (1 + O(\varepsilon)) \|M\|_1.
\end{aligned} \tag{12}$$

This completes the proof. $\square$

3.3 Product of Spread Matrices is Near-Uniform

For a matrix $f : X \times Y \rightarrow \mathbb{R}_{\geq 0}$, define $R_f : X \times Y \rightarrow \mathbb{R}_{\geq 0}$ by

$$(R_f)(x, y) = \mathbb{E}_{y' \sim Y} [f(x, y')].$$

Namely, R_f is the row-averaging of f . We make the following claim.

Lemma 3.3. *Let $f : X \times Y \rightarrow \mathbb{R}_{\geq 0}$ be a matrix with $\|f - R_f\|_{U(2,k)} \geq 6\sqrt{\varepsilon} \cdot \mathbb{E}[f]$. Suppose for every $x \in X$, it holds $\mathbb{E}_y[f(x, y)] \geq (1 - \varepsilon)\mathbb{E}[f]$. Then, for $p \geq \frac{2k}{\varepsilon}$, we have $\|f\|_{U(2,p)} \geq (1 + \varepsilon)\|f\|_1$.*

We mostly use Lemma 3.3 in its contrapositive form: Suppose we can prove a matrix f is such that $\|f\|_{U(2,p)} \leq (1 + \varepsilon)\|f\|_1$. Then, provided with the additional assumption $\mathbb{E}_y[f(x, y)] \geq (1 - \varepsilon)\mathbb{E}[f]$, we may conclude $\|f - R_f\|_{U(2,k)} \leq O(\sqrt{\varepsilon})\|f\|_1$.

To prove Lemma 3.3, we will need the following lemma from [KM23].

Lemma 3.4 (See Proposition D.1 of [KM23]). *Let $\varepsilon \in (0, 1/4)$. Suppose A is a random variable with $\mathbb{E}[A^k] \geq (2\varepsilon)^k$ for some even k and $\mathbb{E}[A^t] \geq 0$ for every $t \geq 1$. Then, for all $p \geq \frac{k}{\varepsilon}$, it holds that $\mathbb{E}[1 + A]^p \geq (1 + \varepsilon)^p$.*

Proof of Lemma 3.3. Consider a random variable A which takes value $\langle (f - R_f)(i, \star), (f - R_f)(j, \star) \rangle$ with probability $\frac{1}{|X|^2}$ for every (i, j) . We know that

$$\mathbb{E}[A^k] = \|f - R_f\|_{U(2,k)}^{2k} \geq (6\sqrt{\varepsilon} \mathbb{E}[f])^{2k},$$

and that for every $t \geq 1$,

$$\begin{aligned}\mathbb{E}[A^t] &= \mathbb{E}_{y_1, \dots, y_t} \mathbb{E}_{x_1, x_2} \left[\prod_{x_i, y_j} (f - R_f)(x_i, y_j) \right] \\ &= \mathbb{E}_{y_1, \dots, y_t} \left(\mathbb{E}_x \left[\prod_{y_j} (f - R_f)(x, y_j) \right] \right)^2 \\ &\geq 0.\end{aligned}$$

Let $\varepsilon' = 2\varepsilon - \varepsilon^2$. Using Lemma 3.4 on $\frac{A}{\mathbb{E}[f]^2(1-\varepsilon')}$, we obtain that $\mathbb{E}[(A + (1 - \varepsilon') \mathbb{E}[f]^2)^p] \geq ((1 + \varepsilon) \mathbb{E}[f]^2)^p$ for every $p \geq k/\varepsilon$. Observe that each row of $(f - R_f)$ sums to zero. Let $J = \mathbf{1}^{X \times Y}$ be the all-one matrix. We calculate

$$\begin{aligned}& \|f - R_f + (1 - \varepsilon) \mathbb{E}[f]J\|_{U(2,p)}^{2p} \\ &= \mathbb{E}_{i,j} \left\langle (f - R_f + (1 - \varepsilon) \mathbb{E}[f]J)(i, \star), \right. \\ &\quad \left. (f - R_f + (1 - \varepsilon) \mathbb{E}[f]J)(j, \star) \right\rangle^p \\ &= \mathbb{E}_A [((1 - \varepsilon') \mathbb{E}[f]^2 + A)^p] \\ &\geq ((1 + \varepsilon) \mathbb{E}[f]^2)^p.\end{aligned} \tag{13}$$

Recall that we have the entry-wise bound of $R_f(x, y) \geq 1 - \varepsilon$. Hence, for every i, j , we have

$$\begin{aligned}& \langle f(i, \star), f(j, \star) \rangle \\ &\geq \left\langle (f - R_f + (1 - \varepsilon) \mathbb{E}[f]J)(i, \star), \right. \\ &\quad \left. (f - R_f + (1 - \varepsilon) \mathbb{E}[f]J)(j, \star) \right\rangle.\end{aligned}$$

This is to say, the random variable $A + (1 - \varepsilon') \mathbb{E}[f]$ is *stochastically dominated* by $\langle f(\mathbf{i}, \star), f(\mathbf{j}, \star) \rangle$.

