

QUASIPOLYNOMIAL BOUNDS FOR THE CORNERS THEOREM

MICHAEL JABER, YANG P. LIU, SHACHAR LOVETT, ANTHONY OSTUNI, AND MEHTAAB SAWHNEY

ABSTRACT. Let G be a finite abelian group and A be a subset of $G \times G$ which is corner-free, meaning that there are no $x, y \in G$ and $d \in G \setminus \{0\}$ such that $(x, y), (x + d, y), (x, y + d) \in A$. We prove that

$$|A| \leq |G|^2 \cdot \exp(-(\log |G|)^{\Omega(1)}).$$

As a consequence, we obtain polynomial (in the input length) lower bounds on the non-deterministic communication complexity of **Exactly-N** in the 3-player Number-on-Forehead model. We also obtain the first “reasonable” lower bounds on the coloring version of the 3-dimensional corners problem and equivalently the deterministic communication complexity of **Exactly-N** in the 4-player Number-on-Forehead model.

CONTENTS

1. Introduction	1
2. Overview of the Proof	6
3. Combinatorial Spreadness and Relative Sifting	13
4. Additional Tools	20
5. The Finite Field Case	22
6. Bohr Sets, Algebraic Spreadness, and Pseudorandomization	35
7. Density Increment by Reduction to the Grid Norm	52
8. Coloring Bounds for 3-Dimensional Corners	69
Appendix A. The Quasirandom Group Case	70
Appendix B. Almost Periodicity	80
References	82

1. INTRODUCTION

The work of Chandra, Furst, and Lipton [14] in 1983 introduced the Number-on-Forehead (NOF) model of communication complexity to model interaction between parties with shared information. The k -NOF model is defined by k players communicating over a shared channel in order to compute a function $f : (\{0, 1\}^n)^k \rightarrow \{0, 1\}$. Each player can see the $k - 1$ inputs of every other player, but they cannot see their own. This model has a number of striking connections in theoretical computer science and combinatorics. For instance lower bounds for $k = \omega(\log n)$ players would imply breakthrough circuit lower bounds [5, 7, 51, 56].

The work of Chandra, Furst, and Lipton [14] was primarily concerned with the **Exactly-N** problem, which is now one of the most studied problems in the NOF model. In **Exactly-N** each player receives a number in $[N] := \{1, 2, \dots, N\}$ (given as $\lceil \log_2 N \rceil$ bits) and the players aim to check if these numbers sum up to N . The authors of [14] observed that the 3-NOF complexity of **Exactly-N** is bounded by $O(\sqrt{\log N})$ via the Behrend construction of sets without three-term arithmetic progressions (3-APs) [8] (observe here that the trivial protocol gives a bound of $O(\log N)$). The authors of [14] in fact observe an equivalence between **Exactly-N** (for three players) and the size of sets $S \subseteq [N]^2$ without *corners*: three points $(x, y), (x + d, y), (x, y + d) \in [N]^2$ with $d \neq 0$. More generally, the

communication complexity of several natural functions is known to be equivalent to central problems in Ramsey theory [14, 46, 63].

Despite these remarkable connections, the present state of knowledge regarding the NOF model is lacking in a number of respects. For instance, despite the fact that optimal separation between randomized and deterministic NOF communication were known nonexplicitly [6] for more than a decade, explicit constructions exhibiting separation for 3-NOF have only been developed recently. In particular, Kelley, Lovett, and Meka [41] exhibited an explicit 3-player function which has a constant cost randomized protocol, but requires $\Omega(n^{1/3})$ bits of communication to compute deterministically. This work relies crucially on finding combinatorial interpretations of ideas stemming from breakthrough work of Kelley and Meka on 3-APs [42]. However, the lower bounds for Exactly-N are not nearly as strong.

The problem of upper bounding the size of corner-free sets falls within a broad program of giving “reasonable” bounds for various results in additive combinatorics which have been proven via ergodic theory. A vast generalization of Szemerédi’s theorem [65, 66] is the multidimensional Szemerédi’s theorem of Furstenberg and Katznelson [23] which states that given a dense subset of $[N]^d$ one may find an arbitrary pattern P . While weak quantitative bounds for this theorem follow via the works of Nagle, Rödl, Schacht, and Skokan [50, 57] and Gowers [25] due to versions of the hypergraph regularity lemma (with a subsequent alternate proof given by Tao [68]), obtaining “reasonable” bounds for this theorem has been seen as a problem of substantial importance (see e.g. [29, Problem 20] or [24]).

Despite the interest in these questions, there are “reasonable” bounds known only for the cases of corners due to Shkredov [61], and for the configuration (x, y) , $(x + d, y)$, $(x + 2d, y)$, $(x, y + d)$ due to Peluse [53]. We note that even considering just corner-free sets has a rich history. Before the work of Furstenberg and Katznelson, Ajtai and Szemerédi [1] were able to prove that corner-free sets of $[N] \times [N]$ have size $o(N^2)$. This proof for instance inspired a portion of the combinatorial proof of the Density Hales–Jewett theorem of Polymath [54]; however due to needing to find growing length progressions via Szemerédi’s theorem the proof gives an exceedingly weak bound. The well-known alternate proof of the corners theorem, due to Solymosi [64] via the regularity/triangle removal lemma, also currently comes with weak bounds.

The previously best known bound for the corners theorem is due to Shkredov [62] who proved that the size of the largest corner-free set is bounded by $N^2 \cdot (\log \log N)^{-c}$ where any $c < 1/73$ is admissible (also see the beautiful exposition by Green over finite fields [30, 31]). We remark that this implies a lower bound of $\Omega(\log \log \log N)$ for Exactly-N in the 3-player NOF model for deterministic protocols.

Via a standard projection argument (see [71, Section 2.4]), bounds on corner-free sets imply bounds on sets free of 3-APs. Given this, many researchers have speculated whether the recent breakthrough of Kelley and Meka [42] could be used to improve bounds for corner-free sets (see e.g. [48] and [40]). Peluse [52, Problem 1.18] even asked whether methods underlying Kelley–Meka could allow one to achieve a savings of one logarithm over Shkredov’s bound. Preliminary work in this direction considered the easier problem of obtaining quasipolynomial bounds for so-called “skew corners” [38, 49]; however the bounds of Shkredov remained unimproved. In this work, we provide quasipolynomial bounds for the corners theorem, achieving a doubly-exponential improvement over the work of Shkredov [62].

Theorem 1.1. *There exists a constant $c > 0$ such that the following holds. Let $(G, +)$ be a finite abelian group. Let $A \subseteq G \times G$ with no $x, y, d \in G$ with $d \neq 0$ such that (x, y) , $(x + d, y)$, $(x, y + d) \in A$. Then*

$$|A| \leq |G|^2 \cdot \exp(-c(\log |G|)^{1/600}).$$

By embedding $[N]$ into the cyclic group $\mathbb{Z}/(4N\mathbb{Z})$ one may obtain a similar conclusion for corner-free subsets of $[N] \times [N]$. Furthermore a trick of Green (see [62, Lemma 6.2]) allows one to prove essentially identical bounds only avoiding corners with $d > 0$.

Note that [Theorem 1.1](#) has the optimal “shape”. Behrend’s construction of 3-AP-free sets [8] can be lifted to corner-free sets of density $\exp(-O(\sqrt{\log N}))$ (see [32, 37, 46] for improvements in the case of corners and [21] for improvements in the 3-AP case). Our techniques appear unable to yield results approaching this lower bound, so we do not attempt to optimize the constant $1/600$.

Interpreting our results in the language of communication complexity via [14], we obtain lower bounds for the **Exactly-N** problem. Because we provide bounds on the density version of the corners problem, we obtain lower bounds against non-deterministic 3-NOF protocols.

Corollary 1.2. *Any non-deterministic 3-NOF protocol computing **Exactly-N** requires $\Omega((\log N)^{\Omega(1)})$ bits of communication.*

We remark that **Exactly-N** in k -NOF for any $k \geq 3$ has a constant communication randomized protocol. Moreover, our results can be combined with a proof of Graham and Solymosi [28] to provide improved bounds for the coloring variant of finding 3-dimensional corners. The following is a consequence of [Theorem 1.1](#).

Corollary 1.3. *For a sufficiently small constant c , and any abelian group G and coloring of $G \times G \times G$ with $c \log \log \log |G|$ colors, there are $x, y, z, d \in G$ with $d \neq 0$ such that (x, y, z) , $(x + d, y, z)$, $(x, y + d, z)$, and $(x, y, z + d)$ are all of the same color.*

The communication complexity translation also carries over to colorings and higher-dimensional corners (see e.g. [35, Appendix A]) to give the following corollary. Here, we obtain a lower bound against deterministic protocols because we obtain a coloring lower bound for 3-dimensional corners.

Corollary 1.4. *Any deterministic 4-NOF protocol computing **Exactly-N** requires $\Omega(\log \log \log \log N)$ bits of communication.*

To the best of our knowledge, this provides the largest separation between the randomized and deterministic k -NOF communication complexity for any $k \geq 4$ for an explicit function. We remark that in order to obtain “reasonable” bounds in [Corollaries 1.3](#) and [1.4](#), it is crucial that we obtain a quasipolynomial bound in [Theorem 1.1](#); for example, an inverse logarithmic-type bound would give a tower type dependence in [Corollaries 1.3](#) and [1.4](#).

Following the influential survey of Green [31], it has become commonplace to consider problems in additive combinatorics in the model setting of finite field vector spaces before considering general abelian groups. Our main result over $\mathbb{F}_2^n$ is the following bound for corners.

Theorem 1.5. *There exists a constant $c > 0$ such that the following holds. Let $A \subseteq \mathbb{F}_2^n \times \mathbb{F}_2^n$ with no $x, y, d \in \mathbb{F}_2^n$ with $d \neq 0$ such that (x, y) , $(x + d, y)$, $(x, y + d) \in A$. Then for $N = 2^n$, it holds that*

$$|A| \leq N^2 \cdot \exp(-c(\log N)^{1/178}).$$

We remark that the previous state of the art was essentially of the same shape as work of Shkredov (due to Lacey and McClain [44]). This is unlike the case of three-term arithmetic progressions and the resolution of the capset problem due to Ellenberg and Gijswijt [20] based on the polynomial method of Croot, Lev and Pach [18]. Furthermore it is now understood that naïve generalizations of the polynomial method technique are unlikely to work for corners [15].

In our setting, it is also convenient to consider the case of BMZ corners [9] for quasirandom groups. In particular, we will consider the largest subset of $G = \text{SL}_2(\mathbb{F}_p)$ (e.g. the set of 2 by 2 matrices over $\mathbb{F}_p$ with determinant 1) and finding subsets which avoid $\{(x, y), (xg, y), (x, gy)\}$ with $g \neq 1_G$.

Theorem 1.6 (Special case of [Theorem A.1](#)). *There exists a constant $c > 0$ such that the following holds. Let $G = \text{SL}_2(\mathbb{F}_p)$ and $A \subseteq G \times G$ with no $x, y, g \in G$ with $g \neq \mathbf{1}_G$ such that $(x, y), (xg, y), (x, gy) \in A$. Then*

$$|A| \leq |G|^2 \cdot \exp(-c(\log |G|)^{1/11}).$$

In this setting, the best known bound previously was $|G|^2 \cdot (\log |G|)^{-\Omega(1)}$ due to Austin [3]. As in the abelian case, [Theorem 1.6](#) immediately implies lower bounds for the NOF communication complexity of Exactly-N over $\text{SL}_2(\mathbb{F}_p)$ (see e.g. [70, Section 5] for a proof, as well as an exposition of Austin's result).

Furthermore we remark that the quasipolynomial bounds on [Theorem 1.5](#) and [Theorem 1.6](#) appear to be at the limit of the density increment method. However, no Behrend-type constructions are known in these settings, and therefore it is plausible that the size of the largest corner-free set in each setting is power saving. Such bounds would imply optimal separations between randomized and deterministic 3-NOF protocols.

Finally we remark that our main result and that of Kelley and Meka [42] demonstrate that a multiplicative density increment strategy gives nearly optimal bounds. While such a strategy can be used for higher complexity patterns, as in Szemerédi's theorem [34, 45], these approaches require passing to comparatively extremely small substructures. However the authors know of no counterexample to finding a multiplicative density increment while paying only a poly-logarithmic dependence in density in the codimension of the associated substructures. Exploring the plausibility of such approaches (and in particular if there are counterexamples which distinguish the higher order cases from the result) are of substantial interest.

1.1. Corners and progressions in nonabelian groups. In this section we record an argument due to Fox (personal communication) which shows that [Theorem 1.1](#) implies quasipolynomial bounds for BMZ and naïve corners in all groups, not necessarily abelian. This also implies similar bounds for Roth's theorem in general groups.

Corollary 1.7. *There exists a constant $c > 0$ such that the following holds. Let G be a finite group and $A \subseteq G \times G$ with no $x, y, g \in G$ with $g \neq \mathbf{1}_G$ such that $(x, y), (xg, y), (x, gy) \in A$ (i.e., a BMZ corner). Then*

$$|A| \leq |G|^2 \cdot \exp(-c(\log |G|)^{1/1200}).$$

Proof. Let $H \subseteq G$ be an abelian subgroup of G of size $|H| \geq 2^{\Omega(\sqrt{\log |G|})}$ – such a subgroup exists by a result of Pyber [55]. For $x, y \in G$ define the set

$$A_{x,y} := \{(h_1, h_2) \in H \times H : (xh_1, h_2y) \in A\}.$$

We claim that $A_{x,y}$ has no corners. If it does, then there are h_1, h_2, h_3 such that

$$(xh_1, h_2y), (xh_1h_3, h_2y), (xh_1, h_3h_2y) \in A,$$

which implies that A contains a BMZ corner. By [Theorem 1.1](#) we conclude that

$$|A_{x,y}| \leq |H|^2 \cdot \exp(-c(\log |H|)^{1/600}) \leq |H|^2 \cdot \exp(-c(\log |G|)^{1/1200}).$$

By averaging, we know that

$$|A| = \frac{|G|^2}{|H|^2} \cdot \mathbb{E}_{x,y \in G} [|A_{x,y}|] \leq |G|^2 \cdot \exp(-c(\log |G|)^{1/1200}). \quad \square$$

Corollary 1.8. *There exists a constant $c > 0$ such that the following holds. Let G be a finite group and $A \subseteq G \times G$ with no $x, y, g \in G$ with $g \neq \mathbf{1}_G$ such that $(x, y), (xg, y), (x, yg) \in A$ (i.e., a naïve corner). Then*

$$|A| \leq |G|^2 \cdot \exp(-c(\log |G|)^{1/1200}).$$

Proof. The proof is nearly identical to that of [Corollary 1.7](#), except one instead defines

$$A_{x,y} := \{(h_1, h_2) \in H \times H : (xh_1, yh_2) \in A\}. \quad \square$$

This implies quasipolynomial bounds for subsets $A \subseteq G$ avoiding nontrivial solutions to $xy = z^2$. This is an analog of the 3-AP question in nonabelian groups.

Corollary 1.9. *There exists a constant $c > 0$ such that the following holds. Let G be a finite group and $A \subseteq G$ with no x, y, z not all equal satisfying $xy = z^2$. Then*

$$|A| \leq |G| \cdot \exp(-c(\log |G|)^{1/1200}).$$

Proof. Define $S = \{(x, y) \in G \times G : x^{-1}y \in A\}$. We claim that S does not contain a corner of the form in [Corollary 1.8](#), i.e., $(x, y), (xg, y), (x, yg) \in S$ for some $x, y, g \in G$ with $g \neq 1_G$. Indeed, otherwise $x^{-1}y, (xg)^{-1}y, x^{-1}yg \in A$, and $x^{-1}yg(xg)^{-1}y = (x^{-1}y)^2$. Thus, [Corollary 1.8](#) implies that

$$|A||G| = |S| \leq |G|^2 \cdot \exp(-c(\log |G|)^{1/1200})$$

as desired. $\square$

The previous best bound for general groups was $|G|^2 \cdot (\log \log |G|)^{-1}$ due to Sanders [\[58\]](#) (which is recovered by the argument of Fox plus Shkredov’s corners bound). For the symmetric group S_n , the previous best bound was $N \cdot e^{-\Omega((\log \log N)^2)}$ for $N = n!$, and is due to Keevash and Lifshitz [\[39\]](#).

1.2. Organization of the remainder of the paper. The remainder of the paper is organized as follows. In [Section 2](#) we give a high-level overview of the main ideas in the proof. In [Section 3](#) we define what it means for a set to be combinatorially spread against rectangles, and prove a *sifting* statement, i.e., that functions with a large grid norm have large density on a large rectangle. Critically, our sifting statement is relative, meaning that it works even when the sets we consider are subsets of a sparse pseudorandom object. In [Section 4](#) we introduce additional additive combinatorial tools needed for our analysis, such as spectral positivity. In [Section 5](#) we prove [Theorem 1.5](#), our corners bound for finite fields, which captures a number of the main conceptual ideas of the general abelian group case (pseudorandomization, density increment), but is technically simpler.

The next two sections of the paper are devoted to proving [Theorem 1.1](#), the corners bound over general abelian groups. We introduce Bohr sets in [Section 6](#), and then define and apply several pseudorandomness properties of subsets of Bohr sets. In addition, we describe how to perform the pseudorandomization procedure in the Bohr set setting. Finally, [Section 7](#) uses a density increment argument to prove [Theorem 1.1](#). [Corollary 1.3](#), our coloring lower bound for 3-dimensional corners, is proven in [Section 8](#). The appendix contains one section ([Appendix A](#)) proving [Theorem 1.6](#), our corners theorem in quasirandom groups, and another stating and applying almost periodicity ([Appendix B](#)).

1.3. Acknowledgements. MJ, SL, and AO thank Russell Impagliazzo and David Zuckerman for helpful conversations and Ilya Shkredov for answering a question about the current state-of-the-art. MJ would like to thank Freddie Manners for useful discussions, Sarah Peluse for her encouragement and collaboration on this question, as well as Amey Bhangale and Surya Teja Gavva for their collaboration on this question at the Simons Institute for the Theory of Computing. MS thanks Tim Gowers, Huy Pham, Ashwin Sah for useful discussions and Sarah Peluse for discussions regarding [Section 8](#). A portion of this work was conducted when MS visited “New Frontiers in Extremal and Probabilistic Combinatorics” at the SwissMAP Research Station.

MJ is supported by NSF Grant CCF-2312573 and a Simons Investigator Award (#409864, David Zuckerman). Part of this research was conducted when YL was a Postdoctoral Member at the IAS, and is based upon work supported by the National Science Foundation under Grant No. DMS-1926686. SL and AO are supported by the Simons Investigator Award #929894 (Shachar Lovett)

and NSF award CCF-2425349. This research was conducted during the period MS served as a Clay Research Fellow.

2. OVERVIEW OF THE PROOF

We now provide an overview of the proof strategy for [Theorem 1.5](#) and then discuss the modifications required for [Theorem 1.1](#). Our result, like several results in additive combinatorics, is established via a density increment argument – either the set A contains the expected number of corners, or has higher density onto some structured sub-instance. We start by reviewing Shkredov’s bound [\[62\]](#) for the corners problem before discussing the novel aspects of this work.

2.1. Shkredov’s corners bound. Consider a subset $A \subseteq \mathbb{F}_2^n \times \mathbb{F}_2^n$ with density α , so $|A| = \alpha \cdot 4^n$. In this case, the expected density of corners is about α^3 for a random set A . The first step in Shkredov’s corners bound is to prove that if A has much fewer than this expected density of corners, then A has a density increment onto a subrectangle. That is, there are subsets $X, Y \subseteq \mathbb{F}_2^n$ such that:

- (1) $|X|, |Y| \geq \alpha^{O(1)} \cdot 2^n$, and
- (2) $|A \cap (X \times Y)| \geq (\alpha + \alpha^{O(1)})|X||Y|$.

This is proven via a Cauchy-Schwarz argument: if A has few corners, then its balanced indicator function, i.e. $\mathbb{1}_A - \alpha$, has a large *box norm*, which is the $(2, 2)$ -grid norm (defined later in [Definition 2.1](#)). This suggests a natural density increment approach: maintain $X \times Y \subseteq \mathbb{F}_2^n \times \mathbb{F}_2^n$ such that the density of A within $X \times Y$ goes up over time. To carry this out, one must establish a statement of the following form: if $A \subseteq X \times Y$ has few corners, then A admits an $\alpha^{O(1)}$ density increment onto some $X' \times Y' \subseteq X \times Y$.

It turns out that proving such a statement requires an additional pseudorandomness assumption on X and Y . In Shkredov’s approach, the pseudorandomness assumption was that the indicator functions of X and Y have very small nontrivial Fourier coefficients – smaller than the densities of X and Y themselves. Because this pseudorandomness guarantee may not hold for the X and Y that are incremented onto, Shkredov uses a pseudorandomization procedure to make X and Y satisfy this property. At a high level, the pseudorandomization procedure repeatedly passes to subspaces of $\mathbb{F}_2^n$ on which the Fourier coefficients of X and Y are large.

We now explain why this approach gets a bound of $N^2/(\log \log N)^c$ for corners, where $N = 2^n$, which is two logarithmic factors off the Behrend-type bounds of $N^2/2^{O((\log N)^c)}$. The first point to understand is what the densities of X and Y are at the end of the density increment procedure. The number of density increment steps is $\alpha^{-O(1)}$, each of which decreases the densities of X and Y by $\alpha^{O(1)}$, so the final densities are $\delta := \exp(-1/\alpha^{O(1)})$. In the pseudorandomization procedure, guaranteeing that all Fourier coefficients are less than δ requires passing to a codimension $\delta^{-O(1)}$ subspace. Thus, we need that $\delta^{-O(1)} \leq n$ so $\alpha = (\log n)^{-c} = (\log \log N)^{-c}$.

This discussion clarifies why Shkredov’s bound is doubly logarithmic: one logarithmic factor stems from the density increment, and the other from pseudorandomization. The remainder of this overview is devoted to explaining our approach for avoiding *both* these losses, and establishing a Behrend-type bound for the corners theorem.

2.2. Hölderizing and XYD containers. In this section we discuss how to avoid the first logarithmic loss in Shkredov’s bound, by improving the density increment obtained from $\alpha^{O(1)}$ to a multiplicative $(1 + \varepsilon)$ for some absolute constant $\varepsilon > 0$. This reduces the number of density increment steps down to $O(\log(1/\alpha))$. The starting point for understanding how to do this is by replacing the initial use of Cauchy–Schwarz in Shkredov [\[62\]](#) with the following Hölder manipulation inspired by [\[42\]](#) (see [\[11, 1. Hölder–lifting\]](#)).

Let A denote a corner-free set of density α (so $|A| = \alpha \cdot 4^n$) and $f_A = \mathbb{1}_A - \alpha$. For an even positive integer k we have that (after substituting d to be $x + y + d$)

$$\begin{aligned} \left| \mathbb{E}_{x,y,d \in \mathbb{F}_2^n} \mathbb{1}_A(x,y) \mathbb{1}_A(x,y+d) f_A(x+d,y) \right|^k &= \left| \mathbb{E}_{x,y,d \in \mathbb{F}_2^n} \mathbb{1}_A(x,y) \mathbb{1}_A(x,x+d) f_A(y+d,y) \right|^k \\ &\leq \left(\mathbb{E}_{x,y \in \mathbb{F}_2^n} \mathbb{1}_A(x,y) \right)^{k-1} \cdot \left(\mathbb{E}_{x,y \in \mathbb{F}_2^n} \mathbb{1}_A(x,y) \cdot \left(\mathbb{E}_{d \in \mathbb{F}_2^n} \mathbb{1}_A(x,x+d) (f_A(y+d,y))^k \right) \right) \\ &\leq \left(\mathbb{E}_{x,y \in \mathbb{F}_2^n} \mathbb{1}_A(x,y) \right)^{k-1} \cdot \left(\mathbb{E}_{x,y \in \mathbb{F}_2^n} \left(\mathbb{E}_{d \in \mathbb{F}_2^n} \mathbb{1}_A(x,x+d) f_A(y+d,y) \right)^k \right). \end{aligned}$$

Via switching the order of summation and a further application of Cauchy–Schwarz, we have that

$$\begin{aligned} &\left(\mathbb{E}_{x,y \in \mathbb{F}_2^n} \left(\mathbb{E}_{d \in \mathbb{F}_2^n} \mathbb{1}_A(x,x+d) f_A(y+d,y) \right)^k \right)^2 \\ &\leq \left(\mathbb{E}_{x_1,x_2 \in \mathbb{F}_2^n} \left(\mathbb{E}_{d \in \mathbb{F}_2^n} \mathbb{1}_A(x_1,x_1+d) \mathbb{1}_A(x_2,x_2+d) \right)^k \right) \cdot \left(\mathbb{E}_{y_1,y_2 \in \mathbb{F}_2^n} \left(\mathbb{E}_{d \in \mathbb{F}_2^n} f_A(y_1+d,y_1) f_A(y_2+d,y_2) \right)^k \right). \end{aligned}$$

Taking $k \approx \log(1/\alpha)$, we find that either

$$\mathbb{E}_{x_1,x_2 \in \mathbb{F}_2^n} \left(\mathbb{E}_{d \in \mathbb{F}_2^n} \mathbb{1}_A(x_1,x_1+d) \mathbb{1}_A(x_2,x_2+d) \right)^k \geq (3\alpha)^{2k}$$

or

$$\mathbb{E}_{y_1,y_2 \in \mathbb{F}_2^n} \left(\mathbb{E}_{d \in \mathbb{F}_2^n} f_A(y_1+d,y_1) f_A(y_2+d,y_2) \right)^k \geq (\alpha/9)^{2k}.$$

We want to express the second case as some property of $\mathbb{1}_A$ more directly. This is done by using a graph theoretic analog of spectral positivity (see [Lemma 4.2](#)), as developed by Kelley, Lovett, and Meka [\[41\]](#), which upgrades the dichotomy to be:

$$\mathbb{E}_{x_1,x_2 \in \mathbb{F}_2^n} \left(\mathbb{E}_{d \in \mathbb{F}_2^n} \mathbb{1}_A(x_1,x_1+d) \mathbb{1}_A(x_2,x_2+d) \right)^k \geq (3\alpha)^{2k} \quad (2.1)$$

or

$$\mathbb{E}_{y_1,y_2 \in \mathbb{F}_2^n} \left(\mathbb{E}_{d \in \mathbb{F}_2^n} \mathbb{1}_A(y_1+d,y_1) \mathbb{1}_A(y_2+d,y_2) \right)^{k'} \geq ((1 + 1/1000)\alpha)^{2k'} \quad (2.2)$$

where $k' \geq k$ but still satisfies $k' \approx \log(1/\alpha)$. Observe that in our analysis to this point we have had A which lives in $\mathbb{F}_2^n \times \mathbb{F}_2^n$.

Equations (2.1) and (2.2) are best interpreted as conditions on certain *grid norms*, as introduced in work of Kelley, Lovett, and Meka [\[41\]](#).

Definition 2.1 (Grid norms). Fix positive integers $k, \ell \geq 1$ and finite sets Ω_1, Ω_2 . Let $A : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$. We define the (k, ℓ) -grid norm of A as

$$\|A\|_{G(k,\ell)} = \left| \mathbb{E}_{\substack{x_1, \dots, x_k \in \Omega_1 \\ y_1, \dots, y_\ell \in \Omega_2}} \prod_{\substack{1 \leq i \leq k \\ 1 \leq j \leq \ell}} A(x_i, y_j) \right|^{1/(k\ell)}.$$

Remark. It is known that the (k, ℓ) -grid norm is a seminorm if k, ℓ are both even positive integers [\[36\]](#).

Intuitively, the grid norm counts the number of copies of the complete bipartite graph $K_{k,\ell}$ in a “graph” given by a function $f : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$. To interpret (2.1) as a grid norm, one can define the function $F : \mathbb{F}_2^n \times \mathbb{F}_2^n \rightarrow [0, 1]$ as $F(x, d) := \mathbb{1}_A(x, x+d)$ and note that (2.1) is exactly equivalent to $\|F\|_{G(2,k)} \geq 3\alpha$. Similarly, defining $G : \mathbb{F}_2^n \times \mathbb{F}_2^n \rightarrow [0, 1]$ as $G(y, d) := \mathbb{1}_A(y+d, y)$, (2.2) is equivalent to $\|G\|_{G(2,k')} \geq (1 + 1/1000)\alpha$. A key result of [\[41\]](#) is a structural theorem regarding functions with large grid norm, which they refer to as *sifting*. Informally, it says that if $\|F\|_{G(k,\ell)} \geq \tau$ for a nonnegative function F , then there is a subrectangle $\Omega'_1 \times \Omega'_2 \subseteq \Omega_1 \times \Omega_2$ on which F has average value at least $(1 - \varepsilon)\tau$, and $|\Omega'_1| \geq (\varepsilon\tau)^{O(k+\ell)} |\Omega_1|$ and $|\Omega'_2| \geq (\varepsilon\tau)^{O(k+\ell)} |\Omega_2|$, i.e., the subrectangle

is not too small. For completeness, we provide a proof of such a sifting statement in [Theorem 3.4](#). Later we will see that this sifting theorem does *not* suffice for our corners bound for two reasons, and instead we require an *asymmetric* and *relative* sifting theorem which is much more challenging to establish (see [Theorem 3.5](#)).

Applying [Theorem 3.4](#), if (2.1) holds, one may find sets $X \subseteq \mathbb{F}_2^n$ and $D \subseteq \mathbb{F}_2^n$ such that

$$\mathbb{E}_{x \in X, d \in D} \mathbb{1}_A(x, x + d) \geq (1 + \Omega(1))\alpha$$

and

$$\mathbb{E}_{x, d} \mathbb{1}_{x \in X} \mathbb{1}_{d \in D} \geq \exp(-O(\log(1/\alpha)^2)).$$

In other words, the set A has larger density on an (X, D) -set, i.e., a set of the form

$$\{(x, y) \in \mathbb{F}_2^n \times \mathbb{F}_2^n : x \in X, x + y \in D\}.$$

Similarly, if (2.2) holds then [Theorem 3.4](#) gives that A has increased density on a (Y, D) -set:

$$\{(x, y) \in \mathbb{F}_2^n \times \mathbb{F}_2^n : y \in Y, x + y \in D\}.$$

This now gives a first crucial difference between our setting and that of Shkredov [62]. While Shkredov’s proof always maintained a set $X \times Y$ as the pseudorandom container to density increment against, we cannot afford to do this as we may be forced to use either (X, D) or (Y, D) increments. Instead, our pseudorandom containers take the form (for a subspace $W \subseteq \mathbb{F}_2^n$)

$$S(X, Y, D) := \{(x, y) \in W \times W : x \in X, y \in Y, x + y \in D\}.$$

We refer to these sets as (X, Y, D) -sets. Thus, in our density increment we maintain a subset $A \subseteq S(X, Y, D)$ whose density increases over the course of the procedure. While this initially seems worrying because there is no obvious guarantee on the size of the set $S(X, Y, D)$, it is not hard to prove that if any of X, Y, D are Fourier-pseudorandom (like in Shkredov’s proof), then the size of such a set is as expected: approximately $|X||Y||D|/|W|$.

It remains to discuss how having a more complicated pseudorandom container affects the remainder of the argument. The main difficulty to establishing bounds of the form $4^n \cdot \exp(-(\log n)^{\Omega(1)})$ (i.e., a single logarithmic improvement over Shkredov’s bound) is in establishing a version of the “sifting” lemma in this setting. The challenge is that in [Theorem 3.4](#) the resulting output rectangle depends on the density/grid norm of the function. Because we are working with (X, Y, D) -sets, when viewed as a subset of $X \times Y$ e.g., the set itself is also extremely sparse. Thus, naïvely applying the sifting lemma of [41] loses factors corresponding to relative density of $S(X, Y, D)$ within $X \times Y$. A key innovation of this paper is therefore proving a version of sifting which does not lose such factors.

2.3. Quasirandom sifting. We now discuss quasirandom sifting which is one of the primary technical innovations of the paper. The necessary quasirandom sifting statement concerns when a function has large $(2, k)$ -grid norm. In this section, we take a graph-theoretic view – if a function has large $(2, k)$ -grid norm then the corresponding graph has “excess” copies of $K_{2,k}$. We prove that this implies that then there is a density increment onto a rectangle.

We first specialize to the case of $K_{2,2}$ which is already nontrivial. In this case, we seek to analyze $f(x, y) : \Omega_1 \times \Omega_2 \rightarrow [0, 1]$ which is bounded by a pseudorandom majorant $0 \leq T(x, y) \leq 1$ (meaning that $f(x, y) \leq T(x, y)$) such that $\mathbb{E}[f] = \alpha\tau$ and $\mathbb{E}[T] = \tau$ and with

$$\mathbb{E}_{\substack{x, x' \in \Omega_1 \\ y, y' \in \Omega_2}} [f(x, y)f(x', y)f(x, y')f(x', y')] \geq (1 + \varepsilon)^4 \alpha^4 \tau^4.$$

We must specify what it means for T to be pseudorandom. Our notion of pseudorandomness will be called *combinatorial spreadness*: the density of T within any somewhat large rectangle is bounded by $(1 + \varepsilon)\tau$ ([Definition 3.1](#)).

We wish to find a large rectangle on which f has density at least $(1 + \Omega(\varepsilon))\alpha\tau$. Crucially, we want the size of the rectangle to not depend on τ , and only on α . The proof of this must necessarily rely on the pseudorandomness of T . In the setting without the pseudorandom majorant T , the proof of sifting in [\[41\]](#) picks a random (x', y') where $f(x', y') = 1$, fixes it, and then notes that $f(x, y)f(x', y)$ gives a rectangle which correlates to f .

However, the density of the rectangle itself is $O(\tau^2)$, because for a fixed (x', y') the set of x such that $f(x, y') = 1$ is $O(\tau)$ due to the presence of the majorant T , and similarly for y .

Motivated by “densification” in works of Conlon, Fox, and Zhao [\[16\]](#), we try to replace the $K_{2,2}$ counts with counts of a larger graph such that there are edges not involved in cycles of length 4. We will see later why this is useful. In order to not lose factors of α , our first step is to apply Hölder’s inequality to obtain that

$$\mathbb{E}_{\substack{x \in \Omega_1 \\ y \in \Omega_2}} [f(x, y)]^{k-1} \cdot \mathbb{E}_{\substack{x \in \Omega_1 \\ y \in \Omega_2}} \left[f(x, y) \left(\mathbb{E}_{\substack{x' \in \Omega_1 \\ y' \in \Omega_2}} f(x', y) f(x, y') f(x', y') \right)^k \right] \geq (1 + \varepsilon)^{4k} \alpha^{4k} \tau^{4k}.$$

Taking $k \approx \log(1/\alpha)/\varepsilon$ and rearranging, we find that

$$\mathbb{E}_{\substack{x \in \Omega_1 \\ y \in \Omega_2}} \left[f(x, y) \left(\mathbb{E}_{\substack{x' \in \Omega_1 \\ y' \in \Omega_2}} f(x', y) f(x, y') f(x', y') \right)^k \right] \geq (1 + \varepsilon)^{4k} \alpha^{4k} \tau^{4k} / (\alpha \cdot \tau)^{k-1} \geq (1 + \varepsilon)^{4k} \cdot \alpha^{3k+1} \tau^{3k+1}.$$

We next apply the fact that $f \leq T$ to obtain

$$\mathbb{E}_{\substack{x \in \Omega_1 \\ y \in \Omega_2}} \left[T(x, y) \left(\mathbb{E}_{\substack{x' \in \Omega_1 \\ y' \in \Omega_2}} f(x', y) f(x, y') f(x', y') \right)^k \right] \geq (1 + \varepsilon)^{4k} \cdot \alpha^{3k+1} \tau^{3k+1}.$$

As T is combinatorially spread, we find that

$$\mathbb{E}_{\substack{x \in \Omega_1 \\ y \in \Omega_2}} \left[\left(\mathbb{E}_{\substack{x' \in \Omega_1 \\ y' \in \Omega_2}} f(x', y) f(x, y') f(x', y') \right)^k \right] \geq (1 + \varepsilon)^{4k} \cdot \alpha^{3k+1} \tau^{3k} \geq (1 + \varepsilon)^{3k} \cdot \alpha^{3k} \tau^{3k}.$$

Observe that through these manipulations we have transformed a C_4 into a graph G' which is a series of k paths of length 3 which are joined at a pair of vertices. The key difference is that this graph has girth 6. By telescoping, there exists an edge e and a subgraph $H \subseteq G'$ such that the count of H is at least $(1 + \varepsilon)\alpha\tau$ times the count of $H \setminus \{e\}$; for the sake of simplicity we assume here that $H = G'$. This give that

$$\begin{aligned} & \mathbb{E}_{\substack{x, x' \in \Omega_1 \\ y, y' \in \Omega_2}} \left[f(x', y) f(x, y') f(x', y') \left(\mathbb{E}_{\substack{x'' \in \Omega_1 \\ y'' \in \Omega_2}} f(x'', y) f(x, y'') f(x'', y'') \right)^{k-1} \right] \\ & \geq (1 + \varepsilon)\alpha\tau \cdot \mathbb{E}_{\substack{x, x' \in \Omega_1 \\ y, y' \in \Omega_2}} \left[f(x, y') f(x', y') \left(\mathbb{E}_{\substack{x'' \in \Omega_1 \\ y'' \in \Omega_2}} f(x'', y) f(x, y'') f(x'', y'') \right)^{k-1} \right]. \end{aligned}$$

Let $R(x, y) = \mathbb{E}_{\substack{x'' \in \Omega_1 \\ y'' \in \Omega_2}} f(x'', y) f(x, y'') f(x'', y'')$ and observe that $R(x, y)$ has mean at least $\alpha^3 \tau^3$ while essentially always being bounded by τ^3 . In particular, observe that $\tau^{-3} R(x, y)$ is a “dense” function. Furthermore let $R_2(x', x) = \mathbb{E}_{y' \in \Omega_2} f(x, y') f(x', y')$; we similarly have that $\tau^{-2} R(x', x)$ is a dense function. This gives that

$$\mathbb{E}_{\substack{x \in \Omega_1 \\ x' \in \Omega_1 \\ y \in \Omega_2}} f(x', y) R_2(x', x) R(x, y)^{k-1} \geq (1 + \varepsilon) \alpha \tau \cdot \mathbb{E}_{\substack{x \in \Omega_1 \\ x' \in \Omega_1 \\ y \in \Omega_2}} R_2(x', x) R(x, y)^{k-1}.$$

Choosing an x for which $R_2(x', x)$ and $R(x, y)$ are appropriately dense then immediately gives the desired conclusion. The extension of the above argument to $K_{2,k}$ involves an iterative combination of Hölder’s inequality and induction on smaller graphs.

This technical manipulation of replacing f by “denser codegree” functions has been described as “densification” in works of Conlon, Fox, and Zhao [16] which proved various graph counting lemmas in the presence of a pseudorandom majorant. An extension of these results to hypergraph was given in work of the same authors [17] to prove a relative Szemerédi theorem; this in turn simplified the proof of a key ingredient to the celebrated Green–Tao theorem [33]. However these results, if applied directly, would lose various factors of α in the density increment which is unacceptable. Intuitively, one way we avoid losing these factors of α is by applying Hölder’s inequality as above instead of Cauchy-Schwarz when doing the densification procedure.

There are several remarks in order for the general case. First observe that the results of Conlon, Fox, and Zhao [16] and more generally counting lemmas with respect to sparse graphs have been phrased with forcing T to be pseudorandom with respect to the cut norm (see Definition A.6). While this would be acceptable for $4^n \cdot \exp(-(\log n)^{\Omega(1)})$ -bounds, it turns out to be insufficient for quasipolynomial bounds. An important observation is that for our proof all that is required is that T is bounded “on large rectangles”; e.g. $\mathbb{E}[A(x)T(x, y)B(y)] \leq (1 + \varepsilon)\tau \cdot \mathbb{E}[A(x)B(y)] + \gamma$ for all 1-bounded functions A and B . This pseudorandomness criterion will prove to be quantitatively superior. We remark however for Theorem 1.6, we only require the cut norm version of the statement (which is satisfied for quasirandom groups). Furthermore, for the quasirandom group application we will not require essentially any of the remaining tools in the sketch. In particular, the use of the quasirandom group assumption allows one to avoid any pseudorandomization in this setting.

One critical feature we have not discussed to this point is that the rectangles we obtain will naturally have asymmetric sizes – note in the statement of Theorem 3.5 that the size of the D side does not have any dependence on k , and is instead just $2^{-O(\log(1/\alpha)^2)}$. To see why this is sensible, let us first consider the simpler case where there is no pseudorandom majorant and f has density α .

Suppose that

$$\mathbb{E} \left[\prod_{i \in [k]} f(x_1, y_i) f(x_2, y_i) \right] \geq (1 + \varepsilon)^{2k} \alpha^{2k}.$$

We will argue informally that there are $g_1(x)$ and $g_2(y)$ such that

$$\begin{aligned} \mathbb{E}[g_1(x) f(x, y) g_2(y)] &\geq (1 + \varepsilon/2) \cdot \alpha \cdot \mathbb{E}[g_1] \cdot \mathbb{E}[g_2] \\ \mathbb{E}[g_1] &\geq \Omega(\varepsilon \alpha)^{O(k)} \\ \mathbb{E}[g_2] &\geq \Omega(\varepsilon \alpha)^{O(1)}. \end{aligned}$$

Think of how g_1 and g_2 are constructed in the proof of sifting (Theorem 3.4), say when the edge (x_1, y_1) is removed. Then g_1 is formed as the intersection of neighborhoods of $k-1$ vertices $y_2, \dots, y_k$, and g_2 is the neighborhood of the vertex x_2 . Thus, if f has density α , then one expects that g_1 has density at least $\alpha^{O(k)}$ and g_2 has density at least $\alpha^{O(1)}$. Observe in particular that $\mathbb{E}[g_2]$ has size independent of k . This has previously gone unnoticed and will be crucial for our analysis.

2.4. Asymmetric increment. We now return to the initial Hölder manipulation in the case where $A \subseteq S(X, Y, D)$ (equivalently, $\mathbb{1}_A(x, y) \leq \mathbb{1}_X(x)\mathbb{1}_Y(y)\mathbb{1}_D(x+y)$) and $\mathbb{E}[f(x, y)] = \alpha \cdot \delta_X \delta_Y \delta_D$, where $\delta_X, \delta_Y, \delta_D$ are the densities of X, Y, D respectively. We have not currently specified the pseudorandomness conditions we will impose on X, Y , and D , but they will guarantee that $\mathbb{E}[\mathbb{1}_X(x)\mathbb{1}_Y(y)\mathbb{1}_D(x+y)] \approx \delta_X \delta_Y \delta_D$.