One may now be tempted to conclude that the p -th moment of $A + (1 - \varepsilon') \mathbb{E}[f]$ gives a lower bound for the p -th moment of $\langle f(\mathbf{i}, \star), f(\mathbf{j}, \star) \rangle$. However, this is generally not true as the random variables can be negative (for example, the random variable $\mathbf{0}$ stochastically dominates $-\mathbf{1}$, but the latter has a larger p -th moment). Nonetheless, since A satisfies that $\mathbb{E}[A^t] \geq 0$ for every $t \geq 1$, we know $\mathbb{E}[(A + (1 - \varepsilon') \mathbb{E}[f]^2)^p] \geq 0$ for every odd p . Therefore, as long as the p -th moment of $A + (1 - \varepsilon') \mathbb{E}[f]^2$ is concerned, the contribution from the “positive” instantiation of the r.v. is no less than the contribution from the negative part. By only considering the non-negative instantiation of $A + (1 - \varepsilon') \mathbb{E}[f]^2$ and using stochastic domination, we may consequently conclude

$$\begin{aligned}\|f\|_{U(2,p)}^{2p} &\geq \mathbb{E}_A [(A + (1 - \varepsilon') \mathbb{E}[f]^2)^p \cdot \mathbf{1}\{A \geq 1\}] \\ &\geq \frac{1}{2} \mathbb{E}_A [(A + (1 - \varepsilon') \mathbb{E}[f]^2)^p] \\ &\geq \frac{1}{2} ((1 + \varepsilon) \mathbb{E}[f]^2)^p \\ &\geq (1 + \varepsilon - O(1/p))^p \mathbb{E}[f]^{2p}.\end{aligned}$$

as desired. For even p , we can work with $p - 1$ and appeal to the monotonicity of norms. $\square$

Decoupling inequality. The rest of the argument is largely similar to [KLM24]. We nonetheless give a self-contained exposition here. First, let $f : X \times Y \rightarrow \mathbb{R}_{\geq 0}, g : Z \times Y \rightarrow \mathbb{R}_{\geq 0}$ be two non-negative matrices that are both spread and smooth. We consider their product $f \circ g : X \times Z \rightarrow \mathbb{R}_{\geq 0}$ defined as

$$(f \circ g)(x, z) = \mathbb{E}_{y \sim Y}[f(x, y) \cdot g(z, y)]. \quad (14)$$

Assuming the spreadness and smoothness of f, g , we claim that $f \circ g$ is close to the matrix $R_f \circ R_g$ (which we note is a rank-1 matrix).

To prove the claim, we use a moment method. For any even $k \geq 1$ of our choice, we can write

$$\begin{aligned} & \|f \circ g - R_f \circ R_g\|_k^k \\ &= \|(f - R_f) \circ (g - R_g)\|_k^k \\ &= \mathbb{E}_{x \sim X, z \sim Z} \langle (f - R_f)(x, *), (g - R_g)(z, *) \rangle^k \\ &= \mathbb{E}_{x, z} \mathbb{E}_{y_1, \dots, y_k} \left(\prod_{1 \leq j \leq k} (f - R_f)(x, y_j) (g - R_g)(z, y_j) \right) \\ &= \mathbb{E}_{y_1, \dots, y_k \sim Y} \left(\mathbb{E}_{x \sim X} \prod_{1 \leq j \leq k} (f - R_f)(x, y_j) \right) \times \\ & \quad \left(\mathbb{E}_{z \sim Z} \prod_{1 \leq j \leq k} (g - R_g)(z, y_j) \right) \\ &\leq \left(\mathbb{E}_{y_1, \dots, y_k \sim Y} \left(\mathbb{E}_{x \sim X} \prod_{1 \leq j \leq k} (f - R_f)(x, y_j) \right)^2 \right)^{1/2} \times \\ & \quad \left(\mathbb{E}_{y_1, \dots, y_k \sim Y} \left(\mathbb{E}_{z \sim Z} \prod_{1 \leq j \leq k} (g - R_g)(z, y_j) \right)^2 \right)^{1/2} \\ &\leq \|f - R_f\|_{U(2, k)}^k \|g - R_g\|_{U(2, k)}^k. \end{aligned} \quad (15)$$

Therefore, as long as we can establish that f, g satisfy the prerequisites of Lemma 3.3, we can then upper bound the deviation $\|f \circ g - R_f \circ R_g\|_k$ by $O(\varepsilon^2 \|f\|_1 \|g\|_1)$, which would then imply that all but a tiny fraction (roughly 2^{-k}) of pairs (x, z) satisfy that $|\langle f(x, *), g(z, *) \rangle - R_f(x) \cdot R_g(z)| \leq O(\varepsilon^2 \|f\|_1 \|g\|_1)$.

Overall, we compile everything we have discussed so far and formulate the following theorem.

Theorem 3.5. *Let $f \in [0, 1]^{X \times Y}$ and $g \in [0, 1]^{Z \times Y}$ be two matrices. Let $d \geq 10$ be an integer. Assume that*

- f, g are $(100d/\varepsilon, \varepsilon)$ -spread,
- $\mathbb{E}_y[f(x, y)] \geq 1 - \varepsilon, \mathbb{E}_y[g(z, y)] \geq 1 - \varepsilon$ for every x, z , and
- $\|f\|_1, \|g\|_1 \geq 2^{-d}$.

Then, it follows that

$$\|f \circ g - R_f \circ R_g\|_d \leq O(\varepsilon \|f\|_1 \cdot \|g\|_1).$$

Proof. From the spreadness and density lower bound of f we obtain that $\|f\|_{U(2,10d/\varepsilon)} \leq (1 + O(\varepsilon))\|f\|_1$ by Lemma 3.2. Using Lemma 3.3 in its contrapositive form we then get $\|f - R_f\|_{U(2,d)} \leq O(\sqrt{\varepsilon}\|f\|_1)$. Repeat the same argument on g . We obtain $\|g - R_g\|_{U(2,d)} \leq O(\sqrt{\varepsilon}\|g\|_1)$. Finally, by the decoupling inequality as shown above, we conclude that

$$\|f \circ g - R_f \circ R_g\|_d \leq \|f - R_f\|_{U(2,d)} \cdot \|g - R_g\|_{U(2,d)} \leq O(\varepsilon \|f\|_1 \|g\|_1),$$

as claimed. $\square$

4 Proof of the Removal Lemma

In this section, we prove Theorem 1.2 by proving the following “fractional cover” version.

Theorem 4.1. *Let $f(x, y), g(x, z), h(y, z)$ be some indicator functions on $\Omega = X \times Y \times Z$. Consider their cylinder intersection $F(x, y, z) = f(x, y) \cdot g(x, z) \cdot h(y, z)$. Let $d \geq 5$ and suppose F is such that*

$$\mathbb{E}_{(x,y,z) \sim \Omega}[F(x, y, z)] \leq 2^{-d}.$$

Then, F is covered by a sum of slice functions:

$$F(x, y, z) \leq \sum_{i=1}^R c_i \cdot s_i(x, y, z),$$

where each s_i is of size at least $2^{-O(d)}$ and

$$\mathbb{E}_{(x,y,z) \sim \Omega} \left[\sum_{i=1}^R c_i s_i(x, y, z) \right] \leq 2^{-\Omega(d^{1/2})}.$$

Rounding. We sketch how to round the fractional cover from Theorem 4.1 to prove Theorem 1.2. Say we have $F \leq \sum_{i=1}^R c_i s_i(x, y, z)$. Consider now a distribution over slice functions where we draw s_i with probability $\frac{c_i}{\sum_{i'} c_{i'}}$. Then, for a given $(x, y, z) \in F$, drawing a random s_i from the distribution can cover (x, y, z) with probability at least $\frac{1}{\sum_{i'} c_{i'}}$. Therefore, drawing $O(\log(|\Omega|) \cdot \sum_{i'} c_{i'})$ slices ensures that one of the slices covers (x, y, z) with probability $1 - \frac{1}{|\Omega|^2}$. So, to obtain an integral covering, we just independently draw $O(\log(|\Omega|) \cdot \sum_{i'} c_{i'})$ functions from the distribution. By a simple union bound, the resulting cover will cover every $(x, y, z) \in F$ with high probability. By Markov’s inequality, we can ensure that the total size of the integral cover is larger than the fractional cover by at most a factor of $O(\log(|\Omega|))$. We conclude by noting that $|\sum_{i'} c_{i'}| \leq 2^{O(d)}$ (this follows from the upper bound of $\mathbb{E}[\sum_{i=1}^R c_i s_i(x, y, z)]$ and the lower bound of $\mathbb{E}[s_i]$). Hence, the number of slices functions in the integral cover is at most $2^{O(d)} \cdot \log(|\Omega|)$.

Proof of Theorem 4.1. We prove the theorem by duality of linear programming. Let $\mathcal{S}$ be the collection of all large slice functions (i.e., slice functions of size $2^{-O(d)}$). We construct a $\Omega \times \mathcal{S}$ matrix $A \in \mathbb{R}^{\Omega \times \mathcal{S}}$. Here, rows and columns are naturally indexed by $(x, y, z) \in \Omega$ and $s \in \mathcal{S}$. The entry indexed by row (x, y, z) and column s is equal to $s(x, y, z)$. We also define the weight of a

slice $s \in \mathcal{S}$ to be $w(s) = \mathbb{E}[s]$. Slightly abusing notation, we understand w as both a $\mathcal{S}$ -dimensional column vector and a function mapping $\mathcal{S}$ to reals. Similarly, identify F with a Ω -dimensional column vector where the (x, y, z) -th entry of F is $F(x, y, z)$. With this setup, let us consider the following linear program:

$$\begin{aligned} \min_{c \in (\mathbb{R}_{\geq 0})^{\mathcal{S}}} \quad & \sum_s c_s \cdot w(s) \\ \text{s.t.} \quad & F \leq A \cdot c. \end{aligned}$$

Intuitively, this is the covering problem considered in Theorem 4.1: we want to choose a weight c_s for each slice function so that the weighted combination of slices covers every non-zero entry of F . Then, the conclusion of Theorem 4.1 is equivalently saying that the optimal value of this problem is less than $2^{-\Omega(d^{1/2})}$.

We can now write down the dual problem:

$$\begin{aligned} \max_{p \in (\mathbb{R}_{\geq 0})^{\Omega}} \quad & \sum_{(x,y,z)} F(x, y, z) \cdot p(x, y, z). \\ \text{s.t.} \quad & w \geq A^{\top} \cdot p. \end{aligned}$$

Let us study this linear program more closely. This problem asks for a packing, subject to the constraint that every slice function s upper bounds the weight placed inside the set indicated by s , and we would like to maximize the amount of weight placed in the set indicated by F . By strong duality, to prove Theorem 4.1, it suffices to show that every feasible solution p to the dual problem has value at most $2^{-\Omega(d^{1/2})}$.