We begin with the Hölder manipulation as in [Section 2.2](#); tracking various support functions we find that

$$\mathbb{E}_{x,y \in \mathbb{F}_2^n} \mathbb{1}_A(x, y) \left(\mathbb{E}_{d \in \mathbb{F}_2^n} \mathbb{1}_A(x, x+d) f_A(y+d, y) \right)^k \geq (c\alpha)^{2k+1} \cdot \delta_X^{k+1} \delta_Y^{k+1} \delta_D^{k+1}$$

where c is a sufficiently small absolute constant. As k is at least a large constant times $\log(1/\alpha)$, we may (at the cost of changing c) instead consider

$$\mathbb{E}_{x,y \in \mathbb{F}_2^n} \mathbb{1}_D(x+y) \left(\mathbb{E}_{d \in \mathbb{F}_2^n} \mathbb{1}_A(x, x+d) f_A(y+d, y) \right)^k \geq (c\alpha)^{2k+1} \cdot \delta_X^{k+1} \delta_Y^{k+1} \delta_D^{k+1}.$$

We now come to one of the main gambits in this paper. In the work of Shkredov, one maintains that the X and Y are Fourier pseudorandom. One may attempt to maintain X, Y, D which are Fourier pseudorandom and this is possible; however the associated pseudorandomization procedure is inherently far too lossy for quasipolynomial bounds. The trick is to give up D (and a factor of δ_D). This is possible if one takes $k \approx \log(1/(\alpha\delta_D))$.

We now apply the appropriate quasirandom sifting theorem ([Theorem 3.5](#)) and we find that there exists g_i such that $g_1(x) \leq \mathbb{1}_X(x)$, $g_2(d) \leq \mathbb{1}_D(d)$, and

$$\begin{aligned} \mathbb{E}[g_1(x)\mathbb{1}_A(x, x+d)g_2(d)] &\geq (1 + \Omega(1)) \cdot \alpha \cdot \delta_Y \cdot \mathbb{E}[g_1] \cdot \mathbb{E}[g_2] \\ \mathbb{E}[g_1(x)] &\geq \delta_X \cdot \exp(-\log(1/(\alpha\delta_D))^{O(1)}) \\ \mathbb{E}[g_2(d)] &\geq \delta_D \cdot \exp(-\log(1/\alpha)^{O(1)}) \end{aligned}$$

or an analogous statement for $\mathbb{1}_A(y+d, y)$. We are crucially using here the asymmetric nature of the density increment. In general, in our argument D will end up being much denser than X, Y .

Heuristically, if pseudorandomization is not too costly, we would have that

$$\begin{aligned} \delta'_D &\leftarrow \delta_D \cdot \exp(-\log(1/\alpha)^{O(1)}), \\ \delta'_X &\leftarrow \delta_X \cdot \exp(-\log(1/(\alpha\delta_D))^{O(1)}), \text{ and} \\ \delta'_Y &\leftarrow \delta_Y \cdot \exp(-\log(1/(\alpha\delta_D))^{O(1)}). \end{aligned}$$

As we will have only $\log(1/\alpha)$ iterations, the densities obtained will always be quasipolynomial. Furthermore, observe that this “dropping” of the D indicator will mean that we will require no pseudorandomness conditions on D . (In our later discussion of the general abelian case, we will need to reimpose such a condition on D .) This asymmetry and the realization to “drop” the indicator of D (e.g. not attempt to account for its density throughout) are crucial ingredients in this work.

2.5. Pseudorandomization into spread components. The one remaining consideration is how to decompose the “dense” set which is output by the previous step and pass to a rectangle where the sides are suitably “spread” (the pseudorandomness property we maintain on X and Y). The work of Shkredov [\[62\]](#) relies on an L^2 -energy increment argument. This however is rather costly; in order to guarantee that the underlying “sides” have Fourier coefficients bounded by δ one needs to pass to a codimension $\delta^{-O(1)}$ set. Furthermore observe that to this point in the argument we have made no use of the tools of almost periodicity which are crucial in the work of Kelley and Meka [\[42\]](#) as well as much of the recent work on Roth’s theorem.

The trick therefore is to relax the notion of pseudorandomness to being “upper bounded” against rectangles. More precisely, we will ensure that if $X : \mathbb{F}_2^n \rightarrow \{0, 1\}$ then for all $A, B : \mathbb{F}_2^n \rightarrow [0, 1]$ that

$$\mathbb{E}[A(x)X(x+y)B(y)] \leq (1 + \varepsilon) \cdot \mathbb{E}[X] \cdot \mathbb{E}[A] \cdot \mathbb{E}[B] + \gamma.$$

The main consequence of the methods of Kelley–Meka [42] (essentially [42, Theorem 4.10]), is that if X is not upper bounded in such a manner, then X has increased density (by a $1 + \Omega(\varepsilon)$ factor) on a subspace of codimension $O_\varepsilon(\log(1/\gamma)^{O(1)})$. This motivates our definition of an algebraically spread set X (Definition 5.1): a set X is *algebraically spread* if it does not admit a density increment onto an affine subspace of small codimension $r = O_\varepsilon(\log(1/\gamma)^{O(1)})$.

This suggests the idea of using a “spread decomposition” to do the pseudorandomization procedure to X and Y . Given a set X , one may either say X is spread and output X or repeatedly pass X to a subspace where it has increased density. This procedure allows one to decompose X into spread pieces and is closely related to various spread decompositions which have notably been used in the context of the sunflower conjecture [2] and by Kupavskii and Zakharov [43] in the context of forbidden intersection problems. In particular, this latter work crucially relies on a decomposition into spread pieces.

The major difficulty in this situation, however, is finding a decomposition which makes both X and Y spread at the same time, while also not causing the density of D and the set A to drop significantly. The ultimate proof is essentially a certain recursive ping-pong; one ensures that X is first so spread that various manipulations on Y will not increase the density of X . Then one splits Y and argues that almost all pieces that X splits into do not drop in density, and are hence still spread. The result of this is a decomposition of $X \times Y$ into subrectangles, where at least half of them have both sides spread. Thus, recursively iterating on the rectangles that are not spread, one decomposes $X \times Y$ into a set of “spread rectangles”, and this allows one to conclude the proof. The details are rather algorithmic/recursive in nature; the main issue whenever one is dealing with spread decomposition in our context is that while one has an upper bound on the density of X , it may drop dramatically on a small fraction of the set. We also remark that our proof could even be used in the case of Shkredov [62] and guarantee a Fourier pseudorandomness condition (with a similar dimension drop as in [62]) while completely avoiding the L^2 -energy increment strategy central to this work.

2.6. Bohr set adaptations and additional pseudorandomization. We now briefly discuss the crucial technical changes required for adapting the proof to general groups. The first technical issue is with the use of the parameterization of (x, y) , $(-y - z, y)$, $(x, -x - z)$ in the context of Bohr sets. (Recall such a parameterization arises from the change of variable $z \rightarrow -x - y - z$ in the definition of a corner.) Observe that if one restricts $x \in B_1$ and $y \in B_2$ where B_1, B_2 are Bohr sets and B_2 is more “narrow” than B_1 , then essentially any restriction on the range of z causes the three coordinates to not “live” on the same range. In this context, we instead consider

$$(x + x', y + y'), (x - y' + z', y + y'), (x + x', y - x' + z')$$

where $x \sim B_1$, $y \sim B_2$, $x' \sim B_3$, $y' \sim B_4$ and $z' \sim B_5$. This essentially corresponds to considering corners in a “narrow” box around (x, y) , and note that now each coordinate individually ranges over the whole domain $B_1 \times B_2$ in an essentially uniform way. The setup of the initial argument is now rather more delicate; one needs to find (x, y) (which we call a *great pair*) such that the three rectangles induced by $(x + x', y + y')$, $(x - y' + z', y + y')$, $(x + x', y - x' + z')$ all have sufficient density and spreadness. This is quite technical, but a certain relatively routine but lengthy argument suffices.

The final technical hurdle is rather more subtle. Observe in the case of Bohr sets that the size of the container is given by $\mathbb{E}_{x \sim B_1} \mathbb{1}_X(x) \mathbb{1}_Y(y) \mathbb{1}_D(x + y)$. We want to enforce pseudorandomness

notions on X, Y, D that guarantee that the size of the container is close to expected. A natural notion to enforce on X, Y is the generalization of algebraic spreadness to Bohr sets: there is no density increment onto a smaller Bohr set of slightly higher rank and smaller radius. Using that X and Y are spread one can prove that this is quite close to $(\mathbb{E}_{x \sim B_1} \mathbb{1}_X(x) \mathbb{1}_D(x+y)) \cdot (\mathbb{E}_{y \in B_2} \mathbb{1}_Y(y))$.

There is no guarantee, however, that $\mathbb{E}_{x \sim B_1} \mathbb{1}_X(x) \mathbb{1}_D(x+y)$ is close to $\mathbb{E}_{x \sim B_1} \mathbb{1}_X(x) \cdot \mathbb{E}_{z \sim B_1} \mathbb{1}_D(z)$. (This is unlike the finite field setting where the initial term factors.) The failure case occurs precisely when D places “most” of its mass where X dips below its mean. Essentially, the minimal condition one can use to obtain our desired guarantee is that D is typically not below its mean; for this the rather weak bound of the form $\mathbb{E}_{z \sim B_1} |\mathbb{E}_{z' \sim B_3} \mathbb{1}_D(z+z') - \delta_D| \leq \varepsilon \cdot \delta_D$ suffices. Observe that if our pseudorandomization procedure could also make D spread this would be sufficient. However, this seems technically quite challenging. Instead, we observe that if this weak ℓ_1 -condition is not true, then $\mathbb{E}_{z \sim B_1} |\mathbb{E}_{z' \sim B_3} \mathbb{1}_D(z+z') - \delta_D| \leq \varepsilon \cdot \delta_D$ fails on a constant fraction of translates of B_3 . This combined with a certain log-potential analysis (see [Lemma 6.21](#)) and an additional recursive layer allows one to handle the necessary pseudorandomization procedure. It appears a rather interesting question to understand what the limit of these “efficient pseudorandomization” procedures are, as these are rather different than the more standard L^2 -based strategies used in the literature.

We remark that much of our analysis in this section is in terms of arithmetic grid norms following the work of Milićević [\[49\]](#) on skew corners.

2.7. Notation. We use standard asymptotic notation for functions on $\mathbb{N}$ throughout. Given functions $f = f(x)$ and $g = g(x)$, we write $f = O(g)$, or $g = \Omega(f)$, to mean that there is a constant C such that $|f(x)| \leq Cg(x)$ for sufficiently large x . Additionally, we write $f = \Theta(g)$ to mean $f = O(g)$ and $g = O(f)$. Subscripts indicate dependence on parameters. Furthermore given a function $f : \mathbb{Z} \rightarrow \mathbb{C}$, we say f is 1-bounded if $\sup_{x \in \mathbb{Z}} |f(x)| \leq 1$. For a finite set Ω and function $f : \Omega \rightarrow \mathbb{C}$, we denote the k -norm of f by

$$\|f\|_k = \left(\mathbb{E}_{x \in \Omega} |f(x)|^k \right)^{1/k}.$$

For a finite abelian group G , we denote the convolution of two functions $f, g : G \rightarrow \mathbb{R}$ by

$$(f * g)(x) = \mathbb{E}_{y \in G} [f(y)g(x-y)].$$

Parameters: α will always be related to the density of the set within the container. Throughout the proof, ε with subscripts, etc. will all denote absolute constants independent of the size of the group G . K, k, ℓ will be used as the size of the grid norms we consider, and will all be $O(\log(1/\alpha)^{O(1)})$. γ will denote a pseudorandomness parameter, and will be $\exp(-\log(1/\alpha)^{O(1)})$. r will be related to the dimension up to which our sets are algebraically spread, and will be $O(\log(1/\alpha)^{O(1)})$.

δ and $\delta_X, \delta_Y, \delta_D$ will denote densities of the sets inducing the container within the ambient subspace. These will all be at least $\exp(-\log(1/\alpha)^{O(1)})$.

In the section on Bohr sets, η will be used to denote the ratio between radii of regular Bohr sets with the same set of frequencies. In particular, we will often consider $B_1 \supseteq B_2 \supseteq B_3 \supseteq \dots$ where B_i are all regular and the radii (denoted as r_i) satisfy $r_{i+1}/r_i \leq \eta$.

3. COMBINATORIAL SPREADNESS AND RELATIVE SIFTING

Throughout our analysis in this paper the sets Ω_i are finite (and therefore have a well-defined uniform measure). We first define the notion of combinatorial spreadness which will be used throughout this section.

Definition 3.1 (Combinatorial spreadness). Let Ω_1, Ω_2 be sets. A subset $T \subseteq \Omega_1 \times \Omega_2$ is (τ, γ) -combinatorially spread if for all functions $f : \Omega_1 \rightarrow [0, 1]$ and $g : \Omega_2 \rightarrow [0, 1]$ it holds that

$$\mathbb{E}_{x \in \Omega_1, y \in \Omega_2} [f(x)g(y)\mathbb{1}_T(x, y)] \leq \tau \mathbb{E}_{x \in \Omega_1} [f(x)] \mathbb{E}_{y \in \Omega_2} [g(y)] + \gamma.$$

Intuitively, one should think that $\mathbb{E}_{x \in \Omega_1, y \in \Omega_2} [\mathbb{1}_T(x, y)]$ is approximately τ . Note we will occasionally refer to a boolean function as (τ, γ) -combinatorially spread if it is the indicator function of a (τ, γ) -combinatorially spread set. If we restrict Definition 3.1 to functions f, g which are $\{0, 1\}$ -valued, it is equivalent to think of combinatorial spreadness as mandating that T has no significant density increment on rectangles of density at least $\Omega(\gamma)$. Later, we will define a version of combinatorial spreadness which is asymmetric; namely, we will have different density requirements on the functions f, g , rather than a single density requirement on the rectangle $f(x)g(y)$. We first note a counting lemma for our notion of combinatorial spreadness.

Lemma 3.2 (Counting lemma). Let Ω_1, Ω_2 be sets, $T : \Omega_1 \times \Omega_2 \rightarrow \{0, 1\}$ be (τ, γ) -combinatorially spread. Let $G = ([k] \cup [\ell], E)$ be a bipartite graph with edge $(i^*, j^*) \in E$. For $(i, j) \in E \setminus \{(i^*, j^*)\}$ let $f_{ij} : \Omega_1 \times \Omega_2 \rightarrow [0, 1]$ be nonnegative 1-bounded functions. Then

$$\mathbb{E}_{\substack{x_1, \dots, x_k \in \Omega_1 \\ y_1, \dots, y_\ell \in \Omega_2}} \left[T(x_{i^*}, y_{j^*}) \prod_{(i, j) \in E \setminus \{(i^*, j^*)\}} f_{ij}(x_i, y_j) \right] \leq \tau \mathbb{E}_{\substack{x_1, \dots, x_k \in \Omega_1 \\ y_1, \dots, y_\ell \in \Omega_2}} \left[\prod_{(i, j) \in E \setminus \{(i^*, j^*)\}} f_{ij}(x_i, y_j) \right] + \gamma.$$

Proof. We can write

$$\begin{aligned} & \mathbb{E}_{\substack{x_1, \dots, x_k \in \Omega_1 \\ y_1, \dots, y_\ell \in \Omega_2}} \left[T(x_{i^*}, y_{j^*}) \prod_{(i, j) \in E \setminus \{(i^*, j^*)\}} f_{ij}(x_i, y_j) \right] \\ &= \mathbb{E}_{\substack{x_i : i \in [k] \setminus \{i^*\} \\ y_j : j \in [\ell] \setminus \{j^*\}}} \left[\prod_{\substack{(i, j) \in E \\ i \neq i^*, j \neq j^*}} f_{ij}(x_i, y_j) \mathbb{E}_{x_{i^*}, y_{j^*}} \left[T(x_{i^*}, y_{j^*}) \prod_{j : (i^*, j) \in E} f_{i^*j}(x_{i^*}, y_j) \prod_{i : (i, j^*) \in E} f_{ij^*}(x_i, y_{j^*}) \right] \right] \\ &\leq \mathbb{E}_{\substack{x_i : i \in [k] \setminus \{i^*\} \\ y_j : j \in [\ell] \setminus \{j^*\}}} \left[\prod_{\substack{(i, j) \in E \\ i \neq i^*, j \neq j^*}} f_{ij}(x_i, y_j) \left(\tau \mathbb{E}_{x_{i^*}, y_{j^*}} \left[\prod_{j : (i^*, j) \in E} f_{i^*j}(x_{i^*}, y_j) \prod_{i : (i, j^*) \in E} f_{ij^*}(x_i, y_{j^*}) \right] + \gamma \right) \right] \\ &\leq \tau \mathbb{E}_{\substack{x_1, \dots, x_k \in \Omega_1 \\ y_1, \dots, y_\ell \in \Omega_2}} \left[\prod_{(i, j) \in E \setminus \{(i^*, j^*)\}} f_{ij}(x_i, y_j) \right] + \gamma, \end{aligned}$$

where the first inequality uses (τ, γ) -combinatorial spreadness of T , and the second uses that the function f_{ij} are 1-bounded. $\square$

We first require the following lemma which states that if one correlates with a bounded product function then one may extract a correlation with a pair of sets. A version of this statement appears as [41, Claim 4.6].

Lemma 3.3. Let $\tau > 0$, $A : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$, and $f_i : \Omega_i \rightarrow [0, 1]$ be such that

$$\mathbb{E}_{x, y} [f_1(x)f_2(y)A(x, y)] \geq \tau \cdot \mathbb{E}_{x, y} [f_1(x)f_2(y)].$$

Then there exist $g_i : \Omega_i \rightarrow \{0, 1\}$ with

$$\mathbb{E}_{x, y} [g_1(x)g_2(y)A(x, y)] \geq \tau \cdot \mathbb{E}_{x, y} [g_1(x)g_2(y)]$$

with $\mathbb{E}[g_i(x)] \geq \mathbb{E}[f_i(x)]/2$.

Proof. Consider $\vec{\alpha} = (\alpha_1, \dots, \alpha_\ell) \in [0, 1]^\ell$ with $\sum \alpha_i = k$. We will show that $\vec{\alpha}$ may be written as a convex combination of $\vec{v} \in \{0, 1\}^\ell$ with $\|\vec{v}\|_1 \in \{[k], \lceil k \rceil\}$. Applying this claim to $\alpha = (f_i(x) : x \in \Omega_i)$, we may decompose

$$f_i(x) = \sum_{j=1}^{t_i} w_j g_{i,j}(x), \quad (3.1)$$

where $g_{i,j} : \Omega_i \rightarrow \{0, 1\}$ with

$$\mathbb{E}[g_{i,j}(x)] \in \left\{ \frac{1}{|\Omega_i|} \cdot \left\lfloor |\Omega_i| \cdot \mathbb{E}[f_i(x)] \right\rfloor, \frac{1}{|\Omega_i|} \left\lceil |\Omega_i| \cdot \mathbb{E}[f_i(x)] \right\rceil \right\}$$

and $w_j > 0$. If $\lfloor \mathbb{E}[f_i(x)] \cdot |\Omega_i| \rfloor = 0$, then we may drop terms with $\mathbb{E}[g_{i,j}(x)] = 0$; note $\mathbb{E}[g_{i,j}(x)] \geq \mathbb{E}[f_i(x)]$ for the remaining terms. Otherwise $\mathbb{E}[f_i(x)] \geq 1/|\Omega_i|$, and we deduce $\mathbb{E}[g_{i,j}(x)] \geq \mathbb{E}[f_i(x)]/2$ for all j . Substituting (3.1) into our original assumption, we find

$$\sum_{j=1}^{t_1} \sum_{j'=1}^{t_2} w_j w_{j'} \mathbb{E}[(A(x, y) - \tau) g_{1,j}(x) g_{2,j'}(y)] \geq 0.$$

At least one of the summands is nonnegative and the result follows.

To prove the initial claim, we proceed by induction on the number of coordinates ℓ which are strictly between zero and one. The case when $\ell = 1$ is trivial. If there are coordinates $\alpha_i + \alpha_j \leq 1$, then by writing

$$(\alpha_i, \alpha_j) = \frac{\alpha_i}{\alpha_i + \alpha_j}(\alpha_i + \alpha_j, 0) + \frac{\alpha_j}{\alpha_i + \alpha_j}(0, \alpha_i + \alpha_j)$$

we may proceed by induction downward. Else if $1 < \alpha_i + \alpha_j \leq 2$, then

$$(\alpha_i, \alpha_j) = \frac{1 - \alpha_i}{2 - \alpha_i - \alpha_j}(\alpha_i + \alpha_j - 1, 1) + \frac{1 - \alpha_j}{2 - \alpha_i - \alpha_j}(1, \alpha_i + \alpha_j - 1),$$

and again we may proceed by induction downward. $\square$

Finally, we require the basic (non-relative) version of sifting. For completeness (and to provide a slightly better bound than is in the literature, e.g., [41, Lemma 4.7]), we provide a proof.

Theorem 3.4 (Sifting). *Let $f : \Omega_1 \times \Omega_2 \rightarrow [0, 1]$ satisfy that $\|f\|_{G(k,\ell)} \geq \alpha$. For any $\varepsilon > 0$, there are functions $g_1 : \Omega_1 \rightarrow [0, 1]$ and $g_2 : \Omega_2 \rightarrow [0, 1]$ such that*

$$\mathbb{E}_{x \in \Omega_1, y \in \Omega_2} [f(x, y) g_1(x) g_2(y)] \geq (1 - \varepsilon) \alpha \mathbb{E}_{x \in \Omega_1} [g_1(x)] \mathbb{E}_{y \in \Omega_2} [g_2(y)],$$

and

$$\mathbb{E}_{x \in \Omega_1} [g_1(x)] \mathbb{E}_{y \in \Omega_2} [g_2(y)] \geq \varepsilon \alpha^{O(k+\ell)}.$$

Proof. If $\|f\|_{G(i,j)} \geq \alpha$ for any positive integers $i \leq k$ and $j \leq \ell$ with $i + j < k + \ell$, then the result follows by induction. Assume otherwise for the remainder of the argument. Let $G^{(0)}, G^{(1)}, \dots, G^{(k+\ell-1)}$ be an increasing sequence of graphs (each adding one edge to the previous) where $G^{(0)}$ is $K_{k-1, \ell-1}$ and $G^{(k+\ell-1)}$ is $K_{k, \ell}$, each with vertex set $V = [k] \cup [\ell]$. Then there is some $0 \leq t \leq k + \ell - 2$ such that if (i^*, j^*) is the edge in $G^{(t+1)} \setminus G^{(t)}$, then

$$\mathbb{E}_{\substack{x_1, \dots, x_k \in \Omega_1 \\ y_1, \dots, y_\ell \in \Omega_2}} \left[f(x_{i^*}, y_{j^*}) \prod_{(i,j) \in E(G^{(t)})} f(x_i, y_j) \right] \geq \alpha \mathbb{E}_{\substack{x_1, \dots, x_k \in \Omega_1 \\ y_1, \dots, y_\ell \in \Omega_2}} \left[\prod_{(i,j) \in E(G^{(t)})} f(x_i, y_j) \right],$$

which can be rearranged to get

$$\begin{aligned} \mathbb{E}_{\substack{x_1, \dots, x_k \in \Omega_1 \\ y_1, \dots, y_\ell \in \Omega_2}} \left[(f(x_{i^*}, y_{j^*}) - (1 - \varepsilon)\alpha) \prod_{(i,j) \in E(G^{(t)})} f(x_i, y_j) \right] &\geq \varepsilon \alpha \mathbb{E}_{\substack{x_1, \dots, x_k \in \Omega_1 \\ y_1, \dots, y_\ell \in \Omega_2}} \left[\prod_{(i,j) \in E(G^{(t)})} f(x_i, y_j) \right] \\ &\geq \varepsilon \alpha \|f\|_{G(k,\ell)}^{k\ell} \geq \varepsilon \alpha^{k\ell+1}. \end{aligned}$$

Let $\overline{E} \subseteq E(G^{(t)})$ be the edges not involving either of i^* or j^* . Note that $\overline{E}$ contains a $K_{k-2, \ell-2}$ as a subgraph. Thus, if $\mathcal{E}$ is the event over $\{x_i : i \in [k] \setminus \{i^*\}, y_j : j \in [\ell] \setminus \{j^*\}\}$ that $\prod_{(i,j) \in \overline{E}} f(x_i, y_j) \neq 0$, then by the assumption in the first line of the proof, we know that $\Pr[\mathcal{E}] \leq \|f\|_{G(k-2, \ell-2)}^{(k-2)(\ell-2)} \leq \alpha^{(k-2)(\ell-2)}$. Thus,

$$\begin{aligned} \varepsilon \alpha^{k\ell+1} &\leq \mathbb{E}_{\substack{x_1, \dots, x_k \in \Omega_1 \\ y_1, \dots, y_\ell \in \Omega_2}} \left[(f(x_{i^*}, y_{j^*}) - (1 - \varepsilon)\alpha) \prod_{(i,j) \in E(G^{(t)})} f(x_i, y_j) \right] \\ &= \mathbb{E}_{x_{i^*}, y_{j^*}} \left[(f(x_{i^*}, y_{j^*}) - (1 - \varepsilon)\alpha) \cdot \right. \\ &\quad \left. \mathbb{E}_{\substack{x_i : i \in [k] \setminus \{i^*\} \\ y_j : j \in [\ell] \setminus \{j^*\}}} \left[\prod_{j : (i^*, j) \in E(G^{(t)})} f(x_{i^*}, y_j) \prod_{i : (i, j^*) \in E(G^{(t)})} f(x_i, y_{j^*}) \prod_{(i,j) \in \overline{E}} f(x_i, y_j) \right] \right] \\ &\leq \Pr[\mathcal{E}] \mathbb{E}_{x_{i^*}, y_{j^*}} \left[(f(x_{i^*}, y_{j^*}) - (1 - \varepsilon)\alpha) \cdot \right. \\ &\quad \left. \mathbb{E}_{\substack{x_i : i \in [k] \setminus \{i^*\} \\ y_j : j \in [\ell] \setminus \{j^*\}}} \left[\prod_{j : (i^*, j) \in E(G^{(t)})} f(x_{i^*}, y_j) \prod_{i : (i, j^*) \in E(G^{(t)})} f(x_i, y_{j^*}) \mid \mathcal{E} \right] \right]. \end{aligned}$$

Thus, there is a choice of $\{x_i : i \in [k] \setminus \{i^*\}, y_j : j \in [\ell] \setminus \{j^*\}\}$ that satisfy $\mathcal{E}$, such that

$$\begin{aligned} \mathbb{E}_{x_{i^*}, y_{j^*}} \left[(f(x_{i^*}, y_{j^*}) - (1 - \varepsilon)\alpha) \prod_{j : (i^*, j) \in E(G^{(t)})} f(x_{i^*}, y_j) \prod_{i : (i, j^*) \in E(G^{(t)})} f(x_i, y_{j^*}) \right] &\geq \Pr[\mathcal{E}]^{-1} \varepsilon \alpha^{k\ell+1} \\ &\geq \varepsilon \alpha^{O(k+\ell)}. \end{aligned}$$

Now we can define $g_1(x) = \prod_{j : (i^*, j) \in E(G^{(t)})} f(x, y_j)$ and $g_2(y) = \prod_{i : (i, j^*) \in E(G^{(t)})} f(x_i, y)$, so that the above equation gives

$$\mathbb{E}_{x,y} [f(x, y) g_1(x) g_2(y)] \geq (1 - \varepsilon)\alpha \mathbb{E}_x [g_1(x)] \mathbb{E}_y [g_2(y)] + \varepsilon \alpha^{O(k+\ell)}.$$

This implies the first conclusion. The second conclusion follows because

$$\mathbb{E}_x [g_1(x)] \mathbb{E}_y [g_2(y)] \geq \mathbb{E}_{x,y} [f(x, y) g_1(x) g_2(y)] \geq \varepsilon \alpha^{O(k+\ell)}. \quad \square$$

Now we are ready to state our main relative sifting statement. There are two elements to note. The first is that the function f is supported on a pseudorandom set T . This allows the expectations of the sifted functions to have *no* dependence on the parameter τ , which one should morally view as the density of T in its universe. Such a conclusion would be impossible to obtain via the standard approach to sifting. Secondly, these resulting functions are asymmetric; only one of them has a dependence on a parameter k . Somewhat surprisingly, leveraging this asymmetry will eventually be critical in our application.

Theorem 3.5 (Relative sifting). *Let $\alpha, \varepsilon, \gamma, \tau \in (0, 1)$ be parameters and k a positive integer, satisfying that*

$$\gamma \leq (\alpha\tau)^{O(\varepsilon^{-2}k \log(1/\alpha)^2 + \varepsilon^{-1}k \log(1/\tau))}.$$

Then the following holds.

Let $T \subseteq \Omega_1 \times \Omega_2$ be (τ, γ) -combinatorially spread, and let $f : \Omega_1 \times \Omega_2 \rightarrow [0, 1]$ be a function supported on T . Suppose that

$$\|f\|_{G(2,k)}^{2k} = \mathbb{E}_{\substack{x_1, x_2 \in \Omega_1 \\ y_1, \dots, y_k \in \Omega_2}} \left[\prod_{i=1}^k f(x_1, y_i) f(x_2, y_i) \right] \geq \alpha^{2k} \tau^{2k}.$$

Then there are functions $g_1 : \Omega_1 \rightarrow [0, 1]$ and $g_2 : \Omega_2 \rightarrow [0, 1]$ such that

$$\mathbb{E}_{x \in \Omega_1, y \in \Omega_2} [f(x, y) g_1(x) g_2(y)] \geq (1 - \varepsilon) \alpha \tau \mathbb{E}_{x \in \Omega_1} [g_1(x)] \mathbb{E}_{y \in \Omega_2} [g_2(y)]$$

and

$$\mathbb{E}_{x \in \Omega_1} [g_1(x)] \geq (\varepsilon \alpha / 2)^{O(\varepsilon^{-1} k^2 \log(1/\alpha))} \quad \text{and} \quad \mathbb{E}_{y \in \Omega_2} [g_2(y)] \geq (\varepsilon \alpha / 2)^{O(\varepsilon^{-1} \log(1/\alpha))}.$$

Remark. In this work we only prove relative sifting for $(2, k)$ norms. It may be of interest to prove relative sifting for (ℓ, k) norms where $\gamma \leq e^{-(\varepsilon^{-1} \log(1/\alpha) k \ell)^{O(1)}}$ is sufficient quasi-randomness and the associated sides have sizes $e^{-(\varepsilon^{-1} \log(1/\alpha) k)^{O(1)}}$, $e^{-(\varepsilon^{-1} \log(1/\alpha) \ell)^{O(1)}}$.

Proof. We proceed by induction on k . Define the function $F : \Omega_1 \times \Omega_1 \rightarrow [0, 1]$ as

$$F(x_1, x_2) := \mathbb{E}_{y_1, \dots, y_{k-1} \in \Omega_2} \left[\prod_{i=1}^{k-1} f(x_1, y_i) f(x_2, y_i) \right].$$

By this definition, we can express the $(2, k)$ -grid norm of f as

$$\|f\|_{G(2,k)}^{2k} = \mathbb{E}_{\substack{x_1, x_2 \in \Omega_1 \\ y \in \Omega_2}} [F(x_1, x_2) f(x_1, y) f(x_2, y)].$$

If $\mathbb{E}_{x \in \Omega_1, y \in \Omega_2} f(x, y) \geq (1 - \varepsilon) \alpha \tau$ then we are done. Otherwise, let $\ell = 100 \lceil \log(1/\alpha) \varepsilon^{-1} \rceil$ and apply Hölder's inequality to get

$$\begin{aligned} \alpha^{2\ell k} \tau^{2\ell k} &\leq \left(\mathbb{E}_{\substack{x_1, x_2 \in \Omega_1 \\ y \in \Omega_2}} [F(x_1, x_2) f(x_1, y) f(x_2, y)] \right)^\ell \\ &\leq \left(\mathbb{E}_{x_2 \in \Omega_1, y \in \Omega_2} f(x_2, y) \right)^{\ell-1} \left(\mathbb{E}_{x_2 \in \Omega_1, y \in \Omega_2} f(x_2, y) \left(\mathbb{E}_{x_1 \in \Omega_1} F(x_1, x_2) f(x_1, y) \right)^\ell \right) \\ &\leq ((1 - \varepsilon) \alpha \tau)^{\ell-1} \left(\mathbb{E}_{x_2 \in \Omega_1, y \in \Omega_2} f(x_2, y) \left(\mathbb{E}_{x_1 \in \Omega_1} F(x_1, x_2) f(x_1, y) \right)^\ell \right). \end{aligned}$$

Using $f \leq T$ and applying the (τ, γ) -combinatorial spreadness of T gives:

$$\begin{aligned} (1 + \varepsilon)^{\ell-1} \alpha^{(2k-1)\ell+1} \tau^{(2k-1)\ell+1} &\leq \mathbb{E}_{x_2 \in \Omega_1, y \in \Omega_2} f(x_2, y) \left(\mathbb{E}_{x_1 \in \Omega_1} F(x_1, x_2) f(x_1, y) \right)^\ell \\ &\leq \tau \mathbb{E}_{x_2 \in \Omega_1, y \in \Omega_2} \left(\mathbb{E}_{x_1 \in \Omega_1} F(x_1, x_2) f(x_1, y) \right)^\ell + \gamma. \end{aligned}$$

For our choice of γ and ℓ , we conclude that

$$\begin{aligned}
(1 + \varepsilon/2)^\ell \alpha^{(2k-1)\ell} \tau^{(2k-1)\ell} &\leq \mathbb{E}_{x_2 \in \Omega_1, y \in \Omega_2} \left(\mathbb{E}_{x_1 \in \Omega_1} F(x_1, x_2) f(x_1, y) \right)^\ell \\
&= \mathbb{E}_{x_1^{(1)}, \dots, x_1^{(\ell)} \in \Omega_1} \left(\mathbb{E}_{x_2 \in \Omega_1} \prod_{i=1}^\ell F(x_1^{(i)}, x_2) \right) \left(\mathbb{E}_{y \in \Omega_2} \prod_{i=1}^\ell f(x_1^{(i)}, y) \right) \\
&\leq \left(\mathbb{E}_{x_1^{(1)}, \dots, x_1^{(\ell)} \in \Omega_1} \mathbb{E}_{y \in \Omega_2} \prod_{i=1}^\ell f(x_1^{(i)}, y) \right)^{\frac{\ell-1}{\ell}} \\
&\quad \cdot \left(\mathbb{E}_{x_1^{(1)}, \dots, x_1^{(\ell)} \in \Omega_1} \left(\mathbb{E}_{y \in \Omega_2} \prod_{i=1}^\ell f(x_1^{(i)}, y) \right) \left(\mathbb{E}_{x_2 \in \Omega_1} \prod_{i=1}^\ell F(x_1^{(i)}, x_2) \right)^\ell \right)^{\frac{1}{\ell}} \\
&\leq \|f\|_{G(\ell,1)}^{\ell-1} \cdot \left(\tau^\ell \mathbb{E}_{x_1^{(1)}, \dots, x_1^{(\ell)} \in \Omega_1} \left(\mathbb{E}_{x_2 \in \Omega_1} \prod_{i=1}^\ell F(x_1^{(i)}, x_2) \right)^\ell + O(\gamma) \right)^{\frac{1}{\ell}} \\
&= \|f\|_{G(\ell,1)}^{\ell-1} \left(\tau^\ell \|F\|_{G(\ell,\ell)}^{\ell^2} + O(\gamma) \right)^{\frac{1}{\ell}}.
\end{aligned}$$

Here, the second inequality is Hölder's inequality, and the third uses that $f \leq T$, T is (τ, γ) -combinatorially spread, and the counting lemma (Lemma 3.2).

We first handle the case where $\|f\|_{G(\ell,1)} \geq \alpha\tau$. Define $d(y) = \mathbb{E}_{x \in \Omega_1} f(x, y)$ and note that $\|f\|_{G(\ell,1)}^\ell = \mathbb{E}_{y \in \Omega_2} d(y)^\ell$. Define $S_1, S_2 \subseteq \Omega_2$ as $S_1 = \{y : d(y) \geq (1 - \varepsilon)\alpha\tau\}$ and $S_2 = \{y : d(y) \geq 2\tau\}$. We want to prove that $|S_1|/|\Omega_2| \geq (\alpha/2)^{O(\ell)}$, as then

$$\mathbb{E}_{x \in \Omega_1, y \in \Omega_2} [f(x, y) \mathbb{1}_{S_1}(y)] \geq (1 - \varepsilon)\alpha\tau \mathbb{E}_{y \in \Omega_2} [\mathbb{1}_{S_1}(y)]$$

by definition. In particular, the theorem would follow by setting $g_1 = 1$ and $g_2 = \mathbb{1}_{S_1}$. To prove this, observe that the combinatorial spreadness of T gives that $|S_2| \leq \frac{\gamma}{\tau} |\Omega_2|$. Thus,

$$\begin{aligned}
(\alpha\tau)^\ell &\leq \|f\|_{G(\ell,1)}^\ell = \mathbb{E}_{y \in \Omega_2} d(y)^\ell \leq ((1 - \varepsilon)\alpha\tau)^\ell + \mathbb{E}_{y \in \Omega_2} \left[\mathbb{1}_{y \in S_1 \setminus S_2} d(y)^\ell \right] + \mathbb{E}_{y \in \Omega_2} \left[\mathbb{1}_{y \in S_2} d(y)^\ell \right] \\
&\leq ((1 - \varepsilon)\alpha\tau)^\ell + \frac{|S_1|}{|\Omega_2|} (2\tau)^\ell + \frac{\gamma}{\tau}.
\end{aligned}$$

Using the choice of γ , this rearranges to $\frac{|S_1|}{|\Omega_2|} \geq (\alpha/2)^{O(\ell)}$ as desired.

Otherwise, we assume $\|f\|_{G(\ell,1)} \leq \alpha\tau$, and this implies that

$$\tau^\ell \|F\|_{G(\ell,\ell)}^{\ell^2} + O(\gamma) \geq (1 + \varepsilon/2)^{\ell^2} \alpha^{(2k-2)\ell^2 + \ell} \tau^{(2k-2)\ell^2 + \ell},$$

which for the choice of γ and ℓ rearranges to

$$\|F\|_{G(\ell,\ell)} \geq (1 + \varepsilon/4)(\alpha\tau)^{2k-2}.$$

The idea is to now use non-relative sifting (Theorem 3.4) to get a correlation of F onto a subset $\Omega'_1 \subseteq \Omega_1$. However, done naïvely this will lose factors of τ in the size of Ω'_1 , which is unacceptable. To remedy this, we first argue that F is basically bounded by $M := \alpha^{-k} \tau^{2k-2}$. Formally, define the function $\tilde{F} : \Omega_1 \times \Omega_1 \rightarrow [0, 1]$ as $\tilde{F}(x_1, x_2) = M^{-1} \min\{F(x_1, x_2), M\}$.

We will prove that

$$\Pr_{x_1, x_2 \in \Omega_1} \left[M \tilde{F}(x_1, x_2) \neq F(x_1, x_2) \right] \leq \gamma^{\Omega\left(\frac{\log(1/\alpha)}{\log(1/\tau)}\right)}. \quad (3.2)$$

It suffices to bound $\Pr_{x_1, x_2 \in \Omega_1} [F(x_1, x_2) > M]$. Let t be the maximum positive integer such that $\gamma \leq \tau^{(2k-2)t}$. Markov's inequality tells us

$$\Pr_{x_1, x_2 \in \Omega_1} [F(x_1, x_2) > M] \leq M^{-t} \mathbb{E}_{x_1, x_2} [F(x_1, x_2)^t].$$

By the definition of F , $f \leq T$, and the counting lemma, we know that

$$\mathbb{E}_{x_1, x_2} [F(x_1, x_2)^t] = \|f\|_{G(2, (k-1)t)}^{(2k-2)t} \leq \tau^{(2k-2)t} + O(\gamma) \leq 2\tau^{(2k-2)t},$$

by applying combinatorial spreadness of T and the choice of γ . Thus,

$$\Pr_{x_1, x_2 \in \Omega_1} [F(x_1, x_2) > M] \leq M^{-t} \cdot 2\tau^{(2k-2)t} \leq 2\alpha^{kt} \leq \gamma^{\Omega\left(\frac{\log(1/\alpha)}{\log(1/\tau)}\right)}.$$

By (3.2) and the triangle inequality (which we may apply, since our choice of ℓ is even, so $\|\cdot\|_{G(\ell, \ell)}$ is a seminorm [36]), we get that

$$\begin{aligned} M\|\tilde{F}\|_{G(\ell, \ell)} &\geq \|F\|_{G(\ell, \ell)} - \|F - M\tilde{F}\|_{G(\ell, \ell)} \\ &\geq (1 + \varepsilon/4)(\alpha\tau)^{2k-2} - \gamma^{\Omega\left(\frac{\log(1/\alpha)}{\ell \log(1/\tau)}\right)} \geq (1 + \varepsilon/8)(\alpha\tau)^{2k-2}, \end{aligned}$$

for the choice of γ . Here, we have used the fact that if a function $G : \Omega_1 \times \Omega_1 \rightarrow [-1, 1]$ is supported on at most a γ' fraction of $\Omega_1 \times \Omega_1$, then $\|G\|_{G(\ell, \ell)} \leq (\gamma')^{1/\ell}$, applied for $G = F - M\tilde{F}$. Indeed,

$$\|G\|_{G(\ell, \ell)}^{\ell^2} = \left| \mathbb{E}_{\substack{x_1, \dots, x_\ell \in \Omega_1 \\ y_1, \dots, y_\ell \in \Omega_1}} \left[\prod_{i=1}^{\ell} \prod_{j=1}^{\ell} G(x_i, y_j) \right] \right| \leq \mathbb{E}_{\substack{x_1, \dots, x_\ell \in \Omega_1 \\ y_1, \dots, y_\ell \in \Omega_1}} \left[\prod_{i=1}^{\ell} |G(x_i, y_i)| \right] \leq (\gamma')^{\ell}.$$

Thus,

$$\|\tilde{F}\|_{G(\ell, \ell)} \geq M^{-1}(1 + \varepsilon/8)(\alpha\tau)^{2k-2} = (1 + \varepsilon/8)\alpha^{3k-2}.$$

By Theorem 3.4 and Lemma 3.3, we conclude that there are functions $g_1 : \Omega_1 \rightarrow \{0, 1\}$ and $g_2 : \Omega_1 \rightarrow \{0, 1\}$ such that

$$\mathbb{E}_{x \in \Omega_1} [g_i(x)] \geq \left(\frac{\varepsilon\alpha^{3k-2}}{2} \right)^{O(\ell)} \quad \text{for } i = 1, 2,$$

and

$$\mathbb{E}_{x_1, x_2 \in \Omega_1} [\tilde{F}(x_1, x_2)g_1(x_1)g_2(x_2)] \geq \alpha^{3k-2} \mathbb{E}_{x \in \Omega_1} [g_1(x)] \mathbb{E}_{x \in \Omega_1} [g_2(x)].$$

Thus,

$$\begin{aligned} \mathbb{E}_{x_1, x_2 \in \Omega_1} [F(x_1, x_2)g_1(x_1)g_2(x_2)] &\geq M\alpha^{3k-2} \mathbb{E}_{x \in \Omega_1} [g_1(x)] \mathbb{E}_{x \in \Omega_1} [g_2(x)] \\ &= (\alpha\tau)^{2k-2} \mathbb{E}_{x \in \Omega_1} [g_1(x)] \mathbb{E}_{x \in \Omega_1} [g_2(x)]. \end{aligned}$$

Note that F , viewed as a matrix on $\Omega_1 \times \Omega_1$, is symmetric and positive semi-definite. Thus if we define $h_i(x) = g_i(x)/\mathbb{E}[g_i]$ then

$$\begin{aligned} 0 &\leq \mathbb{E}_{x_1, x_2 \in \Omega_1} [F(x_1, x_2)(h_1(x_1) - h_2(x_1))(h_1(x_2) - h_2(x_2))] \\ &= \mathbb{E}_{x_1, x_2 \in \Omega_1} [F(x_1, x_2)h_1(x_1)h_1(x_2)] + \mathbb{E}_{x_1, x_2 \in \Omega_1} [F(x_1, x_2)h_2(x_1)h_2(x_2)] - 2 \mathbb{E}_{x_1, x_2 \in \Omega_1} [F(x_1, x_2)h_1(x_1)h_2(x_2)] \end{aligned}$$

gives us that either

$$\begin{aligned} \mathbb{E}_{x_1, x_2 \in \Omega_1} [F(x_1, x_2)g_1(x_1)g_1(x_2)] &\geq (\alpha\tau)^{2k-2} \mathbb{E}_{x \in \Omega_1} [g_1(x)]^2, \quad \text{or} \\ \mathbb{E}_{x_1, x_2 \in \Omega_1} [F(x_1, x_2)g_2(x_1)g_2(x_2)] &\geq (\alpha\tau)^{2k-2} \mathbb{E}_{x \in \Omega_1} [g_2(x)]^2. \end{aligned}$$

Without loss of generality, assume the former holds. Let

$$\Omega'_1 = \{x \in \Omega_1 : g_1(x) = 1\}$$

and define $f' : \Omega'_1 \times \Omega_2 \rightarrow [0, 1]$ as the restriction of f , i.e., $f'(x, y) = f(x, y)$. Then by definition,

$$\|f'\|_{G(2, k-1)}^{2k-2} = \mathbb{E}_{x_1 \in \Omega'_1, x_2 \in \Omega'_1} [F(x_1, x_2)] \geq (\alpha\tau)^{2k-2}.$$

Now, the result follows by induction on f' and the lower bound we have on $|\Omega'_1|/|\Omega_1| = \mathbb{E}_{x \in \Omega_1} [g_1(x)]$. $\square$

4. ADDITIONAL TOOLS

In this section we reproduce several combinatorial tools from the works of Kelley-Meka and Kelley-Lovett-Meka, e.g. [41, Theorem 2.8]. The main result is a key spectral positivity input that we need. We remark that there is a minor issue in the spectral positivity argument as given in [41, Theorem 2.8] and hence we provide complete proofs here. These give graph theoretic interpretations of two of the key steps in breakthrough work of Kelley and Meka on 3-term arithmetic progressions [42].