Suppose for the sake of contradiction that p is a feasible solution with too large a value. We can assume W.L.O.G. that $p(x) = 0$ whenever $F(x) = 0$. Indeed, for any feasible p , it is easy to see that $(p \cdot F)$ is also feasible and attains the same value. To employ the pseudorandomness machinery, we would like interpret p as a probability density function. Note that p is non-negative by definition. After re-scaling p by $\bar{p} = \frac{p}{\|p\|_1}$, we obtain a probability density function $\bar{p}$ and let $\mathcal{D}$ be the distribution described by $\bar{p}$. Note that $\mathcal{D}$ is supported only on the support of F . Moreover, assuming $\sum_{x,y,z} p(x, y, z) \geq 2^{-c \cdot d^{1/2}}$ for some absolute $c > 0$, we have $\|p\|_1 \geq |\Omega|^{-1} \cdot 2^{-c \cdot d^{1/2}}$. Thus, for any slice function s it holds that

$$\mathbb{E}_{\mathcal{D}}[s] \leq \frac{|s|}{|\Omega|} \mathbb{E}_{\mathcal{D}}[\bar{p}] \leq \frac{1}{|\Omega|} \cdot \sum_{(x,y,z) \in s} p(x, y, z) \cdot |\Omega| \cdot 2^{c\sqrt{d}} \leq 2^{c\sqrt{d}} \mathbb{E}_{\mathcal{U}}[s].$$

In the following, we call any distribution $\mathcal{D}$ satisfying the condition above a $2^{cd^{1/2}}$ -*evasive* distribution to slice functions.

Definition 4.2. Let $k, d \in \mathbb{N}$. A distribution $\mathcal{D}$ supported on $X \times Y \times Z$ is (k, d) -*evasive*, if for every slice function s of density $\|s\|_1 \geq 2^{-d}$, it holds that $\mathbb{E}_{w \sim \mathcal{D}}[s(w)] \leq 2^k \mathbb{E}_{w \sim \mathcal{U}}[s(w)]$. A set $S \subseteq X \times Y \times Z$ is (k, d) -*evasive*, if the uniform distribution over S is (k, d) -*evasive*.

It has been clear from the above discussion that, to rule out the existence of a high-value dual solution, it suffices to prove the following proposition.

Proposition 4.3. There is an absolute constant $c > 0$ for which the following is true. Let F be a cylinder interaction. Suppose $\mathcal{D}$ is a $(cd^{1/2}, d/c)$ -*evasive* distribution supported on F . Then, it holds that $\mathbb{E}[F] > 2^{-d} \cdot |\Omega|$.

4.1 Proof of Proposition 4.3

This subsection is devoted to the proof of Proposition 4.3.

In this subsection, we slightly abuse notation by using p to denote the density function of $\mathcal{D}$. By Identity (1) (namely, $\mathbb{E}_{w \sim \mathcal{D}}[s(w)] = \mathbb{E}_{w \sim s}[p(w)] \cdot \mathbb{E}_{w \sim \mathcal{U}}[s(w)]$), the condition of the proposition can be equivalently written as $\mathbb{E}_{w \sim s}[p(w)] \leq 2^{cd^{1/2}}$ for every “large” slice function.

Density increment step. Let $t = c \cdot d^{1/2}$ where $c > 0$ is a small but absolute constant. Also let $\varepsilon > 0$ be a small constant. We would like to find a subcube C^* such that the distribution $\mathcal{D}$ becomes spread in the subcube. An iterative greedy approach taken by [KLM24] is the following: initialize with $C^* = \Omega$, whenever there is a “large” subcube in which $\mathcal{D}$ has a noticeably higher density, pass C^* to the said subcube. Iterate this until no new “density increment” step can be made.

It turns out that any greedy approach based on “locally” adjusting the solution can be conveniently carried out via a potential analysis: one defines a suitable potential function Φ measuring the “quality” of the solution. Then, one proves that the potential function strictly improves throughout the greedy procedure. When the procedure halts at a solution C^* , the solution achieves a local maximal for the potential. Moreover, once one has the potential function in hand, it is no longer necessary to explicitly track the greedy procedure. Instead, one can just *define* C^* to be a local optimum of the potential.

We will follow this plan. In particular, let C^* be a subcube maximizing the following optimization problem:

$$C^* = \arg \max_{C: |C| \geq 2^{-\frac{1}{\varepsilon}t^2}|\Omega|} \left\{ \Phi(C) := \mathbb{E}_{w \sim C}[p(w)] \cdot \left(\frac{|C|}{|\Omega|} \right)^{1/t} \right\}. \quad (16)$$

Intuitively, we restrict to a smaller cube C and try to balance between two terms: the “penalty” term of $\left(\frac{|C|}{|\Omega|} \right)^{1/t}$ and the “reward” term of $\mathbb{E}_{w \sim C}[p(w)]$. We observe some properties regarding C^* . First, since $C = \Omega$ is a feasible solution to the problem, we have

$$\Phi(C^*) \geq \Phi(\Omega) = \mathbb{E}_{\mathcal{U}}[p] = 1. \quad (17)$$

Since every subcube C with $|C| \geq 2^{-\frac{1}{\varepsilon}t^2}|\Omega|$ is also a large slice function, we have $\mathbb{E}_{w \sim C}[p(w)] \leq 2^t$. That is, the “gain” from $\mathbb{E}_{w \sim C^*}[p(w)]$ is upper bounded by 2^t . Hence, it follows that $\left(\frac{|C^*|}{|\Omega|} \right)^{1/t} \geq 2^{-t}$ or equivalently $|C^*| \geq 2^{-t^2}|\Omega|$.