We now give the key spectral positivity argument. The first step is the following unbalancing inequality (see [42, Proposition D.1]). We include the proof for completeness.

Lemma 4.1. *Let $\varepsilon \in (0, 1/10)$, k be a positive integer, and $p = 6\lceil k/\varepsilon \rceil$. Let X be a real random variable such that $\mathbb{E}[X^k] \geq \varepsilon^k$ and $\mathbb{E}[X^r] \geq 0$ for all $r \in \mathbb{Z}^{\geq 0}$. Then*

$$\mathbb{E}[(X+1)^p] \geq (1 + \varepsilon/2)^p.$$

Proof. Using $\mathbb{E}[X^r] \geq 0$ for all $r \in \mathbb{Z}^{\geq 0}$ and Hölder's inequality, we have

$$\begin{aligned} \mathbb{E}[(X+1)^p] &= \sum_{\ell=0}^p \binom{p}{\ell} \mathbb{E}[X^\ell] \geq \sum_{\ell \equiv 0 \pmod 2} \binom{p}{\ell} \cdot \mathbb{E}[X^\ell] \\ &\geq \sum_{\substack{\ell \equiv 0 \pmod 2 \\ \ell \geq k}} \binom{p}{\ell} \cdot (\mathbb{E}[X^k])^{\ell/k} \geq \sum_{\substack{\ell \equiv 0 \pmod 2 \\ \ell \geq k}} \binom{p}{\ell} \cdot \varepsilon^\ell \\ &= (1 + \varepsilon)^p \cdot \sum_{\substack{\ell \equiv 0 \pmod 2 \\ \ell \geq k}} \binom{p}{\ell} \cdot \left(\frac{\varepsilon}{1 + \varepsilon}\right)^\ell \cdot \left(\frac{1}{1 + \varepsilon}\right)^{p-\ell}. \end{aligned}$$

Let $\varepsilon' = \varepsilon/(1 + \varepsilon)$ and $\text{Bin}(n, q)$ denote a binomial random variable with n independent trials where each has success probability q . Via chasing definitions, we have that

$$\sum_{\substack{\ell \equiv 0 \pmod 2 \\ \ell \geq k}} \binom{p}{\ell} \cdot \left(\frac{\varepsilon}{1 + \varepsilon}\right)^\ell \left(\frac{1}{1 + \varepsilon}\right)^{p-\ell} = \Pr_{X \sim \text{Bin}(p, \varepsilon')} [X \geq k \wedge X \equiv 0 \pmod 2].$$

For the sake of simplicity, we omit that $X \sim \text{Bin}(p, \varepsilon')$ below. Note that

$$\begin{aligned} \Pr[X \equiv 0 \pmod 2] &= \frac{1}{2} + \frac{\Pr[X \equiv 0 \pmod 2] - \Pr[X \equiv 1 \pmod 2]}{2} = \frac{1}{2} + \frac{\sum_{\ell=0}^p (-1)^\ell \binom{p}{\ell} (\varepsilon')^\ell (1 - \varepsilon')^{p-\ell}}{2} \\ &= \frac{1}{2} + \frac{(1 - 2\varepsilon')^p}{2} \geq \frac{1}{2}. \end{aligned}$$

Thus

$$\Pr[X \geq k \wedge X \equiv 0 \pmod 2] \geq \frac{1}{2} - \Pr[X < k].$$

Observe that $\mathbb{E}[X] = p \cdot \varepsilon' \geq 4k$. Furthermore we have that

$$\text{Var}[X] = p \cdot \varepsilon' \cdot (1 - \varepsilon') \leq \frac{7k}{(1 + \varepsilon)^2} \leq 7k.$$

Thus by Chebyshev's inequality, we have that

$$\Pr[X < k] \leq \frac{\text{Var}[X]}{(\mathbb{E}[X] - k)^2} \leq \frac{7k}{(3k)^2} \leq \frac{1}{k} \leq \frac{1}{6}.$$

Thus we have that

$$\mathbb{E}[(X + 1)^p] \geq \frac{1}{3} \cdot (1 + \varepsilon)^p \geq (1 + \varepsilon/2)^p \cdot \frac{(1 + \varepsilon/3)^{6/\varepsilon}}{3} \geq (1 + \varepsilon/2)^p. \quad \square$$

We now complete the proof of the necessary spectral positivity argument.

Lemma 4.2. *Let $\varepsilon \in (0, 1/10)$, k be an even positive integer, and $p = 36\lceil k/\varepsilon^4 \rceil$. Consider $A : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}^+$, and define $\alpha := \|A\|_{G(1,1)} = \mathbb{E}_{x \in \Omega_1, y \in \Omega_2}[A(x, y)]$. Suppose that*

$$\|A - \alpha\|_{G(2,k)} \geq \varepsilon\alpha \quad \text{and} \quad \inf_x \mathbb{E}_y[A(x, y) - \alpha] \geq -\varepsilon^2\alpha/36.$$

Then

$$\|A\|_{G(2,p)} \geq (1 + \varepsilon^2/36)\alpha.$$

Proof. By normalizing, we may assume that $\alpha = 1$. We prove that if

$$\|A\|_{G(2,p)} \leq 1 + \varepsilon^2 \quad \text{and} \quad \inf_x \mathbb{E}_y[(A(x, y) - 1)] \geq -\varepsilon^2$$

then

$$\|A - 1\|_{G(2,k)} \leq 6\varepsilon.$$

Let $q = 6\lceil k/\varepsilon^2 \rceil$. We first handle the case that

$$\|A - 1\|_{G(1,q)} \geq 7\varepsilon^2.$$

In this case, observe that

$$\|A - 1\|_{G(1,q)}^q = \mathbb{E}_x(\mathbb{E}_y A(x, y) - 1)^q;$$

therefore if we define $F(x) = \mathbb{E}_y A(x, y) - 1$ then $\|F\|_q \geq 7\varepsilon^2$. Let

$$F(x) = \max(F(x), 0) + \min(F(x), 0) =: F^+(x) + F^-(x)$$

and observe that $|F^-(x)| \leq \varepsilon^2$ by our earlier assumption. Hence by the triangle inequality $\|F^+\|_q \geq 6\varepsilon^2$ and by definition $F^+ \geq 0$. Thus via [Lemma 4.1](#), we have that

$$\|F^+ + 1\|_p \geq 1 + 3\varepsilon^2$$

and therefore by the triangle inequality again,

$$\|F + 1\|_p \geq 1 + 2\varepsilon^2.$$

However by the monotonicity of grid norms (see, e.g., [\[41, Claim 4.2\]](#))

$$1 + 2\varepsilon^2 \leq \|F + 1\|_p = \|A\|_{G(1,p)} \leq \|A\|_{G(2,p)},$$

yielding the desired contradiction.

Thus we may assume for the remainder of the proof that

$$\|A - 1\|_{G(1,q)} \leq 7\varepsilon^2.$$

Observe that

$$\mathbb{E}_y A(x, y) A(x', y) = \mathbb{E}_y [(A(x, y) - 1)(A(x', y) - 1) + 1 + (A(x, y) - 1) + (A(x', y) - 1)]. \quad (4.1)$$

Define

$$B(x, x') := \mathbb{E}_y [(A(x, y) - 1)(A(x', y) - 1)]$$

and observe that for any integer $r \geq 1$ it satisfies $\mathbb{E}[B^r] \geq 0$ since

$$\mathbb{E}_{x, x'} [B(x, x')^r] = \mathbb{E}_{y_1, \dots, y_r} \mathbb{E}_{x, x'} \prod_{j=1}^r (A(x, y_j) - 1)(A(x', y_j) - 1) = \mathbb{E}_{y_1, \dots, y_r} \left(\mathbb{E}_x \prod_{j=1}^r (A(x, y_j) - 1) \right)^2 \geq 0.$$

Now assume by contradiction that

$$\|A - 1\|_{G(2, k)} = \mathbb{E}_{x, x'} [B(x, x')^k]^{1/(2k)} > 6\varepsilon.$$

That is, we have $\|B\|_k \geq 36\varepsilon^2$. We may apply [Lemma 4.1](#) to obtain that

$$\|B + 1\|_q = \mathbb{E}_{x, x'} [(B(x, x') + 1)^q]^{1/q} \geq 1 + 18\varepsilon^2.$$

Via [\(4.1\)](#) and the triangle inequality (as $\|A - 1\|_{G(1, q)} \leq 7\varepsilon^2$), we obtain that

$$\begin{aligned} \|A\|_{G(2, q)}^2 &= \mathbb{E}_{x, x'} [\mathbb{E}_y (A(x, y)A(x', y))^q]^{1/q} \\ &= \mathbb{E}_{x, x'} [\mathbb{E}_y ((A(x, y) - 1)(A(x', y) - 1) + 1 + (A(x, y) - 1) + (A(x', y) - 1))^q]^{1/q} \\ &\geq \|B + 1\|_q - 2\|A - 1\|_{G(1, q)} \geq 1 + 18\varepsilon^2 - 2(7\varepsilon^2) \geq 1 + 4\varepsilon^2. \end{aligned}$$

By monotonicity of grid norms, we get

$$\|A\|_{G(2, p)} \geq \|A\|_{G(2, q)} \geq (1 + 4\varepsilon^2)^{1/2} > 1 + \varepsilon^2$$

which gives the desired contradiction. $\square$

A fact we use several times throughout the work is the “reverse Markov inequality”.

Fact 4.3. *Let $\rho, \gamma \in (0, 1)$. Let V be a random variable with $V \leq (1 + \rho)\mathbb{E}[V]$. Then*

$$\Pr[V \leq (1 - \gamma)\mathbb{E}[V]] \leq \frac{\rho}{\gamma + \rho}.$$

Proof. Define $V' = (1 + \rho)\mathbb{E}[V] - V$. Observe that $V' \geq 0$ and $\mathbb{E}[V'] = \rho\mathbb{E}[V]$. Thus

$$\Pr[V \leq (1 - \gamma)\mathbb{E}[V]] = \Pr[V' \geq (\gamma + \rho)\mathbb{E}[V]] \leq \frac{\rho}{\gamma + \rho}. \quad \square$$

5. THE FINITE FIELD CASE

5.1. General setup. Throughout our analysis we will consider a linear subspace $W \subseteq \mathbb{F}_2^n$. (Although in occasional convenient contexts, we will let W be an affine subspace.) We will additionally consider subsets $X, Y, D \subseteq W$, and define our *container set* to be

$$S(X, Y, D) := \{(x, y) \in W \times W : x \in X, y \in Y, x + y \in D\} \subseteq W \times W.$$

The sets X, Y will sometimes be pseudorandom in the analysis, according to the definition of *algebraically spread* in [Definition 5.1](#). At the start of the density increment phase, both X and Y will be (r, ε_s) -algebraically spread for well-chosen parameters r and ε_s . At the start of the pseudorandomization phase, we will only guarantee that at least one of X, Y is (r, ε_s) -algebraically spread, and after the pseudorandomization, both X, Y will be (r, ε_s) -algebraically spread.

5.2. Algebraic spreadness. We start by defining a notion of algebraic pseudorandomness.

Definition 5.1 (Algebraic spreadness). Let $W \subseteq \mathbb{F}_2^n$ be an affine subspace. We say that a subset $X \subseteq W$ is (r, ε) -algebraically spread within W if for all affine subspaces $W' \subseteq W$ satisfying $\dim(W') \geq \dim(W) - r$, it holds that

$$\frac{|X \cap W'|}{|W'|} \leq (1 + \varepsilon) \frac{|X|}{|W|}.$$

The following results of Kelley and Meka [42] will prove useful throughout the subsection.

Lemma 5.2 ([42, Theorem 4.10]). Let $\varepsilon, \tau, \gamma \in (0, 1/2)$, $r \geq \Omega(\varepsilon^{-7} \log(1/\tau) \log(1/\gamma)^7)$ a large enough integer, and $A \subseteq \mathbb{F}_2^n$ of size $|A| = \tau |\mathbb{F}_2^n|$. If A is $(r, \varepsilon/8)$ -algebraically spread, then for all $B, C \subseteq \mathbb{F}_2^n$ of size at least $\gamma |\mathbb{F}_2^n|$, we have

$$\langle \mathbb{1}_B * \mathbb{1}_C, \mathbb{1}_A \rangle \leq (1 + \varepsilon) \mathbb{E}[\mathbb{1}_A] \mathbb{E}[\mathbb{1}_B] \mathbb{E}[\mathbb{1}_C].$$

Lemma 5.3 ([42, Proposition 2.16]). Let $\varepsilon, \tau \in (0, 1/2)$, $k \geq 1$ an integer, $r \geq \Omega(\varepsilon^{-7} \log(1/\tau)^4 k^4)$ a sufficiently large integer, and $A, B \subseteq \mathbb{F}_2^n$ of size at least $\tau |\mathbb{F}_2^n|$. If A and B are both $(r, \varepsilon/8)$ -algebraically spread, then

$$\left\| \frac{\mathbb{1}_A}{\mathbb{E}[\mathbb{1}_A]} * \frac{\mathbb{1}_B}{\mathbb{E}[\mathbb{1}_B]} - 1 \right\|_k \leq \varepsilon.$$

We require a lemma which relates algebraic spreadness to combinatorial spreadness in the case of corners.

Lemma 5.4. Let $\varepsilon, \tau, \gamma \in (0, 1/2)$, and $r \geq \Omega(\varepsilon^{-7} \log(1/\tau) \log(1/\gamma)^7)$ a large enough integer. Let $W \subseteq \mathbb{F}_2^n$ be a linear subspace, and let $D \subseteq W$ with $\tau = \frac{|D|}{|W|}$. Define $T \subseteq W \times W$ as $T := \{(x, y) : x + y \in D\}$. If D is $(r, \varepsilon/8)$ -algebraically spread within W for $r \geq \Omega(\varepsilon^{-7} \log(1/\tau) \log(1/\gamma)^7)$, then T is $((1 + \varepsilon)\tau, \gamma)$ -combinatorially spread.

Proof. Suppose that T is not $((1 + \varepsilon)\tau, \gamma)$ -combinatorially spread. Then there exist functions $f, g : W \rightarrow [0, 1]$ such that

$$\mathbb{E}_{x, y \in W} [f(x)g(y)\mathbb{1}_T(x, y)] > (1 + \varepsilon)\tau \cdot \mathbb{E}_{x \in W} [f(x)] \cdot \mathbb{E}_{x \in W} [g(x)] + \gamma > (1 + \varepsilon)\tau \cdot \mathbb{E}_{x \in W} [f(x)] \cdot \mathbb{E}_{x \in W} [g(x)].$$

Note that f and g are 1-bounded, and therefore for the first inequality to hold we must have that $\mathbb{E}[f] \geq \gamma$ and $\mathbb{E}[g] \geq \gamma$.

By applying Lemma 3.3, there exist boolean functions $F, G : W \rightarrow \{0, 1\}$ such that

$$\mathbb{E}_{x, y \in W} [F(x)G(y)\mathbb{1}_T(x, y)] > (1 + \varepsilon)\tau \cdot \mathbb{E}_{x \in W} [F(x)] \cdot \mathbb{E}_{x \in W} [G(x)]$$

where $\mathbb{E}[F] \geq \gamma/2$ and $\mathbb{E}[G] \geq \gamma/2$. Recalling that $\mathbb{1}_T(x, y) = \mathbb{1}_D(x + y)$, we have that

$$\mathbb{E}_{x, y \in W} [F(x)G(y)\mathbb{1}_D(x + y)] > (1 + \varepsilon)\tau \cdot \mathbb{E}_{x \in W} [F(x)] \cdot \mathbb{E}_{x \in W} [G(x)].$$

This however contradicts Lemma 5.2. □

An upper bound on the size of $S(X, Y, D)$ follows from one of X or Y being algebraically spread.

Lemma 5.5. Let $X, Y, D \subseteq W$ for a linear subspace $W \subseteq \mathbb{F}_2^n$, and let $\delta = |W|^{-3}|X||Y||D|$. If X or Y is $(r, \varepsilon/8)$ -algebraically spread for $r \geq \Omega(\varepsilon^{-7} \log(1/\delta)^8)$, then $|S(X, Y, D)| \leq (1 + \varepsilon)\delta|W|^2$.

Proof. We focus on the case where X is spread; the Y case is identical. Observe that

$$\begin{aligned} |S(X, Y, D)| &= \sum_{x, y \in W} \mathbb{1}_X(x) \mathbb{1}_Y(y) \mathbb{1}_D(x + y) = |W|^2 \mathbb{E}_{x \in W} \left[\mathbb{1}_X(x) \mathbb{E}_{y \in W} [\mathbb{1}_Y(y) \mathbb{1}_D(x + y)] \right] \\ &= |W|^2 \langle \mathbb{1}_X, \mathbb{1}_Y * \mathbb{1}_D \rangle. \end{aligned}$$

We conclude by observing that [Lemma 5.2](#) yields $\langle \mathbb{1}_X, \mathbb{1}_Y * \mathbb{1}_D \rangle \leq (1 + \varepsilon) \mathbb{E}[\mathbb{1}_X] \mathbb{E}[\mathbb{1}_Y] \mathbb{E}[\mathbb{1}_D] = (1 + \varepsilon)\delta$. $\square$

If additionally both X and Y are algebraically spread, then the size of $S(X, Y, D)$ is also lower-bounded.

Lemma 5.6. *Let $X, Y, D \subseteq W$ for a linear subspace $W \subseteq \mathbb{F}_2^n$, and let $\delta = |W|^{-3} |X| |Y| |D|$. If both X and Y are $(r, \varepsilon/16)$ -algebraically spread for $r \geq \Omega(\varepsilon^{-7} \log(1/\delta)^8)$, then it holds that*

$$(1 - \varepsilon)\delta |W|^2 \leq |S(X, Y, D)| \leq (1 + \varepsilon)\delta |W|^2.$$

Proof. As above, we have $|S(X, Y, D)| = |W|^2 \langle \mathbb{1}_X * \mathbb{1}_Y, \mathbb{1}_D \rangle$. Then Hölder's inequality with $p = \lceil \log(1/\delta) \rceil$ and [Lemma 5.3](#) give

$$\begin{aligned} ||S(X, Y, D)| - \delta |W|^2| &= |W|^2 |\langle \mathbb{1}_X * \mathbb{1}_Y, \mathbb{1}_D \rangle - \delta| \\ &\leq |W|^2 \|\mathbb{1}_X * \mathbb{1}_Y - \mathbb{E}[\mathbb{1}_X] \mathbb{E}[\mathbb{1}_Y]\|_p \cdot (\mathbb{E}[\mathbb{1}_D])^{1-1/p} \leq \varepsilon \delta |W|^2. \quad \square \end{aligned}$$

5.3. Algebraic pseudorandomization. In the subsequent subsections, we will show that under certain conditions, which include the algebraic spreadness of X and Y , A admits a density increment onto either some $S(X', Y, D')$ or $S(X, Y', D')$. However, the density increment may spoil the algebraic spreadness of (say) X' . Thus, we require a procedure to pseudorandomize X' . This is precisely the goal of this section. Now we state the algebraic pseudorandomization theorem.

Theorem 5.7. *Let $X, Y, D \subseteq W$ for a linear subspace $W \subseteq \mathbb{F}_2^n$. Define $\delta_X = |X|/|W|$, $\delta_Y = |Y|/|W|$, $\delta_D = |D|/|W|$, and $\delta = \delta_X \delta_Y \delta_D$. Let $\varepsilon > 0$ and $r \geq \Omega(\varepsilon^{-8} \log(1/(\varepsilon \delta))^8 + \varepsilon^{-8} \log(1/\delta_D)^{16})$ be an integer, such that at least one of X or Y is $(r, \varepsilon/8)$ -algebraically spread. Let $A \subseteq S(X, Y, D)$ with $|A| \geq \alpha \delta |W|^2$. Then there is a linear subspace $W' \subseteq W$ with shifts $x, y \in W$ and sets $X' \subseteq W' + x$, $Y' \subseteq W' + y$, $D' \subseteq W' + x + y$, and $A' := A \cap S(X', Y', D')$ satisfying:*

- (1) $\dim(W') \geq \dim(W) - O(r \varepsilon^{-2} \log(1/(\varepsilon \alpha \delta))^2 \log(1/(\varepsilon \alpha \delta_D)) + r \varepsilon^{-2} \log(1/(\varepsilon \alpha \delta_D))^5)$,
- (2) $\frac{|D'|}{|W'|} \geq (\varepsilon \alpha / 2) \delta_D$,
- (3) $\frac{|X'| |Y'|}{|W'|^2} \geq 2^{-O(\log(1/(\varepsilon \alpha \delta_D))^2)} \delta_X \delta_Y$,
- (4) X' and Y' are (r, ε) -algebraically spread in $W' + x$ and $W' + y$, respectively, and
- (5) $|A'| \geq (1 - 5\varepsilon) \alpha \delta_{X'} \delta_{Y'} \delta_{D'} |W'|^2$ where $\delta_{X'} = \frac{|X'|}{|W'|}$, $\delta_{Y'} = \frac{|Y'|}{|W'|}$, and $\delta_{D'} = \frac{|D'|}{|W'|}$.

The first step towards proving [Theorem 5.7](#) is to note that we can almost completely partition $X \times Y$ into subrectangles $X' \times Y'$ which are all algebraically spread (inside their respective subspaces) and cover all but an η -fraction of $X \times Y$. For this, it is useful to observe that for any set $X \subseteq W$, we can find a relatively large subset within an affine subspace that is algebraically spread.

Claim 5.8. *Let r be an integer, $W \subseteq \mathbb{F}_2^n$ be a linear subspace, and $\varepsilon > 0$. For $X \subseteq W$ with $\delta_X := |X|/|W|$, there is an affine subspace $W' \subseteq W$ such that $X' := X \cap W'$ is (r, ε) -algebraically spread within W' , $\dim(W') \geq \dim(W) - O(r \varepsilon^{-1} \log(1/\delta_X))$, and $\frac{|X'|}{|W'|} \geq \delta_X$.*

Proof. We proceed iteratively. Initialize $W^{(0)} = W$. If $X \cap W^{(0)}$ is (r, ε) -algebraically spread within $W^{(0)}$, we are done. Otherwise, there must exist an affine subspace $W^{(1)} \subseteq W^{(0)}$ with

$\dim(W^{(1)}) \geq \dim(W^{(0)}) - r$ satisfying

$$\frac{|X \cap W^{(1)}|}{|W^{(1)}|} > (1 + \varepsilon) \frac{|X \cap W^{(0)}|}{|W^{(0)}|}.$$

We now repeat this process with $W^{(1)}$. After i iterations, the density of $X \cap W^{(i)}$ within $W^{(i)}$ is at least $(1 + \varepsilon)^i \delta_X \geq \delta_X$ and $\dim(W^{(i)}) \geq \dim(W^{(0)}) - ri$. As the density may not exceed 1, we are guaranteed to obtain our desired affine subspace after $O(\varepsilon^{-1} \log(1/\delta_X))$ iterations. $\square$

Now we state and prove the key partitioning lemma over finite fields.

Lemma 5.9. *Let r be an integer, $W \subseteq \mathbb{F}_2^n$ be a linear subspace, and $\varepsilon, \eta \in (0, 1/10)$. Additionally, let $X, Y \subseteq W$, and define $\delta_X = |X|/|W|$ and $\delta_Y = |Y|/|W|$. Then there is a positive integer T and for all $i = 1, \dots, T$, a subspace $V_i \subseteq W$, points $x_i, y_i \in W/V_i$, and subsets $X_i \subseteq V_i + x_i, Y_i \subseteq V_i + y_i$ satisfying:*

- (1) $\dim(V_i) \geq \dim(W) - O(r\varepsilon^{-2} \log(1/(\delta_X \delta_Y))^2 \log(1/\eta) + r\varepsilon^{-2} \log(1/\eta)^5)$ for $i = 1, \dots, T$.
- (2) $X_i \times Y_i$ and $X_j \times Y_j$ are disjoint for $1 \leq i \neq j \leq T$.
- (3) $|X \times Y \setminus \bigcup_{i=1}^T X_i \times Y_i| \leq \eta |X| |Y|$.
- (4) $\frac{|X_i| |Y_i|}{|V_i|^2} \geq 2^{-O(\log(1/\eta)^2)} \delta_X \delta_Y$ for all $i = 1, 2, \dots, T$.
- (5) X_i and Y_i are (r, ε) -algebraically spread within $V_i + x_i$ and $V_i + y_i$, respectively, for $i = 1, \dots, T$.

Towards proving Lemma 5.9, we first provide a simpler one round partitioning result.

Lemma 5.10. *Let r be an integer, $W \subseteq \mathbb{F}_2^n$ be a linear subspace, and $\varepsilon, \eta \in (0, 1/10)$. Additionally, let $X, Y \subseteq W$, and define $\delta_X = |X|/|W|$ and $\delta_Y = |Y|/|W|$. Then there is a positive integer T , subset $\mathcal{G} \subseteq [T]$, and for all $i = 1, \dots, T$, a subspace $V_i \subseteq W$, points $x_i, y_i \in W/V_i$, and subsets $X_i \subseteq V_i + x_i, Y_i \subseteq V_i + y_i$ satisfying:*

- (1) $\dim(V_i) \geq \dim(W) - O(r\varepsilon^{-2} \log(1/(\delta_X \delta_Y \eta))^2)$ for $i = 1, \dots, T$.
- (2) $X_i \times Y_i$ and $X_j \times Y_j$ are disjoint for $1 \leq i \neq j \leq T$.
- (3) $|X \times Y \setminus \bigcup_{i=1}^T X_i \times Y_i| \leq \eta^2 |X| |Y|$.
- (4) $\frac{|X_i|}{|V_i|} \geq \frac{\eta^4}{100} \delta_X$ and $\frac{|Y_i|}{|V_i|} \geq \frac{\eta^2}{10} \delta_Y$ for all $i = 1, 2, \dots, T$.
- (5) X_i and Y_i are (r, ε) -algebraically spread within $V_i + x_i$ and $V_i + y_i$, respectively, for all $i \in \mathcal{G}$.
- (6) $\sum_{i \in \mathcal{G}} |X_i| |Y_i| \geq \frac{1}{2} |X| |Y|$.

Proof. Let $X^{(0)} = X$ and perform the following algorithm. For $t = 0, 1, \dots$: if $|X^{(t)}| \leq \frac{\eta^2}{10} \delta_X |W|$, then terminate. Otherwise, let $X_t \subseteq X^{(t)}$ be $(r', \varepsilon/5)$ -algebraically spread within an affine subspace $W_t + x_t$ for $r' = \Omega(r\varepsilon^{-1} \log(1/(\delta_Y \eta)))$, as given by Claim 5.8. Note that $|X_t| \geq \frac{\eta^2}{10} \delta_X |W_t|$ and

$$\dim(W_t) \geq \dim(W) - O(r'\varepsilon^{-1} \log(1/(\delta_X \eta))).$$

Now let $X^{(t+1)} = X^{(t)} \setminus X_t$. Let T be the total number of iterations, so that

$$X = X^{(T)} \cup X_0 \cup X_1 \cup \dots \cup X_{T-1},$$

where $|X^{(T)}| \leq \frac{\eta^2}{10} |X|$.

Next, fix some $t \in \{0, 1, \dots, T-1\}$. For $x \in W/W_t$ define $Y_{t,x} := Y \cap (W_t + x)$. Note that over $x \in W/W_t$, the $Y_{t,x}$ form a partition of Y . Now for each $x \in W/W_t$ further partition

$$Y_{t,x} = Y_{t,x}^{(T'_{t,x})} \cup Y_{t,x,0} \cup \dots \cup Y_{t,x,T'_{t,x}-1},$$

using the algorithm in the first paragraph, where $|Y_{t,x}^{(T'_{t,x})}| \leq \frac{\eta^2}{10} \delta_Y |W_t|$ and each $Y_{t,x,t'}$ for $0 \leq t' \leq T'_{t,x} - 1$ is (r, ε) -algebraically spread within some affine subspace $W_{t,x,t'} + y_{t,x,t'}$ where

$$\dim(W_{t,x,t'}) \geq \dim(W_t) - O(r\varepsilon^{-1} \log(1/(\delta_Y \eta))).$$

In particular, the $Y_{t,x,t'}$ satisfy $\frac{|Y_{t,x,t'}|}{|W_{t,x,t'}|} \geq \frac{\eta^2}{10} \delta_Y$. Finally, for each $y \in W_t/W_{t,x,t'}$ define $X_{t,x,t',y} := X_t \cap (W_{t,x,t'} + x_t + y)$, and note that for every fixed t, x, t' that $X_{t,x,t',y}$ partition X_t over $y \in W_t/W_{t,x,t'}$.

Now we define the pieces $X_i \times Y_i$ in the lemma statement and $\mathcal{G}$. The pieces $X_i \times Y_i$ are all pieces

$$X_{t,x,t',y} \times Y_{t,x,t'} \subseteq (W_{t,x,t'} + x_t + y) \times (W_{t,x,t'} + y_{t,x,t'})$$

where $\frac{|X_{t,x,t',y}|}{|W_{t,x,t'}|} \geq \frac{\eta^4}{100} \delta_X$. This combined with the density lower bound $\frac{|Y_{t,x,t'}|}{|W_{t,x,t'}|} \geq \frac{\eta^2}{10} \delta_Y$ gives guarantee (4). Finally, $\mathcal{G}$ consists of those pieces with $\frac{|X_{t,x,t',y}|}{|W_{t,x,t'}|} \geq (1 - 3\varepsilon/5) \frac{|X_t|}{|W_t|}$.

Let us now verify the remaining conclusions. (1) follows because

$$\begin{aligned} \dim(W_{t,x,t'}) &\geq \dim(W_t) - O(r\varepsilon^{-1} \log(1/(\delta_Y \eta))) \\ &\geq \dim(W) - O(r\varepsilon^{-2} \log(1/(\delta_X \delta_Y \eta))) - O(r\varepsilon^{-1} \log(1/(\delta_Y \eta))), \end{aligned}$$

as desired. (2) follows by construction. To check (3), first define $\mathcal{I}$ to be the set of tuples (t, x, t', y) with $\frac{|X_{t,x,t',y}|}{|W_{t,x,t'}|} < \frac{\eta^4}{100} \delta_X$. Note that by the construction

$$|X \times Y| - \sum_i |X_i \times Y_i| = |X^{(T)}| |Y| + \sum_{t,x} |X_t| |Y_{t,x}^{(T'_{t,x})}| + \sum_{(t,x,t',y) \in \mathcal{I}} |X_{t,x,t',y}| |Y_{t,x,t'}|.$$

We bound this term by term. By construction, $|X^{(T)}| |Y| \leq \frac{\eta^2}{10} |X| |Y|$. Also, $|Y_{t,x}^{(T'_{t,x})}| \leq \frac{\eta^2}{10} \delta_Y |W_t|$, so $\sum_{x \in W/W_t} |Y_{t,x}^{(T'_{t,x})}| \leq \frac{\eta^2}{10} \delta_Y |W|$. Overall,

$$\sum_{t,x} |X_t| |Y_{t,x}^{(T'_{t,x})}| \leq \sum_t |X_t| \cdot \frac{\eta^2}{10} \delta_Y |W| \leq \frac{\eta^2}{10} |X| |Y|. \quad (5.1)$$

For the final term, we can write

$$\begin{aligned} \sum_{(t,x,t',y) \in \mathcal{I}} |X_{t,x,t',y}| |Y_{t,x,t'}| &\leq \sum_{\substack{t,x,t' \\ y \in W_t/W_{t,x,t'}}} |Y_{t,x,t'}| \cdot \frac{\eta^4}{100} \delta_X |W_{t,x,t'}| \\ &= \sum_{t,x,t'} |Y_{t,x,t'}| \cdot \frac{\eta^4}{100} \delta_X |W_t| \leq \sum_t |Y_t| \cdot \frac{\eta^4}{100} \delta_X |W_t| \\ &\leq \sum_t |Y_t| \cdot \frac{\eta^4}{100} \cdot \frac{10}{\eta^2} |X_t| \leq \frac{\eta^2}{10} |X| |Y|. \end{aligned}$$

Here, the second to last inequality follows because $\frac{|X_t|}{|W_t|} \geq \frac{\eta^2}{10} \delta_X$. Combining these verifies item (3).

To verify (5), note that each $Y_{t,x,t'} \subseteq W_{t,x,t'} + y_{t,x,t'}$ is (r, ε) -algebraically spread by construction. Additionally, note that for any affine subspace $V \subseteq W_{t,x,t'}$ with $\dim(V) \geq \dim(W_{t,x,t'}) - r$ it holds that

$$\frac{|X_{t,x,t',y} \cap V|}{|V|} \leq \frac{|X_t \cap V|}{|V|} \leq (1 + \varepsilon/5) \frac{|X_t|}{|W_t|},$$

because $\dim(V) \geq \dim(W_t) - O(r\varepsilon^{-1} \log(1/(\delta_Y \eta)))$ and X_t is $(r', \varepsilon/5)$ -algebraically spread inside $W_t + x_t$ for $r' = \Omega(r\varepsilon^{-1} \log(1/(\delta_Y \eta)))$ by construction. Thus if $\frac{|X_{t,x,t',y}|}{|W_{t,x,t'}|} \geq (1 - 3\varepsilon/5) \frac{|X_t|}{|W_t|}$ (as is true for the tuples in $\mathcal{G}$), then $X_{t,x,t',y}$ is (r, ε) -algebraically spread.

For item (6), first consider fixing some t, x, t' , and let $\delta_t = \frac{|X_t|}{|W_t|}$ to simplify notation. Note that $\mathbb{E}_{y \in W_t/W_{t,x,t'}} \left[\frac{|X_{t,x,t',y}|}{|W_{t,x,t'}|} \right] = \delta_t$, and that $\frac{|X_{t,x,t',y}|}{|W_{t,x,t'}|} \leq (1 + \varepsilon/5)\delta_t$ for all y , as argued above. Thus by the reverse Markov inequality (Fact 4.3), $\Pr_{y \in W_t/W_{t,x,t'}} \left[\frac{|X_{t,x,t',y}|}{|W_{t,x,t'}|} \geq (1 - 3\varepsilon/5)\delta_t \right] \geq \frac{3}{4}$. Hence,

$$\sum_{\substack{y \in W_t/W_{t,x,t'} \\ \frac{|X_{t,x,t',y}|}{|W_{t,x,t'}|} \geq (1-3\varepsilon/5)\delta_t}} |X_{t,x,t',y}| \geq \frac{3}{4} \frac{|W_t|}{|W_{t,x,t'}|} (1 - 3\varepsilon/5)\delta_t |W_{t,x,t'}| \geq \frac{2}{3} |X_t|.$$

We conclude the total size of pieces from $\mathcal{G}$ is at least

$$\sum_{t,x,t'} \frac{2}{3} |X_t| |Y_{t,x,t'}| = \sum_t \frac{2}{3} |X_t| |Y| - \sum_{t,x} \frac{2}{3} |X_t| |Y_{t,x}^{(T',x)}| \geq \frac{1}{2} |X| |Y|,$$

where we have used that $\sum_t |X_t| \geq (1 - \eta^2) |X|$ and $\sum_{t,x} |X_t| |Y_{t,x}^{(T',x)}| \leq \frac{\eta^2}{10} |X| |Y|$ from (5.1). $\square$

Iterating Lemma 5.10 proves Lemma 5.9.

Proof of Lemma 5.9. Perform the following algorithm given $X \times Y \subseteq W \times W$. Consider $X_i \times Y_i \subseteq (V_i + x_i) \times (V_i + y_i)$ for $i = 1, \dots, T$ as given in Lemma 5.10. Now, for each $i \in [T] \setminus \mathcal{G}$ invoke Lemma 5.10 recursively. Note that if $X \subseteq W$ is spread, then $X + x \subseteq W + x$ is spread with the same parameters. Thus, even though Lemma 5.10 is stated in terms of linear subspaces W , by shifting X, Y we can apply the lemma to $(X_i + x_i) \times (Y_i + y_i) \subseteq V_i \times V_i$ to obtain a partition, then shift all of the sets back by x_i, y_i , respectively. Terminate when the recursion depth is $O(\log(1/\eta))$, and remove any remaining pieces from the partition.

Let's check that all the conclusions of Lemma 5.9 hold. By conclusion (4) of Lemma 5.10 and iterating for $O(\log(1/\eta))$ recursive layers gives that $|X_i|/|V_i| \geq 2^{-O(\log(1/\eta)^2)} \delta_X$ and $|Y_i|/|V_i| \geq 2^{-O(\log(1/\eta)^2)} \delta_Y$. Thus (4) holds. (1) holds because the recursion depth is $O(\log(1/\eta))$ combined with conclusion (1) of Lemma 5.10 – the additional $O(r\varepsilon^{-2} \log(1/\eta)^5)$ terms is due to the fact that δ_X and δ_Y may drop by $2^{-O(\log(1/\eta)^2)}$ during the algorithm. (2) and (5) hold by construction. (3) holds because conclusion (3) of Lemma 5.10 implies that at most $\eta^2 |X| |Y|$ total size of pieces is thrown out as each recursive layer, so $O(\eta^2 \log(1/\eta) |X| |Y|)$ total. Additionally, by conclusion (5) of Lemma 5.10 the total size of pieces being recursed on after $O(\log(1/\eta))$ layers is only $2^{-O(\log(1/\eta))} |X| |Y| \leq \eta^2 |X| |Y|$. So the total amount size of pieces that are either thrown out or not partitioned is at most $O(\eta^2 \log(1/\eta) |X| |Y|) \leq \eta |X| |Y|$ as desired. $\square$

Now we apply Lemma 5.9 to prove Theorem 5.7.