Now, write $C^* = X^* \times Y^* \times Z^*$. Let p_{C^*} be the conditional distribution of p onto C^* . By (17), we have that $p_{C^*}(w) = \frac{p(w)}{\mathbb{E}_{w \sim C^*}[p(w)]} < p(w)$ for every $w \in C^*$. We define the projection of p_{C^*} onto the three two-marginals (a.k.a. faces):

$$\begin{aligned} f(x, y) &:= \mathbb{E}_z[p_{C^*}(x, y, z)], \\ g(x, z) &:= \mathbb{E}_y[p_{C^*}(x, y, z)], \\ h(y, z) &:= \mathbb{E}_x[p_{C^*}(x, y, z)]. \end{aligned}$$

Note that by construction, we get $\|f\|_1 = \mathbb{E}_{(x,y) \sim X^* \times Y^*}[f(x, y)] = \mathbb{E}_{(x,y,z) \in C^*}[p_{C^*}(x, y, z)] = 1$ and similarly we have $\|g\|_1 = \|h\|_1 = 1$.

The faces are spread. We now show that each face is spread with respect to large rectangles. We take f as a demonstrating example here and note that similar arguments apply equally well to g and h . Our main claim is that, for any rectangle $X' \times Y'$ of density $\delta = \frac{|X' \times Y'|}{|X^* \times Y^*|}$, as long as $\delta > 2^{-t^2}$, it holds that

$$\mathbb{E}_{(x,y) \sim X' \times Y'} [f(x, y)] \leq \delta^{-1/t} \mathbb{E}_{(x,y) \sim X^* \times Y^*} [f(x, y)]. \quad (18)$$

To briefly justify this, assume otherwise and observe that $(X' \times Y' \times Z^*)$ has density at least 2^{-2t^2} , and it attains a higher value for the optimization problem of (16). This contradicts the optimality of C^* . Note that (18) is almost the condition of Definition 3.1, with the caveat that we only consider $X' \times Y'$ of size at least δ . Before we argue for rectangles $X' \times Y'$ with size smaller than 2^{-t^2} , we first truncate f and enforce an ℓ_∞ condition for it.

Truncating the faces: Step 1. We consider the ℓ_∞ -truncated version of these functions:

$$\begin{aligned} \bar{f}(x, y) &= \min(f(x, y), 2^t), \\ \bar{g}(x, z) &= \min(g(x, z), 2^t), \\ \bar{h}(y, z) &= \min(h(y, z), 2^t). \end{aligned}$$

We claim that $\|\bar{f} - f\|_{L^1(X^* \times Y^*)} \leq 2^{-\Omega(t^2/c)}$, and similar conditions hold for g and h .

Indeed, consider

$$S := \{(x, y) : f(x, y) > 2^t\} \times Z^*.$$

Note that S is a slice. If $|S| > 2^{-t^2/c}|\Omega|$, it would follow that

$$\begin{aligned} \mathbb{E}_{w \sim p} [S(w)] &= \Pr_{w \sim p} [w \in C^*] \mathbb{E}_{w \sim p_{C^*}} [S(w)] \\ &= \frac{|C^*|}{|\Omega|} \mathbb{E}_{w \sim C^*} [p(w)] \mathbb{E}_{w \sim p_{C^*}} [S(w)] \quad (\text{By (1)}) \\ &= \frac{|C^*|}{|\Omega|} \mathbb{E}_{w \sim C^*} [p(w)] \cdot \frac{|S|}{|C^*|} \mathbb{E}_{w \sim S} [p_{C^*}(w)] \quad (\text{By (1)}) \\ &= \frac{|C^*|}{|\Omega|} \cdot \frac{|S|}{|C^*|} \mathbb{E}_{w \sim S} [p(w)] \\ &> \frac{|S|}{|\Omega|} \cdot 2^t \\ &= 2^t \cdot \Pr_{w \sim \Omega} [S(w)]. \end{aligned}$$

This contradicts our assumption on p . Therefore we must have $|S| < 2^{-t^2/c}|\Omega|$.

Now we know that S must be too small so that it is not subject to the evasiveness condition. Still, we can include some extra pairs (x, y) into S to get a larger slice function $\bar{S} = S_{xy} \times Z^*$ such that

$|S'| \approx 2^{-t^2/c} |\Omega|$. Noting that $f(x, y) = \bar{f}(x, y)$ for every $(x, y) \notin S_{xy}$, we proceed as

$$\begin{aligned}
& \|\bar{f} - f\|_{L^1(X^* \times Y^*)} \\
&= \mathbb{E}_{(x,y) \sim X^* \times Y^*} [|f(x, y) - \bar{f}(x, y)|] \\
&= \frac{|S_{xy}|}{|X^* \times Y^*|} \mathbb{E}_{(x,y) \sim S_{xy}} [|f(x, y) - \bar{f}(x, y)|] \\
&\leq \frac{|S_{xy}|}{|X^* \times Y^*|} \mathbb{E}_{(x,y) \sim S_{xy}} [f(x, y)] \\
&= \frac{|S_{xy}|}{|X^* \times Y^*|} \mathbb{E}_{(x,y,z) \sim S} [p_{C^*}(x, y, z)] \\
&\leq \frac{|S_{xy}|}{|X^* \times Y^*|} \mathbb{E}_{(x,y,z) \sim S} [p(x, y, z)] \quad (\text{Recall } p_{C^*} \leq p \text{ pointwise}) \\
&\leq 2^{t-t^2/c}. \quad (\text{By evasiveness})
\end{aligned}$$

Truncating the faces: Step 2. To apply Theorem 3.5, we also need to enforce the left-min-degree condition on $\bar{f}$ and $\bar{h}$. Toward that goal, we simply remove all $x \in X^*$ such that $\mathbb{E}_y[\bar{f}(x, y)] < (1 - \varepsilon)\|\bar{f}\|_1$ and similarly all $z \in Z^*$ such that $\mathbb{E}_y[\bar{h}(z, y)] < (1 - \varepsilon)\|\bar{h}\|_1$.