Proof of Theorem 5.7. Set $\eta = \varepsilon \alpha \delta_D / 16$ and let $\{V_i, X_i, x_i, Y_i, y_i\}_{i \in [T]}$ be as given by Lemma 5.9 with the choice $\varepsilon \rightarrow \varepsilon/16$. We will show some choice of i satisfies the conclusions of the theorem; thus, (1), (3), (4) follow by construction. In order to verify (2) and (5), we want to compare the sizes of $A \cap S(X_i, Y_i, D)$ and $S(X_i, Y_i, D)$, while also comparing these to the “original” densities δ_X, δ_Y , and most importantly δ_D . This motivates us to consider the quantity

$$\begin{aligned} & \sum_{i \in [T]} (|A \cap S(X_i, Y_i, D)| - (1 - 4\varepsilon) \alpha |S(X_i, Y_i, D)| - \varepsilon \alpha \delta_D |X_i| |Y_i|) \\ & \geq |A| - \eta |X| |Y| - (1 - 4\varepsilon) \alpha |S(X, Y, D)| - \varepsilon \alpha \delta_D |X| |Y| \\ & \geq \alpha \delta |W|^2 - \varepsilon \alpha \delta_D |X| |Y| - (1 - 4\varepsilon) (1 + \varepsilon) \alpha \delta |W|^2 - \varepsilon \alpha \delta_D |X| |Y| \geq 0, \end{aligned}$$

where we have used that $\delta = \delta_X \delta_Y \delta_D$, and $|S(X, Y, D)| \leq (1 + \varepsilon) \delta |W|^2$ by [Lemma 5.5](#) because one of X or Y is $(r, \varepsilon/8)$ -algebraically spread. By averaging, there exists an index i for which

$$|A \cap S(X_i, Y_i, D)| \geq (1 - 4\varepsilon) \alpha |S(X_i, Y_i, D)| + \varepsilon \alpha \delta_D |X_i| |Y_i|.$$

Note that if $X_i \subseteq V_i + x_i$ and $Y_i \subseteq V_i + y_i$ then $S(X_i, Y_i, D) = S(X_i, Y_i, D_i)$ for $D_i := D \cap (V_i + x_i + y_i)$. Thus

$$|A \cap S(X_i, Y_i, D_i)| \geq (1 - 4\varepsilon) \alpha |S(X_i, Y_i, D_i)| \geq (1 - 5\varepsilon) \alpha \frac{|X_i| |Y_i| |D_i|}{|V_i|},$$

where the final inequality follows from [Lemma 5.6](#) as both X_i, Y_i are $(r, \varepsilon/16)$ -algebraically spread, which gives (5). Additionally

$$\varepsilon \alpha \delta_D |X_i| |Y_i| \leq |A \cap S(X_i, Y_i, D_i)| \leq |S(X_i, Y_i, D_i)| \leq 2 |X_i| |Y_i| \frac{|D_i|}{|V_i|},$$

because X_i, Y_i are $(r, \varepsilon/16)$ -algebraically spread. Thus, $\frac{|D_i|}{|V_i|} \geq \varepsilon \alpha \delta_D / 2$, and so (2) is satisfied. $\square$

5.4. Von Neumann lemma. In this subsection, we give suitable conditions under which a set $A \subseteq S(X, Y, Z)$ contains roughly as many corners as a random set of the same density. More specifically, let δ_X denote the density of some set X within a subspace W , and similarly for δ_Y, δ_D . Suppose $A \subseteq S(X, Y, D)$ has size $|A| = \alpha \delta_X \delta_Y \delta_D |W|^2$. For a typical choice of X, Y, D , the size of $S(X, Y, D)$ will be roughly $\delta_X \delta_Y \delta_D |W|^2$ (see [Lemma 5.6](#)), so one should morally view α as the density of A within its container.

Define the trilinear form $\Phi(f_1, f_2, f_3) := \mathbb{E}_{x, y, z \in W} [f_1(x, y) f_2(y + z, y) f_3(x, x + z)]$. Observe that by the change of variable $z \rightarrow x + y + z$, $\Phi(f_1, f_2, f_3) = \mathbb{E}_{x, y, z \in W} [f_1(x, y) f_2(x + z, y) f_3(x, y + z)]$, so that $\Phi(\mathbb{1}_A, \mathbb{1}_A, \mathbb{1}_A)$ counts the number of corners in A (up to normalization). We will show that if A is sufficiently pseudorandom, then $\Phi(\mathbb{1}_A, \mathbb{1}_A, \mathbb{1}_A)$ is approximately as large as $\alpha^3 \delta_X^2 \delta_Y^2 \delta_D^2$, which is the number of corners in randomly chosen A, X, Y, Z of the same densities. Concretely, we will require two grid norms related to $\mathbb{1}_A$ be bounded, as well as requiring that A has no columns which are too sparse.

Lemma 5.11. *Let $W \subseteq \mathbb{F}_2^n$ be a linear subspace, and let $X, Y, D \subseteq W$ be subsets of size $|X| = \delta_X |W|, |Y| = \delta_Y |W|, |D| = \delta_D |W|$. Additionally, let $A \subseteq S(X, Y, D)$ be a subset of size $|A| = \alpha \delta_X \delta_Y \delta_D |W|^2$. Let $\varepsilon \in (0, 1/10)$ and let $p = \Omega(\log(1/(\alpha \delta_D))/\varepsilon^4)$ be a positive integer. Define the functions $F_1 : Y \times D \rightarrow \{0, 1\}$ where $F_1(y, z) = \mathbb{1}_A(y + z, y)$ and $F_2 : X \times D \rightarrow \{0, 1\}$ where $F_2(x, z) = \mathbb{1}_A(x, x + z)$. Suppose the following conditions hold:*

- (1) $\|F_1\|_{G(2, p)} < (1 + \varepsilon^2/36) \alpha \delta_X$,
- (2) $\|F_2\|_{G(2, p)} < 2 \alpha \delta_Y$,
- (3) For all $y \in Y$, we have $\mathbb{E}_{x \in W} [\mathbb{1}_A(x, y)] \geq (1 - \varepsilon^2/36) \alpha \delta_X \delta_D$.

Then,

$$\Phi(\mathbb{1}_A, \mathbb{1}_A, \mathbb{1}_A) \geq (1 - 4\varepsilon) \alpha^3 \delta_X^2 \delta_Y^2 \delta_D^2.$$

Proof. For clarity, let $f := \mathbb{1}_A$ denote the indicator function for A . Additionally, let $S = S(W, Y, D) = \{(y + z, y) \in W \times W : y \in Y, z \in D\}$, and let $g = f - \alpha \delta_X \mathbb{1}_S$. Then, we have

$$\Phi(f, f, f) = \Phi(f, g, f) + \alpha \delta_X \cdot \Phi(f, \mathbb{1}_S, f).$$

We will proceed by lower bounding the second term. Afterwards, we will upper bound the magnitude of the first term, showing it is ultimately dominated by the second.

Using the observation that whenever $f(x, y) = 1$ and $f(x, x + z) = 1$ we must have $x \in X, y \in Y, z \in D$, we can lower bound

$$\Phi(f, \mathbb{1}_S, f) = \mathbb{E}_{x, y, z \in W} [f(x, y) \mathbb{1}_S(y + z, y) f(x, x + z)]$$

$$\begin{aligned}
&= \mathbb{E}_{x,y,z \in W} [f(x,y) \mathbb{1}_Y(y) \mathbb{1}_D(z) f(x, x+z)] \\
&= \mathbb{E}_{x,y,z \in W} [f(x,y) f(x, x+z)] \\
&= \mathbb{E}_{x \in W} \left(\mathbb{E}_{y \in W} f(x,y) \right)^2 \geq \delta_X^{-1} \left(\mathbb{E}_{x,y \in W} f(x,y) \right)^2 = \alpha^2 \delta_X \delta_Y^2 \delta_D^2,
\end{aligned}$$

where we used the Cauchy-Schwarz inequality in the last line. In particular,

$$\alpha \delta_X \cdot \Phi(f, \mathbb{1}_S, f) \geq \alpha^3 \delta_X^2 \delta_Y^2 \delta_D^2.$$

Thus, it remains to bound the magnitude of $\Phi(f, g, f)$. Let $k = 2 \lceil \log(1/(\alpha \delta_D)) \rceil$. Hölder's inequality gives

$$\begin{aligned}
\left| \mathbb{E}_{x,y,z \in W} f(x,y) g(y+z,y) f(x, x+z) \right| &= \delta_X \delta_Y \delta_D \left| \mathbb{E}_{x \in X, y \in Y, z \in D} f(x,y) g(y+z,y) f(x, x+z) \right| \\
&\leq \delta_X \delta_Y \delta_D \left(\mathbb{E}_{x \in X, y \in Y} f(x,y) \right)^{\frac{k-1}{k}} \\
&\quad \cdot \left(\mathbb{E}_{x \in X, y \in Y} \left| \mathbb{E}_{z \in D} g(y+z,y) f(x, x+z) \right|^k \right)^{1/k} \\
&= \delta_X \delta_Y \delta_D (\alpha \delta_D)^{1-\frac{1}{k}} \cdot \left(\mathbb{E}_{\substack{x \in X, y \in Y \\ z_1, \dots, z_k \in D}} \prod_{i=1}^k g(y+z_i, y) f(x, x+z_i) \right)^{1/k}.
\end{aligned}$$

By our choice of k , we have $(\alpha \delta_D)^{-1/k} \leq 2$, so it suffices to bound the second factor by $2\varepsilon \alpha^2 \delta_X \delta_Y$. Define the function $G : Y \times D \rightarrow [-1, 1]$ as $G(y, z) = g(y+z, y)$. Now, apply the Cauchy-Schwarz inequality to get

$$\begin{aligned}
\left(\mathbb{E}_{\substack{x \in X, y \in Y \\ z_1, \dots, z_k \in D}} \prod_{i=1}^k g(y+z_i, y) f(x, x+z_i) \right)^{1/k} &\leq \left(\mathbb{E}_{\substack{y_1, y_2 \in Y \\ z_1, \dots, z_k \in D}} \prod_{i=1}^k g(y_1+z_i, y_1) g(y_2+z_i, y_2) \right)^{1/(2k)} \\
&\quad \cdot \left(\mathbb{E}_{\substack{x_1, x_2 \in X \\ z_1, \dots, z_k \in D}} \prod_{i=1}^k f(x_1, x_1+z_i) f(x_2, x_2+z_i) \right)^{1/(2k)} \\
&= \left(\mathbb{E}_{\substack{y_1, y_2 \in Y \\ z_1, \dots, z_k \in D}} \prod_{i=1}^k G(y_1, z_i) G(y_2, z_i) \right)^{1/(2k)} \\
&\quad \cdot \left(\mathbb{E}_{\substack{x_1, x_2 \in X \\ z_1, \dots, z_k \in D}} \prod_{i=1}^k F_2(x_1, z_i) F_2(x_2, z_i) \right)^{1/(2k)} \\
&= \|G\|_{G(2,k)} \cdot \|F_2\|_{G(2,k)}.
\end{aligned}$$

The second hypothesis, along with the fact that grid norms are monotonic, bounds the second factor by $2\alpha \delta_Y$.

We will finish the proof by bounding the first factor by $\varepsilon \alpha \delta_X$. To do this, we will use the third item of the hypothesis as well as [Lemma 4.2](#) to reduce this quantity to the first hypothesis. First, note that

$$\|F_1\|_{G(1,1)} = (\delta_Y \delta_D)^{-1} \mathbb{E}_{y,z \in W} [f(y+z, y)] = \alpha \delta_X$$

by the definition of α . Thus $G = F_1 - \|F_1\|_{G(1,1)}$ by the definition of g . Observe the third item of the hypothesis guarantees lower-boundedness on the rows of G ; namely for all $y \in Y$, we have that

$$\mathbb{E}_{z \in D} [G(y, z)] = \delta_D^{-1} \mathbb{E}_{z \in W} [g(y + z, y)] \geq (1 - \varepsilon^2/36)\alpha\delta_X - \alpha\delta_X = -\varepsilon^2\alpha\delta_X/36.$$

Combining with the first item of the hypothesis, the contrapositive of [Lemma 4.2](#) gives that

$$\|G\|_{G(2,k)} < \varepsilon\alpha\delta_X,$$

as desired. $\square$

5.5. Density increment. Now we prove that if the assumptions of [Lemma 5.11](#) are not satisfied, we can pass to a subset $X' \times Y' \times Z'$ where (essentially) the density of A in $S(X', Y', Z')$ increases by a constant factor. There are two ways to obtain this density increment: an appropriate grid norm is large, or there are too many sparse rows.

Let α be such that $|A| = \alpha\delta_X\delta_Y\delta_D|W|^2$, where as before we morally view α as the density of A within $S(X, Y, D)$. To start, let $\tilde{\varepsilon} > 0$ be a parameter and define $L = \{y \in Y : \mathbb{E}_x[\mathbb{1}_A(x, y)] < (1 - \tilde{\varepsilon})\alpha\delta_X\delta_D\} \subseteq Y$ to be the set of sparse rows of A . We will later set $\tilde{\varepsilon} = \varepsilon^2/36$, motivated by [Lemma 5.11](#). First we obtain a density increment in the case where L is large.

Lemma 5.12. *Let $X, Y, D \subseteq W$ and $\varepsilon_L \in (0, 1/2)$. If $|L| \geq \varepsilon_L|Y|$, then there is some subset $Y' \subseteq Y$ of density $\delta_{Y'} \geq \delta_Y/2$ such that*

$$\frac{|A \cap S(X, Y', D)|}{\delta_X\delta_{Y'}\delta_D|W|^2} \geq (1 + \tilde{\varepsilon}\varepsilon_L/2)\alpha.$$

Proof. Let L' be an arbitrary subset of L of size $|L'| = \varepsilon_L|Y|$ and set $Y' = Y \setminus L'$. We have

$$\begin{aligned} |A \cap S(X, Y', D)| &\geq |A| - |L'| |W| (1 - \tilde{\varepsilon})\alpha\delta_X\delta_D = \alpha\delta_X\delta_Y\delta_D|W|^2 - (1 - \tilde{\varepsilon})\varepsilon_L\alpha\delta_X\delta_Y\delta_D|W|^2 \\ &= (1 - (1 - \tilde{\varepsilon})\varepsilon_L)\alpha\delta_X\delta_Y\delta_D|W|^2. \end{aligned}$$

Thus,

$$\frac{|A \cap S(X, Y', D)|}{\delta_X\delta_{Y'}\delta_D|W|^2} \geq \frac{1 - (1 - \tilde{\varepsilon})\varepsilon_L}{1 - \varepsilon_L}\alpha \geq (1 + \tilde{\varepsilon}\varepsilon_L/2)\alpha.$$

This completes the proof. $\square$

The next lemma says that if one of the grid norms considered in [Lemma 5.11](#) is large, then A admits a density increment.

Lemma 5.13. *Let k be a positive integer, $\varepsilon \in (0, 1)$, and set $\varepsilon_s = \Theta(\varepsilon)$ sufficiently small. Let $X, Y, D \subseteq W$ and $A \subseteq S(X, Y, D)$. Define the functions $F_1 : Y \times D \rightarrow \{0, 1\}$ where $F_1(y, z) = \mathbb{1}_A(y + z, y)$ and $F_2 : X \times D \rightarrow \{0, 1\}$ where $F_2(x, z) = \mathbb{1}_A(x, x + z)$. Suppose X and Y are both (r, ε_s) -algebraically spread for*

$$r \geq \Omega(\varepsilon^{-7} \log(1/(\delta_X\delta_Y)) \log(1/(\delta_X\delta_Y\delta_D\gamma)))^7,$$

where

$$\gamma \leq (\alpha\delta_X\delta_Y)^{O(\varepsilon^{-2}k \log(1/\alpha)^2 + \varepsilon^{-1}k \log(1/(\delta_X\delta_Y)))}.$$

Then if $\|F_1\|_{G(2,k)} \geq (1 + \varepsilon/32)\alpha\delta_X$, then there are $Y' \subseteq Y$ and $D' \subseteq D$ with $|Y'| \geq (\varepsilon\alpha/2)^{O(\varepsilon^{-1}k^2 \log(1/\alpha))}|Y|$ and $|D'| \geq (\varepsilon\alpha/2)^{O(\varepsilon^{-1} \log(1/\alpha))}|D|$, and

$$|A \cap S(X, Y', D')| \geq (1 + \Omega(\varepsilon))\alpha\delta_X|Y'||D'|.$$

Similarly, if $\|F_2\|_{G(2,k)} \geq 2\alpha\delta_Y$, then there are $X' \subseteq X$ and $D' \subseteq D$ with $|X'| \geq (\varepsilon\alpha/2)^{O(\varepsilon^{-1}k^2 \log(1/\alpha))}|X|$ and $|D'| \geq (\varepsilon\alpha/2)^{O(\varepsilon^{-1} \log(1/\alpha))}|D|$, and

$$|A \cap S(X', Y, D')| \geq (1 + \Omega(\varepsilon))\alpha\delta_Y|X'||D'|.$$

Proof. We only prove the former assertion, as the latter one has an identical proof. Set $\tau = (1 + 8\varepsilon_s)\delta_X$. We would like to argue that the set $T := \{(y, z) \in Y \times D : y + z \in X\}$ is (τ, γ) -combinatorially spread. We will start with the set $T' \subseteq W \times W$, which we define to be $T' := \{(y + z, y) : y, z \in W, y + z \in X\} = S(X, W, W)$. By Lemma 5.4 as well as the (r, ε_s) -algebraic spreadness of X for r large enough, we know that T' is $(\tau, \delta_Y \delta_D \gamma)$ -combinatorially spread. This is enough to imply the desired combinatorial spreadness of T , since

$$\begin{aligned} \mathbb{E}_{y \in Y, z \in D} [\mathbb{1}_T(y, z) g_1(y) g_2(z)] &= (\delta_Y \delta_D)^{-1} \mathbb{E}_{y, z \in W} [\mathbb{1}_{T'}(y + z, y) g_1(y) g_2(z)] \\ &\leq (\delta_Y \delta_D)^{-1} \left(\tau \mathbb{E}_{y \in W} [g_1] \mathbb{E}_{z \in W} [g_2] + \delta_Y \delta_D \gamma \right) \\ &= \tau \mathbb{E}_{y \in Y} [g_1] \mathbb{E}_{z \in D} [g_2] + \gamma. \end{aligned}$$

If we write our grid norm assumption in terms of τ , we have

$$\|F_1\|_{G(2, k)} \geq (1 + \varepsilon/32) \alpha \delta_X = \left(\frac{1 + \varepsilon/32}{1 + 8\varepsilon_s} \right) \cdot \alpha \tau \geq (1 + \varepsilon/64) \alpha \tau.$$

Thus applying Theorem 3.5 and Lemma 3.3 with sufficiently small ε_s , gives that there are functions $g_1 : Y \rightarrow \{0, 1\}$ and $g_2 : D \rightarrow \{0, 1\}$ satisfying:

$$\begin{aligned} \mathbb{E}_{y \in Y, d \in D} [F_1(y, d) g_1(y) g_2(d)] &\geq (1 - \varepsilon_s) (1 + \varepsilon/64) \alpha \tau \mathbb{E}_{y \in Y} [g_1(y)] \mathbb{E}_{d \in D} [g_2(d)] \\ &\geq (1 + \varepsilon/128) \alpha \tau \mathbb{E}_{y \in Y} [g_1(y)] \mathbb{E}_{d \in D} [g_2(d)] \\ &\geq (1 + \varepsilon/128) \alpha \delta_X \mathbb{E}_{y \in Y} [g_1(y)] \mathbb{E}_{d \in D} [g_2(d)] \end{aligned}$$

and

$$\mathbb{E}_{y \in Y} [g_1(y)] \geq (\varepsilon \alpha / 2)^{O(\varepsilon^{-1} k^2 \log(1/\alpha))} \quad \text{and} \quad \mathbb{E}_{d \in D} [g_2(d)] \geq (\varepsilon \alpha / 2)^{O(\varepsilon^{-1} \log(1/\alpha))}.$$

Letting Y' and D' be the indicator functions of g_1 and g_2 respectively completes the proof. $\square$

5.6. Obtaining spreadness. In this section we will use a density increment algorithm to reach a state $A \subseteq S(X, Y, D)$ where the conditions of the von Neumann Lemma (Lemma 5.11) are satisfied, and both X, Y are algebraically spread. A useful definition is the notion of asymmetric combinatorial spreadness. Intuitively, this says that the function f does not admit a density increment onto somewhat larger subrectangles, where we control the densities of the rows and columns separately.

Definition 5.14 (Asymmetric combinatorial spreadness). We say that a function $f : \Omega_1 \times \Omega_2 \rightarrow [0, 1]$ is (s, t, ε) -combinatorially spread if for all functions $g_1 : \Omega_1 \rightarrow \{0, 1\}$ and $g_2 : \Omega_2 \rightarrow \{0, 1\}$ with

$$\mathbb{E}[g_1(x)] \geq 2^{-s} \quad \text{and} \quad \mathbb{E}[g_2(y)] \geq 2^{-t},$$

it holds that

$$\mathbb{E}[f(x, y) g_1(x) g_2(y)] \leq (1 + \varepsilon) \mathbb{E}[f] \mathbb{E}[g_1] \mathbb{E}[g_2].$$

We will apply this definition with s much larger than t . This corresponds to the fact that in our proof δ_D is much larger than δ_X and δ_Y throughout. Note also that unlike Definition 3.1, the g_i are simply subsets of Ω_i . Now, we specialize the above definition to define what it means for a set A to be combinatorially spread within the container $S(X, Y, D)$.

Definition 5.15. Let $A \subseteq \mathbb{F}_2^n \times \mathbb{F}_2^n$, $W \subseteq \mathbb{F}_2^n$ be a linear subspace, and $X, Y, D \subseteq W$. Let f be the indicator function of $A \cap S(X, Y, D)$. We say that A is (s, t, ε) -combinatorially spread in a container $S(X, Y, D)$ if the functions $F_1 : Y \times D \rightarrow \{0, 1\}$, $F_2 : X \times D \rightarrow \{0, 1\}$ defined as

$$F_1(y, d) = f(y + d, y) \quad \text{and} \quad F_2(x, d) = f(x, x + d)$$

are (s, t, ε) -combinatorially spread.

Our main lemma says that if one repeatedly does a density increment followed by pseudorandomization, then we reach a state where A is combinatorially spread, both X, Y are algebraically spread, and A has lower-bounded rows. Additionally, the densities of X, Y, D have not dropped too much, and the dimension of the subspace we are working in has not decreased significantly.

Lemma 5.16. *Let $r, s, t \geq 1$ and $\varepsilon \in (0, 1/64)$. Suppose that $A \subseteq \mathbb{F}_2^n \times \mathbb{F}_2^n$ has size $|A| = \alpha 4^n$. Then, there exists a subspace $W \subseteq \mathbb{F}_2^n$ along with points $x, y \in \mathbb{F}_2^n/W$ and sets $X \subseteq W + x, Y \subseteq W + y, D \subseteq W + x + y$ with sizes $|X| = \delta_X |W|, |Y| = \delta_Y |W|$, and $|D| = \delta_D |W|$ satisfying:*

- (1) X, Y are $(r, 5\varepsilon^{1/2})$ -algebraically spread in $W + x, W + y$, respectively,
- (2) $\delta_D \geq \bar{\delta}_D := \exp(-O(t\varepsilon^{-1} \log(1/\alpha) + \varepsilon^{-2} \log(1/\alpha)^2))$,
- (3) $\min\{\delta_X, \delta_Y\} \geq \bar{\delta} := \exp(-O(s\varepsilon^{-1} \log(1/\alpha) + \varepsilon^{-1} \log(1/(\varepsilon\alpha\bar{\delta}_D))^2 \log(1/\alpha)))$,
- (4) $\dim(W) \geq n - O(r\varepsilon^{-3} \log(1/(\varepsilon\alpha\bar{\delta}_D\bar{\delta}))^2 \log(1/(\varepsilon\alpha\bar{\delta}_D)) \log(1/\alpha) + r\varepsilon^{-3} \log(1/(\varepsilon\alpha\bar{\delta}_D))^5 \log(1/\alpha))$,
- (5) A is $(s, t, 5\varepsilon^{1/2})$ -combinatorially spread within the container $S(X, Y, D)$ (see [Definition 5.15](#)),
- (6) The rows of $A \cap S(X, Y, D)$ are lower bounded; namely, for all $y \in Y$,

$$\mathbb{E}_{x \in W} [\mathbb{1}_{A \cap S(X, Y, D)}(x, y)] \geq (1 - 2\varepsilon^{1/2})\alpha^* \delta_X \delta_D$$

$$\text{where } \alpha^* = \frac{|A \cap S(X, Y, D)|}{\delta_X \delta_Y \delta_D |W|^2} \text{ and } \alpha^* \geq \alpha,$$

as long as $r \geq \Omega(\varepsilon^{-8} \log(1/\bar{\delta})^8)$ large enough.

Proof. Perform the following algorithm, beginning with $X, Y, D = \mathbb{F}_2^n$. As long as A is not $(s+1, t, \varepsilon)$ -combinatorially spread in its container $S(X, Y, D)$, iteratively restrict to containers $S(X_1, Y, D_1)$ or $S(X, Y_1, D_1)$ to obtain a $(1 + \varepsilon)$ density increment (as is guaranteed by [Definition 5.15](#)), followed by applying [Theorem 5.7](#) with $\varepsilon \rightarrow 8\varepsilon_s$ for sufficiently small $\varepsilon_s := \Theta(\varepsilon)$ at each step to reestablish algebraic pseudorandomness of X, Y . At the end, remove all rows from A that violate (6).

We start by analyzing how X, Y, D change in one iteration of the algorithm, before the row removal phase. If A is not $(s+1, t, \varepsilon)$ -combinatorially spread, without loss of generality we can find subsets $X_1 \subseteq X$ and $D_1 \subseteq D$ with $|X_1| \geq 2^{-s-1}|X|$ and $|D_1| \geq 2^{-t}|D|$ where

$$\mathbb{E}_{x \in X, d \in D} [\mathbb{1}_{X_1}(x) \mathbb{1}_{D_1}(d) f(x, x+d)] \geq (1 + \varepsilon) \mathbb{E}_{x \in X, d \in D} [f(x, x+d)] \cdot \mathbb{E}_{x \in X} [\mathbb{1}_{X_1}] \cdot \mathbb{E}_{d \in D} [\mathbb{1}_{D_1}],$$

or equivalently after normalizing,

$$\mathbb{E}_{x \in X_1, d \in D_1} [f(x, x+d)] \geq (1 + \varepsilon) \mathbb{E}_{x \in X, d \in D} [f(x, x+d)].$$

In other words,

$$\frac{|A \cap S(X_1, Y, D_1)|}{|X_1||D_1|} \geq (1 + \varepsilon) \frac{|A \cap S(X, Y, D)|}{|X||D|}.$$

Let $\alpha_1 = \frac{|A \cap S(X_1, Y, D_1)|}{\delta_{X_1} \delta_Y \delta_{D_1} |W|^2}$, and note that $\alpha_1 \geq (1 + \varepsilon) \frac{|A \cap S(X, Y, D)|}{\delta_X \delta_Y \delta_D |W|^2}$ by the above inequality. Set $\delta = \delta_{X_1} \delta_Y \delta_{D_1}$, and recall that $\varepsilon_s = \Theta(\varepsilon)$ small enough. At this point, we will also require that $r \geq \Omega(\varepsilon_s^{-8} (\log(1/(\varepsilon\alpha_1\delta_{D_1}))^2 + \log(1/\delta))^8)$ large enough. Now, we apply [Theorem 5.7](#) with $A \cap S(X_1, Y, D_1)$ ¹, r , and ε_s to obtain sets X', Y', D' along with shifts x, y so that

- (1) $\dim(W') \geq n - O(r\varepsilon_s^{-2} \log(1/(\varepsilon_s\alpha\delta))^2 \log(1/(\varepsilon_s\alpha_1\delta_{D_1})) + r\varepsilon_s^{-2} \log(1/(\varepsilon_s\alpha_1\delta_{D_1}))^5)$,
- (2) $\frac{|D'|}{|W'|} \geq \varepsilon_s \alpha_1 \delta_{D_1} / 2$,

¹In general, if working in a container $S(X, Y, D)$ with $X \subseteq W + x, Y \subseteq W + y, D \subseteq W + x + y$, we can shift A by (x, y) to work in a container which is contained in $W \times W$ as prescribed by [Theorem 5.7](#), followed by shifting the new container back by (x, y) . The only conclusion which could be affected is (4), but this is not an issue since if $X \subseteq W$ is algebraically spread, then $X + x \subseteq W + x$ is algebraically spread with the same parameters.

- (3) $\frac{|X'| |Y'|}{|W'|^2} \geq e^{-O(\log(1/(\varepsilon \alpha_1 \delta_{D_1})))^2} \delta_{X_1} \delta_{Y_1}$,
(4) X' and Y' are $(r, \varepsilon_s/16)$ -algebraically spread in $W' + x, W' + y$, respectively, and
(5) $|A \cap S(X', Y', D')| \geq (1 - 5\varepsilon_s) \alpha_1 \delta_{X'} \delta_{Y'} \delta_{D'} |W'|^2$ where $\delta_{X'} = \frac{|X'|}{|W'|}$, $\delta_{Y'} = \frac{|Y'|}{|W'|}$, and $\delta_{D'} = \frac{|D'|}{|W'|}$.

For $\varepsilon_s \leq O(\varepsilon)$ small enough, (5) and our earlier lower bound on α_1 give

$$\frac{|A \cap S(X', Y', D')|}{\delta_{X'} \delta_{Y'} \delta_{D'} |W'|^2} \geq (1 + \varepsilon/2) \frac{|A \cap S(X, Y, D)|}{\delta_X \delta_Y \delta_D |W|^2}.$$

If we can show that $|S(X', Y', D')|$ is roughly $\delta_{X'} \delta_{Y'} \delta_{D'} |W'|^2$, then we can use the left-hand side as a measure of progress, since the density of A in $S(X', Y', D')$ cannot exceed 1. Thus, it will follow that this process ends in at most $O(\varepsilon^{-1} \log(1/\alpha))$ many iterations, since $\alpha = \frac{|A \cap S(X, Y, D)|}{\delta_X \delta_Y \delta_D |W|^2}$. By (2) and (3), we know that

$$\delta_{X'} \delta_{Y'} \delta_{D'} \geq \varepsilon_s \alpha_1 e^{-O(\log(1/(\varepsilon \alpha_1 \delta_{D_1})))^2} \delta_{X_1} \delta_{Y_1} \delta_{D_1}.$$

Thus, our choice of r guarantees that $r \geq \Omega(\varepsilon_s^{-8} \log(1/(\delta_{X'} \delta_{Y'} \delta_{D'}))^8)$. In general, we have chosen r sufficiently large to satisfy such an inequality for all iterations. Hence, [Lemma 5.6](#) implies that $|S(X', Y', D')| \leq (1 + \varepsilon_s) \delta_{X'} \delta_{Y'} \delta_{D'} |W'|^2$. In particular,

$$1 \geq \frac{|A \cap S(X', Y', D')|}{|S(X', Y', D')|} = \frac{|A \cap S(X', Y', D')|}{\delta_{X'} \delta_{Y'} \delta_{D'} |W'|^2} \cdot \frac{\delta_{X'} \delta_{Y'} \delta_{D'} |W'|^2}{|S(X', Y', D')|} \geq \frac{|A \cap S(X', Y', D')|}{\delta_{X'} \delta_{Y'} \delta_{D'} |W'|^2} \cdot \frac{1}{1 + \varepsilon_s},$$

so $|A \cap S(X', Y', D')|/\delta_{X'} \delta_{Y'} \delta_{D'} |W'|^2$ cannot exceed $1 + \varepsilon_s$, and at each iteration this quantity increases by a factor of $1 + \varepsilon/2$. Thus, if we iterate this process for $O(\varepsilon^{-1} \log(1/\alpha))$ many iterations, we obtain sets X^*, Y^*, D^* along with a subspace W^* and shifts x^*, y^* so that A is $(s+1, t, \varepsilon)$ -combinatorially spread in the container $S(X^*, Y^*, D^*)$. Additionally, X^*, Y^* are (r, ε_s) -algebraically spread in $W^* + x^*, W^* + y^*$ by construction, which verifies (1).

We now verify the size lower bounds on X^*, Y^*, D^* as well as the dimension bound on W^* . First, we lower bound δ_{D^*} . At every iteration, the density of D is decreasing by a factor of at most 2^{-t} due to the combinatorial density increment, and then by an additional factor of at most $\varepsilon_s \alpha/2$ due to the application of [Theorem 5.7](#), which gives

$$\frac{|D^*|}{|W^*|} \geq (\varepsilon_s \alpha 2^{-t})^{O(\varepsilon^{-1} \log(1/\alpha))} \geq \exp(-O(t\varepsilon^{-1} \log(1/\alpha) + \varepsilon^{-2} \log(1/\alpha)^2)) =: \bar{\delta}_D$$

for our choice of ε_s . Similarly, X, Y decrease in density by a factor of at most 2^{-s-1} due to the combinatorial density increment, and then by an additional factor of $e^{-O(\log(1/(\varepsilon \alpha \bar{\delta}_D)))^2}$ due to the application of [Theorem 5.7](#). Note that $\log(1/(\varepsilon \alpha \bar{\delta}_D)) \leq O(t\varepsilon^{-1} \log(1/\alpha) + \varepsilon^{-2} \log(1/\alpha)^2)$, yielding

$$\begin{aligned} \frac{|X^*|}{|W^*|}, \frac{|Y^*|}{|W^*|} &\geq \left(2^{-O(s + \log(1/(\varepsilon \alpha \bar{\delta}_D)))^2}\right)^{O(\varepsilon^{-1} \log(1/\alpha))} \\ &\geq \exp(s\varepsilon^{-1} \log(1/\alpha) + \varepsilon^{-1} \log(1/(\varepsilon \alpha \bar{\delta}_D))^2 \log(1/\alpha)) =: \bar{\delta}. \end{aligned}$$

Finally, we can verify the dimension bound on W^* . At each iteration, the application of [Theorem 5.7](#) decreases the dimension of W by

$$O(r\varepsilon_s^{-2} \log(1/(\varepsilon_s \alpha \delta^*))^2 \log(1/\varepsilon \alpha \bar{\delta}_D) + r\varepsilon_s^{-2} \log(1/\varepsilon \alpha \bar{\delta}_D)^5)$$

for $\delta^* = \bar{\delta}_D \bar{\delta}^2$. This gives

$$\dim(W^*) \geq n - O(r\varepsilon^{-3} \log(1/\alpha) \log(1/(\varepsilon \alpha \delta^*))^2 \log(1/(\varepsilon \alpha \bar{\delta}_D)) + r\varepsilon^{-3} \log(1/\alpha) \log(1/(\varepsilon \alpha \bar{\delta}_D))^5).$$

Notice also that on the last iteration of the algorithm, we need spreadness for $r \geq \varepsilon^{-8} \log(1/\delta^*)^8$ in order to control the size of $S(X^*, Y^*, D^*)$, which accounts for the lower bound on r in the theorem statement.

To complete the proof of the lemma, we must guarantee that $A \cap S(X^*, Y^*, D^*)$ has lower-bounded rows. To achieve this, we will simply remove all of the rows which violate (6). More formally, let f be the indicator of $A \cap S(X^*, Y^*, D^*)$ and $\alpha^* = |A \cap S(X^*, Y^*, D^*)|/(\delta_{X^*} \delta_{Y^*} \delta_{D^*} |W^*|^2)$, and let $L = \{y \in Y^* : \mathbb{E}_{x \in W^*} [f(x, y)] < (1 - \varepsilon^{1/2}) \alpha^* \delta_{X^*} \delta_{D^*}\}$. We may assume $|L| \leq 4\varepsilon^{1/2} |Y^*|$, as otherwise [Lemma 5.12](#) implies there exists a subset $Y^+ \subseteq Y^*$ with $\delta_{Y^+} \geq \delta_{Y^*}/2$ such that

$$\frac{|A \cap S(X^*, Y^+, D^*)|}{\delta_{X^*} \delta_{Y^+} \delta_{D^*} |W^*|^2} \geq (1 + \varepsilon) \alpha^*,$$

which contradicts the (s, t, ε) -combinatorial spreadness of A in $S(X^*, Y^*, D^*)$. Thus, we define $Y^+ := Y \setminus L$ so that $|Y^+| \geq (1 - 4\varepsilon^{1/2}) |Y^*|$. Define

$$\alpha^+ := \frac{|A \cap S(X^*, Y^+, D^*)|}{\delta_{X^*} \delta_{Y^+} \delta_{D^*} |W^*|^2}.$$

Note that $\alpha^+ \geq \alpha^*$ since we only deleted sparse rows. Additionally, we have $\alpha^+ \leq (1 + \varepsilon) \alpha^+$ by $(s+1, t, \varepsilon)$ -combinatorial spreadness. Thus, in terms of the new density α^+ , all of the columns with $y \in Y^+$ satisfy

$$\mathbb{E}_{x \in G} [\mathbb{1}_{S(X^*, Y^+, D^*)}(x, y)] \geq (1 - \varepsilon^{1/2}) \alpha^* \delta_{X^*} \delta_{Y^*} \geq (1 - 2\varepsilon^{1/2}) \alpha^+ \delta_{X^*} \delta_{D^*}.$$

Consider the container $S(X^*, Y^+, D^*)$; the lemma will follow if we can show that Y^+ is (r, ε) -algebraically spread in $W^* + y^*$, and that A is (s, t, ε) -combinatorially spread in this new container. Firstly, we have $\delta_{Y^+} \geq (1 - 4\varepsilon^{1/2}) \delta_{Y^*}$. The algebraic spreadness of Y^* implies that its density on any codimension r subspace of W^* is bounded by $(1 + \varepsilon_s) \delta_{Y^*}$. Thus, it follows that Y^+ is $(r, \varepsilon_s + 4\varepsilon^{1/2})$ -algebraically spread in $W^* + y^*$. We conclude by showing that A is $(s, t, 5\varepsilon^{1/2})$ -combinatorially spread in the container $S(X^*, Y^+, D^*)$. For clarity, define $F_1 : Y^* \times D^* \rightarrow \{0, 1\}$ and $F_2 : X^* \times D^* \rightarrow \{0, 1\}$ with $F_1(y, d) = f(y + d, y)$ and $F_2(x, d) = f(x, x + d)$. The (s, t, ε) -combinatorial spreadness of F_2 follows easily, since for any $X' \subseteq X^*$ with size $|X'| \geq 2^{-(s+1)} |X^*|$ and $D' \subseteq D^*$ with size $|D'| \geq 2^{-t} |D^*|$, we have

$$|A \cap S(X', Y^+, D')| \leq |A \cap S(X', Y^*, D')| \leq (1 + \varepsilon) \alpha^* \delta_{X'} \delta_{Y^*} \delta_{D'} |W^*|^2.$$

This combined with the fact that $\delta_{Y^+} \geq (1 - 4\varepsilon^{1/2}) \delta_{Y^*}$ and $\alpha^+ \geq \alpha^*$ gives

$$|A \cap S(X', Y^+, D')| \leq (1 + 5\varepsilon^{1/2}) \alpha^+ \delta_{X'} \delta_{Y^+} \delta_{D'} |W^*|^2,$$

which is equivalent to F_2 being $(s+1, t, 5\varepsilon^{1/2})$ -combinatorially spread. Showing spreadness of F_1 is only slightly more subtle. For $Y' \subseteq Y^+$ with size $|Y'| \geq 2^{-(s+1)} |Y^*|$ and $D' \subseteq D^*$ with size $|D'| \geq 2^{-t} |D^*|$, we again have

$$|A \cap S(X^*, Y', D')| \leq (1 + \varepsilon) \alpha^* \delta_{X^*} \delta_{Y'} \delta_{D'} |W^*|^2.$$

Since $|Y^+| \geq |Y^*|/2$, this means that Y' must have density at least $2 \cdot 2^{-s-1} = 2^{-s}$ in Y^+ , which means F_1 is $(s, t, 5\varepsilon^{1/2})$ -combinatorially spread. $\square$

5.7. Completing the proof. In this short section, we combine the previous pieces we have developed to establish [Theorem 1.5](#).

Proof of Theorem 1.5. Let $A \subseteq \mathbb{F}_2^n \times \mathbb{F}_2^n$ be a set of size $|A| = \alpha 4^n$. The proof proceeds by restricting A to a large container $S(X, Y, D)$ where A is combinatorially spread using [Lemma 5.16](#). Then, we will argue that combinatorial spreadness is enough to ensure bounded grid norms using [Lemma 5.13](#). At that point, we can apply [Lemma 5.11](#) to show that A contains many corners.

Let ε be a sufficiently small constant, and let

$$r = O(\log(1/\alpha)^{156}), \quad s = O(\log(1/\alpha)^8), \quad t = O(\log(1/\alpha)^2)$$

for large enough implicit constants. By applying [Lemma 5.16](#) with $r, s, t, \varepsilon_s = O(\varepsilon^4)$ small enough, there exists a subspace $W \subseteq \mathbb{F}_2^n$ along with shifts $x, y \in \mathbb{F}_2^n$ and sets $X \subseteq W + x, Y \subseteq W + y, D \subseteq W + x + y$ satisfying the following properties:

- (1) X, Y are (r, ε) -algebraically spread in $W + x, W + y$, respectively.
- (2) $|D| \geq 2^{-O(\log(1/\alpha)^3)}|W|$
- (3) $|X|, |Y| \geq 2^{-O(\log(1/\alpha)^9)}|W|$
- (4) $\dim(W) \geq n - O(r \log(1/\alpha)^{22}) \geq n - O(\log(1/\alpha)^{178})$.
- (5) The set A is $(s, t, O(\varepsilon^2))$ -combinatorially spread in the container $S(X, Y, D)$.
- (6) The rows of $A \cap S(X, Y, D)$ are lower bounded; namely for all $y \in Y$,

$$\mathbb{E}_{x \in W} [\mathbb{1}_{A \cap S(X, Y, D)}(x, y)] \geq (1 - O(\varepsilon^2))\alpha^* \delta_X \delta_D$$

where $\alpha^* = \frac{|A \cap S(X, Y, D)|}{\delta_X \delta_Y \delta_D |W|^2}$ and $\alpha^* \geq \alpha$.

Let f denote the indicator of $A \cap S(X, Y, D)$, and define the functions $F_1: Y \times D \rightarrow \{0, 1\}$ with $F_1(y, d) = f(y, y + d)$ and $F_2: X \times D$ with $F_2(x, d) = f(x, x + d)$. We will argue that the combinatorial spreadness of A in the container $S(X, Y, D)$ implies certain grid norms are bounded. In particular, we will show that

$$\|F_1\|_{G(2, k)} \leq (1 + \varepsilon^2/36)\alpha^* \delta_X \quad \text{and} \quad \|F_2\|_{G(2, k)} \leq 2\alpha^* \delta_Y$$

for $k = O(\log(\alpha^* \delta_D)/\varepsilon^4) = O(\log(1/\alpha)^3)$ large enough. Without loss of generality, assume the first assumption does not hold. If we can verify that X is sufficiently algebraically spread, then [Lemma 5.13](#) implies that there are $Y' \subseteq Y$ and $D' \subseteq D$ with $|Y'| \geq (\varepsilon\alpha/2)^{O(k^2 \log(1/\alpha))}|Y|$ and $|D'| \geq (\varepsilon\alpha/2)^{O(\log(1/\alpha))}|D|$, and

$$|A \cap S(X, Y', D')| \geq (1 + \Omega(\varepsilon^2))\alpha \delta_X |Y'| |D'|.$$

This, however contradicts the $(O(\log(1/\alpha)^8), O(\log(1/\alpha)^2), O(\varepsilon^2))$ -combinatorial spreadness of A in the container $S(X, Y, D)$. To see why X is sufficiently spread, note that we need

$$r \geq \Omega(\log(1/(\delta_X \delta_Y)) \log(1/(\delta_X \delta_Y \delta_D \gamma))^7) \geq \Omega(\log(1/\alpha)^9 \log(1/\gamma)^7)$$

for

$$\gamma \leq (\alpha \delta_X \delta_Y)^{O(k \log(1/\alpha)^2 + k \log(1/(\alpha \delta_X \delta_Y)))} \leq (\alpha \delta_X \delta_Y)^{O(\log(1/\alpha)^{12})} \leq 2^{-O(\log(1/\alpha)^{21})}.$$

Our choice of r suffices, since

$$r \geq \Omega(\log(1/\alpha)^9 \log(1/\gamma)^7) \geq \Omega(\log(1/\alpha)^{156}).$$

Thus, the conditions of [Lemma 5.11](#) are met, which implies that

$$\Phi(f, f, f) \geq (1 - 4\varepsilon)\alpha^3 \delta_X^2 \delta_Y^2 \delta_D^2.$$

This implies that A contains at least

$$(1 - 4\varepsilon)\alpha^3 \delta_X^2 \delta_Y^2 \delta_D^2 |W|^3 \geq 2^{-O(\log(1/\alpha)^9)} |W|^3 \geq 2^{-O(\log(1/\alpha)^{178})} |\mathbb{F}_2^n|^3$$

many corners. □

6. BOHR SETS, ALGEBRAIC SPREADNESS, AND PSEUDORANDOMIZATION

The vast majority of the remainder of the body of the paper is devoted to establishing [Theorem 1.1](#), an improved corners bound over general abelian groups. The proof in many regards closely follows that in the finite field model setting, but as is standard one is forced to work with Bohr sets throughout the analysis.