We actually only remove very few $x \in X^*$ and $z \in Z^*$ in this process. To see this, we note that the original f satisfies a stronger min-degree assumption: namely we have $\mathbb{E}_y[f(x, y)] \geq (1 - \varepsilon/2) \mathbb{E}[f]$. Indeed, suppose this fails at a row x . Let us consider $X' = X^* \setminus \{x\}$ and calculate the average of f on $X' \times Y$:

$$\begin{aligned}
& \mathbb{E}_{(x,y) \sim X' \times Y} [f(x, y)] \\
&> \frac{|X^*|}{|X^*| - 1} \left(1 - \frac{(1 - \varepsilon/2)}{|X^*|} \right) \\
&= \frac{|X^*| - 1 + \varepsilon/2}{|X^*| - 1} \\
&= 1 + \frac{\varepsilon/2}{|X^*| - 1}.
\end{aligned}$$

Thus, it is clearly seen that

$$\begin{aligned}
& \mathbb{E}_{(x,y) \sim X' \times Y} [f(x, y)] \cdot \left(\frac{|X'|}{|X^*|} \right)^{1/t} \\
&\geq \left(\left(1 + \frac{\varepsilon/2}{|X^*| - 1} \right)^t \cdot \left(1 - \frac{1}{|X^*|} \right) \right)^{1/t} > 1.
\end{aligned}$$

This shows that f is not spread, as witnessed by $X' \times Y$, a contradiction, contradicting to (18).

We have shown that $\bar{f}$ is obtained from f by removing at most $2^{t-t^2/c} \mathbb{E}[f]$ “mass” (measured in ℓ_1) from f . On the one hand, the decrease of $\mathbb{E}[\bar{f}]$, as compared with $\mathbb{E}[f]$, is negligible. On the other hand, one needs to remove $\frac{\varepsilon}{2} \mathbb{E}[f]$ of the “mass” from a row x to compromise the min-degree condition on that row. Thus, a simple application of Markov’s inequality implies that at most $2^{-\frac{t^2}{2c}} \cdot |X^*|$ rows are removed in Step 2.

Scaling back. After the two steps, say we currently have functions $\bar{f}, \bar{g}, \bar{h}$ and the ambient space $\bar{X} \times Y^* \times \bar{Z}$ ($\bar{X}, \bar{Z}$ may be ever so slightly smaller than X^*, Z^*). We scale up $\bar{f}, \bar{g}, \bar{h}$ and obtain:

$$\tilde{f} = \bar{f} / \|\bar{f}\|_1, \quad \tilde{g} = \bar{g} / \|\bar{g}\|_1, \quad \tilde{h} = \bar{h} / \|\bar{h}\|_1.$$

Note that $\|\bar{f}\|_1$ is understood as $\mathbb{E}_{(x,y) \sim \bar{X} \times Y^*}[\bar{f}(x,y)]$ (and similarly for $\|\bar{g}\|_1, \|\bar{h}\|_1$).

We now prove that the functions $\tilde{f}, \tilde{g}, \tilde{h}$ are almost as spread as f, g, h . We prove the claim for $\tilde{f}$ as the claims for $\tilde{g}, \tilde{h}$ are similarly established. The only property we need here is that $\tilde{f}, \tilde{g}$ and $\tilde{h}$ are extremely close to f, g, h in ℓ_1 distance. Take $\tilde{f}$ as a demonstrating example: for every $X' \times Y'$ of density $\delta > 2^{-t^2/4c}$, we have

$$\begin{aligned} \mathbb{E}_{w \sim X' \times Y'}[\tilde{f}(w)] &\leq \|f - \tilde{f}\|_1 \cdot \frac{|X^* \times Y^*|}{|X' \times Y'|} + \mathbb{E}_{w \sim X' \times Y'}[f(w)] \\ &\leq 2^{-\frac{t^2}{2c} + \frac{t^2}{4c} + O(t)} + \delta^{-1/t} \mathbb{E}_{w \sim X^* \times Y^*}[f(w)] \\ &\leq \delta^{-1/2t} \mathbb{E}_{w \sim X^* \times Y^*}[\tilde{f}(w)]. \end{aligned}$$

Also, for $X' \times Y'$ of density $\delta < 2^{-t^2/4c}$, we have $\|\tilde{f}|_{X' \times Y'}\|_1 \cdot \delta^{1/t} \leq \|\tilde{f}\|_\infty \cdot \delta^{1/t} \leq \|\tilde{f}\|_1$. These two cases combined imply that $\|\tilde{f}\|_{\tilde{t}, \tilde{t}} \leq (1 + \varepsilon)\|f\|_1$.

Recall that $\|\tilde{f}\|_1 = 1$ and $\|\tilde{f}\|_\infty \leq 2^t$. We now make use of Lemma 3.2 and conclude that

$$\|\tilde{f}\|_{U(k,2)} \leq (1 + \varepsilon)\|\tilde{f}\|_1$$

for some $k = \Theta(t/\varepsilon)$.