6.1. Bohr sets. We now recall various standard material regarding Bohr sets. The influence of Bohr sets in additive combinatorics stems from seminal work of Bourgain [12]. For a textbook treatment, we refer the reader to [69, Section 4.4]. We first define a Bohr set of a finite abelian group G .

Definition 6.1 (Bohr set). Let $\varepsilon \in \mathbb{R}^+$, G be a finite abelian group, and $\Theta = (\Theta_1, \dots, \Theta_d)$ where $\Theta_i \in \widehat{G}$ are additive homomorphisms from G to $\mathbb{R}/\mathbb{Z}$. We define the *Bohr set*

$$\Lambda = \Lambda_{\Theta, \varepsilon} = \bigcap_{i=1}^d \left\{ x \in G : \|\Theta_i(x)\|_{\mathbb{R}/\mathbb{Z}} \leq \varepsilon \right\},$$

where $\|x\|_{\mathbb{R}/\mathbb{Z}} = \min_{z \in \mathbb{Z}} |x - z|$.

For any real number $c > 0$, we define the dilated Bohr set

$$c\Lambda_{\Theta, \varepsilon} = \Lambda_{\Theta, c\varepsilon}.$$

We will refer to d as the dimension of the Bohr set and ε as the radius. We denote the radius of a Bohr set B as $\nu(B)$.

We first require that the size of a Bohr set is lower bounded in terms of its radius and dimension. This appears as [69, Lemma 4.20].

Lemma 6.2. *Let $\Lambda = \Lambda_{\Theta, \varepsilon}$ be a Bohr set of dimension d and radius ε . Then $|\Lambda| \geq \varepsilon^d |G|$.*

For the vast majority of our analysis we will operate with regular Bohr sets as introduced by Bourgain [12]. Heuristically, regular Bohr sets are those such that altering the radius parameter causes the size of the underlying Bohr set to vary in a predictable manner.

Definition 6.3 (Regular). A Bohr set $\Lambda = \Lambda_{\Theta, \varepsilon}$ of dimension d is *regular* if for all $|c| \leq 1/(100d)$, we have that

$$1 - 100d|c| \leq \frac{|(1+c)\Lambda|}{|\Lambda|} \leq 1 + 100d|c|.$$

A crucial feature of regular Bohr sets is that they may be constructed easily (at worst at the cost of passing to radius half the size). This was established by Bourgain [12] (see also [69, Lemma 4.25]).

Lemma 6.4. *For any Bohr set Λ , there exists $\alpha \in [1/2, 1]$ such that $\alpha\Lambda$ is regular.*

Furthermore we additionally have that regular Bohr sets are essentially “shift-invariant” when shifted by elements in a smaller Bohr set. We will use such inequalities repeatedly and without substantial comment.

Lemma 6.5. *Let f be a 1-bounded function and Λ be a regular Bohr set of dimension d . If $|c| \leq 1/(100d)$ and $n' \in c\Lambda$, then*

$$\mathbb{E}_{n \in \Lambda} f(n) = \mathbb{E}_{n \in \Lambda} f(n + n') + O(cd).$$

Proof. The above claim follows as

$$\left| \mathbb{E}_{n \in \Lambda} f(n) - \mathbb{E}_{n \in \Lambda} f(n + n') \right| \leq 2 \mathbb{E}_{n \in \Lambda} \mathbb{1}[n + n' \in (1+c)\Lambda \setminus \Lambda] = O(cd). \quad \square$$

To simplify the presentation in the remainder of the paper, we define a (d, η) -small sequence of Bohr sets.

Definition 6.6 (Small & exact sequences). We say that $B_1, B_2, \dots$ is a (d, η) -small sequence of Bohr sets if all B_i have the same set of d frequencies (so that $\text{rank}(B_i) = d$), are all regular, and $\nu(B_{i+1})/\nu(B_i) \leq \eta$ for $i = 1, 2, \dots$. We say that the sequence is (d, η) -exact if additionally $\nu(B_{i+1})/\nu(B_i) \in [\eta/2, \eta]$ for all $i = 1, 2, \dots$.

Most of our lemma statements will involve a (d, η) -small sequence of Bohr sets as input hypotheses, and may output a (d', η) -exact sequence of Bohr sets. By [Lemma 6.4](#), any Bohr set B_1 with $\text{rank}(B_1) = d$ can be extended to a (d, η) -exact sequence of Bohr sets of arbitrary length.

We next require certain specialized Gowers grid norms which will be used throughout the analysis. These were introduced in the work of Milićević [\[49\]](#).

Definition 6.7 (Gowers grid norm). Fix integers $k, \ell \geq 1$. Consider a triplet of Bohr sets B_1, B_2 , and B_3 and a function $f : G \rightarrow \mathbb{R}$. We define the (B_1, B_2, B_3, k, ℓ) -Gowers grid norm to be

$$\|f\|_{(B_1, B_2, B_3, k, \ell)}^{k\ell} = \mathbb{E}_{\substack{x \sim B_1 \\ y_1, \dots, y_k \sim B_2 \\ z_1, \dots, z_\ell \sim B_3}} \prod_{\substack{i \in [k] \\ j \in [\ell]}} f(x + y_i + z_j).$$

When the Bohr sets are clear from context, we will refer to this quantity more simply as the (k, ℓ) -Gowers grid norm.

Often, we will abuse terminology and refer to the Gowers grid norm of a set X to mean the Gowers grid norm of its indicator function $\mathbb{1}_X$.

We will require that the norms increase under passing to finer Bohr sets. The key inequality for this result will be the following Gowers–Hölder inequality which appears as [\[22, Lemma 2.2\]](#).

Lemma 6.8. Fix integers $k, \ell \geq 1$ and finite sets X, Y . Consider functions $f_{ij} : X \times Y \rightarrow \mathbb{R}^{\geq 0}$. Then we have that

$$\mathbb{E}_{\substack{x_1, \dots, x_k \in X \\ y_1, \dots, y_\ell \in Y}} \prod_{\substack{i \in [k] \\ j \in [\ell]}} f_{ij}(x_i, y_j) \leq \prod_{\substack{i \in [k] \\ j \in [\ell]}} \|f_{ij}\|_{G(k, \ell)} = \prod_{\substack{i \in [k] \\ j \in [\ell]}} \left(\mathbb{E}_{\substack{x_1, \dots, x_k \in X \\ y_1, \dots, y_\ell \in Y}} \prod_{\substack{i' \in [k] \\ j' \in [\ell]}} f_{ij}(x_{i'}, y_{j'}) \right)^{1/(k\ell)}.$$

Proof. Observe via repeated application of rearranging and Hölder’s inequality that

$$\begin{aligned} \mathbb{E}_{\substack{x_1, \dots, x_k \in X \\ y_1, \dots, y_\ell \in Y}} \prod_{\substack{i \in [k] \\ j \in [\ell]}} f_{ij}(x_i, y_j) &= \mathbb{E}_{x_1, \dots, x_k \in X} \prod_{j \in [\ell]} \left(\mathbb{E}_{y_1, \dots, y_\ell \in Y} \prod_{i \in [k]} f_{ij}(x_i, y_j) \right) \\ &\leq \prod_{j \in [\ell]} \left(\mathbb{E}_{x_1, \dots, x_k \in X} \left(\mathbb{E}_{y_1, \dots, y_\ell \in Y} \prod_{i \in [k]} f_{ij}(x_i, y_j) \right)^\ell \right)^{1/\ell} \\ &= \prod_{j \in [\ell]} \left(\mathbb{E}_{\substack{x_1, \dots, x_k \in X \\ y_1, \dots, y_\ell \in Y}} \prod_{\substack{i \in [k] \\ j' \in [\ell]}} f_{ij}(x_i, y_{j'}) \right)^{1/\ell} \\ &= \prod_{j \in [\ell]} \left(\mathbb{E}_{y_1, \dots, y_\ell \in Y} \prod_{i \in [k]} \mathbb{E}_{x_1, \dots, x_k \in X} \prod_{j' \in [\ell]} f_{ij}(x_i, y_{j'}) \right)^{1/\ell} \\ &\leq \prod_{\substack{i \in [k] \\ j \in [\ell]}} \left(\mathbb{E}_{y_1, \dots, y_\ell \in Y} \left(\mathbb{E}_{x_1, \dots, x_k \in X} \prod_{j' \in [\ell]} f_{ij}(x_i, y_{j'}) \right)^k \right)^{1/(k\ell)} \\ &= \prod_{\substack{i \in [k] \\ j \in [\ell]}} \left(\mathbb{E}_{\substack{x_1, \dots, x_k \in X \\ y_1, \dots, y_\ell \in Y}} \prod_{\substack{i' \in [k] \\ j' \in [\ell]}} f_{ij}(x_{i'}, y_{j'}) \right)^{1/(k\ell)}. \quad \square \end{aligned}$$

The following approximate monotonicity of Gowers grid norms in [Definition 6.7](#) will be crucial in our analysis.

Lemma 6.9. Let $\eta \in (0, 1/(100d))$ and fix integers $k, \ell \geq 1$. Let $B_1, B_2, \dots$ be a (d, η) -small sequence of Bohr sets. Then for a function $f : G \rightarrow [0, 1]$ we have that

$$\|f\|_{(B_1, B_2, B_3, k, \ell)} \leq \|f\|_{(B_1, B_4, B_5, k, \ell)} + O((\eta d)^{1/(k\ell)}).$$

Proof. Observe that it suffices to establish that

$$\|f\|_{(B_1, B_2, B_3, k, \ell)}^{k\ell} \leq \|f\|_{(B_1, B_4, B_5, k, \ell)}^{k\ell} + O(k\ell \cdot \eta d).$$

Next using Lemma 6.5, we have that

$$\begin{aligned} \|f\|_{(B_1, B_2, B_3, k, \ell)}^{k\ell} &= \mathbb{E}_{\substack{x \sim B_1 \\ y_1, \dots, y_k \sim B_2 \\ z_1, \dots, z_\ell \sim B_3}} \prod_{\substack{i \in [k] \\ j \in [\ell]}} f(x + y_i + z_j) \\ &\leq \mathbb{E}_{\substack{x \sim B_1 \\ y_1, \dots, y_k \sim B_2 \\ z_1, \dots, z_\ell \sim B_3 \\ y'_1, \dots, y'_k \sim B_4 \\ z'_1, \dots, z'_\ell \sim B_5}} \prod_{\substack{i \in [k] \\ j \in [\ell]}} f(x + y_i + z_j + y'_i + z'_j) + O(k\ell \cdot \eta d). \end{aligned}$$

Taking the expectation on x, y_i, z_j outside and defining $f_{i,j}(y, z) = f(x + y_i + z_j + y + z)$ we observe that the final quantity is exactly as in Lemma 6.8. Therefore we have that

$$\begin{aligned} \mathbb{E}_{\substack{x \sim B_1 \\ y_1, \dots, y_k \sim B_2 \\ z_1, \dots, z_\ell \sim B_3 \\ y'_1, \dots, y'_k \sim B_4 \\ z'_1, \dots, z'_\ell \sim B_5}} \prod_{\substack{i \in [k] \\ j \in [\ell]}} f(x + y_i + z_j + y'_i + z'_j) &\leq \mathbb{E}_{\substack{x \sim B_1 \\ y_1, \dots, y_k \sim B_2 \\ z_1, \dots, z_\ell \sim B_3}} \prod_{\substack{i \in [k] \\ j \in [\ell]}} \left(\mathbb{E}_{\substack{y'_1, \dots, y'_k \sim B_4 \\ z'_1, \dots, z'_\ell \sim B_5}} \prod_{\substack{i' \in [k] \\ j' \in [\ell]}} f(x + y_i + z_j + y'_{i'} + z'_{j'}) \right)^{1/(k\ell)} \\ &\leq \prod_{\substack{i \in [k] \\ j \in [\ell]}} \left(\mathbb{E}_{\substack{x \sim B_1 \\ y_1, \dots, y_k \sim B_2 \\ z_1, \dots, z_\ell \sim B_3}} \mathbb{E}_{\substack{y'_1, \dots, y'_k \sim B_4 \\ z'_1, \dots, z'_\ell \sim B_5}} \prod_{\substack{i' \in [k] \\ j' \in [\ell]}} f(x + y_i + z_j + y'_{i'} + z'_{j'}) \right)^{1/(k\ell)} \\ &\leq \prod_{\substack{i \in [k] \\ j \in [\ell]}} \left(\mathbb{E}_{x \sim B_1} \mathbb{E}_{\substack{y'_1, \dots, y'_k \sim B_4 \\ z'_1, \dots, z'_\ell \sim B_5}} \prod_{\substack{i' \in [k] \\ j' \in [\ell]}} f(x + y'_{i'} + z'_{j'}) + O(k\ell \cdot \eta d) \right)^{1/(k\ell)} \\ &= \mathbb{E}_{x \sim B_1} \mathbb{E}_{\substack{y'_1, \dots, y'_k \sim B_4 \\ z'_1, \dots, z'_\ell \sim B_5}} \prod_{\substack{i' \in [k] \\ j' \in [\ell]}} f(x + y'_{i'} + z'_{j'}) + O(k\ell \cdot \eta d). \end{aligned}$$

Here we first applied Lemma 6.8, then Hölder's inequality, and then Lemma 6.5. $\square$

The next several lemmas are somewhat analogous to the ones in the finite field section bounding the size of the container set $S(X, Y, D) = \{(x, y) : x \in X, y \in Y, x + y \in D\}$ under various spreadness conditions on X, Y, D (see Lemmas 5.5 and 5.6). These lemmas are substantially more complicated in the setting of general abelian groups. To explain why, we first briefly describe the setup. We have Bohr sets B_1, B_2 with the same frequencies and $\nu(B_2)/\nu(B_1) \leq \eta$. We will assume that $X, D \subseteq B_1$ and $Y \subseteq B_2$.

We first prove that if one of X or D has a bounded Gowers grid norm, then the size of the container $S(X, Y, D)$ is upper bounded. Below, one should think of g as the indicator function of either X or D .

Lemma 6.10. Fix an even integer $K \geq 1$. Let $B_1, B_2, \dots$ be a (d, η) -small sequence of Bohr sets. Consider a triplet of functions $f_1 : B_1 \rightarrow [0, 1]$, $f_2 : B_2 \rightarrow [0, 1]$, and $g : B_1 \rightarrow [0, 1]$, and suppose that $\|g\|_{(B_1, B_3, B_4, K, K)} \leq \tau$, where $\tau \in [0, 1]$.

Then we have that

$$\mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} f_1(x)f_2(y)g(x+y) \leq \tau \cdot (1+3\varepsilon) \cdot \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} f_1(x)f_2(y) + O(e^{-\Omega(\varepsilon K)} + \tau^{-K^2} \eta d).$$

Proof. Observe by Lemma 6.5 that

$$\mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} f_1(x)f_2(y)g(x+y) \leq \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} f_1(x+x')f_2(y+y')g(x+y+x'+y') + O(\eta d).$$

For fixed $x \in B_1, y \in B_2$, we can apply Hölder's inequality twice to get that

$$\begin{aligned} & \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} f_1(x+x')f_2(y+y')g(x+y+x'+y') \\ & \leq \left(\mathbb{E}_{x' \sim B_3} f_1(x+x') \right)^{(K-1)/K} \cdot \left(\mathbb{E}_{x' \sim B_3} \left(\mathbb{E}_{y' \sim B_4} f_2(y+y')g(x+y+x'+y') \right)^K \right)^{1/K} \\ & \leq \left(\mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} f_1(x+x')f_2(y+y') \right)^{(K-1)/K} \cdot \left(\mathbb{E}_{\substack{x'_1, \dots, x'_K \sim B_3 \\ y'_1, \dots, y'_K \sim B_4}} \prod_{\substack{i \in [K] \\ j \in [K]}} g(x+y+x'_i+y'_j) \right)^{1/K^2} \\ & \leq (1+\varepsilon) \cdot \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} f_1(x+x')f_2(y+y') \cdot \left(\mathbb{E}_{\substack{x'_1, \dots, x'_K \sim B_3 \\ y'_1, \dots, y'_K \sim B_4}} \prod_{\substack{i \in [K] \\ j \in [K]}} g(x+y+x'_i+y'_j) \right)^{1/K^2} + e^{-\Omega(\varepsilon K)}, \end{aligned}$$

where the bound in the last line is because either $\mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} f_1(x+x')f_2(y+y') \leq e^{-\Omega(\varepsilon K)}$, which

makes the bound trivial, or $\left(\mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} f_1(x+x')f_2(y+y') \right)^{-1/K} \leq 1+\varepsilon$.

For $x \in B_1, y \in B_2$ define the event

$$\mathcal{E}(x, y) = \mathbb{1} \left[\left(\mathbb{E}_{\substack{x'_1, \dots, x'_K \sim B_3 \\ y'_1, \dots, y'_K \sim B_4}} \prod_{\substack{i \in [K] \\ j \in [K]}} g(x+y+x'_i+y'_j) \right)^{1/K^2} \geq \tau \cdot (1+\varepsilon) \right]$$

and observe that

$$\begin{aligned} \mathbb{E}[\mathcal{E}(x, y)] & \leq (\tau \cdot (1+\varepsilon))^{-K^2} \cdot \mathbb{E}_{\substack{x \sim B_1, y \sim B_2 \\ x'_1, \dots, x'_K \sim B_3 \\ y'_1, \dots, y'_K \sim B_4}} \prod_{\substack{i \in [K] \\ j \in [K]}} g(x+y+x'_i+y'_j) \\ & \leq e^{-\Omega(\varepsilon K^2)} \tau^{-K^2} (\|g\|_{(B_1, B_3, B_4, K, K)}^{K^2} + O(\eta d)) \leq e^{-\Omega(\varepsilon K^2)} + O(\tau^{-K^2} \cdot \eta \cdot d). \end{aligned}$$

Combining this exceptional set bound with the earlier Hölder argument we have

$$\begin{aligned} & \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} f_1(x+x')f_2(y+y')g(x+y+x'+y') \\ & \leq \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} (1 - \mathcal{E}(x, y)) \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} f_1(x+x')f_2(y+y')g(x+y+x'+y') + e^{-\Omega(\varepsilon K^2)} + O(\tau^{-K^2} \cdot \eta \cdot d) \\ & \leq \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} \tau \cdot (1+3\varepsilon) \cdot \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} f_1(x+x')f_2(y+y') + e^{-\Omega(\varepsilon K)} + O(\tau^{-K^2} \cdot \eta \cdot d) \\ & = \tau \cdot (1+3\varepsilon) \cdot \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} f_1(x)f_2(y) + e^{-\Omega(\varepsilon K)} + O(\tau^{-K^2} \cdot \eta \cdot d) \end{aligned}$$

as desired. □

Our next lemma gives an estimate of the container size in the setting when X, Y have bounded Gowers grid norms. We note that this estimate is *not* $\mathbb{E}[\mathbb{1}_X] \mathbb{E}[\mathbb{1}_Y] \mathbb{E}[\mathbb{1}_D]$, and is instead $\mathbb{E}[\mathbb{1}_X \mathbb{1}_D] \mathbb{E}[\mathbb{1}_Y]$. Later, we give more conditions under which $\mathbb{E}[\mathbb{1}_X \mathbb{1}_D] \approx \mathbb{E}[\mathbb{1}_X] \mathbb{E}[\mathbb{1}_D]$. Below, we encourage the reader to think of f_1, f_2, g as the indicator functions of X, Y, D respectively.

Lemma 6.11. *Let $\varepsilon \in (0, 1/100)$ and fix an even integer $K \geq 2$. Let $B_1, B_2, \dots$ be a (d, η) -small sequence of Bohr sets. Consider a triplet of functions $f_1 : B_1 \rightarrow [0, 1]$, $f_2 : B_2 \rightarrow [0, 1]$, and $g : B_1 \rightarrow [0, 1]$. Suppose that $\|f_1\|_{(B_1, B_4, B_5, K, K)} \leq (1 + \varepsilon) \cdot \mathbb{E}[f_1]$, $\|f_2\|_{(B_2, B_4, B_5, K, K)} \leq (1 + \varepsilon) \cdot \mathbb{E}[f_2]$, $\eta \leq \varepsilon^3 \cdot (\mathbb{E}[f_1] \cdot \mathbb{E}[f_2] \cdot \mathbb{E}[g])^{O(K^2)}$, and $K \geq 100\varepsilon^{-8} \log(2/(\mathbb{E}[f_1] \cdot \mathbb{E}[f_2] \cdot \mathbb{E}[g]))$.*

Then we have that

$$\left| \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} f_1(x) f_2(y) g(x+y) - \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} [f_1(x) g(x+y)] \cdot \mathbb{E}_{y \sim B_2} [f_2(y)] \right| \\ = O(\varepsilon^{1/2}) \cdot \mathbb{E}[f_1] \mathbb{E}[f_2] \mathbb{E}[g] + O((\eta d)^{1/(2K)} + e^{-\Omega(\varepsilon^8 K)}).$$

Proof. Observe that

$$\begin{aligned} \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} f_1(x) f_2(y) g(x+y) &= \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} f_1(x-y) f_2(y) g(x) + O(\eta d) \\ &= \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} f_1(x-y) \tilde{f}_2(y) g(x) + \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} [f_1(x) g(x+y)] \cdot \mathbb{E}_{y \sim B_2} [f_2(y)] + O(\eta d) \\ &= \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2 \\ z \sim B_3}} f_1(x-y) \tilde{f}_2(y+z) g(x+z) + \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} [f_1(x) g(x+y)] \cdot \mathbb{E}_{y \sim B_2} [f_2(y)] + O(\eta d) \end{aligned}$$

where $\tilde{f}_2(y) = f_2(y) - \mathbb{E}_{y' \sim B_2} [f_2(y')]$.

We let L be a positive integer to be chosen later. Now observe that

$$\begin{aligned} \left| \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2 \\ z \sim B_3}} f_1(x-y) \tilde{f}_2(y+z) g(x+z) \right| &\leq \mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} g(x+z) \left| \mathbb{E}_{y \sim B_2} f_1(x-y) \tilde{f}_2(y+z) \right| \\ &\leq \left(\mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} g(x+z) \right)^{(L-1)/L} \cdot \left(\mathbb{E}_{\substack{x \sim B_1 \\ y_1, \dots, y_L \sim B_2}} \mathbb{E}_{z \sim B_3} \prod_{j=1}^L f_1(x-y_j) \tilde{f}_2(y_j+z) \right)^{1/L} \\ &\leq (1 + \varepsilon) \cdot \mathbb{E}_{y \sim B_2} g(y) \cdot \left(\mathbb{E}_{\substack{x_1, x_2 \sim B_1 \\ y_1, \dots, y_L \sim B_2}} \prod_{j=1}^L f_1(x_1 - y_j) f_1(x_2 - y_j) \right)^{1/(2L)} \\ &\quad \cdot \left(\mathbb{E}_{\substack{z_1, z_2 \sim B_3 \\ y_1, \dots, y_L \sim B_2}} \prod_{j=1}^L \tilde{f}_2(y_j + z_1) \tilde{f}_2(y_j + z_2) \right)^{1/(2L)} + O(\eta d + e^{-\Omega(\varepsilon L)}), \end{aligned}$$

where the final inequality follows from $\delta^{\frac{L-1}{L}} \leq (1 + \varepsilon)\delta + e^{-\Omega(\varepsilon L)}$, and an application of Cauchy-Schwarz. In order to complete our analysis it suffices to bound the last two terms:

$$\left(\mathbb{E}_{\substack{x_1, x_2 \sim B_1 \\ y_1, \dots, y_L \sim B_2}} \prod_{j=1}^L f_1(x_1 - y_j) f_1(x_2 - y_j) \right)^{1/(2L)} \text{ and } \left(\mathbb{E}_{\substack{z_1, z_2 \sim B_3 \\ y_1, \dots, y_L \sim B_2}} \prod_{j=1}^L \tilde{f}_2(y_j + z_1) \tilde{f}_2(y_j + z_2) \right)^{1/(2L)}.$$

We now derive a bound on the second term involving $\tilde{f}_2$ – the bound on the first term involving f_1 is similar. For the second term, observe that

$$\min_{z \in B_3} \mathbb{E}_{y \sim B_2} [\tilde{f}_2(y + z)] \geq -O(\eta d).$$

Furthermore observe that by Hölder's inequality that $\|f_2\|_{(B_2, B_4, B_5, K, 2)} \leq \|f_2\|_{(B_2, B_4, B_5, K, K)} \leq (1 + \varepsilon) \cdot \mathbb{E}_{y \sim B_2, z \sim B_3} [f_2(y + z)] + O(\eta d)$. Moreover,

$$\begin{aligned} & \mathbb{E}_{\substack{z_1, z_2 \sim B_3 \\ y_1, \dots, y_K \sim B_2}} \prod_{j=1}^K f_2(y_j + z_1) f_2(y_j + z_2) \\ & \leq \mathbb{E}_{\substack{z_1, z_2 \sim B_3 \\ y_1, \dots, y_K \sim B_2 \\ y' \sim B_4 \\ z' \sim B_5}} \prod_{j=1}^K f_2(y_j + z_1 + y' + z') f_2(y_j + z_2 + y' + z') + O(K^2 \eta d) \\ & \leq \mathbb{E}_{\substack{z_1, z_2 \sim B_3 \\ y_1, \dots, y_K \sim B_2}} \prod_{i \in [2]} \left(\mathbb{E}_{\substack{z'_1, z'_2 \sim B_5 \\ y'_1, \dots, y'_K \sim B_4}} \prod_{\ell \in [2]} f_2(y_j + z_i + y'_r + z'_\ell) \right)^{1/(2K)} + O(K^2 \eta d) \\ & \leq \left(\prod_{\substack{i \in [2] \\ j \in [K]}} \mathbb{E}_{\substack{z_1, z_2 \sim B_3 \\ y_1, \dots, y_K \sim B_2}} \mathbb{E}_{\substack{z'_1, z'_2 \sim B_5 \\ y'_1, \dots, y'_K \sim B_4}} \prod_{\ell \in [2]} f_2(y_j + z_i + y'_r + z'_\ell) \right)^{1/(2K)} + O(K^2 \eta d) \\ & \leq \mathbb{E}_{\substack{y \sim B_2 \\ z \sim B_3 \\ z'_1, z'_2 \sim B_5 \\ y'_1, \dots, y'_K \sim B_4}} \prod_{\substack{\ell \in [2] \\ r \in [K]}} f_2(y + z + y'_r + z'_\ell) + O(K^2 \eta d) \\ & \leq \|f_2\|_{(B_2, B_4, B_5, K, 2)}^{2K} + O(K^2 \eta d). \end{aligned}$$

Recall that we assumed that $\|f_2\|_{(B_2, B_4, B_5, K, 2)} \leq (1 + \varepsilon) \mathbb{E}[f_2]$. Also, note that

$$\mathbb{E}_{y \in B_3} \mathbb{E}_{x \in B_2} \tilde{f}_2(x + y) \geq -O(\eta d) \geq -\varepsilon^2/1000 \cdot \mathbb{E}[f_2].$$

Therefore [Lemma 4.2](#) applies, and for $L = 10^{-2} \lfloor K \cdot \varepsilon^4 \rfloor$ we get

$$\left(\mathbb{E}_{\substack{z_1, z_2 \sim B_3 \\ y_1, \dots, y_L \sim B_2}} \prod_{j=1}^L \tilde{f}_2(y_j + z_1) \tilde{f}_2(y_j + z_2) \right)^{1/(2L)} = O(\varepsilon^{1/2}) \cdot \mathbb{E}[f_2] + O(K^2 \eta d).$$

An identical analysis for f_1 yields that

$$\left(\mathbb{E}_{\substack{x_1, x_2 \sim B_1 \\ y_1, \dots, y_K \sim B_2}} \prod_{j=1}^K f_1(x_1 - y_j) f_1(x_2 - y_j) \right) \leq \|f_1\|_{(B_1, B_4, B_5, K, 2)}^{2K} + O(K^2 \eta d).$$

Combining these two bounds gives the desired result. $\square$

We will also consider the case where the both X and D have bounded Gowers grid norms. In this case, the size of the container is indeed as is expected. Below one should view f_1, f_2 as the indicator functions of X, D .

Lemma 6.12. *Let $\varepsilon \in (0, 1/100)$ and fix an even integer $K \geq 2$. Let $B_1, B_2, \dots$ be a (d, η) -small sequence of Bohr sets. Consider a triplet of functions $f_1 : B_1 \rightarrow [0, 1]$, $f_2 : B_1 \rightarrow [0, 1]$, and*

$g : B_2 \rightarrow [0, 1]$. Suppose that $\|f_1\|_{(B_1, B_4, B_5, K, K)} \leq (1 + \varepsilon) \cdot \mathbb{E}[f_1]$, $\|f_2\|_{(B_1, B_4, B_5, K, K)} \leq (1 + \varepsilon) \cdot \mathbb{E}[f_2]$, $\eta \leq \varepsilon^3 \cdot (\mathbb{E}[f_1] \cdot \mathbb{E}[f_2] \cdot \mathbb{E}[g])^{O(K^2)}$, and $K \geq 100\varepsilon^{-8} \log(2/(\mathbb{E}[f_1] \cdot \mathbb{E}[f_2] \cdot \mathbb{E}[g]))$.

Then we have that

$$\left| \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} f_1(x)g(y)f_2(x+y) - \mathbb{E}[f_1] \mathbb{E}[f_2] \mathbb{E}[g] \right| = O(\varepsilon^{1/2}) \cdot \mathbb{E}[f_1] \mathbb{E}[f_2] \mathbb{E}[g] + O((\eta d)^{1/(2K)} + e^{-\Omega(\varepsilon^8 K)}).$$

Proof. Observe that

$$\begin{aligned} \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} f_1(x)g(y)f_2(x+y) &= \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} f_1(x)g(y)(f_2(x+y) - \mathbb{E}[f_2]) + \mathbb{E}[f_1] \mathbb{E}[f_2] \mathbb{E}[g] \\ &= \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2 \\ z \sim B_3}} f_1(x+z)g(y-z)\tilde{f}_2(x+y) + \mathbb{E}[f_1] \mathbb{E}[f_2] \mathbb{E}[g] + O(\eta d) \end{aligned}$$

where $\tilde{f}_2(z) = f_2(z) - \mathbb{E}_{z' \sim B_1}[f_2(z')]$.

We let L be a positive integer to be chosen later. Now observe that

$$\begin{aligned} \left| \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2 \\ z \sim B_3}} f_1(x+z)g(y-z)\tilde{f}_2(x+y) \right| &\leq \mathbb{E}_{\substack{y \sim B_2 \\ z \sim B_3}} g(y-z) \left| \mathbb{E}_{x \sim B_1} f_1(x+z)\tilde{f}_2(x+y) \right| \\ &\leq \left(\mathbb{E}_{\substack{y \sim B_2 \\ z \sim B_3}} g(y-z) \right)^{(L-1)/L} \cdot \left(\mathbb{E}_{\substack{y \sim B_2 \\ z \sim B_3}} \mathbb{E}_{x_1, \dots, x_L \sim B_2} \prod_{j=1}^L f_1(x_j+z)\tilde{f}_2(x_j+y) \right)^{1/L} \\ &\leq (1 + \varepsilon) \cdot \mathbb{E}_{y \sim B_2} g(y) \cdot \left(\mathbb{E}_{\substack{z_1, z_2 \sim B_3 \\ x_1, \dots, x_L \sim B_1}} \prod_{j=1}^L f_1(x_j+z_1)f_1(x_j+z_2) \right)^{1/(2L)} \\ &\quad \cdot \left(\mathbb{E}_{\substack{y_1, y_2 \sim B_2 \\ x_1, \dots, x_L \sim B_1}} \prod_{j=1}^L \tilde{f}_2(x_j+y_1)\tilde{f}_2(x_j+y_2) \right)^{1/(2L)} + O(\eta d + e^{-\Omega(\varepsilon L)}). \end{aligned}$$

In order to complete our analysis it suffices to bound the last two terms:

$$\left(\mathbb{E}_{\substack{z_1, z_2 \sim B_3 \\ x_1, \dots, x_L \sim B_1}} \prod_{j=1}^L f_1(x_j+z_1)f_1(x_j+z_2) \right)^{1/(2L)} \text{ and } \left(\mathbb{E}_{\substack{y_1, y_2 \sim B_2 \\ x_1, \dots, x_L \sim B_1}} \prod_{j=1}^L \tilde{f}_2(x_j+y_1)\tilde{f}_2(x_j+y_2) \right)^{1/(2L)}.$$

We now consider the second term involving $\tilde{f}_2$ where we will need to invoke spectral positivity. For this term, observe that

$$\min_{y \in B_2} \mathbb{E}_{x \sim B_1} [\tilde{f}_2(x+y)] \geq -O(\eta d).$$

Furthermore observe that by Hölder's inequality that $\|f_2\|_{(B_2, B_4, B_5, K, 2)} \leq \|f_2\|_{(B_2, B_4, B_5, K, K)} \leq (1 + \varepsilon) \cdot \mathbb{E}_{y \sim B_2} \mathbb{E}_{z \sim B_3}[f_2(y+z)] + O(\eta d)$. Moreover,

$$\begin{aligned} &\mathbb{E}_{\substack{y_1, y_2 \sim B_2 \\ x_1, \dots, x_K \sim B_1}} \prod_{j=1}^K f_2(x_j+y_1)f_2(x_j+y_2) \\ &\leq \mathbb{E}_{\substack{y_1, y_2 \sim B_2 \\ x_1, \dots, x_K \sim B_1 \\ x' \sim B_4 \\ y' \sim B_5}} \prod_{j=1}^K f_2(x_j+y_1+x'+y')f_2(x_j+y_2+x'+y') + O(K^2 \eta d) \end{aligned}$$

$$\begin{aligned}
&\leq \mathbb{E}_{\substack{y_1, y_2 \sim B_2 \\ x_1, \dots, x_K \sim B_1}} \prod_{\substack{i \in [2] \\ j \in [K]}} \left(\mathbb{E}_{\substack{y'_1, y'_2 \sim B_5 \\ x'_1, \dots, x'_K \sim B_4}} \prod_{\substack{\ell \in [2] \\ r \in [K]}} f_2(y_i + x_j + y'_r + x'_\ell) \right)^{1/(2K)} + O(K^2 \eta d) \\
&\leq \left(\prod_{\substack{i \in [2] \\ j \in [K]}} \mathbb{E}_{\substack{y_1, y_2 \sim B_2 \\ x_1, \dots, x_K \sim B_1}} \mathbb{E}_{\substack{y'_1, y'_2 \sim B_5 \\ x'_1, \dots, x'_K \sim B_4}} \prod_{\substack{\ell \in [2] \\ r \in [K]}} f_2(y_i + x_j + y'_r + x'_\ell) \right)^{1/(2K)} + O(K^2 \eta d) \\
&\leq \mathbb{E}_{\substack{y \sim B_2 \\ x \sim B_1 \\ z'_1, z'_2 \sim B_5 \\ x'_1, \dots, x'_K \sim B_4}} \prod_{\substack{\ell \in [2] \\ r \in [K]}} f_2(y + x + y'_r + x'_\ell) + O(K^2 \eta d) \\
&\leq \|f_2\|_{(B_2, B_4, B_5, K, 2)}^{2K} + O(K^2 \eta d).
\end{aligned}$$

Therefore by [Lemma 4.2](#), if we set $L = 10^{-2} \lfloor K \cdot \varepsilon^4 \rfloor$ then

$$\left(\mathbb{E}_{\substack{y_1, y_2 \sim B_2 \\ x_1, \dots, x_L \sim B_1}} \prod_{j=1}^L \tilde{f}_2(x_j + y_1) \tilde{f}_2(x_j + y_2) \right)^{1/(2L)} = O(\varepsilon^{1/2}) \cdot \mathbb{E}[f_2] + O((\eta d)^{1/(2L)}).$$

An identical analysis for f_1 yields that

$$\left(\mathbb{E}_{\substack{z_1, z_2 \sim B_3 \\ x_1, \dots, x_K \sim B_1}} \prod_{j=1}^K f_1(x_j + z_1) f_1(x_j + z_2) \right) \leq \|f_1\|_{(B_1, B_4, B_5, K, 2)}^{2K} + O(K^2 \eta d).$$

Combining these two bounds gives the desired result. $\square$

In this setting we also require a weaker notion of spreadness which we call (B_1, B_2, ε) ℓ_1 -spreadness. Informally, a set $D \subseteq B_1$ satisfies the ℓ_1 -spreadness property if, when B_1 is averaged on B_2 shifts, then for almost all the shifts, the density of D doesn't deviate too far from its global density.

Definition 6.13 (ℓ_1 -spreadness). Let B_1, B_2 be regular Bohr sets with the same frequencies and $\nu(B_2) \leq \nu(B_1)$. We say that a function $f : B_1 \rightarrow [0, 1]$ is (B_1, B_2, ε) ℓ_1 -spread if

$$\mathbb{E}_{x \sim B_1} \left| \mathbb{E}_{y \sim B_2} [f(x + y)] - \mathbb{E}[f] \right| \leq \varepsilon \cdot \mathbb{E}[f].$$

We say that a set $D \subseteq B_1$ is ℓ_1 -spread if its indicator function is, according to [Definition 6.13](#). The next lemma proves that $\mathbb{E}[f_1 g] \approx \mathbb{E}[f_1] \mathbb{E}[g]$, which is useful given [Lemma 6.11](#), holds as long as f_1 has small Gowers grid norm and g is ℓ_1 -spread, which is a much weaker hypothesis on g than having small Gowers grid norm.

Lemma 6.14. Let $\varepsilon \in (0, 1/100)$ and fix an even integer $K \geq 2$. Let $B_1, B_2, \dots$ be a (d, η) -small sequence of Bohr sets. Consider a pair of functions $f : B_1 \rightarrow [0, 1]$ and $g : B_1 \rightarrow [0, 1]$. Suppose that $\|f\|_{(B_1, B_3, B_4, K, K)} \leq (1 + \varepsilon) \cdot \mathbb{E}[f]$, g is (B_1, B_4, ε) ℓ_1 -spread, $\eta \leq \varepsilon^3 \cdot (\mathbb{E}[f] \cdot \mathbb{E}[g])^{O(K^2)} \cdot d^{-O(1)}$, and $K \geq 100\varepsilon^{-8} \log(2/(d \cdot \mathbb{E}[f] \cdot \mathbb{E}[g]))$.

Then we have that

$$\left| \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} f(x)g(x + y) - \mathbb{E}_{x \sim B_1} [f(x)] \cdot \mathbb{E}_{x \sim B_1} [g(x)] \right| = O(\varepsilon^{1/2}) \cdot \mathbb{E}[f] \mathbb{E}[g] + O((\eta d)^{1/(2K)} + e^{-\Omega(\varepsilon^8 K)}).$$

Proof. For the upper bound, first write

$$\mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} f(x)g(x + y) \leq \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} f(x + y)g(x) + O(\eta d).$$

Now apply [Lemma 6.10](#) with f_1 as g , $f_2 = 1$, and the g in [Lemma 6.10](#) as f_1 with $\tau = \|f\|_{(B_1, B_3, B_4, K, K)}$ (and the same η and K), which gives the desired result.

For the lower bound, define

$$\mathcal{E}_1(x, y) = \mathbb{1} \left[\left(\mathbb{E}_{\substack{x'_1, \dots, x'_K \sim B_3 \\ y'_1, \dots, y'_K \sim B_4}} \prod_{i \in [K]} f(x + y + x'_i + y'_i) \geq ((1 + 2\varepsilon) \cdot \mathbb{E}[f]) \right)^{K^2} \right].$$

To bound $\mathbb{E}[\mathcal{E}_1(x, y)]$, note that

$$\begin{aligned} & \mathbb{1} \left[\mathbb{E}_{\substack{x'_1, \dots, x'_K \sim B_3 \\ y'_1, \dots, y'_K \sim B_4}} \prod_{i \in [K]} f(x + y + x'_i + y'_i) \geq ((1 + 2\varepsilon) \cdot \mathbb{E}[f])^{K^2} \right] \\ & \leq ((1 + 2\varepsilon) \mathbb{E}[f])^{-K^2} \mathbb{E}_{\substack{x \in B_1 \\ y \in B_2}} \mathbb{E}_{\substack{x'_1, \dots, x'_K \sim B_3 \\ y'_1, \dots, y'_K \sim B_4}} \prod_{i \in [K]} f(x + y + x'_i + y'_i) \\ & \leq ((1 + 2\varepsilon) \mathbb{E}[f])^{-K^2} \left(\|f\|_{B_1, B_3, B_4, K, K} + O(\eta d) \right) \leq e^{-\Omega(\varepsilon K^2)} + O(\mathbb{E}[f]^{-K^2} \eta d), \end{aligned}$$

because $\|f\|_{B_1, B_3, B_4, K, K} \leq (1 + \varepsilon) \mathbb{E}[f]$.