Lower-bounding the cylinder intersection. We are finally ready to lower-bound the size of the cylinder intersection F . Let $\bar{C} = \bar{X} \times Y^* \times \bar{Z}$. Recall we have the bound that $\|\tilde{f}\| \leq 2^t \cdot (1 + o(1))$ (and similarly for $\tilde{g}$ and $\tilde{h}$). We thus get the following pointwise bound:

$$F(x, y, z) \geq \frac{1}{4 \cdot 2^{3t}} \tilde{f}(x, y) \cdot \tilde{g}(x, z) \cdot \tilde{h}(y, z) \quad (19)$$

Write $(\tilde{f} \circ \tilde{h})(x, z) = \mathbb{E}_{y \sim Y^*}[\tilde{f}(x, y) \cdot \tilde{h}(z, y)]$. Applying Theorem 3.5, we conclude that

$$\|\tilde{f} \circ \tilde{h} - R_{\tilde{f}} \circ R_{\tilde{h}}\|_k \leq O(\varepsilon^2 \|\tilde{f}\|_1 \|\tilde{h}\|_1).$$

Therefore, we have

$$\begin{aligned} &\mathbb{E}_{(x,y,z) \sim \bar{C}}[\tilde{f}(x, y) \cdot \tilde{g}(x, z) \cdot \tilde{h}(y, z)] \\ &= \langle \tilde{f} \circ \tilde{h}, g \rangle \\ &\geq \langle R_{\tilde{f}} \circ R_{\tilde{h}}, g \rangle - \|\tilde{f} \circ \tilde{h} - R_{\tilde{f}} \circ R_{\tilde{h}}\|_k \cdot \|g\|_{k^*} \\ &\geq (1 - O(\varepsilon)) \|\tilde{f}\|_1 \|\tilde{g}\|_1 \|\tilde{h}\|_1 - O(\varepsilon^2 \|\tilde{f}\|_1 \|\tilde{g}\|_1 \|\tilde{h}\|_1) \\ &\geq (1 - O(\varepsilon)) \|\tilde{f}\|_1 \|\tilde{g}\|_1 \|\tilde{h}\|_1. \end{aligned}$$

Plug this back in (19), we get

$$\mathbb{E}_{(x,y,z) \sim \bar{C}}[F(x, y, z)] \geq \Omega(2^{-3t}).$$

Combining this with the bound $|\bar{C}| \geq \Omega(2^{-t^2}|\Omega|)$, the proof is complete.

5 Number-on-Forehead Communication Lower Bound

We describe a couple of consequences of Theorem 1.2. The first is a quantitative improvement of the main pseudorandomness result from [KLM24].

Theorem 5.1. *There is an absolute constant $c > 0$ for which the following is true. Let $t \in \mathbb{N}$, and let $D \subseteq [N]^3$ be a set that is $(c\sqrt{t}, \frac{t}{c})$ -evasive w.r.t. slice functions. For any cylinder intersection F of density $\mathbb{E}[F(x, y, z)] \leq 2^{-t}$, it holds that*

$$\mathbb{E}_{(x,y,z) \sim D}[F(x, y, z)] \leq 2^{-\Omega(\sqrt{t})}.$$

Proof. By Theorem 4.1, we can write $F \leq \sum_{i=1}^R c_i \cdot s_i(x, y, z)$ where each s_i is of size at least $2^{-O(t)}$ and $\sum_i c_i \mathbb{E}[s_i] \leq 2^{-\Omega(\sqrt{t})}$. Then, by the evasiveness of D and assuming $c > 0$ is small enough, we obtain:

$$\begin{aligned} \mathbb{E}_{w \sim D}[F(w)] &\leq \sum_{i=1}^R c_i \mathbb{E}_{w \sim D}[s_i(w)] \\ &\leq \sum_{i=1}^R c_i \cdot 2^{c\sqrt{t}} \cdot \mathbb{E}_{w \sim [N]^3}[s_i(w)] \\ &\leq 2^{-\Omega(\sqrt{t})}, \end{aligned}$$

as claimed. $\square$

We cite a construction of evasive sets from [KLM24].

Theorem 5.2 (See Lemma 2.16 and Lemma 2.17 of [KLM24]). *Let $q \geq 1$ be a prime power and $\mathbb{F}_q$ be the finite field of order q . Let $k \geq 1$ and $N = q^k$. Identify $[N]$ with $\mathbb{F}_q^k$. Then, the set*

$$D = \{(x, y, z) \in [N]^3 : \langle x, y \rangle = \langle x, z \rangle = \langle y, z \rangle\}$$

is $(O(1), \Omega(k \log(q)))$ -evasive against slice functions. Here $\langle x, y \rangle := \sum_{i=1}^k x_i y_i \in \mathbb{F}_q$ denotes the inner product between x and y .

We sketch the proof of Theorem 5.2 below and refer the readers to [KLM24] for the complete proof of a stronger statement. Note that here we only need to establish the evasiveness of D against slice functions. In [KLM24], a much stronger pseudorandomness property of D is proved.

Proof sketch. Let $S = S_{xy} \times S_z \subseteq (X \times Y) \times Z$ be a relatively large slice function (i.e., of size at least $q^{-ck} \cdot N^3$ where $c > 0$ is a small but absolute constant). The evasiveness condition on S is equivalent to $\frac{|S \cap D|}{|D|} \leq O\left(\frac{|S|}{|\Omega|}\right)$. So, to prove Theorem 5.2 we must upper-bound $\frac{|S \cap D|}{|S|}$ by roughly $\frac{|D|}{|\Omega|} \approx q^{-2}$. We may decompose $S_{xy} = \bigsqcup_{c \in \mathbb{F}_q} S_{xy}^{(c)}$ where $S_{xy}^{(c)} := \{(x, y) \in S_{xy} : \langle x, y \rangle = c\}$ and write $|S \cap D| = \sum_c |\{(x, y, z) \in S_{xy}^{(c)} \times S_z : \langle x, z \rangle = \langle y, z \rangle = c\}|$. For each $c \in \mathbb{F}_q$, we upper-bound the term by applying counting lemmas between z and x, y (Alternatively, one can appeal to the property that Inner Product is a good extractor). $\square$

We prove Theorem 1.3 as a corollary below.

Proof of Theorem 1.3. The randomized NOF communication upper bound can be obtained by running a (randomized) equality check among three players.

To see the deterministic (or non-deterministic) NOF lower bound, it is well-known that if D is solved by a non-deterministic NOF protocol with complexity $c \log^{1/2}(N)$ where $c > 0$ is sufficiently small, then there is a cylinder intersection $F \subseteq D$ that covers at least $2^{-c \log^{1/2}(N)}$ points of D . Namely, for $t = c^2 \log(N)$, we have

$$\mathbb{E}_{(x,y,z) \sim D} [F(x,y,z)] > 2^{-\sqrt{t}}.$$

Using Theorem 5.1 in its contrapositive form, this implies that $\frac{|F|}{|\Omega|} \geq 2^{-t} \geq N^{-O(c)}$. For $c > 0$ small and k (as in Theorem 5.2) large enough, we have $\frac{|F|}{|\Omega|} \geq N^{-O(c)} \gg N^{-2/k} \approx \frac{|D|}{|\Omega|}$. This implies that F cannot be a subset of D , a contradiction. $\square$

Acknowledgment

We thank Anthony Ostuni and Amit Chakrabarti for valuable comments on an earlier version of the paper. We are also grateful to the anonymous FOCS reviewers for their useful feedback.

References

- [AFK⁺24] Amir Abboud, Nick Fischer, Zander Kelley, Shachar Lovett, and Raghu Meka. New graph decompositions and combinatorial boolean matrix multiplication algorithms. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 935–943, 2024.
- [Baz09] Louay MJ Bazzi. Polylogarithmic independence can fool DNF formulas. *SIAM Journal on Computing*, 38(6):2220–2272, 2009.
- [Beh46] Felix A Behrend. On sets of integers which contain no three terms in arithmetical progression. *Proceedings of the National Academy of Sciences*, 32(12):331–332, 1946.
- [BS23a] Thomas F Bloom and Olof Sisask. An improvement to the kelley-meka bounds on three-term arithmetic progressions. *arXiv preprint arXiv:2309.02353*, 2023.
- [BS23b] Thomas F Bloom and Olof Sisask. The kelley-meka bounds for sets free of three-term arithmetic progressions. *Essential Number Theory*, 2(1):15–44, 2023.
- [DETT10] Anindya De, Omid Etesami, Luca Trevisan, and Madhur Tulsiani. Improved pseudo-random generators for depth 2 circuits. In *International Workshop on Randomization and Approximation Techniques in Computer Science*, pages 504–517. Springer, 2010.
- [EHPS24] Christian Elsholtz, Zach Hunter, Laura Proske, and Lisa Sauermann. Improving behrend’s construction: Sets without arithmetic progressions in integers and over finite fields. *arXiv preprint arXiv:2406.12290*, 2024.
- [Elk10] Michael Elkin. An improved construction of progression-free sets. In *Proceedings of the twenty-first annual ACM-SIAM symposium on Discrete Algorithms*, pages 886–905. SIAM, 2010.

- [FHHK24] Yuval Filmus, Hamed Hatami, Kaave Hosseini, and Esty Kelman. Sparse graph counting and kelley-meka bounds for binary systems. In *2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 1559–1578. IEEE, 2024.
- [Fox11] Jacob Fox. A new proof of the graph removal lemma. *Annals of Mathematics*, pages 561–579, 2011.
- [G⁺08] Loukas Grafakos et al. *Classical fourier analysis*, volume 2. Springer, 2008.
- [GW10] Ben Green and Julia Wolf. A note on Elkin’s improvement of Behrend’s construction. In *Additive number theory*, pages 141–144. Springer, 2010.
- [Hun24] Zach Hunter. New lower bounds for $r\text{-}3(n)$. *arXiv preprint arXiv:2401.16106*, 2024.
- [KLM24] Zander Kelley, Shachar Lovett, and Raghu Meka. Explicit separations between randomized and deterministic number-on-forehead communication. In *Proceedings of the 56th Annual ACM Symposium on Theory of Computing*, pages 1299–1310, 2024.
- [KM23] Zander Kelley and Raghu Meka. Strong bounds for 3-progressions. In *2023 IEEE 64th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 933–973. IEEE, 2023.
- [O’B11] Kevin O’Bryant. Sets of integers that do not contain long arithmetic progressions. *the electronic journal of combinatorics*, 18(P59):1, 2011.
- [Raz09] Alexander Razborov. A simple proof of Bazzi’s theorem. *ACM Transactions on Computation Theory (TOCT)*, 1(1):1–5, 2009.
- [Tal17] Avishay Tal. Tight bounds on the Fourier spectrum of AC0. In *32nd Computational Complexity Conference (CCC 2017)*, pages 15–1. Schloss Dagstuhl–Leibniz-Zentrum für Informatik, 2017.