Furthermore observe that if $\mathcal{E}(x, y) = 0$, then monotonicity of the (K, K) -Gowers grid norms ([Lemma 6.9](#)) gives that $\mathbb{E}_{z \sim B_3} f(x + y + z) \leq (1 + 3\varepsilon) \cdot \mathbb{E}[f] + O(\eta d)$. Now define

$$\mathcal{E}_2(x, y, y') = \mathbb{1} \left[\mathbb{E}_{z \sim B_3} f(x + y + z) \leq (1 - \varepsilon^{1/2}) \cdot \mathbb{E}[f] \right] + \mathbb{1} \left[\mathbb{E}_{z \sim B_3} g(x + y' + z) \leq (1 - \varepsilon^{1/2}) \cdot \mathbb{E}[g] \right].$$

[Fact 4.3](#) (applied to $\min(\mathbb{E}_{z \sim B_3} f(x + y' + z), (1 + 3\varepsilon) \cdot \mathbb{E}[f])$) implies that

$$\mathbb{E}_{x \sim B_1, y \sim B_2} \left[\mathbb{1} \left[\mathbb{E}_{z \sim B_3} f(x + y + z) \leq (1 - \varepsilon^{1/2}) \cdot \mathbb{E}[f] \right] \right] \leq O(\varepsilon^{1/2} + \eta d),$$

and the (B_1, B_4, ε) ℓ_1 -spreadness of g and Markov's inequality imply that

$$\mathbb{E}_{x \sim B_1, y' \sim B_2} \left[\mathbb{1} \left[\mathbb{E}_{z \sim B_3} g(x + y' + z) \leq (1 - \varepsilon^{1/2}) \cdot \mathbb{E}[g] \right] \right] \leq O(\varepsilon^{1/2} + \eta d).$$

In total, we have concluded that

$$\mathbb{E}[\mathcal{E}_2(x, y, y')] \leq O(\varepsilon^{1/2} + \eta d).$$

Thus

$$\begin{aligned} \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} f(x)g(x + y) & \geq \mathbb{E}_{\substack{x \sim B_1 \\ y, y' \sim B_2 \\ z, z' \sim B_3}} f(x + y' + z)g(x + y + z') - O(\eta d) \\ & \geq \mathbb{E}_{\substack{x \sim B_1 \\ y, y' \sim B_2}} (1 - \mathcal{E}_2(x, y, y')) \left(\mathbb{E}_{z \sim B_3} f(x + y' + z) \right) \left(\mathbb{E}_{z' \sim B_4} g(x + y + z') \right) - O(\eta d) \\ & \geq (1 - O(\varepsilon^{1/2})) \cdot \mathbb{E}[f] \cdot \mathbb{E}[g] - O(\eta d) \end{aligned}$$

which implies the desired result. $\square$

6.2. Algebraic spreadness and Gowers grid norms. We now give the required analog of “algebraic” spreadness in the case of Bohr sets.

Definition 6.15 (Algebraic spreadness). Let $B \subseteq G$ be a regular Bohr set. We say that $X \subseteq B$ is (r, η_s, ε) -algebraically spread within B if for all regular Bohr sets $B' \subseteq B$ with $\text{rank}(B') \leq \text{rank}(B) + r$ and radius $\nu(B') \geq \eta_s \nu(B)$, and for all $x \in G$ it holds that

$$\frac{|(X - x) \cap B'|}{|B'|} \leq (1 + \varepsilon) \frac{|X|}{|B|}.$$

The key technical input from the work of Kelley and Meka [42] is converting between algebraic spreadness and the guarantee that a function f does not correlate with “convolutions”. The statement we need appears with a careful proof in the work of Filmus, Hatami, Hosseini, and Kelman [22] but not as a standalone statement. We reproduce a proof (assuming some known almost periodicity results) in [Appendix B](#).

Theorem 6.16. Let $\eta, \varepsilon \in (0, 1/2)$ and $k \geq 1$. Let $B_1, B_2, \dots$ be a (d, η) -small sequence of Bohr sets. Let $f_i : B_i \rightarrow [0, 1]$ and $g : B_1 \rightarrow [0, 1]$.

Suppose that $\mathbb{E}_{x \sim B_i}[f_i(x)] \geq 2^{-k}$ and $\mathbb{E}_{x \sim B_1}[g(x)] \geq 2^{-k}$ and $\eta \leq O(d^{-2} \cdot 2^{-O(k^2 \log(1/\varepsilon)/\varepsilon)})$. Furthermore suppose that

$$\mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} [f_1(x)f_2(y)g(x+y)] \geq (1 + \varepsilon) \cdot \mathbb{E}[g] \cdot \mathbb{E}[f_1] \cdot \mathbb{E}[f_2].$$

Then there exists a regular Bohr set $B' \subseteq B_1$ with radius r' and dimension $d + d'$ such that

$$r' \geq r_2 \cdot \varepsilon \cdot d^{-4} \cdot \eta \cdot 2^{-O(k^2 \log(1/\varepsilon)/\varepsilon)} \text{ and } d' = O(k^8 \cdot \varepsilon^{-9})$$

and $x^* \in B_1$ such that

$$\mathbb{E}_{x \sim B'} [g(x + x^*)] \geq (1 + \varepsilon/2) \cdot \mathbb{E}_{x \sim B_1} [g(x)].$$

The last remaining technical issue for this section is connecting algebraic spreadness ([Definition 6.15](#)) with the Gowers grid norms ([Definition 6.7](#)). This is essentially an immediate consequence of sifting.

Theorem 6.17. Fix an integer $K \geq 1$, $\eta \in (0, 1/2)$, and $\tau \in (0, 1)$. Let $B_1, B_2, \dots$ be a (d, η) -small sequence of Bohr sets. Define $f : G \rightarrow [0, 1]$ such that

$$\|f\|_{(B_1, B_2, B_3, K, K)} \geq \tau.$$

Then for any $\varepsilon > 0$ there exists $g_1 : B_2 \rightarrow [0, 1]$, $g_2 : B_3 \rightarrow [0, 1]$, and $x \in B_1$ such that

$$\mathbb{E}_{\substack{y \sim B_2 \\ z \sim B_3}} [g_1(y)g_2(z)f(x + y + z)] \geq (1 - \varepsilon) \cdot \tau \cdot \mathbb{E}_{\substack{y \sim B_2 \\ z \sim B_3}} [g_1(y)g_2(z)]$$

and $\mathbb{E}[g_i] \geq \varepsilon \cdot \tau^{O(K)}$.

Proof. From our assumption, we immediately have

$$\sup_{x \in B_1} \mathbb{E}_{\substack{y_1, \dots, y_K \sim B_2 \\ z_1, \dots, z_K \sim B_3}} \prod_{\substack{i \in [K] \\ j \in [K]}} f(x + y_i + z_j) \geq \mathbb{E}_{\substack{x \sim B_1 \\ y_1, \dots, y_K \sim B_2 \\ z_1, \dots, z_K \sim B_3}} \prod_{\substack{i \in [K] \\ j \in [K]}} f(x + y_i + z_j) \geq \tau^{K^2}.$$

Fix x achieving the supremum and define $F(y, z) = f(x + y + z)$. We may apply [Theorem 3.4](#) and obtain $g_1 : B_2 \rightarrow [0, 1]$ and $g_2 : B_3 \rightarrow [0, 1]$ with $\mathbb{E}[g_i] \geq \varepsilon \cdot \tau^{O(K)}$ and

$$\mathbb{E}_{\substack{y \sim B_2 \\ z \sim B_3}} [g_1(y)g_2(z)f(x + y + z)] \geq (1 - \varepsilon) \cdot \tau \cdot \mathbb{E}_{\substack{y \sim B_2 \\ z \sim B_3}} [g_1(y)g_2(z)]. \quad \square$$

6.3. Pseudorandomization. We note the easy claim that every set contains a somewhat large subset that is spread within a smaller Bohr set. This is a bit more complicated than the analogous [Claim 5.8](#) because we need to work to make the sets that we increment onto also regular Bohr sets.

Claim 6.18. *Let r be an integer, $B \subseteq G$ be a regular Bohr set with $d = \text{rank}(B)$, and $\varepsilon, \eta_s \in (0, 1)$. For $X \subseteq B$ with $\delta := |X|/|B|$ there is a regular Bohr set $B' \subseteq B$ and $x \in G$ such that $X' := X \cap (x + B')$ is (r, η_s, ε) -algebraically spread within $x + B'$, $\text{rank}(B') \leq d + O(r\varepsilon^{-1} \log(1/\delta))$, $\frac{|X'|}{|B'|} \geq \delta$, and $\nu(B') \geq \nu(B) \cdot (\varepsilon\delta\eta_s/(2d'))^{O(\varepsilon^{-1} \log(1/\delta))}$, where $d' := d + O(r\varepsilon^{-1} \log(1/\delta))$.*

Proof. We proceed iteratively. Initialize $B^{(0)} = B$. If $X \cap B^{(0)}$ is (r, η_s, ε) -algebraically spread within $B^{(0)}$, we are done. Otherwise, there must exist $x \in G$ and a Bohr set $B^{(0)'} \subseteq B^{(0)}$ with $\text{rank}(B^{(0)'}) \leq \text{rank}(B^{(0)}) + r$ and $\nu(B^{(0)'}) \geq \eta_s \cdot \nu(B^{(0)})$ satisfying

$$\frac{|X \cap (x + B^{(0)'})|}{|B^{(0)'}|} \geq (1 + \varepsilon) \frac{|X \cap B^{(0)}|}{|B^{(0)}|}.$$

Now, we want to simply iterate the argument, but we must be a bit careful because $B^{(0)'}$ is not necessarily regular (and we wish to output a regular Bohr set). Instead let $B^{(1)} = \eta' B^{(0)'}$ for some $\eta' \in [\frac{\varepsilon\delta}{2Cd'}, \frac{\varepsilon\delta}{Cd'}]$ for sufficiently large C so that $B^{(1)}$ is regular. Let $X^{(0)'} = X \cap (x + B^{(0)'})$. Note that

$$\mathbb{E}_{y \in B^{(0)'}} |X^{(0)'} \cap (y + x + B^{(1)})| = \frac{|B^{(1)}|}{|B^{(0)'}|} |X^{(0)'}| \pm O(\eta' d' |B^{(1)}|).$$

Thus there is some $y \in B^{(0)'}$ so that

$$\frac{|X^{(0)'} \cap (y + x + B^{(1)})|}{|B^{(1)}|} \geq \frac{|X^{(0)'}|}{|B^{(0)'}|} - O(\eta' d') \geq (1 + \varepsilon/2)\delta.$$

Now set $X^{(1)} := X^{(0)'} \cap (y + x + B^{(1)})$ and iterate the same process starting with $X^{(1)} \subseteq y + x + B^{(1)}$. The number of iterations is at most $O(\varepsilon^{-1} \log(1/\delta))$. The conclusions follow because after i iterations we know that $\text{rank}(B^{(i)}) \leq d + ri$, $\nu(B^{(i)}) \geq (\varepsilon\delta\eta_s/(2Cd'))^i \nu(B)$, and $\frac{|X^{(i)}|}{|B^{(i)}|} \geq (1 + \varepsilon/2)^i \delta$. $\square$

In the setting of general abelian groups we will pseudorandomize all three sets X , Y , and D . X and Y will be made to be algebraically spread (see [Definition 6.15](#)), while D will be made to be ℓ_1 -spread (see [Definition 6.13](#)) – we do not see how to guarantee that D is also algebraically spread. This contrasts with the situation in the finite field setting where we only made X and Y algebraically spread. We make D ℓ_1 -spread in order to establish that the size of $S(X, Y, D)$ is close to expected (see [Lemmas 6.11](#) and [6.14](#)).

Our proof proceeds in two phases. First, we will prove a lemma which allows us to pseudorandomize both X and Y , analogous to how [Theorem 5.7](#) is proven. Then we will do a very particular partitioning algorithm to D to ensure ℓ_1 -spreadness. This might ruin the algebraic spreadness of X and Y on a small fraction of pieces, which we then fix with recursion.

To start, we prove the analogue of [Lemma 5.9](#), which said that we can partition $X \times Y$ into $X_i \times Y_i$, all of which are spread and not too small, except for a very small fraction. In this setting, we will not obtain a true partition, but instead a distribution over $X_i \times Y_i$ that on average covers each element of $X \times Y$ once (again, minus some small fraction). We also note that to interact with the density increment argument in [Section 7](#) that the initial conditions for the lemma are a bit different. While in [Lemma 5.9](#) we started with X, Y, D such that at least one of X or Y was algebraically spread, in this new setting we instead just have that X is upper bounded in (K, K) -Gowers grid norm. The argument of [Lemma 5.9](#) still carries over even with these changes.

Lemma 6.19. Let r be a positive integer and $\varepsilon, \eta, \eta_s, \beta > 0$. Let $B_1, B_2, \dots$ be a (d, η) -small sequence of Bohr sets. Let $X \subseteq B_1$, $Y \subseteq B_2$, and let $\delta_X = |X|/|B_1|$ and $\delta_Y = |Y|/|B_2|$. Define

$$g := O(\varepsilon^{-1} \log(1/(\delta_X \delta_Y))) + \varepsilon^{-1} \log(1/\beta)^2,$$

$d' = \text{rank}(B_1) + O(r g^2 \log(1/\beta))$, and assume that $\eta \leq (\delta_X \delta_Y / d')^{O(1)} 2^{-O(\log(1/\beta)^2)}$. Then there is a measure μ over rectangles $X_i \times Y_i \subseteq X \times Y$ such that:

- (1) Each $X_i \times Y_i \in \text{supp}(\mu)$ satisfies that $(X_i - x_i)$ is (r, η_s, ε) -spread in $B_1^{(i)}$ for some x_i and $(Y_i - y_i)$ is (r, η_s, ε) -spread in $B_2^{(i)}$ for some y_i , where $B_2^{(i)} \subseteq B_1^{(i)}$ are regular Bohr sets with the same frequencies and $\nu(B_2^{(i)})/\nu(B_1^{(i)}) \in [\eta/2, \eta]$.
- (2) Additionally, $\text{rank}(B_1^{(i)}) \leq \text{rank}(B_1) + O(r g^2 \log(1/\beta))$.
- (3) Also, $\nu(B_1^{(i)}) \geq \eta_s^{O(g^2 \log(1/\beta))} \nu(B_1)$.
- (4) For any 1-bounded function $f : X \times Y \rightarrow [-1, 1]$ it holds that

$$\mathbb{E}_{x \in X, y \in Y} f(x, y) = \mathbb{E}_{X_i \times Y_i \sim \mu} \mathbb{E}_{x \in X_i, y \in Y_i} f(x, y) \pm O(\beta).$$

- (5) For any $X_i \times Y_i \in \text{supp}(\mu)$ it holds that $\frac{|X_i||Y_i|}{|B_1^{(i)}||B_2^{(i)}|} \geq e^{-O(\log(1/\beta)^2)} \delta_X \delta_Y$.

Again, we prove this by first establishing a one round partitioning statement which is then used recursively.

Lemma 6.20. Let r be a positive integer and $\varepsilon, \eta, \eta_s, \beta > 0$. Let $B_1, B_2, \dots$ be a (d, η) -small sequence of Bohr sets. Let $X \subseteq B_1$, $Y \subseteq B_2$, and let $\delta_X = |X|/|B_1|$ and $\delta_Y = |Y|/|B_2|$. Define $g := O(\varepsilon^{-1} \log(1/(\delta_X \delta_Y \beta)))$, $d' = \text{rank}(B_1) + O(r g^2)$, and assume that $\eta \leq (\beta \delta_X \delta_Y / d')^{O(1)}$. Then there is a measure μ over rectangles $X_i \times Y_i \subseteq X \times Y$, and an event $\mathcal{E}(X_i, Y_i)$ such that:

- (1) Each $X_i \times Y_i \in \text{supp}(\mu)$ satisfies that $(X_i - x_i) \subseteq B_1^{(i)}$ for some x_i and $(Y_i - y_i) \subseteq B_2^{(i)}$ for some y_i , where $B_2^{(i)} \subseteq B_1^{(i)}$ are regular Bohr sets with the same frequencies and $\nu(B_2^{(i)})/\nu(B_1^{(i)}) \in [\eta/2, \eta]$.
 - (2) Additionally, $\text{rank}(B_1^{(i)}) \leq \text{rank}(B_1) + O(r g^2)$.
 - (3) Also, $\nu(B_1^{(i)}) \geq \eta_s^{O(g^2)} \nu(B_1)$.
 - (4) For any 1-bounded function $f : X \times Y \rightarrow [-1, 1]$ it holds that
- $$\mathbb{E}_{x \in X, y \in Y} f(x, y) = \mathbb{E}_{X_i \times Y_i \sim \mu} \mathbb{E}_{x \in X_i, y \in Y_i} f(x, y) \pm O(\beta^2).$$
- (5) For any $X_i \times Y_i \in \text{supp}(\mu)$ it holds that $\frac{|X_i||Y_i|}{|B_1^{(i)}||B_2^{(i)}|} \geq \beta^{O(1)} \delta_X \delta_Y$.
 - (6) If $\mathcal{E}$ holds then $X_i - x_i \subseteq B_1^{(i)}$ and $Y_i - y_i \subseteq B_2^{(i)}$ are (r, η_s, ε) -algebraically spread.
 - (7) $\mathbb{E}_{X_i \times Y_i \sim \mu} [\mathcal{E}(X_i, Y_i)] \geq 1/2$.

Proof. Let $X^{(0)} = X$ and perform the following algorithm. For $t = 0, 1, \dots$: if $|X^{(t)}| \leq \frac{\beta^2}{10} |X|$, then terminate. Otherwise, let $X_t \subseteq X^{(t)}$ be $(r', \eta'_s, \varepsilon/5)$ -algebraically spread within $x_t + \mathcal{B}_t$ for $r' = \Omega(r \varepsilon^{-1} \log(1/(\delta_Y \beta)))$ and $\eta'_s = \eta_s^{O(\varepsilon^{-1} \log(1/(\delta_Y \beta)))}$, as given by Claim 6.18. Note that

$$\text{rank}(\mathcal{B}_t) \leq \text{rank}(B_1) + O(r' \varepsilon^{-1} \log(1/(\delta_X \beta))).$$

Now let $X^{(t+1)} = X^{(t)} \setminus X_t$. Let T be the total number of iterations, so that

$$X = X^{(T)} \cup X_0 \cup X_1 \cup \dots \cup X_{T-1}.$$

Now fix a $t \in \{0, 1, \dots, T-1\}$. Let $\mathcal{B}'_t = \eta' \mathcal{B}_t$ be regular for some $\eta' \in [\eta^2/2, \eta^2]$. Now for $x \in B_2$ define $Y_{t,x} := Y \cap (x + \mathcal{B}'_t)$, and partition

$$Y_{t,x} = Y_{t,x}^{(T'_{t,x})} \cup Y_{t,x,0} \cup \dots \cup Y_{t,x,T'_{t,x}-1},$$

using the algorithm in the first paragraph, where $|Y_{t,x}^{(T',x)}| \leq \frac{\beta^2}{10} \delta_Y |\mathcal{B}'_t|$ and each $Y_{t,x,t'}$ for $0 \leq t' \leq T'_{t,x} - 1$ is (r, η_s, ε) -algebraically spread within some shifted regular Bohr set $y_{t,x,t'} + \mathcal{B}'_{t,x,t'}$ where

$$\text{rank}(\mathcal{B}'_{t,x,t'}) \leq \text{rank}(\mathcal{B}_t) + O(r\varepsilon^{-1} \log(1/(\delta_Y \beta))).$$

Let $\mathcal{B}_{t,x,t'}$ be a regular Bohr set with $\mathcal{B}'_{t,x,t'} = \eta'' \mathcal{B}_{t,x,t'}$ for $\eta'' \in [\eta/2, \eta]$. Finally, for each $y \in \mathcal{B}_t$ define $X_{t,x,t',y} := X_t \cap (x_t + y + \mathcal{B}_{t,x,t'})$ – note that for every fixed t, x, t' that the average of the indicator functions of $X_{t,x,t',y}$ is approximately the indicator function of X_t in ℓ_1 error, by Lemma 6.5 up to scaling.

Now we define the distribution μ over the pieces $X_i \times Y_i$ and the event $\mathcal{E}$. The pieces $X_i \times Y_i$ are all pieces

$$X_{t,x,t',y} \times Y_{t,x,t'} \subseteq (x_t + y + \mathcal{B}_{t,x,t'}) \times (y_{t,x,t'} + \mathcal{B}'_{t,x,t'})$$

where $\frac{|X_{t,x,t',y}|}{|\mathcal{B}_{t,x,t'}|} \geq \frac{\beta^4}{100} \delta_X$. The probability mass in μ of such a piece is defined as: $\frac{|X_{t,x,t',y}| |Y_{t,x,t'}|}{|X| |Y| |\mathcal{B}_t| |\mathcal{B}_{t,x,t'}|}$. μ may not have total probability mass 1, so scale it appropriately so that it does – later we argue that the scaling is only on the order of $1 \pm O(\beta^2)$. Finally, $\mathcal{E}$ consists of the pieces with $\frac{|X_{t,x,t',y}|}{|\mathcal{B}_{t,x,t'}|} \geq (1 - 3\varepsilon/5) \frac{|X_t|}{|\mathcal{B}_t|}$.

Let us now verify all the conclusions. (1) follows because by construction. (2) follows because

$$\begin{aligned} \text{rank}(\mathcal{B}_{t,x,t'}) &\leq \text{rank}(\mathcal{B}_t) + O(r\varepsilon^{-1} \log(1/(\delta_Y \beta))) \\ &\leq \text{rank}(\mathcal{B}_1) + O(r'\varepsilon^{-1} \log(1/(\delta_X \beta))) + O(r\varepsilon^{-1} \log(1/(\delta_Y \beta))), \end{aligned}$$

as desired. (3) follows similarly. Towards checking (4), first define $\mathcal{I}$ to be the set of tuples (t, x, t', y) with $\frac{|X_{t,x,t',y}|}{|\mathcal{B}_{t,x,t'}|} \leq \frac{\beta^4}{100} \delta_X$. For a 1-bounded function $f : X \times Y \rightarrow \mathbb{R}$ we first write

$$\mathbb{E}_{x \in X, y \in Y} f(x, y) = \sum_{t=0}^{T-1} \frac{|X_t|}{|X|} \mathbb{E}_{x \in X_t, y \in Y} f(x, y) \pm \frac{|X^{(T)}|}{|X|}.$$

Also, $|X^{(T)}|/|X| \leq \beta^2/10$ by construction. By Lemma 6.5 we know that

$$\mathbb{E}_{x \in X_t, y \in Y} f(x, y) = \mathbb{E}_{x \in \mathcal{B}_2} \delta_Y^{-1} \frac{|Y_{t,x}|}{|\mathcal{B}'_t|} \mathbb{E}_{x_1 \in X_t, y_1 \in Y_{t,x}} f(x_1, y_1) \pm \eta d' \delta_Y^{-1}.$$

Continuing, we can bound

$$\delta_Y^{-1} \frac{|Y_{t,x}|}{|\mathcal{B}'_t|} \mathbb{E}_{x_1 \in X_t, y_1 \in Y_{t,x}} f(x_1, y_1) = \sum_{t'=0}^{T'_{t,x}-1} \delta_Y^{-1} \frac{|Y_{t,x,t'}|}{|\mathcal{B}'_t|} \mathbb{E}_{x_1 \in X_t, y_1 \in Y_{t,x,t'}} f(x_1, y_1) \pm \delta_Y^{-1} \frac{|Y_{t,x}^{(T')}|}{|\mathcal{B}'_t|}.$$

Now $\delta_Y^{-1} \frac{|Y_{t,x}^{(T')}|}{|\mathcal{B}'_t|} \leq \frac{\beta^2}{10}$ by definition, so that error is acceptable. Finally, we can write

$$\mathbb{E}_{x_1 \in X_t, y_1 \in Y_{t,x,t'}} f(x_1, y_1) = \mathbb{E}_{y \in \mathcal{B}_t} \frac{|\mathcal{B}_t|}{|X_t|} \frac{|X_{t,x,t',y}|}{|\mathcal{B}_{t,x,t'}|} \mathbb{E}_{x_1 \in X_{t,x,t',y}, y_1 \in Y_{t,x,t'}} f(x_1, y_1) + O\left(\eta d' \frac{|\mathcal{B}_t|}{|X_t|}\right).$$

Note that $\frac{|\mathcal{B}_t|}{|X_t|} \leq O(\beta^{-2} \delta_X^{-1})$ by definition, so this error is acceptable. Finally, we have to bound the error contribution from tuples $(t, x, t', y) \notin \mathcal{I}$. Because $\frac{|\mathcal{B}_t|}{|X_t|} \leq \frac{10}{\beta^2 \delta_X}$, the total contribution is bounded by

$$\mathbb{E}_y \left[\mathbb{1}_{(t,x,t',y) \notin \mathcal{I}} \frac{|\mathcal{B}_t|}{|X_t|} \frac{|X_{t,x,t',y}|}{|\mathcal{B}_{t,x,t'}|} \right] \leq \frac{10}{\beta^2 \delta_X} \frac{\beta^4 \delta_X}{100} \leq \beta^2/10.$$

Thus the total accumulated error is $O(\beta^2 + \eta d' \delta_X^{-1} \delta_Y^{-1} \beta^{-2}) = O(\beta^2)$ as desired. This also argues that the scaling of μ that needs to be done is $1 + O(\beta^2)$ as claimed earlier.

(5) follows by by construction.

To verify (6), note that each $Y_{t,x,t'} \subseteq y_{t,x,t'} + \mathcal{B}'_{t,x,t'}$ is (r, η_s, ε) -algebraically spread by construction. Note that for any shifted Bohr set $b + B$ with $\text{rank}(B) \leq \text{rank}(\mathcal{B}'_{t,x,t'}) + r$ it holds that

$$\frac{|X_{t,x,t',y} \cap (b + B)|}{|B|} \leq (1 + \varepsilon/5) \frac{|X_t|}{|\mathcal{B}_t|},$$

because $\text{rank}(B) \leq \text{rank}(\mathcal{B}_t) + r'$ and X_t was (r', η, ε) -algebraically spread by construction. Thus if $\frac{|X_{t,x,t',y}|}{|\mathcal{B}_{t,x,t'}|} \geq (1 - 3\varepsilon/5) \frac{|X_t|}{|\mathcal{B}_t|}$ (as is true for the tuples in $\mathcal{E}$), then $X_{t,x,t',y} \subseteq x_t + y + \mathcal{B}_{t,x,t'}$ is (r, η_s, ε) -algebraically spread.

For item (7), first consider fixing some t, x, t' , and let $\delta_t = \frac{|X_t|}{|\mathcal{B}_t|}$ to simplify notation. Note that

$$\mathbb{E}_{y \in \mathcal{B}_t} \left[\frac{|X_{t,x,t',y}|}{|\mathcal{B}_{t,x,t'}|} \right] = \delta_t \pm O(\eta d' \delta_X^{-1} \delta_Y^{-1} \beta^{-2}) \geq (1 - \varepsilon/10) \delta_t$$

as above, and that $\frac{|X_{t,x,t',y}|}{|\mathcal{B}_{t,x,t'}|} \leq (1 + \varepsilon/5) \delta_t$ for all y , as argued above. Thus, by the reverse Markov inequality (Fact 4.3) we have

$$\Pr_{y \in \mathcal{B}_t} \left[\frac{|X_{t,x,t',y}|}{|\mathcal{B}_{t,x,t'}|} \geq (1 - 3\varepsilon/5) \delta_t \right] \geq 5/8.$$

Thus,

$$\mathbb{E}_{y \in \mathcal{B}_t: \frac{|X_{t,x,t',y}|}{|\mathcal{B}_{t,x,t'}|} \geq (1-3\varepsilon/5)\delta_t} |X_{t,x,t',y}| \geq \frac{5}{8} (1 - 3\varepsilon/5) \delta_t |\mathcal{B}_{t,x,t'}| \geq \frac{3}{5} |X_t| \cdot \frac{|\mathcal{B}_{t,x,t'}|}{|\mathcal{B}_t|}.$$

Combining this with the fact with item (4) (that at most $O(\beta^2 + \eta d' \delta_X^{-1} \delta_Y^{-1} \beta^{-2})$ mass is thrown out) completes the proof. $\square$

Iterating Lemma 6.20 proves Lemma 6.19.

Proof of Lemma 6.19. Perform the following algorithm given $X \times Y \subseteq B_1 \times B_2$. Return $X_i \times Y_i \subseteq (x_i + B_1^{(i)}) \times (y_i + B_2^{(i)})$ for $i = 1, \dots, T$ as given in Lemma 6.20. Now, for each $i \in [T]$ such that $\mathcal{E}(X_i, Y_i)$ does not hold, invoke Lemma 6.20 recursively. Terminate when the recursion depth is $O(\log(1/\beta))$, and remove any remaining pieces from the partition.

Let's check that all the conclusions of Lemma 6.19 hold. First note that for all pieces $X' \times Y'$ in the partition that $\delta_{X'} \delta_{Y'} \geq 2^{-O(\log(1/\beta)^2)} \delta_X \delta_Y$. So these are the parameters δ_X and δ_Y that one applies Lemma 6.20 with. Now, (1) holds by construction. (2), (3), (5) hold by applying (2), (3), (5) of Lemma 6.20 respectively for recursion depth $O(\log(1/\beta))$ – note that

$$g \leq O(\varepsilon^{-1} \log(1/(\delta_{X'} \delta_{Y'}))) \leq O(\varepsilon^{-1} \log(1/(\delta_X \delta_Y))) + \varepsilon^{-1} \log(1/\beta^2).$$

(4) holds by using item (4) of Lemma 6.20, plus that after $O(\log(1/\beta))$ rounds the total size of the “bad event” (not in $\mathcal{E}$) is at most $2^{-O(\log(1/\beta))} \leq \beta^{O(1)}$. $\square$

Applying Lemma 6.19 gives a way to split $S(X, Y, D)$ into $S(X_i, Y_i, D)$ where $X_i \times Y_i$ are a partition of measure of $X \times Y$ (minus some small error). Say that $X_i \subseteq x_i + B_1^{(i)}$ and $Y_i \subseteq y_i + B_2^{(i)}$ as in Lemma 6.19. Then $S(X_i, Y_i, D) \approx S(X_i, Y_i, D_i)$ where $D_i := D \cap (x_i + y_i + B_1^{(i)})$. While X_i and Y_i are algebraically spread, D_i may not be ℓ_1 -spread, which we wanted to enforce. Thus, we design a scheme to partition D_i into ℓ_1 -spread pieces, while not affecting the average measure of X_i and Y_i on those pieces.

Lemma 6.21. *Let $X, D \subseteq B_1$ and $Y \subseteq B_2$ where $B_1, B_2, \dots$ form a (d, η) -small sequence of Bohr sets. Assume that X and Y are (r, η_s, ε) -algebraically spread. Let $\beta \in (0, 1)$ be a parameter, where $\eta \leq (\varepsilon\beta/d)^{O(1)}$. There is a distribution ν over $(x_i + B_1^{(i)}) \times (y_i + B_2^{(i)})$ satisfying the following properties:*

- (1) $B_1^{(i)}, B_2^{(i)}, \dots$ form a (d, η) -exact sequence of Bohr sets.
- (2) $\nu(B_1^{(i)}) \geq \eta^{O(\log(1/(\varepsilon\alpha\beta))^2)} \nu(B_1)$.
- (3) For all $(x_i + B_1^{(i)}) \times (y_i + B_2^{(i)}) \sim \mu$, it holds that either:
 - (a) $D_i := D \cap (x_i + y_i + B_1^{(i)})$ is $(B_1^{(i)}, B_9^{(i)}, \varepsilon)$ ℓ_1 -spread, or
 - (b) $|D_i|/|B_1^{(i)}| \leq \beta^3$.
- (4) For any 1-bounded function $f : B_1 \times B_2 \rightarrow [-1, 1]$ it holds that

$$\mathbb{E}_{x \in B_1, y \in B_2} [f(x, y)] = \mathbb{E}_{(x_i + B_1^{(i)}) \times (y_i + B_2^{(i)}) \sim \mu} \mathbb{E}_{x \sim x_i + B_1^{(i)} \atop y \sim y_i + B_2^{(i)}} [f(x, y)] + O(\beta^3).$$

- (5) The probability over $(x_i + B_1^{(i)}) \times (y_i + B_2^{(i)}) \sim \mu$ that $X_i := X \cap (x_i + B_1^{(i)})$ and $Y_i := Y \cap (y_i + B_2^{(i)})$ are both $(r, \eta'_s, 10\varepsilon)$ -algebraically spread for $\eta'_s := \eta^{-O(\log(1/(\varepsilon\alpha\beta))^2)} \eta_s$ is at least $1/2$.

Proof. We first describe how to do one round of partitioning to D , which we then recurse heavily on.

One round partitioning: Let $\gamma = \beta^6$. We define a sequence of distributions $\mu^{(0)}, \mu^{(1)}, \dots$ over shifted regular Bohr sets $x + B' \subseteq B_1$. $\mu^{(0)}$ is defined to have all mass on B_1 . Let t be a time step, and for $x + B' \sim \mu^{(t)}$ say that $x + B'$ is *bad* if $\frac{|D \cap (x + B')|}{|B'|} \geq \gamma$ and $D \cap (x + B')$ is not (B', B'', ε) ℓ_1 -spread where B'' has the same frequencies as B' and $\nu(B'')/\nu(B') \in [\eta^8/500, \eta^8]$. If the probability over $x + B' \sim \mu$ that $x + B'$ is bad is at least $1/2$, define $\mu^{(t+1)}$ as follows. For the not bad $x + B' \sim \mu$, keep the same mass in $\mu^{(t)}$. Otherwise, replace the mass of $x + B'$ with $x + x' + B''$ where $x' \sim B'$ is uniform.

We prove that this process terminates within $T = O(\varepsilon^{-2} \log(1/\gamma))$ steps. To prove this, consider the potential function

$$\Phi^{(t)} := \mathbb{E}_{x + B' \sim \mu^{(t)}} \left[-\log \left(\gamma + \frac{|D \cap (x + B')|}{|B'|} \right) \right].$$

Note that $\Phi^{(0)} \geq -\log(\gamma + 1) \geq -\gamma$ and $\Phi^{(t)} \leq \log(1/\gamma)$ always. We will use the following inequality in our analysis: for any $\delta_0, \delta_1 > 0$ it holds that

$$-\log \delta_1 = -\log \delta_0 - \log \left(1 + \left(\frac{\delta_1}{\delta_0} - 1 \right) \right) \geq -\log \delta_0 + 1 - \frac{\delta_1}{\delta_0} + \Omega \left(\min \left\{ 1, \left(\frac{\delta_1}{\delta_0} - 1 \right)^2 \right\} \right). \quad (6.1)$$

We now lower bound $\Phi^{(t+1)}$. Consider a bad $x + B'$ and let $\delta := \frac{|(x + B') \cap D|}{|B'|}$. For $x' \sim B'$ define $\delta_{x'} := \frac{|(x + x' + B'') \cap D|}{|B''|}$ – we know that $\mathbb{E}_{x'} [\delta_{x'}] = \delta \pm O(\eta d)$. Additionally, because $D \cap (x + B')$ was not (B', B'', ε) ℓ_1 -spread we know that

$$\mathbb{E}_{x'} [\mathbb{1}_{\delta_{x'} < \delta} (\delta - \delta_{x'})] \geq \varepsilon \delta / 2 - O(\eta d). \quad (6.2)$$

Combining this with (6.1) gives that

$$\begin{aligned} \mathbb{E}_{x'} [-\log(\gamma + \delta_{x'})] + \log(\gamma + \delta) &\geq \mathbb{E}_{x'} \left[1 - \frac{\gamma + \delta_{x'}}{\gamma + \delta} + \Omega \left(\min \left\{ 1, \frac{(\delta - \delta_{x'})^2}{(\gamma + \delta)^2} \right\} \right) \right] \\ &\geq -O(\eta d \cdot \gamma^{-1}) + \Omega(\varepsilon^2) \geq \Omega(\varepsilon^2), \end{aligned}$$

where we have used that $\gamma + \delta \leq 2\delta$ and (6.2). Thus by definition $\Phi^{(t+1)} \geq \Phi^{(t)} + \Omega(\varepsilon^2)$ (at least $1/2$ of $x + B' \sim \mu^{(t)}$ was bad), so the total number of iterations is bounded by $T = O(\varepsilon^{-2} \log(1/\gamma))$ as desired.

Recursive partitioning: Given $\mu^{(T)}$ where the one round partitioning scheme terminated, consider each bad $x + B' \sim \mu^{(T)}$. For each of these, apply the one round partitioning scheme to it recursively. Do this until the recursion depth is $O(\log(1/\gamma))$.

Constructing μ : We now describe the distribution μ . Let μ' be the distribution over $x + B'$ after doing the recursive partitioning described in the above paragraph. For $x + B' \sim \mu'$ put the following mass in μ : pick $y \sim B_2$ uniformly, and put $(x - y + B') \times (y + B'')$ in μ , where B'' is chosen to be regular, with the same frequencies as B' , and $\nu(B'')/\nu(B') \in [\eta/2, \eta]$.

Verifying conclusions: We prove that the construction described satisfies all the desired conclusions. (1) follows by construction. (2) follows because the recursion depth of the recursive partitioning is $O(\log(1/\beta))$ and the number of iterations in the one round partition is $O(\varepsilon^{-2} \log(1/\beta))$, each of which drops the radius by $\eta^{O(1)}$. (3) follows by construction, after throwing out the bad pieces (of which the total mass is at most β^{100} , because of the recursion).

We now verify (4). It suffices to verify (4) under a single partition of B_1 in the one-round partitioning scheme, as we can then apply induction. This amounts to verifying that

$$\mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} \mathbb{E}_{\substack{x' \in x - y + B_9 \\ y' \in y + B_{10}}} f(x', y') = \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2 \\ x' \sim B_9 \\ y' \sim B_{10}}} f(x - y + x', y + y') = \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} f(x, y) + O(\eta d),$$

as desired. Because there are $O(\log(1/\beta)^2)$ layers of recursion and the total size of bad pieces is at most β^4 by the recursion, the result follows.

(5) follows because the expectation density of X_i is $\delta_X + O(\eta d)$ by (4) applied to $f = \mathbb{1}_X$. Thus the spreadness of X and the reverse Markov inequality (Fact 4.3) imply that the probability over μ that $|X_i|/|B_1^{(i)}| \geq (1 - 8\varepsilon)\delta_X$ is at least $3/4$. The same reasoning applies to Y , and thus (5) follows by a union bound. $\square$

We now have the necessary tools to start with $A \subseteq S(X, Y, D)$ and pass to a subset $A' \subseteq S(X', Y', D')$ where X', Y' are algebraically spread and D' is ℓ_1 -spread. The benefit of this is that for example, the size of the container $S(X', Y', D')$ is close to what is expected.

Theorem 6.22. *Let $X, D \subseteq B_1$ and $Y \subseteq B_2$ where $B_1, B_2, \dots$ form a (d, η) -small sequence of Bohr sets. Define $\delta_Y = |Y|/|B_2|$, $\delta_D = |D|/|B_1|$, and $\tilde{\delta}_X := \|X\|_{(B_1, B_2, B_3, K, K)}$. Let $A \subseteq S(X, Y, D)$ with $|A| \geq \alpha \tilde{\delta}_X \delta_Y \delta_D |B_1| |B_2|$. Then there are shifted Bohr sets $x + B'_1 \subseteq B_1$ and $y + B'_2 \subseteq B_2$ where $B'_1, B'_2, \dots$ form a (d', η) -exact sequence of Bohr sets, and X', Y', D' satisfying:*

- (1) $X' \subseteq X \cap (x + B'_1)$, $Y' \subseteq Y \cap (y + B'_2)$, and $D' \subseteq D \cap (x + y + B'_1)$,
- (2) $d' \leq \text{rank}(B_1) + O(\eta g^2 \log(1/(\varepsilon \alpha \delta_D))^2)$, and $\nu(B'_1) \geq \eta_s^{O(g^2 \log(1/(\varepsilon \alpha \delta_D))^2)} \nu(B_1)$, for

$$g = O(\log(1/(\delta_X \delta_Y)) + \log(1/(\varepsilon \alpha \delta_D))^3),$$

- (3) X' and Y' are $(r, \eta_s, \varepsilon_s)$ -algebraically spread,
- (4) D' is $(B'_1, B'_2, \varepsilon_s)$ ℓ_1 -spread,
- (5) $|D'|/|B'_1| \geq \Omega(\varepsilon \alpha \delta_D)$, and $\frac{|X'| |Y'|}{|B'_1| |B'_2|} \geq 2^{-O(\log(1/(\varepsilon_s \alpha \delta_D))^3)} \tilde{\delta}_X \delta_Y$, and
- (6) $A' := A \cap S(X', Y', D')$ satisfies $|A'| \geq (1 - \varepsilon) \alpha |S(X', Y', D')|$.

Proof. We define a sequence of distributions $\mu^{(0)}, \dots, \mu^{(T)}$ over triples (X_i, Y_i, D_i) such that there are $B_1^{(i)}, B_2^{(i)}$ and x_i, y_i such that $X_i \subseteq x_i + B_1^{(i)}$, $Y_i \subseteq y_i + B_2^{(i)}$, and $D_i \subseteq x_i + y_i + B_1^{(i)}$. Define $\mu^{(0)}$ to be identically (X, Y, D) .

At a time step t , first define $\mu^{(t)'}$ to be the result of replacing each $(X', Y', D') \sim \mu^{(t)}$ which does not satisfy (3) in [Theorem 6.22](#) with the distribution given by applying [Lemma 6.19](#) with the choice $\beta = (\varepsilon\alpha\delta_D)^{O(1)}$, where $\delta_D = |D|/|B_1|$ (the original density of D), but throwing out any subpiece where the densities of X'' or Y'' has decreased from that of X' or Y' by more than a $\beta^{O(1)}$ factor (recall that it cannot increase significantly because X', Y' are algebraically spread). Formally, [Lemma 6.19](#) gives a distribution over subsets $X'' \times Y'' \subseteq X' \times Y'$ where $X'' \subseteq x'' + B_1''$ and $Y'' \subseteq y'' + B_2''$ for some regular Bohr sets B_1'', B_2'' . In this case, put the triple (X'', Y'', D'') in the distribution where $D'' := D' \cap (x'' + y'' + B_1'')$.

Now define $\mu^{(t+1)}$ by replacing each $(X', Y', D') \sim \mu^{(t)'}$ which does not satisfy (4) in [Theorem 6.22](#) with the distribution given by [Lemma 6.21](#).

Consider $\mu^{(T)}$ for $T = O(\log(1/(\varepsilon_s\alpha\delta_D)))$. We want to extract a piece (X', Y', D') where X', Y', D' are all not too small, and $A \cap S(X', Y', D')$ is relatively large. This motivates the calculation

$$\begin{aligned} \mathbb{E}_{(X', Y', D') \sim \mu^{(T)}} & \left[\frac{1}{|X'| |Y'|} \left(|A \cap S(X', Y', D')| - (1 - \varepsilon)\alpha |S(X', Y', D')| - \frac{\varepsilon}{2}\alpha\delta_D |X'| |Y'| \right) \right] \\ & \geq \frac{1}{|X| |Y|} |A| - \frac{1}{|X| |Y|} (1 - \varepsilon)\alpha |S(X, Y, D)| - \frac{\varepsilon}{2}\alpha\delta_D - O(\beta T) - 2^{-T} \\ & \geq \frac{\varepsilon}{3}\alpha\delta_D. \end{aligned}$$

Here we have repeatedly used item (4) of [Lemma 6.19](#) and [Lemma 6.21](#). Also, we threw out at most $\beta^{O(1)}$ extra mass from small pieces when going from $\mu^{(t)'}$ to $\mu^{(t+1)}$. Finally, the 2^{-T} is because the number of iterations is $T = O(\log(1/(\varepsilon_s\alpha\delta_D)))$, and the amount of mass in level T that is still *bad* is at most 2^{-T} by item (5) of [Lemma 6.21](#). Thus there is a piece (X', Y', D') where

$$|A \cap S(X', Y', D')| \geq (1 - \varepsilon)\alpha |S(X', Y', D')| + \frac{\varepsilon}{3}\alpha\delta_D |X'| |Y'|.$$

Let us verify the hypotheses for this choice of X', Y', D' . Indeed, (1), (2), (3), (4), (6) follow by construction. For (5), the bound on $\frac{|X'| |Y'|}{|B_1'| |B_2'|}$ follows by [Lemma 6.19](#) item (5). For D' , note that

$$\frac{\varepsilon}{3}\alpha\delta_D |X'| |Y'| \leq |S(X', Y', D')| \leq 2\delta_{D'} |X'| |Y'|$$

where we have used [Lemma 6.10](#) and the spreadness of X' and Y' . □

7. DENSITY INCREMENT BY REDUCTION TO THE GRID NORM

7.1. Density increment. We now carry out the requisite density increment argument over general abelian groups. The analysis here corresponds in a certain sense to combining [Lemma 5.11](#), [Lemma 5.12](#), and [Lemma 5.13](#); however the analysis is rather more technical in the case of general abelian groups.

The first technical issue is that the parametrization used for corners over finite fields

$$(x', y'), (x', -x' + z'), \text{ and } (-y' + z', y')$$

directly is unusable over general group. This is due to the fact that we are assuming our set of corners lives $B_1 \times B_2$ and thus any naive parameterization fails to give that each coordinate varies over $B_1 \times B_2$ uniformly. This is fixed via parameterizing corners in the form

$$(x + x', y + y'), (x + x', y - x' + z'), \text{ and } (x - y' + z', y + y').$$

In this parametrization we will always have that $x \sim B_1$, $y \sim B_2$, $x' \sim B_3$, $y' \sim B_4$, and $z' \sim B_5$.

In our analysis $B_1, B_2, \dots$ will be a (d, η) -exact sequence of Bohr sets. We will have 3 majorants, corresponding to rows, columns, and diagonals which satisfy

$$\mathbb{E}_{x \sim B_1} [\mathbb{1}_X(x)] = \delta_X, \quad \mathbb{E}_{y \sim B_2} [\mathbb{1}_Y(y)] = \delta_Y, \quad \text{and} \quad \mathbb{E}_{z \sim B_1} [\mathbb{1}_D(z)] = \delta_D.$$

Furthermore we will assume that the indicator of our set $f : G \times G \rightarrow \{0, 1\}$ satisfies

$$\mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} f(x, y) = \alpha \cdot \delta_X \delta_Y \delta_D.$$

We additionally assume that our set possesses few corners; precisely we will assume that

$$\sup_{\substack{x \in B_1 \ x' \sim B_3 \\ y \in B_2 \ y' \sim B_4 \\ z' \sim B_5}} \mathbb{E} f(x + x', y + y') f(x + x', y - x' + z') f(x - y' + z', y + y') \leq 2^{-3} \cdot \alpha^3 \cdot \delta_X^2 \delta_Y^2 \delta_D^2.$$

We fix a set of constant size parameters

$$1/C \ll \varepsilon_s \ll \varepsilon_R \ll \varepsilon_L \ll \varepsilon,$$

where $\ll$ means “much less”, so that for example ε_L is chosen to be sufficiently small in terms of ε , etc. Also, by “constant size” we mean that these will be chosen to be absolute constants independent of α or $\delta_X, \delta_Y, \delta_D$.

We define $K = C \lceil \log(4/(\delta_X \delta_Y \delta_D)) \rceil$ and assume that $\eta \leq e^{-O(K^3)} \cdot d^{-O(1)}$. The crucial input pseudorandomness condition on the container functions (X, Y, D) will be that

$$\begin{aligned} \|\mathbb{1}_X\|_{(B_1, B_8, B_9, K, K)} &\leq (1 + \varepsilon_s) \delta_X \\ \|\mathbb{1}_Y\|_{(B_2, B_8, B_9, K, K)} &\leq (1 + \varepsilon_s) \delta_Y \\ \mathbb{E}_{x \sim B_1} \mid \mathbb{E}_{z \sim B_{10}} \mathbb{1}_D(x + z) - \delta_D \mid &\leq \varepsilon_s \cdot \delta_D. \end{aligned}$$

Intuitively, this means that X, Y are algebraically spread and that D is ℓ_1 -spread. This asymmetry between X, Y , and D will persist throughout much of the analysis and cause substantial headaches.

The first step in our analysis is noting that when passing to smaller Bohr sets various grid norm guarantees are likely to persist.

Lemma 7.1. *Fix $i \in \{3, 4, 5, 6, 7\}$ and for $x \in B_1$ denote*

$$\mathcal{E}_{X,i}(x) = \mathbb{1} \left[\|\mathbb{1}_X(x + \cdot)\|_{(B_i, B_8, B_9, K, K)} \geq (1 + 2\varepsilon_s) \delta_X \right].$$

Denote $\mathcal{E}_{Y,i}$ analogously. Then

$$\mathbb{E}_{x \sim B_1} [\mathcal{E}_{X,i}(x)] \leq e^{-\Omega(\varepsilon_s K^2)}$$

and analogously for $\mathcal{E}_{Y,i}$.

Proof. We may observe that

$$\begin{aligned} \|\mathbb{1}_X\|_{(B_1, B_8, B_9, K, K)}^{K^2} &= \mathbb{E}_{\substack{x \sim B_1 \\ y_1, \dots, y_K \sim B_8 \\ z_1, \dots, z_K \sim B_9}} \prod_{\substack{j \in [K] \\ \ell \in [K]}} \mathbb{1}_X(x + y_j + z_\ell) \\ &= \mathbb{E}_{\substack{x \sim B_1, x' \sim B_i \\ y_1, \dots, y_K \sim B_8 \\ z_1, \dots, z_K \sim B_9}} \prod_{\substack{j \in [K] \\ \ell \in [K]}} \mathbb{1}_X(x + x' + y_j + z_\ell) \pm O(K^2 \eta d) \\ &= \mathbb{E}_{x \sim B_1} \|\mathbb{1}_X(x + \cdot)\|_{(B_i, B_8, B_9, K, K)}^{K^2} \pm O(K^2 \eta d) \\ &\geq \mathbb{E}_{x \sim B_1} \mathcal{E}_{X,i}(x) \cdot ((1 + 2\varepsilon_s) \cdot \delta_X)^{K^2} - O(K^2 \eta d). \end{aligned}$$

Recall that we assume that $\|\mathbb{1}_X\|_{(B_1, B_8, B_9, K, K)} \leq (1 + \varepsilon_s)\delta_X$. Rearranging and using $\eta \leq e^{-O(K^3)}$. $d^{-O(1)}$ we immediately have the desired bound. $\square$

We define $(x, y) \in B_1 \times B_2$ to be *well-conditioned* if

$$x \in \bigcap_{4 \leq j \leq 7} \text{supp}(1 - \mathcal{E}_{X,i}) \text{ and } y \in \bigcap_{4 \leq j \leq 7} \text{supp}(1 - \mathcal{E}_{Y,i}).$$

Next we define different types of dense rectangles, which are cheap ways to obtain a density increment. Otherwise, we are able to guarantee that on most of the space that the densities of the objects we care about such as X, Y, D are relatively unchanged when passing to the smaller Bohr sets.

We say (x, y) gives a *dense* rectangle of Type 1 if (x, y) is well-conditioned and

$$\begin{aligned} \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} f(x + x', y + y') &\geq (1 + \varepsilon_R)\alpha\delta_X\delta_Y \cdot \mathbb{E}_{z \sim B_3} \mathbb{1}_D(x + y + z) \\ \mathbb{E}_{z \sim B_3} \mathbb{1}_D(x + y + z) &\geq \varepsilon_s^2 \cdot \alpha^2 \cdot \delta_D. \end{aligned}$$

We say (x, y) gives a *dense* rectangle of Type 2 if (x, y) is well-conditioned and

$$\begin{aligned} \mathbb{E}_{\substack{x' \sim B_3 \\ z' \sim B_5}} f(x + x', y - x' + z') &\geq (1 + \varepsilon_R)\alpha\delta_X\delta_Y \cdot \mathbb{E}_{z' \sim B_5} \mathbb{1}_D(x + y + z') \\ \mathbb{E}_{z' \sim B_5} \mathbb{1}_D(x + y + z') &\geq \varepsilon_s^2 \cdot \alpha^2 \cdot \delta_D. \end{aligned}$$

We say (x, y) gives a *dense* rectangle of Type 3 if (x, y) is well-conditioned and

$$\begin{aligned} \mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} f(x - y' + z', y + y') &\geq (1 + \varepsilon_R)\alpha\delta_X\delta_Y \cdot \mathbb{E}_{z' \sim B_5} \mathbb{1}_D(x + y + z') \\ \mathbb{E}_{z' \sim B_5} \mathbb{1}_D(x + y + z') &\geq \varepsilon_s^2 \cdot \alpha^2 \cdot \delta_D. \end{aligned}$$

We now prove that if (x, y) gives a dense rectangle this is sufficient to obtain a density increment.

Lemma 7.2. *Suppose that (x, y) gives a dense rectangle of Type 1, 2, or 3. Then there exists $i \in \{3, 4, 5, 6, 7\}$ and (x^*, y^*) such that the following all hold:*

$$\mathbb{E}_{\substack{x' \sim B_i \\ y' \sim B_{i+1}}} f(x^* + x', y^* + y') \geq (1 + \varepsilon_R/2) \cdot \alpha\delta_X\delta_Y \cdot \mathbb{E}_{z' \sim B_i} \mathbb{1}_D(x^* + y^* + z')$$

$$\mathbb{E}_{z' \sim B_i} \mathbb{1}_D(x^* + y^* + z') \geq \varepsilon_s^4 \cdot \alpha^3 \cdot \delta_D$$

$$\|\mathbb{1}_X(x^* + \cdot)\|_{(B_i, B_8, B_9, K, K)} \leq (1 + 3\varepsilon_s)\delta_X$$

$$\|\mathbb{1}_Y(y^* + \cdot)\|_{(B_{i+1}, B_8, B_9, K, K)} \leq (1 + 3\varepsilon_s)\delta_Y.$$

Proof. We observe that the Type 1 case is immediate by definition. The Type 2 and Type 3 cases are essentially identical; we handle the Type 2 case.

The input condition to our lemma implies that

$$\mathbb{E}_{\substack{x' \sim B_3 \\ z' \sim B_5}} f(x + x', y - x' + z') \geq (1 + 3\varepsilon_R/4)\alpha\delta_X\delta_Y \cdot \mathbb{E}_{z' \sim B_5} \mathbb{1}_D(x + y + z') + \delta_X\delta_Y \cdot \varepsilon_s^3\alpha^2\delta_D.$$

Combining the above with [Lemma 6.5](#) and the choice of η now gives

$$\mathbb{E}_{\substack{x' \sim B_3 \\ z' \sim B_5}} \left(\mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} f(x + x' + x'', y - x' + z' + y'') - (1 + 2\varepsilon_R/3)\alpha\delta_X\delta_Y \cdot \mathbb{1}_D(x + y + z' + x'' + y'') \right)$$

$$\geq \delta_X \delta_Y \cdot \varepsilon_s^3 \alpha^2 \delta_D.$$

Let

$$\begin{aligned} \mathcal{E}(x', z') &= \mathbb{1} \left[\|\mathbb{1}_X(x + x' + \cdot)\|_{(B_6, B_8, B_9, K, K)} \leq (1 + 3\varepsilon_s) \delta_X \right] \\ &\quad \cdot \mathbb{1} \left[\|\mathbb{1}_Y(y - x' + z' + \cdot)\|_{(B_7, B_8, B_9, K, K)} \leq (1 + 3\varepsilon_s) \delta_X \right]. \end{aligned}$$

Then via the proof of [Lemma 7.1](#), we have that

$$\begin{aligned} \mathbb{E}_{\substack{x' \sim B_3 \\ z' \sim B_5}} \mathcal{E}(x', z') &\left(\mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} f(x + x' + x'', y - x' + z' + y'') - (1 + 3\varepsilon_R/5) \alpha \delta_X \delta_Y \cdot \mathbb{1}_D(x + y + z' + x'' + y'') \right) \\ &\geq \delta_X \delta_Y \cdot \varepsilon_s^3 \alpha^2 \delta_D. \end{aligned}$$

Let

$$\mathcal{E}_2(x', z') = \mathbb{1} \left[\mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} D(x + y + z' + x'' + y'') \geq \varepsilon_s^4 \cdot \alpha^3 \delta_D \right].$$

Via [Lemma 6.10](#), we have that

$$\begin{aligned} \mathbb{E}_{\substack{x' \sim B_3 \\ z' \sim B_5}} \mathcal{E}(x', z') \mathcal{E}_2(x', z') &\left(\mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} f(x + x' + x'', y - x' + z' + y'') \right. \\ &\quad \left. - (1 + 3\varepsilon_R/5) \alpha \delta_X \delta_Y \cdot \mathbb{1}_D(x + y + z' + x'' + y'') \right) > 0. \end{aligned}$$

Taking x' and z' such that $\mathcal{E}(x', z') \mathcal{E}_2(x', z') = 1$ and the inner expression is strictly positive then completes the proof. $\square$

A major technical issue when handling the case of well-conditioned rectangles is that one has no matching lower bounds on the density of X , Y , and D . We now use the reverse Markov inequality ([Fact 4.3](#)) to ensure that the density is preserved with at least “99%” probability.

Lemma 7.3. *For $i \in \{3, 4, 5, 6, 7\}$, define*

$$\mathcal{E}_{\text{Small}, X, i}(x) = \mathbb{1} \left[\mathbb{E}_{x' \sim B_i} \mathbb{1}_X(x + x') \leq (1 - \varepsilon_s^{1/2}) \delta_X \right].$$

Define $\mathcal{E}_{\text{Small}, Y, i}$ analogously. Then we have that

$$\mathbb{E}_{x \sim B_1} [\mathcal{E}_{\text{Small}, X, i}(x)] = O(\varepsilon_s^{1/2})$$

and similarly for $\mathcal{E}_{\text{Small}, Y, i}$.

Furthermore define

$$\mathcal{E}_{\text{Special}, D, i}(z) = \mathbb{1} \left[\mathbb{E}_{z' \sim B_i} \left| \mathbb{E}_{z'' \sim B_{10}} \mathbb{1}_D(z + z' + z'') - \delta_D \right| \geq \varepsilon_s^{1/2} \cdot \delta_D \right].$$

Then

$$\mathbb{E}_{z \sim B_1} [\mathcal{E}_{\text{Special}, D, i}(z)] = O(\varepsilon_s^{1/2}).$$

Proof. We first handle the case of X (with Y being completely analogous) and then handle the case of D .

For bounding $\mathcal{E}_{\text{Small}, X, i}$, observe that

$$\mathbb{E}_{x' \sim B_i} \mathbb{1}_X(x + x') \leq \mathbb{E}_{\substack{x' \sim B_i \\ x_1 \sim B_8 \\ x_2 \sim B_9}} \mathbb{1}_X(x + x' + x_1 + x_2) + O(\eta d) \leq \|\mathbb{1}_X(x + \cdot)\|_{(B_i, B_8, B_9, K, K)} + O(\eta d).$$

By applying [Lemma 7.1](#), we have that

$$\mathbb{E}_{x \sim B_1} (1 - \mathcal{E}_{X,i}(x)) \mathbb{E}_{x' \sim B_i} \mathbb{1}_X(x + x') \geq (1 - \varepsilon_s) \delta_X,$$

and for x such that $\mathcal{E}_{X,i}(x) = 0$ we have that $\mathbb{E}_{x' \sim B_i} \mathbb{1}_X(x + x') \leq (1 + 3\varepsilon_s) \delta_X$.

Applying [Fact 4.3](#) to $(1 - \mathcal{E}_{X,i}(x)) \mathbb{E}_{x' \sim B_i} \mathbb{1}_X(x + x')$, we have that

$$\Pr_{x \sim B_1} \left[(1 - \mathcal{E}_{X,i}(x)) \mathbb{E}_{x' \sim B_i} \mathbb{1}_X(x + x') \leq (1 - \varepsilon_s^{1/2}) \cdot \delta_X \right] = O(\varepsilon_s^{1/2}).$$

This combined with [Lemma 7.1](#) immediately gives the desired result for $\mathcal{E}_{\text{Small},X,i}$.

We now handle the event for D . The result follows via Markov's inequality and the bound that

$$\mathbb{E}_{\substack{z \sim B_1 \\ z' \sim B_i}} \left| \mathbb{E}_{z'' \sim B_{10}} \mathbb{1}_D(z + z' + z'') - \delta_D \right| \leq 2\varepsilon_s \delta_D,$$

giving the desired result. $\square$

We now define the notion of a great pair. We say (x, y) is a *great pair* if (x, y) is well-conditioned, $x \in \bigcap_{4 \leq j \leq 7} \text{supp}(1 - \mathcal{E}_{\text{Small},X,j})$, $y \in \bigcap_{4 \leq j \leq 7} \text{supp}(1 - \mathcal{E}_{\text{Small},Y,j})$, and $x + y \in \bigcap_{4 \leq j \leq 7} \text{supp}(1 - \mathcal{E}_{\text{Special},D,j})$ and that

$$\begin{aligned} \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} f(x + x', y + y') &\geq (1 - \varepsilon_R^{1/2}) \alpha \delta_X \delta_Y \delta_D \\ \mathbb{E}_{\substack{x' \sim B_3 \\ z' \sim B_5}} f(x + x', y - x' + z') &\geq (1 - \varepsilon_R^{1/2}) \alpha \delta_X \delta_Y \delta_D \\ \mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} f(x - y' + z', y + y') &\geq (1 - \varepsilon_R^{1/2}) \alpha \delta_X \delta_Y \delta_D. \end{aligned}$$

The existence of a “great pair” will allow us to completely localize our analysis to within a certain subset of rectangles where we have appropriately lower-bounded marginals. We will derive the existence of a “great pair” via an averaging argument and noting that the mass in “misbehaving” pairs is bounded via the condition of having no dense rectangles.

Lemma 7.4. *If f has no dense rectangles (of Type 1, 2, or 3), then there exists a great pair (x, y) .*

Proof. By assumption, we have that

$$\mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} f(x, y) = \alpha \delta_X \delta_Y \delta_D.$$

Define

$$\mathcal{E}_1(x, y) = \prod_{i \in \{3,4,5,6,7\}} (1 - \mathcal{E}_{X,i}(x))(1 - \mathcal{E}_{Y,i}(y)).$$

Via [Lemma 7.1](#), we immediately have that

$$\begin{aligned} \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} \mathcal{E}_1(x, y) \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ z' \sim B_5}} [f(x + x', y + y') + f(x + x', y - x' + z') + f(x - y' + z', y + y')] \\ = 3\alpha \delta_X \delta_Y \delta_D + O(\eta d) + O(e^{\Omega(-\varepsilon_s K^2)}). \end{aligned}$$

Let $G : B_1 \rightarrow [0, 1]$ be such that $\mathbb{E}[G(z)] \leq \tau$. Then

$$\begin{aligned} \mathbb{E}_{z \sim B_1} G(z) \mathbb{E}_{z' \sim B_i} \mathbb{1}_D(z + z') &= \mathbb{E}_{z \sim B_1} G(z) \mathbb{E}_{\substack{z' \sim B_i \\ z'' \sim B_9}} \mathbb{1}_D(z + z' + z'') + O(\eta d) \\ &\leq \mathbb{E}_{\substack{z \sim B_1 \\ z' \sim B_i}} G(z) \left| \mathbb{E}_{z'' \sim B_9} \mathbb{1}_D(z + z' + z'') - \delta_D \right| + \delta_X \cdot \tau + O(\eta d) \\ &\leq (\tau + \varepsilon_s) \cdot \delta_D + O(\eta d). \end{aligned}$$

We next define

$$\mathcal{E}_2(x, y) = \prod_{4 \leq j \leq 7} (1 - \mathcal{E}_{\text{Special}, D, j}(x + y)).$$

Via combining [Lemma 7.3](#) and that there are no dense rectangles, we have that

$$\begin{aligned} \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} \mathcal{E}_1(x, y) \mathcal{E}_2(x, y) \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ z' \sim B_5}} [f(x + x', y + y') + f(x + x', y - x' + z') + f(x - y' + z', y + y')] \\ \geq 3(1 - \varepsilon_R^{3/4}) \alpha \delta_X \delta_Y \delta_D. \end{aligned}$$

We are using implicitly here that if the density of D is especially small on the relevant shift of a Bohr set (e.g. $\leq \varepsilon_s^3 \cdot \alpha^2 \cdot \delta_D$) then [Lemma 6.10](#) removed the contribution.

Finally we may define

$$\mathcal{E}_3(x, y) = \prod_{4 \leq j \leq 7} (1 - \mathcal{E}_{\text{Small}, X, j}(x)) \cdot \prod_{4 \leq j \leq 7} (1 - \mathcal{E}_{\text{Small}, Y, j}(y)).$$

Here by the assumption of no dense rectangle and [Lemma 7.3](#), we immediately have that

$$\begin{aligned} \mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2}} \mathcal{E}_1(x, y) \mathcal{E}_2(x, y) \mathcal{E}_3(x, y) \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ z' \sim B_5}} [f(x + x', y + y') + f(x + x', y - x' + z') + f(x - y' + z', y + y')] \\ \geq 3(1 - 2\varepsilon_R^{3/4}) \alpha \delta_X \delta_Y \delta_D. \end{aligned}$$

Furthermore by the assumption of no dense rectangle, if $\mathcal{E}_1(x, y) \mathcal{E}_2(x, y) \mathcal{E}_3(x, y)$ hold then

$$\begin{aligned} \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} [f(x + x', y + y')] &\leq (1 + 2\varepsilon_R) \cdot \delta_X \delta_Y \delta_Z \\ \mathbb{E}_{\substack{x' \sim B_3 \\ z' \sim B_5}} [f(x + x', y - x' + z')] &\leq (1 + 2\varepsilon_R) \cdot \delta_X \delta_Y \delta_Z \\ \mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} [f(x - y' + z', y + y')] &\leq (1 + 2\varepsilon_R) \cdot \delta_X \delta_Y \delta_Z. \end{aligned}$$

The existence of a great pair then follows via the reverse Markov inequality ([Fact 4.3](#)). $\square$

For the remainder of the analysis in this section we will fix a pair (x, y) which is “great”.

The next step in our analysis is analogous to the “degree-regularization” procedures in the finite field section where low density rows are removed (see [Lemma 5.12](#)). We define

$$Y_-(y') = \mathbb{1}_Y(y + y') \mathbb{1} \left[\mathbb{E}_{z' \sim B_5} f(x - y' + z', y + y') \geq (1 - \varepsilon_L) \cdot \alpha \cdot \delta_X \delta_D \right].$$

We say that f is *poor* in the third coordinate with respect to the direction Y if

$$\mathbb{E}_{y' \sim B_4} [(\mathbb{1}_Y(y + y') - Y_-(y'))] \geq \varepsilon_L \cdot \delta_Y.$$

We now derive a density increment provided that f is poor in the third coordinate with respect to the direction Y .

Lemma 7.5. *Suppose that f is poor in the third coordinate with respect to the direction Y . Then there exists (x^*, y^*) and a function $g : B_7 \rightarrow \{0, 1\}$ such that:*

$$\begin{aligned} \mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} g(y'') \cdot f(x^* + x'', y^* + y'') &\geq (1 + \varepsilon_L^2/2) \cdot \alpha \delta_X \cdot \mathbb{E}_{y'' \sim B_7} g(y'') \cdot \mathbb{E}_{x'' \sim B_6} \mathbb{1}_D(x^* + y^* + x'') \\ \mathbb{E}_{y'' \sim B_7} [g(y'')] &\geq \varepsilon_L^2 \cdot \alpha^2 \cdot \delta_Y \\ \mathbb{E}_{x'' \sim B_6} \mathbb{1}_D(x^* + y^* + x'') &\geq \varepsilon_L^2 \cdot \alpha^2 \cdot \delta_D \\ g(y'') &\leq \mathbb{1}_Y(y^* + y'') \\ \|\mathbb{1}_X(x^* + \cdot)\|_{(B_6, B_8, B_9, K, K)} &\leq (1 + 3\varepsilon_s) \delta_X \\ \|\mathbb{1}_Y(y^* + \cdot)\|_{(B_7, B_8, B_9, K, K)} &\leq (1 + 3\varepsilon_s) \delta_Y. \end{aligned}$$

Proof. By assumption we have that

$$\mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} f(x - y' + z', y + y') \geq (1 - \varepsilon_R^{1/2}) \cdot \alpha \cdot \delta_X \delta_Y \delta_D.$$

We define $g_1 : B_4 \rightarrow \{0, 1\}$ such that $g_1(y') \leq Y_-(y')$ and $\mathbb{E}_{y' \sim B_4} g_1(y') = \varepsilon_L \cdot \delta_Y$. We have that

$$\mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} f(x - y' + z', y + y') (\mathbb{1}_Y(y + y') - g_1(y')) \geq (1 + 3\varepsilon_L^2/4) \cdot \alpha \cdot \delta_X \delta_D \cdot \mathbb{E}_{y' \sim B_4} [\mathbb{1}_Y(y + y') - g_1(y')].$$

This implies that if $g_2(y') = \mathbb{1}_Y(y + y') - g_1(y')$ then

$$\mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} \mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} f(x - y' + z' + x'', y + y' + y'') g_2(y' + y'') \geq (1 + 5\varepsilon_L^2/8) \cdot \alpha \cdot \delta_X \delta_D \cdot \mathbb{E}_{y' \sim B_4} [g_2(y')].$$

We now set

$$\begin{aligned} \mathcal{E}_1(y', z') &= \mathbb{1}[\|\mathbb{1}_X(x - y' + z' + \cdot)\|_{(B_6, B_8, B_9, K, K)} \leq (1 + 3\varepsilon_s) \delta_X \\ &\quad \vee \|\mathbb{1}_Y(y + y' + \cdot)\|_{(B_7, B_8, B_9, K, K)} \leq (1 + 3\varepsilon_s) \delta_Y]. \end{aligned}$$

Via the analysis in [Lemma 7.2](#), we immediately have that

$$\mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} \mathcal{E}_1(y', z') \mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} f(x - y' + z' + x'', y + y' + y'') g_2(y' + y'') \geq (1 + 9\varepsilon_L^2/16) \cdot \alpha \cdot \delta_X \delta_D \cdot \mathbb{E}_{y' \sim B_4} [g_2(y')].$$

We next define

$$\mathcal{E}_2(z') = \mathbb{1}[\mathbb{E}_{z'' \sim B_6} \mathbb{1}_D(x + y + z' + z'') \leq \alpha^2 \cdot \varepsilon_L^2 \cdot \delta_D].$$

Observe that

$$\begin{aligned} &\mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} \mathcal{E}_1(y', z') \mathcal{E}_2(z') \mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} f(x - y' + z' + x'', y + y' + y'') g_2(y' + y'') \\ &\leq \mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} \mathcal{E}_1(y', z') \mathcal{E}_2(z') \mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} \mathbb{1}_X(x - y' + z' + x'') \mathbb{1}_Y(y + y' + y'') \mathbb{1}_D(x + y + z' + x'' + y'') \\ &\leq 2\alpha^2 \cdot \varepsilon_L^2 \cdot \delta_X \delta_Y \delta_D. \end{aligned}$$

Thus we may assume that

$$\begin{aligned} & \mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} \mathcal{E}_1(y', z')(1 - \mathcal{E}_2(z')) \mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} f(x - y' + z' + x'', y + y' + y'') g_2(y' + y'') \\ & \geq (1 + 17\varepsilon_L^2/32) \cdot \alpha \cdot \delta_X \cdot \mathbb{E}_{z' \sim B_5} \mathbb{1}_D(x + y + z') \cdot \mathbb{E}_{y' \sim B_4} [g_2(y')]. \end{aligned}$$

We then define

$$\mathcal{E}_3(y') = \mathbb{1}[\mathbb{E}_{y'' \sim B_7} g_2(y' + y'') \leq \varepsilon_L^3 \cdot \alpha^2 \cdot \delta_Y].$$

We may observe that

$$\begin{aligned} & \mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} \mathcal{E}_1(y', z')(1 - \mathcal{E}_2(z')) \mathcal{E}_3(y') \mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} f(x - y' + z' + x'', y + y' + y'') g_2(y' + y'') \\ & \leq \mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} \mathcal{E}_1(y', z') \mathcal{E}_3(y') \mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} \mathbb{1}_X(x - y' + z' + x'') \mathbb{1}_D(x + y + z' + x'' + y'') g_2(y' + y''). \end{aligned}$$

We are now in position to use [Lemma 6.10](#) on the inner sum. Applying [Lemma 6.10](#) and using the definition of $\mathcal{E}_3$ (and using $\mathcal{E}_1$ to guarantee spreadness), we find that

$$\begin{aligned} & \mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} \mathcal{E}_1(y', z') \mathcal{E}_3(y') \mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} \mathbb{1}_X(x - y' + z' + x'') \mathbb{1}_D(x + y + z' + x'' + y'') g_2(y' + y'') \\ & \leq 2\alpha^2 \cdot \varepsilon_L^3 \cdot \delta_X \delta_Y \delta_D. \end{aligned}$$

Thus we have that

$$\begin{aligned} & \mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} \mathcal{E}_1(y', z')(1 - \mathcal{E}_2(z'))(1 - \mathcal{E}_3(y')) \left(\mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} f(x - y' + z' + x'', y + y' + y'') g_2(y' + y'') \right. \\ & \quad \left. - (1 + \varepsilon_L^2/2) \cdot \alpha \cdot \delta_X \cdot g_2(y' + y'') \mathbb{1}_D(x + y + z' + x'' + y'') \right) > 0. \end{aligned}$$

Choosing y' and z' such that the expression in the brackets is positive while $(1 - \mathcal{E}_1(y', z'))(1 - \mathcal{E}_2(y', z'))(1 - \mathcal{E}_3(y')) \neq 0$ gives the result. $\square$

In a completely analogous manner we may define that f is poor in the third coordinate in the Z direction, that f is poor in the second coordinate in the X or Z direction, or that f is poor in the first coordinate in the X or Y direction. At various points in the proof we will require that f is not poor in certain senses and we will derive these as the analysis proceeds and then handle the analog of [Lemma 7.5](#).

Recall that

$$Y_-(y') = \mathbb{1}_Y(y + y') \mathbb{1}[\mathbb{E}_{z' \sim B_5} f(x - y' + z', y + y') \geq (1 - \varepsilon_L) \cdot \alpha \cdot \delta_X \delta_D]$$

and define

$$f^-(y', z') = f(x - y' + z', y + y') \cdot Y_-(y').$$

Let α^* be such that

$$\alpha^* \delta_X \mathbb{E}_{y' \sim B_4} Y_-(y') \cdot \mathbb{E}_{z' \sim B_5} \mathbb{1}_D(x + y + z') = \mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} f^-(y', z');$$

observe by construction that $\alpha^* = \alpha(1 \pm \varepsilon_L^{1/2})$. Finally we define

$$h(y', z') = \alpha^* \delta_X Y_-(y') \mathbb{1}_D(x + y + z').$$

We will consider the difference between the following pair of counting operators. We will consider corners of the form

$$\begin{aligned} & \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ z' \sim B_5}} (f(x+x', y+y')L^-(y'))(f(x+x', y-x'+z'))(f(x-y'+z', y+y')Y_-(y')) \\ &= \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ z' \sim B_5}} (f(x+x', y+y')Y_-(y'))(f(x+x', y-x'+z'))(f^-(y', z')) \end{aligned}$$

and compare this to

$$\mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ z' \sim B_5}} (f(x+x', y+y')Y_-(y'))(f(x+x', y-x'+z'))(h(y', z')).$$

Due to the assumption that our set has few corners, we immediately have that

$$\begin{aligned} & \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ z' \sim B_5}} (f(x+x', y+y')Y_-(y'))(f(x+x', y-x'+z'))(f(x-y'+z', y+y')Y_-(y')) \\ & \leq \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ z' \sim B_5}} f(x+x', y+y')f(x+x', y-x'+z')f(x-y'+z', y+y') \leq 2^{-3} \cdot \alpha^3 \delta_X^2 \delta_Y^2 \delta_Z^2. \end{aligned}$$

On the other hand, we have that

$$\begin{aligned} & \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ z' \sim B_5}} (f(x+x', y+y')Y_-(y'))(f(x+x', y-x'+z'))(h(y', z')) \\ &= \alpha^* \delta_X \cdot \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ z' \sim B_5}} (f(x+x', y+y')Y_-(y'))(f(x+x', y-x'+z')) \\ &= \alpha^* \delta_X \cdot \mathbb{E}_{x' \sim B_3} \left(\mathbb{E}_{y' \sim B_4} f(x+x', y+y')Y_-(y') \right) \left(\mathbb{E}_{z' \sim B_5} f(x+x', y-x'+z') \right). \end{aligned}$$

We first observe that as (x, y) is not poor in the x direction in the second coordinate; for all but $O(\varepsilon_L \delta_X |B_3|)$ many $x' \in \text{supp}(X(x+\cdot))$ we have that

$$\mathbb{E}_{z' \sim B_5} f(x+x', y-x'+z') \geq (1-\varepsilon_L) \cdot \alpha \cdot \delta_Y \delta_D.$$

In the case where (x, y) is poor in the direction x in the second coordinate, we may use [Lemma 7.8](#) (with $g_2 = \mathbb{1}_D(x+y+\cdot)$ and g_1 being the indicator of the complement of the sparse x coordinates).

Thus there exists $X_- : B_3 \rightarrow \{0, 1\}$ with $X_-(\cdot) \leq \mathbb{1}_X(x+\cdot)$ and $\mathbb{E}_{x' \sim B_3} [X_-(x')] \geq (1-O(\varepsilon_L))\delta_X$ such that our count is lower bounded by

$$\alpha^2 \delta_X \delta_Y \delta_D \cdot (4/5) \cdot \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} f(x+x', y+y')Y_-(y')X_-(x').$$

Now observe that if

$$\mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} f(x+x', y+y')(\mathbb{1}_Y(y+y') - Y_-(y')) \geq \frac{1}{10} \cdot \alpha \delta_X \delta_Y \delta_D,$$

then we have a massive density increment as $\mathbb{E}_{y' \sim B_4}[(\mathbb{1}_Y(y+y') - Y_-(y'))] \leq \varepsilon_L \cdot \delta_Y$. This case immediately gives a density increment of the required type.

Analogously we have that if

$$\mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} f(x+x', y+y')(\mathbb{1}_X(x+x') - X_-(x')) \geq \frac{1}{10} \cdot \alpha \delta_X \delta_Y \delta_D$$

we have a massive density increment as $\mathbb{E}_{x' \sim B_3}[(\mathbb{1}_X(x+x') - X_-(x'))] \leq \varepsilon_L \cdot \delta_X$. We handle processing this density increment at the end.

Now observe that

$$\begin{aligned} X_-(x')Y_-(y') &\geq \mathbb{1}_X(x+x')\mathbb{1}_Y(y+y') - \mathbb{1}_Y(y+y')(\mathbb{1}_X(x+x') - X_-(x')) \\ &\quad - \mathbb{1}_X(x+x')(\mathbb{1}_Y(y+y') - Y_-(y')), \end{aligned}$$

and this is $(\mathbb{1}_X(x+x') - X_-(x'))(\mathbb{1}_Y(y+y') - Y_-(y')) \geq 0$. Thus if we do not have a massive density increment, we have that

$$\mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} f(x+x', y+y')X_-(x')Y_-(y') \geq \alpha \cdot \delta_X \delta_Y \delta_Z \cdot 3/4;$$

observe here that $1 - 2(1/10) = 4/5 > 3/4$. Thus we obtain a lower bound (or a suitable density increment) of the form

$$\mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ z' \sim B_5}} (f(x+x', y+y')Y_-(y'))(f(x+x', y-x'+z'))(h(y', z')) \geq \frac{3}{4} \cdot \alpha^3 \delta_X^2 \delta_Y^2 \delta_D^2.$$

Therefore we may assume that

$$\left| \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ z' \sim B_5}} (f(x+x', y+y')Y_-(y'))(f(x+x', y-x'+z'))(f^-(y', z') - h(y', z')) \right| \geq \frac{\alpha^3 \delta_X^2 \delta_Y^2 \delta_D^2}{2}.$$

We now handle the one deferred density increment.

Lemma 7.6. *Suppose that there exists $g : B_3 \rightarrow \{0, 1\}$ with $\mathbb{E}_{x' \sim B_3}[g(x')] = \varepsilon_L \cdot \delta_X$, $g(x') \leq \mathbb{1}_X(x+x')$, and*

$$\mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} f(x+x', y+y')g(x') \geq 2^{-4} \cdot \alpha \delta_X \delta_Y \delta_Z.$$

Then there exists (x^, y^*) and a function $h : B_5 \rightarrow \{0, 1\}$ such that:*

$$\mathbb{E}_{\substack{x'' \sim B_5 \\ y'' \sim B_4}} h(x'') \cdot f(x^* + x'', y^* + y'') \geq 2 \cdot \alpha \delta_Y \cdot \mathbb{E}_{x'' \sim B_5} h(x'') \cdot \mathbb{E}_{y'' \sim B_4} \mathbb{1}_D(x^* + y^* + y'');$$

$$\mathbb{E}_{x'' \sim B_5} [h(x'')] \geq \varepsilon_L^2 \cdot \alpha^2 \cdot \delta_D \cdot \delta_X;$$

$$\mathbb{E}_{y'' \sim B_4} \mathbb{1}_D(x^* + y^* + y'') \geq \varepsilon_L^2 \cdot \alpha^2 \cdot \delta_D;$$

$$h(x'') \leq \mathbb{1}_X(x^* + x'');$$

$$\|\mathbb{1}_Y(y^* + \cdot)\|_{(B_4, B_8, B_9, K, K)} \leq (1 + 3\varepsilon_s) \delta_Y.$$

Proof. Observe that the input relation may be written as

$$\mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ x'' \sim B_5}} f(x+x'+x'', y+y')g(x'+x'') \geq 2^{-4} \cdot \alpha \delta_X \delta_Y \delta_D - O(\eta d).$$

The crucial claim to prove is that

$$\mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ x'' \sim B_5}} g(x' + x'') \mathbb{1}_D(x + y + x' + x'' + y') = (1 \pm \varepsilon_s^{\Omega(1)}) \cdot \mathbb{E}_{x' \sim B_3} [g(x')] \cdot \delta_D.$$

To prove this, it suffices to note that

$$\begin{aligned} \left| \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ x'' \sim B_5}} g(x' + x'') (\mathbb{1}_D(x + y + x' + x'' + y') - \delta_D) \right| &= \left| \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ x'' \sim B_5}} g(x' + x'') (\mathbb{1}_D(x + y + x' + y') - \delta_D) \right| \\ &= \left| \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ x'' \sim B_5 \\ y'' \sim B_6}} g(x' + x'') (\mathbb{1}_D(x + y + x' + y' + y'') - \delta_D) \right| \\ &\leq \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} \left| \mathbb{E}_{x'' \sim B_5} \mathbb{1}_X(x' + x'') \right| \cdot \left| \mathbb{E}_{y'' \sim B_6} \mathbb{1}_D(x + y + x' + y' + y'') - \delta_D \right| \\ &\leq 2\delta_X \cdot \mathbb{E}_{x' \sim B_3} \cdot \left| \mathbb{E}_{y'' \sim B_6} \mathbb{1}_D(x + y + x' + y'') - \delta_D \right| + \varepsilon_s \delta_X \delta_D \leq 4\varepsilon_s \cdot \delta_X \delta_D. \end{aligned}$$

Thus we may write

$$\mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ x'' \sim B_5}} f(x + x' + x'', y + y') g(x' + x'') \geq 3\alpha \cdot \delta_Y \cdot \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ x'' \sim B_5}} g(x' + x'') \mathbb{1}_D(x + y + x' + x'' + y') + 2^{-5} \cdot \alpha \delta_X \delta_Y \delta_D.$$

We define

$$\mathcal{E}_1(x') = \mathbb{1} \left[\mathbb{E}_{x'' \sim B_5} g(x' + x'') \leq 2^{-7} \cdot \alpha^2 \cdot \delta_D \cdot \delta_X \right].$$

Then

$$\begin{aligned} \mathbb{E}_{x' \sim B_3} \mathcal{E}_1(x') \mathbb{E}_{\substack{y' \sim B_4 \\ x'' \sim B_5}} f(x + x' + x'', y + y') g(x' + x'') \\ \leq \mathbb{E}_{x' \sim B_3} \mathcal{E}_1(x') \mathbb{E}_{\substack{y' \sim B_4 \\ x'' \sim B_5}} \mathbb{1}_Y(y + y') g(x' + x'') \leq 2^{-6} \cdot \alpha^2 \cdot \delta_D \delta_X \delta_Y. \end{aligned}$$

Thus

$$\begin{aligned} \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} (1 - \mathcal{E}_1(x')) \mathbb{E}_{\substack{y' \sim B_4 \\ x'' \sim B_5}} f(x + x' + x'', y + y') g(x' + x'') \\ \geq 3\alpha \cdot \delta_Y \cdot \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ x'' \sim B_5}} g(x' + x'') \mathbb{1}_D(x + y + x' + x'' + y') + 2^{-6} \cdot \alpha \delta_X \delta_Y \delta_D. \end{aligned}$$

We define

$$\mathcal{E}_2(x') = \mathbb{1} \left[\|\mathbb{1}_X(x' + \cdot)\|_{(B_5, B_8, B_9, K, K)} \leq (1 + 3\varepsilon_s) \cdot \delta_X \right]$$

and we have that

$$\begin{aligned} \mathbb{E}_{x' \sim B_3} (1 - \mathcal{E}_1(x')) \mathcal{E}_2(x') \mathbb{E}_{\substack{y' \sim B_4 \\ x'' \sim B_5}} f(x + x' + x'', y + y') g(x' + x'') \\ \geq 3\alpha \cdot \delta_Y \cdot \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ x'' \sim B_5}} g(x' + x'') \mathbb{1}_D(x + y + x' + x'' + y') + 2^{-7} \cdot \alpha \delta_X \delta_Y \delta_D. \end{aligned}$$

We finally define

$$\mathcal{E}_3(x') = \mathbb{1} \left[\mathbb{E}_{\substack{y' \sim B_4 \\ x'' \sim B_5}} \mathbb{1}_D(x + y + x' + y' + x'') \leq 2^{-9} \cdot \alpha^2 \cdot \delta_D \right].$$

Then

$$\begin{aligned} & \mathbb{E}_{x' \sim B_3} \mathcal{E}_2(x') \mathcal{E}_3(x') \mathbb{E}_{\substack{y' \sim B_4 \\ x'' \sim B_5}} f(x + x' + x'', y + y') g(x' + x'') \\ & \leq \mathbb{E}_{x' \sim B_3} \mathcal{E}_2(x') \mathcal{E}_3(x') \mathbb{E}_{\substack{y' \sim B_4 \\ x'' \sim B_5}} \mathbb{1}_D(x + x' + x'' + y + y') \mathbb{1}_X(x' + x'') \mathbb{1}_Y(y + y') \leq 2^{-8} \cdot \alpha^2 \cdot \delta_X \delta_Y \delta_D. \end{aligned}$$

Thus

$$\begin{aligned} & \mathbb{E}_{x' \sim B_3} (1 - \mathcal{E}_1(x')) \mathcal{E}_2(x') (1 - \mathcal{E}_3(x')) \left(\mathbb{E}_{\substack{y' \sim B_4 \\ x'' \sim B_5}} f(x + x' + x'', y + y') g(x' + x'') \right. \\ & \quad \left. - 3\alpha \cdot \delta_Y \cdot g(x' + x'') \mathbb{1}_D(x + y + x' + x'' + y') \right) > 0 \end{aligned}$$

and choosing x' such that $(1 - \mathcal{E}_1(x')) \mathcal{E}_2(x') (1 - \mathcal{E}_3(x')) = 0$ and the internal bracket is positive gives the result. $\square$

We now let $k_* = 2^{20} \lceil \log(1/(\alpha \delta_D)) \rceil$. Observe that by Hölder's inequality, we have

$$\begin{aligned} & \left| \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4 \\ z' \sim B_5}} (f(x + x', y + y') Y_-(y')) (f(x + x', y - x' + z')) (f^-(y', z') - h(y', z')) \right|^{k_*} \\ & \leq \left(\mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} f(x + x', y + y') Y_-(y') \right)^{k_* - 1} \\ & \quad \cdot \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} f(x + x', y + y') Y_-(y') \left(\mathbb{E}_{z' \sim B_5} (f(x + x', y - x' + z')) (f^-(y', z') - h(y', z')) \right)^{k_*} \\ & \leq (2\alpha \delta_X \delta_Y \delta_D)^{k_* - 1} \cdot \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} \left(\mathbb{E}_{z' \sim B_5} (f(x + x', y - x' + z')) (f^-(y', z') - h(y', z')) \right)^{k_*}. \end{aligned}$$

Via rearranging, this implies that

$$4^{-k_*} \alpha^{2k_* + 1} (\delta_X \delta_Y \delta_D)^{k_* + 1} \leq \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} \left(\mathbb{E}_{z' \sim B_5} (f(x + x', y - x' + z')) (f^-(y', z') - h(y', z')) \right)^{k_*}.$$

Due to the choice of k_* , this in fact implies that

$$8^{-k_*} \alpha^{2k_*} (\delta_X \delta_Y)^{k_* + 1} \delta_D^{k_*} \leq \mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} \left(\mathbb{E}_{z' \sim B_5} (f(x + x', y - x' + z')) (f^-(y', z') - h(y', z')) \right)^{k_*}.$$

We remark here that this final step is crucial; we will be able to absorb the loss of a few factors of δ_D but will not be able to absorb corresponding losses in δ_X and δ_Y .

We next observe that

$$\begin{aligned} & \left(\mathbb{E}_{\substack{x' \sim B_3 \\ y' \sim B_4}} \left(\mathbb{E}_{z' \sim B_5} (f(x + x', y - x' + z')) (f^-(y', z') - h(y', z')) \right)^{k_*} \right)^2 \\ & = \mathbb{E}_{z'_1, \dots, z'_{k_*} \sim B_5} \left(\mathbb{E}_{x' \sim B_3} \prod_{j=1}^{k_*} f(x + x', y - x' + z'_j) \right) \cdot \left(\mathbb{E}_{y' \sim B_4} \prod_{j=1}^{k_*} (f^-(y', z'_j) - h(y', z'_j)) \right) \end{aligned}$$

$$\leq \left(\mathbb{E}_{z'_1, \dots, z'_{k_*} \sim B_5} \mathbb{E}_{x'_1, x'_2 \sim B_3} \prod_{j=1}^{k_*} f(x + x'_1, y - x'_1 + z'_j) f(x + x'_2, y - x'_2 + z'_j) \right) \\ \cdot \left(\mathbb{E}_{z'_1, \dots, z'_{k_*} \sim B_5} \mathbb{E}_{y'_1, y'_2 \sim B_4} \prod_{j=1}^{k_*} (f^-(y'_1, z'_j) - h(y'_1, z'_j))(f^-(y'_2, z'_j) - h(y'_2, z'_j)) \right).$$

Therefore we either have that at least one of the following holds:

$$2^{k_*} \alpha^{2k_*} \delta_X^{2k_*} \delta_Y^2 \delta_D^{k_*} \leq \left(\mathbb{E}_{z'_1, \dots, z'_{k_*} \sim B_5} \mathbb{E}_{x'_1, x'_2 \sim B_3} \prod_{j=1}^{k_*} f(x + x'_1, y - x'_1 + z'_j) f(x + x'_2, y - x'_2 + z'_j) \right) \\ 128^{-k_*} \alpha^{2k_*} \delta_X^2 \delta_Y^{2k_*} \delta_D^{k_*} \leq \left(\mathbb{E}_{z'_1, \dots, z'_{k_*} \sim B_5} \mathbb{E}_{y'_1, y'_2 \sim B_4} \prod_{j=1}^{k_*} (f^-(y'_1, z'_j) - h(y'_1, z'_j))(f^-(y'_2, z'_j) - h(y'_2, z'_j)) \right).$$

We will proceed under the assumption that the latter case holds; the former is strictly simpler.

Observe that $f^-(y, z) - h(y, z)$ is supported only on $\text{supp}(Y_-(\cdot)) \times \text{supp}(\mathbb{1}_D(x + y + \cdot))$. Furthermore for $y \in \text{supp}(Y_-(\cdot))$, we have that

$$\mathbb{E}_z f^-(y, z) \geq (1 - \varepsilon_L) \cdot \alpha \cdot \delta_X \delta_D.$$

Therefore we exactly have the required setup to apply [Lemma 4.2](#). Letting $p = 2^{1000} \cdot k_*$, we have that

$$(1 + 1/128)^{2p} \alpha^{2p} \delta_X^2 \delta_Y^{2p} \delta_D^p \leq \mathbb{E}_{y'_1, y'_2 \sim B_4} \mathbb{E}_{z'_1, \dots, z'_p \sim B_5} \prod_{j=1}^p f^-(y'_1, z'_j) f^-(y'_2, z'_j).$$

We now seek to apply [Theorem 3.5](#). In order to do so, we will need to properly bound the size of the associated container coming from (X, Y_-, D) . We claim that

$$\mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} \mathbb{1}_X(x - y' + z') Y_-(y') \mathbb{1}_D(x + y + z') = (1 \pm \varepsilon_L^{\Omega(1)}) \cdot \delta_X \delta_Y \delta_D.$$

To prove this, first observe that we have that

$$0 \leq \mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} \mathbb{1}_X(x - y' + z') (\mathbb{1}_Y(y + y') - Y_-(y')) \mathbb{1}_D(x + y + z') \leq \varepsilon_L^{\Omega(1)} \cdot \delta_X \delta_Y$$

where the first is by definitions and the second via [Lemma 6.10](#).

Therefore it suffices to bound

$$\mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} \mathbb{1}_X(x - y' + z') \mathbb{1}_Y(y + y') \mathbb{1}_D(x + y + z') = (1 + \varepsilon_s^{\Omega(1)}) \delta_X \delta_Y \delta_D.$$

This however is an immediate consequence of [Lemma 6.12](#).

We next observe that one may obtain suitable upper bounds on

$$\mathbb{E}_{\substack{y' \sim B_3 \\ z' \sim B_5}} [A(y') B(z') \mathbb{1}_X(x - y' + z')]$$

via [Lemma 6.10](#) for any functions $A : B_3 \rightarrow [0, 1]$ and $B : B_5 \rightarrow [0, 1]$.

Therefore we may apply [Theorem 3.5](#). We obtain functions $g_1(y') \leq \mathbb{1}_Y(y + y')$ and $g_2(z') \leq \mathbb{1}_D(x + y + z')$ such that:

$$\mathbb{E}_{y' \sim B_4} [g_1(y')] \geq \delta_Y \cdot (\alpha/2)^{O(k_*^2 \log(1/\alpha))} = \delta_Y \cdot (1/2)^{O(\log(1/(\alpha \delta_D))^2 \log(1/\alpha)^2)};$$

$$\mathbb{E}_{z' \sim B_5} [g_2(z')] \geq \delta_D \cdot (\alpha/2)^{O(\log(1/\alpha))} = \delta_D \cdot (1/2)^{O(\log(1/\alpha)^2)};$$

$$\mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} [f(x - y' + z', y + y') g_1(y') g_2(z')] \geq (1 + 2\varepsilon) \cdot \alpha \delta_X \cdot \mathbb{E}_{y' \sim B_4} [g_1(y')] \cdot \mathbb{E}_{z' \sim B_5} [g_2(z')].$$

The above density increment is on a tilted rectangle. We now obtain the desired density increment via a further averaging argument.

Lemma 7.7. *Suppose that (x, y) is a good pair. Furthermore consider $g_1 : B_4 \rightarrow \{0, 1\}$ and $g_2 : B_5 \rightarrow \{0, 1\}$ such that*

$$\mathbb{E}_{y' \sim B_4} [g_1(y')] \geq \delta_Y \cdot (1/2)^{O(\log(1/(\alpha\delta_D))^2 \log(1/\alpha)^2)};$$

$$\mathbb{E}_{z' \sim B_5} [g_2(z')] \geq \delta_D \cdot (1/2)^{O(\log(1/\alpha)^2)};$$

$$g_1(y') \leq \mathbb{1}_Y(y + y');$$

$$g_2(z') \leq \mathbb{1}_D(x + y + z');$$

$$\mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} f(x - y' + z', y + y') g_1(y') g_2(z') \geq (1 + 2\varepsilon) \cdot \alpha \delta_X \cdot \mathbb{E}_{y' \sim B_4} [g_1(y')] \cdot \mathbb{E}_{z' \sim B_5} [g_2(z')].$$

Then there exist x^* , y^* , and $h_1 : B_7 \rightarrow \{0, 1\}$ and $h_2 : B_6 \rightarrow \{0, 1\}$ such that:

$$h_1(y') \leq \mathbb{1}_Y(y^* + y');$$

$$h_2(z') \leq \mathbb{1}_D(x^* + y^* + z');$$

$$\mathbb{E}_{y' \sim B_7} [h_1(y')] \geq \delta_Y \cdot (\alpha/2)^{O(k_*^2 \log(1/\alpha))} = \delta_Y \cdot (1/2)^{O(\log(1/(\alpha\delta_D))^2 \log(1/\alpha)^2)};$$

$$\mathbb{E}_{z' \sim B_6} [h_2(z')] \geq \delta_D \cdot (1/2)^{O(\log(1/\alpha)^2)};$$

$$\mathbb{E}_{\substack{x' \sim B_6 \\ y' \sim B_7}} f(x^* + x', y^* + y') h_1(y') h_2(x' + y') \geq (1 + \varepsilon) \cdot \alpha \delta_X \cdot \mathbb{E}_{y' \sim B_7} [h_1(y')] \cdot \mathbb{E}_{z' \sim B_6} [h_2(z')];$$

$$\|\mathbb{1}_X(x^* + \cdot)\|_{(B_6, B_8, B_9, K, K)} \leq (1 + 4\varepsilon_s) \cdot \delta_X.$$

Proof. Let $z' = z' + x'' + y''$ and $y' = y' + y''$ with $x'' \sim B_6$ and $y'' \sim B_7$. We then have that

$$\begin{aligned} & \mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5 \\ x'' \sim B_6 \\ y'' \sim B_7}} f(x - y' + z' + x'', y + y' + y'') g_1(y' + y'') g_2(z' + x'' + y'') \\ & \geq (1 + 2\varepsilon) \cdot \alpha \delta_X \cdot \mathbb{E}_{y' \sim B_4} [g_1(y')] \cdot \mathbb{E}_{z' \sim B_5} [g_2(z')] - O(\eta d). \end{aligned}$$

We define

$$\mathcal{E}_1(y', z') = \mathbb{1}[\|\mathbb{1}_X(x - y' + z' + \cdot)\|_{(B_6, B_8, B_9, K, K)} \leq (1 + 4\varepsilon_s) \cdot \delta_X]$$

and by the proof of [Lemma 7.1](#) we have that

$$\begin{aligned} & \mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} \mathcal{E}_1(y', z') \mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} f(x - y' + z' + x'', y + y' + y'') g_1(y' + y'') g_2(z' + x'' + y'') \\ & \geq (1 + 2\varepsilon) \cdot \alpha \delta_X \cdot \mathbb{E}_{y' \sim B_4} [g_1(y')] \cdot \mathbb{E}_{z' \sim B_5} [g_2(z')] - O(\eta d). \end{aligned}$$

We then define

$$\mathcal{E}_2(y', z') = \mathbb{1}\left[\mathbb{E}_{x'' \sim B_6} g_2(z' + x'') \leq 2^{-3} \varepsilon^2 \alpha^2 \cdot \mathbb{E}_{z \sim B_5} [g_2(z)]\right].$$

Observe that

$$\begin{aligned}
& \mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} \mathcal{E}_1(y', z') \mathcal{E}_2(y', z') \mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} f(x - y' + z' + x'', y + y' + y'') g_1(y' + y'') g_2(z' + x'' + y'') \\
& \leq \mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} \mathcal{E}_1(y', z') \mathcal{E}_2(y', z') \mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} \mathbb{1}_X(x - y' + z' + x'') g_1(y' + y'') g_2(z' + x'' + y'') \\
& \leq 2 \mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} \mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} \delta_X \cdot \left(\varepsilon^2 \alpha^2 / 8 \cdot \mathbb{E}_{z' \sim B_5} [g_2(z')] \right) \cdot g_1(y' + y'') \\
& \leq 2^{-2} \varepsilon^2 \cdot \alpha^2 \cdot \delta_X \cdot \mathbb{E}_{y' \sim B_4} [g_1(y')] \cdot \mathbb{E}_{z' \sim B_5} [g_2(z')]
\end{aligned}$$

where we have used [Lemma 6.10](#) and the definition of $\mathcal{E}_2(y', z')$.

We may similarly define

$$\mathcal{E}_3(y', z') = \mathbb{1} \left[\mathbb{E}_{y'' \sim B_7} g_1(y' + y'') \leq 2^{-3} \varepsilon^2 \alpha^2 \cdot \delta_D^2 \cdot \mathbb{E}_{z \sim B_4} [g_1(z)] \right].$$

Using the same analysis, except replacing the final internal bound on the g_2 expectation by g_1 , we obtain that

$$\begin{aligned}
& \mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} \mathcal{E}_1(y', z') \mathcal{E}_3(y', z') \mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} f(x - y' + z' + x'', y + y' + y'') g_1(y' + y'') g_2(z' + x'' + y'') \\
& \leq \mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} \mathcal{E}_1(y', z') \mathcal{E}_3(y', z') \mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} \mathbb{1}_X(x - y' + z' + x'') g_1(y' + y'') \\
& \leq 2^{-2} \varepsilon^2 \cdot \alpha^2 \cdot \delta_X \delta_D \cdot \mathbb{E}_{y' \sim B_4} [g_1(y')] \cdot \mathbb{E}_{z' \sim B_5} [g_2(z')].
\end{aligned}$$

This implies that

$$\begin{aligned}
& \mathbb{E}_{\substack{y' \sim B_4 \\ z' \sim B_5}} \mathcal{E}_1(y', z') (1 - \mathcal{E}_2(y', z')) (1 - \mathcal{E}_3(y', z')) \\
& \cdot \left(\mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} f(x - y' + z' + x'', y + y' + y'') g_1(y' + y'') g_2(z' + x'' + y'') \right. \\
& \quad \left. - (1 + 9\varepsilon/8) \cdot \alpha \cdot \delta_X \cdot \mathbb{E}_{\substack{x'' \sim B_6 \\ y'' \sim B_7}} [g_1(y' + y'') g_2(z' + x'')] \right) > 0.
\end{aligned}$$

The result then follows by choosing a (y', z') pair such that the LHS is strictly positive. (Observe this forces $\mathcal{E}_1(y', z')(1 - \mathcal{E}_2(y', z'))(1 - \mathcal{E}_3(y', z')) = 1$ which are precisely the conditions we require.) $\square$

We briefly end this section with a discussion of the case of

$$2^{k_*} \alpha^{2k_*} \delta_X^{2k_*} \delta_Y^2 \delta_D^{k_*} \leq \left(\mathbb{E}_{\substack{z'_1, \dots, z'_{k_*} \sim B_5}} \mathbb{E}_{\substack{x'_1, x'_2 \sim B_3}} \prod_{j=1}^{k_*} f(x + x'_1, y - x'_1 + z'_j) f(x + x'_2, y - x'_2 + z'_j) \right)$$

where certain slight differences arise. Observe that in this case there is no need to invoke spectral positivity. The analysis to obtain the appropriate upper-regularity on the container is exactly as before; the critical issue is that the density increment coming from [Theorem 3.5](#) now has a different form.

We obtain functions $g_1(x') \leq \mathbb{1}_X(x + x')$ and $g_2(z') \leq \mathbb{1}_D(x + y + z')$ such that:

$$\mathbb{E}_{x' \sim B_3} [g_1(x')] \geq \delta_X \cdot (\alpha/2)^{O(k_*^2 \log(1/\alpha))} = \delta_X \cdot (1/2)^{O(\log(1/(\alpha \delta_D))^2 \log(1/\alpha)^2)};$$

$$\begin{aligned} \mathbb{E}_{z' \sim B_5} [g_2(z')] &\geq \delta_D \cdot (\alpha/2)^{O(\log(1/\alpha))} = \delta_D \cdot (1/2)^{O(\log(1/\alpha)^2)}; \\ \mathbb{E}_{\substack{x' \sim B_3 \\ z' \sim B_5}} [f(x+x', y-x'+z')g_1(x')g_2(z')] &\geq (1+2\varepsilon) \cdot \alpha\delta_Y \cdot \mathbb{E}_{x' \sim B_3} [g_1(x')] \cdot \mathbb{E}_{z' \sim B_5} [g_2(z')]. \end{aligned}$$

We now derive an increment on an untilted rectangle. A key differing feature is that x will now range in a smaller Bohr set than y .

Lemma 7.8. *Suppose that (x, y) is a good pair. Furthermore consider $g_1 : B_3 \rightarrow \{0, 1\}$ and $g_2 : B_5 \rightarrow \{0, 1\}$ such that*

$$\begin{aligned} \mathbb{E}_{x' \sim B_3} [g_1(x')] &\geq \delta_X \cdot (1/2)^{O(\log(1/(\alpha\delta_D))^2 \log(1/\alpha)^2)}; \\ \mathbb{E}_{z' \sim B_5} [g_2(z')] &\geq \delta_D \cdot (1/2)^{O(\log(1/\alpha)^2)}; \\ g_1(x') &\leq \mathbb{1}_X(x+x'); \\ g_2(z') &\leq \mathbb{1}_D(x+y+z'); \\ \mathbb{E}_{\substack{x' \sim B_3 \\ z' \sim B_5}} [f(x+x', y-x'+z')g_1(x')g_2(z')] &\geq (1+2\varepsilon) \cdot \alpha\delta_Y \cdot \mathbb{E}_{x' \sim B_3} [g_1(x')] \cdot \mathbb{E}_{z' \sim B_5} [g_2(z')]. \end{aligned}$$

Then there exists (x^*, y^*) and $h_1 : B_6 \rightarrow \{0, 1\}$ and $h_2 : B_5 \rightarrow \{0, 1\}$ such that:

$$\begin{aligned} \mathbb{E}_{x' \sim B_6} [h_1(x')] &\geq \delta_X \cdot (1/2)^{O(\log(1/(\alpha\delta_D))^2 \log(1/\alpha)^2)}; \\ \mathbb{E}_{z' \sim B_5} [h_2(z')] &\geq \delta_D \cdot (1/2)^{O(\log(1/\alpha)^2)}; \\ h_1(x') &\leq \mathbb{1}_X(x^*+x'); \\ h_2(z') &\leq \mathbb{1}_D(x^*+y^*+z'); \\ \mathbb{E}_{\substack{x' \sim B_6 \\ z' \sim B_5}} [f(x^*+x', y^*+y')h_1(x')h_2(z')] &\geq (1+\varepsilon) \cdot \alpha\delta_Y \cdot \mathbb{E}_{x' \sim B_3} [h_1(x')] \cdot \mathbb{E}_{z' \sim B_5} [h_2(z')]; \end{aligned}$$

$$\|Y(y^* + \cdot)\|_{(B_6, B_8, B_9, K, K)} \leq (1+4\varepsilon_s) \cdot \delta_Y.$$

Proof. Observe that the final condition implies that

$$\mathbb{E}_{\substack{x' \sim B_3 \\ z' \sim B_5 \\ \ell \sim B_6}} [f(x+x'+\ell, y-x'+z')g_1(x'+\ell)g_2(z'+\ell)] \geq (1+2\varepsilon) \cdot \alpha\delta_Y \cdot \mathbb{E}_{x' \sim B_3} [g_1(x')] \cdot \mathbb{E}_{z' \sim B_5} [g_2(z')] - O(\eta d).$$

We define

$$\mathcal{E}_1(x') = \mathbb{1} \left[\|\mathbb{1}_Y(y - x' + \cdot)\|_{(B_5, B_8, B_9, K, K)} \leq (1+3\varepsilon_s)\delta_Y \right]$$

and by the proof of [Lemma 7.1](#) we have that

$$\begin{aligned} \mathbb{E}_{x' \sim B_3} \mathcal{E}_1(x') \mathbb{E}_{\substack{z' \sim B_5 \\ \ell \sim B_6}} [f(x+x'+\ell, y-x'+z')g_1(x'+\ell)g_2(z'+\ell)] \\ \geq (1+2\varepsilon) \cdot \alpha\delta_Y \cdot \mathbb{E}_{x' \sim B_3} [g_1(x')] \cdot \mathbb{E}_{z' \sim B_5} [g_2(z')] - O(\eta d). \end{aligned}$$

We now define

$$\mathcal{E}_2(x') = \mathbb{1} \left[\mathbb{E}_{\ell \sim B_6} [g_1(x'+\ell)] \leq \varepsilon^2 \cdot 2^{-O(\log(1/\alpha)^3)} \cdot \mathbb{E}_{x'' \sim B_3} [g_1(x'')] \right].$$

We now have that

$$\begin{aligned}
& \mathbb{E}_{x' \sim B_3} \mathcal{E}_1(x') \mathcal{E}_2(x') \mathbb{E}_{\substack{z' \sim B_5 \\ \ell \sim B_6}} [f(x + x' + \ell, y - x' + z') g_1(x' + \ell) g_2(z' + \ell)] \\
& \leq \mathbb{E}_{x' \sim B_3} \mathcal{E}_1(x') \mathcal{E}_2(x') \mathbb{E}_{\substack{z' \sim B_5 \\ \ell \sim B_6}} [\mathbb{1}_Y(y - x' + z') g_1(x' + \ell) \mathbb{1}_D(x + y + z' + \ell)] \\
& \leq 2\delta_Y \delta_D \cdot \mathbb{E}_{x' \sim B_3} \mathcal{E}_1(x') \mathcal{E}_2(x') \mathbb{E}_{\substack{z' \sim B_5 \\ \ell \sim B_6}} [g_1(x' + \ell)] + O(e^{-\Omega(\varepsilon_s K)}) \\
& \leq 2\delta_Y \delta_D \cdot \varepsilon^2 \cdot 2^{-O(\log(1/\alpha)^3)} \cdot \mathbb{E}_{x'' \sim B_3} [g_1(x'')] + O(e^{-\Omega(\varepsilon_s K)}) + O(\eta d).
\end{aligned}$$

Here we have applied [Lemma 6.10](#). Therefore we have that

$$\begin{aligned}
& \mathbb{E}_{x' \sim B_3} \mathcal{E}_1(x') (1 - \mathcal{E}_2(x')) \left(\mathbb{E}_{\substack{z' \sim B_5 \\ \ell \sim B_6}} [f(x + x' + \ell, y - x' + z') g_1(x' + \ell) g_2(z' + \ell)] \right. \\
& \quad \left. - (1 + 7\varepsilon/4) \cdot \alpha \delta_X \cdot g_1(x' + \ell) g_2(z' + \ell) \right) > 0.
\end{aligned}$$

Thus there exists x' with $\mathcal{E}_1(x')(1 - \mathcal{E}_2(x')) = 1$ such that

$$\mathbb{E}_{\substack{z' \sim B_5 \\ \ell \sim B_6}} [f(x + x' + \ell, y - x' + z') g_1(x' + \ell) g_2(z' + \ell)] \geq (1 + 7\varepsilon/4) \cdot \alpha \delta_X \cdot \mathbb{E}_{\substack{z' \sim B_5 \\ \ell \sim B_6}} [g_1(x' + \ell) g_2(z' + \ell)].$$

This immediately gives the desired result. $\square$

7.2. Completing the proof. We now tie together the final loose ends; this material is little more than chaining various lemmas to complete the proof. We will in fact prove the stronger counting version of [Theorem 1.1](#), which we will need later to prove [Corollaries 1.3](#) and [1.4](#).

Theorem 7.9. *Let G be a finite abelian group and $A \subseteq G \times G$ with $|A| = \alpha|G|^2$. Then*

$$\Pr_{x, y, d \in G} [(x, y), (x, y + d), (x + d, y) \in A] \geq 2^{-O(\log(1/\alpha)^{600})}.$$

Proof. Initially let $|A| = \alpha|G|^2$. We will maintain Bohr sets B_1, B_2 of the same frequencies and $\nu(B_2)/\nu(B_1) \leq \eta$, and $X \subseteq x + B_1$, $Y \subseteq y + B_2$, and $D \subseteq x + y + B_1$. The density of $A \cap S(X, Y, D)$ within $S(X, Y, D)$ will increase over the course of the procedure. Throughout this proof, let $\delta_X = |X|/|B_1|$, $\delta_Y = |Y|/|B_2|$, and $\delta_D = |D|/|B_1|$.

We will start describing how large various parameters are over the course of the density increment. We will prove that $\delta_D \geq 2^{-O(\log(1/\alpha)^3)}$ and $\delta_X, \delta_Y \geq 2^{-O(\log(1/\alpha)^{10})}$. Also we will choose $K = O(\log(1/\alpha)^{30})$, $\eta = 2^{-O(\log(1/\alpha)^{100})}$, and $\eta_s = \eta^C$ for sufficiently large constant C . Finally we will choose $r = O(\log(1/\alpha)^{350})$.

Before a density increment, we will ensure the following pseudorandomness properties on X, Y, D . We will maintain that X, Y are $(r, \eta_s, \varepsilon_s)$ -algebraically spread, and that D is $(B_1, B_9, \varepsilon_s)$ ℓ_1 -spread. Let us argue that this implies that $\|\mathbb{1}_X\|_{(B_1, B_8, B_9, K, K)} \leq (1 + O(\varepsilon_s))\delta_X$ and similarly $\|\mathbb{1}_Y\|_{(B_2, B_8, B_9, K, K)} \leq (1 + O(\varepsilon_s))\delta_Y$. Indeed, if the former fails, applying [Theorem 6.17](#) and then [Theorem 6.16](#) gives that X admits a density increment onto a Bohr set with rank increase at most $O(\log(1/\delta_X^K)^8) \leq O(\log(1/\alpha)^{350})$.

To apply the results in the section we must check that [Lemma 6.10](#) provides enough combinatorial spreadness to apply [Theorem 3.5](#). This holds for the choice of K and η . Thus, we may apply the results in this section to conclude that either:

$$\Pr_{x, d \in B_1, y \in B_2} [(x, y), (x + d, y), (x, y + d) \in A] \geq \alpha^3 \delta_X^2 \delta_Y^2 \delta_D^2 / 10,$$

or A admits a density increment of the form described in [Lemma 7.7](#) or [Lemma 7.8](#). It is worth pointing out that in [Lemma 7.8](#), that now Y, D are in the larger Bohr set and X is in the smaller one, but this can be handled by switching the roles of X and Y in the analysis each time this happens. In this density increment, the size of δ_D drops by $2^{-O(\log(1/\alpha)^2)}$ as stated, so the total drop is $2^{-O(\log(1/\alpha)^3)}$ over $O(\log(1/\alpha))$ steps. The drop of δ_X and δ_Y is $2^{-O(\log(1/(\alpha\delta_D))^2 \log(1/\alpha)^2)} = 2^{-O(\log(1/\alpha)^8)}$ so a total of $2^{-O(\log(1/\alpha)^9)}$ over $O(\log(1/\alpha))$ steps. The radii of the Bohr sets B_1, B_2 also drop by $\eta^{O(1)}$ as stated.

After a density increment we apply [Theorem 6.22](#) to pseudorandomize X, Y, D . Let us first discuss how this affects δ_D and δ_X, δ_Y . δ_D drops by an additional $O(\varepsilon\alpha)$ per step, which does not affect the lower bound. δ_X and δ_Y drop by $2^{-O(\log(1/\delta_D)^3)} = 2^{-O(\log(1/\alpha)^9)}$, for a total of $2^{-O(\log(1/\alpha)^{10})}$ over $O(\log(1/\alpha))$ steps.

Now we track how the rank and radius of B_1, B_2 change. Note in [Theorem 6.22](#) that $g = O(\log(1/\alpha)^9)$, so the rank of B_1 increases by $O(rg^2 \log(1/\delta_D)^2) = O(\log(1/\alpha)^{400})$. The radius decreases of $\eta_s^{O(g^2 \log(1/\delta_D)^2)} = 2^{-O(\log(1/\alpha)^{150})}$. Thus, the number of corners in A at the end is at least

$$\alpha^3 \delta_X^2 \delta_Y^2 \delta_D^2 |B_1| |B_2| \geq \nu(B_2)^{O(\text{rank}(B_2))} |G|^2 \geq 2^{-O(\log(1/\alpha)^{600})} |G|^2,$$

where we have applied [Lemma 6.2](#). □

8. COLORING BOUNDS FOR 3-DIMENSIONAL CORNERS

In this section we prove [Corollary 1.3](#). The proof requires the definition of a cylinder intersection, which is known to be related to the corners problem and communication complexity.

Definition 8.1 (Cylinder intersection). Let G be an abelian group and $S_{XY}, S_{YZ}, S_{XZ} \subseteq G \times G$. Then the cylinder intersection $\mathcal{I}(S_{XY}, S_{YZ}, S_{XZ})$ is defined as

$$G \times G \times G \supseteq \mathcal{I}(S_{XY}, S_{YZ}, S_{XZ}) := \{(x, y, z) \in G \times G \times G : (x, y) \in S_{XY}, (y, z) \in S_{YZ}, (x, z) \in S_{XZ}\}.$$

[Corollary 1.3](#) follows by inducting on the following statement. In the statement below, one should think of $*$ as representing points that are uncolored, where the number of uncolored points is at most $\tilde{O}(|G|^2)$.

Lemma 8.2. *Let $A = \mathcal{I}(S_{XY}, S_{YZ}, S_{XZ}) \subseteq G \times G \times G$ be a cylinder intersection. Let $f : A \rightarrow [L] \cup \{*\}$ be a coloring of A , and let $U = |f^{-1}(*)|$. If A contains no monochromatic 3D corners with colors in $[L]$, then there are subsets $A' \subseteq A$ and $S'_{XY} \subseteq S_{XY}, S'_{YZ} \subseteq S_{YZ}, S'_{XZ} \subseteq S_{XZ}$ such that:*

- (1) $A' = \mathcal{I}(S'_{XY}, S'_{YZ}, S'_{XZ})$ is a cylinder intersection.
- (2) There is a color $c \in [L]$ such that $|A' \cap f^{-1}(\{c, *\})| \leq U + |G|^2$. Informally, almost all elements of A' are in one of $L - 1$ colors.
- (3) There is a universal constant C such that for $\delta := \frac{|A| - U}{L|G|^3}$, it holds that

$$\frac{|A'|}{|G|^3} \geq e^{-O(\log(1/\delta)^C)}.$$

Proof. By the Pigeonhole principle, there is a color $c \in [L]$ and $g \in G$ such that the set

$$T = \{(x, y, z) \in G \times G \times G : (x, y, z) \in A, f((x, y, z)) = c, x + y + z = g\}$$

is large, specifically that $|T| \geq \frac{|A| - U}{L|G|}$. Let S'_{XY}, S'_{YZ} , and S'_{XZ} be the projections of T onto the XY , XZ , and YZ faces, formally

$$S'_{XY} := \{(x, y) \in G \times G : \exists z \in G \text{ such that } (x, y, z) \in T\},$$

and similarly for S'_{YZ} and S'_{XZ} . Because $T \subseteq A$ we know that $S'_{XY} \subseteq S_{XY}, S'_{YZ} \subseteq S_{YZ}, S'_{XZ} \subseteq S_{XZ}$ and thus $A' \subseteq A$.

We now check item (2). This amounts to checking that if a point in A' is colored c , then it must lie in T , and then noting that $|T| \leq |G|^2$. Indeed, assume that $(x, y, z) \in A'$, so that $(x, y, g - x - y)$, $(g - y - z, y, z)$, $(x, g - x - z, z)$ are in T . If $x + y + z \neq g$, then (x, y, z) forms a 3D corner with these points (for $d = g - x - y - z$). Thus, (x, y, z) cannot have color c .

To prove (3) we first prove the claim that the size of A' is at least the number of 2D corners in $S'_{XY} \subseteq G \times G$. Indeed, assume that S'_{XY} contains a 2D corner (x, y) , $(x, y + d)$, $(x + d, y) \in S'_{XY}$. Then $(x, y, g - x - y)$, $(x, y + d, g - x - y - d)$, $(x + d, y, g - x - y - d) \in T$ by the definition of S'_{XY} , so $(x, y, g - x - y - d) \in A'$. Each triple (x, y, d) generates a distinct point, so the claim is proven. Because $|S'_{XY}| = |T| \geq \delta |G|^2$, (3) follows by [Theorem 7.9](#) (the counting version of [Theorem 1.1](#)). $\square$

From here we can conclude the proof of [Corollary 1.3](#).

Proof of Corollary 1.3. Suppose by contradiction there exists a coloring of $G \times G \times G$ with $L = c \log \log \log |G|$ colors so that there are no monochromatic 3D corners. We will iteratively apply [Lemma 8.2](#) to restrict our attention to increasingly smaller subsets of $G \times G \times G$ with additional colors removed from consideration (by replacing any occurrences of them with $*$). Our contradiction will arise from running out of colors before coloring all points in $G \times G \times G$.

Let $A_0 = G \times G \times G$. We will define A_t for $t = 1, \dots, L$, and let $\delta_t = |A_t|/|G|^3$, so that $\delta_0 = 1$. We will maintain the invariant that $\delta_t \geq |G|^{-1/2}$. With this invariant and the choice of L , choosing A_{t+1} to be the A' in [Lemma 8.2](#) for $A = A_t$ gives that

$$\delta_{t+1} = \frac{|A_{t+1}|}{|G|^3} \geq e^{-O(\log(1/\delta)^C)} \geq e^{-O(\log(2L/\delta_t)^C)} \quad (8.1)$$

for $\delta = \frac{|A_t| - |f^{-1}(*)|}{|G|^3}$. Then, we have used that $\delta \geq \frac{\delta_t}{2L}$ because $\delta_t \geq |G|^{-1/2}$ and $|f^{-1}(*)| \leq L|G|^2$ by induction using item (2). Iterating (8.1) gives that $\log(1/\delta_L) \leq L^{C^L}$, so $\delta_L \geq e^{-L^{C^L}}$. For $L = c \log \log \log |G|$ for sufficiently small c , we know that $\delta_L \geq |G|^{-1/2}$, thus establishing that the desired invariant holds throughout. This is a contradiction because we have no colors left. $\square$

It is worth emphasizing again that we were only able to obtain “reasonable” bounds for the 3D corners problem in the coloring setting because of our quasipolynomial bounds for the density version of 2D corners. With the analogous version of the above argument, one can check that even inverse logarithmic bounds for density 2D corners would only yield tower type bounds for coloring 3D corners.

APPENDIX A. THE QUASIRANDOM GROUP CASE

In this section, we provide a full proof of the following generalized version of [Theorem 1.6](#). Note that several parts of the presentation follow almost verbatim from their analogs in [Section 5](#).

Theorem A.1. *There exists a constant $c > 0$ such that the following holds. Let G be a Q -quasirandom group (as in [Definition A.2](#)) and $A \subseteq G \times G$ with no $x, y, g \in G$ with $g \neq \mathbf{1}_G$ such that $(x, y), (xg, y), (x, gy) \in A$. Then*

$$|A| \leq |G|^2 \cdot \exp(-c(\log Q)^{1/11}).$$

Before proceeding to the details, we review some additional background and useful results about quasirandom groups.

A.1. Background. The notion of quasirandom groups was introduced by Gowers [\[26\]](#) in studying product-free sets. Quasirandom groups enjoy the property that for any two large sets $A, B \subseteq G$, the distribution obtained by taking uniform random samples $a \in A$, $b \in B$ and outputting ab is close to uniform in an L_2 -sense. Of course, abelian groups fail to satisfy the mixing property mentioned

above. In $\mathbb{F}_2^n$ for instance, a subspace V of codimension 1 is very dense, but $V + V = V$ is far from being uniform. We now present the formal definition.

Definition A.2 (Q -Quasirandom group). Let $Q \geq 1$. A finite group G is Q -*quasirandom* if every nontrivial irreducible representation over $\mathbb{C}$ has dimension at least Q .

One can show that every irreducible representation of a finite group G must have dimension at most $\sqrt{|G|}$. We record this fact for later use, which can be found in most introductory texts on representation theory of finite groups. For example, this follows from Proposition 5 in [60, Chapter 2.4].

Fact A.3. *Every complex irreducible representation of a finite group G has dimension at most $\sqrt{|G|}$.*

There are choices of groups G which are Q -quasirandom for $Q = |G|^{\Omega(1)}$. Contrast this with abelian groups, where every irreducible representation has dimension 1. A common example of a quasirandom group is

$$\mathrm{SL}_2(\mathbb{F}_p) = \left\{ \begin{pmatrix} a & b \\ c & d \end{pmatrix} : a, b, c, d \in \mathbb{F}_p, ad - bc = 1 \right\},$$

the set of 2×2 matrices over the finite field $\mathbb{F}_p$ for p prime with determinant 1. A classical result of Frobenius shows that every nontrivial irreducible representation of $\mathrm{SL}_2(\mathbb{F}_p)$ has dimension at least $\frac{p-1}{2} \sim |G|^{1/3}$. (For exposition of these facts and additional background on quasirandom groups, see e.g. [67].)

Quasirandom groups have seen various application in constructing pseudorandom objects. The Ramanujan graphs of Lubotzky, Phillips, and Sarnak [47] are built from Cayley graphs of $\mathrm{PSL}_2(\mathbb{F}_p)$ for specific choices of p . (The group $\mathrm{PSL}_2(\mathbb{F}_p)$ is obtained by quotienting $\mathrm{SL}_2(\mathbb{F}_p)$ by its center.) Bourgain and Gamburd [13] showed that random Cayley graphs on $\mathrm{SL}_2(\mathbb{F}_p)$ are expanders. The quasirandomness property was useful for arguing about the multiplicity of eigenvalues of the Cayley graph. Outside of graph theory, ideas from the study of quasirandom groups led to an optimal inapproximability result for k -LIN over non-abelian groups [10]. A recent work of Derksen, Lee, and Viola [19] building on [27] proved k -NOF lower bounds for computing an “interleaved product” of elements coming from a quasirandom group. In particular, their bounds are best understood in the regime where k is growing, and they match that of [5].

For a finite group G , denote the convolution of functions $f, g : G \rightarrow \mathbb{R}$ by $(f * g)(x) = \mathbb{E}_{y \in G} f(y)g(y^{-1}x)$. The most useful property of quasirandom groups is that convolutions of dense sets mix. More formally, we have the following:

Theorem A.4 ([4, 26]). *Let G be a Q -quasirandom group and $f, g : G \rightarrow \mathbb{R}$, and assume at least one has mean zero. Then*

$$\|f * g\|_2 \leq Q^{-1/2} \|f\|_2 \|g\|_2.$$

There are various proofs of this fact [4, 26, 67] which use representation theory. Over abelian groups, convolutions and L_2 -norms have pleasant interpretations when working in the Fourier basis. Over nonabelian groups, a similar theory exists which can be formulated using representation theory.

Theorem A.4 can be used to count solutions to equations in quasirandom groups. [26] did exactly this to bound the size of the largest product-free set in a quasirandom group. Namely, we can count solutions (x, y) where $x \in X, y \in Y, xy \in Z$. We have the following:

Corollary A.5 ([4, 26]). *Let G be a Q -quasirandom group, and let $f, g, h : G \rightarrow [-1, 1]$. Then,*

$$\left| \mathbb{E}_{x, y \in G} [f(x)g(y)h(xy)] - \mathbb{E}[f] \mathbb{E}[g] \mathbb{E}[h] \right| \leq Q^{-1/2} \|f\|_2 \|g\|_2 \|h\|_2.$$

Proof. Let $\tilde{f} = f - \mathbb{E}[f]$. We have

$$\begin{aligned}
\mathbb{E}_{x,y \in G} [f(x)g(y)h(xy)] &= \mathbb{E}_{z \in G} [(f * g)(z)h(z)] \\
&= \mathbb{E}[f] \mathbb{E}[g] \mathbb{E}[h] + \mathbb{E}_{z \in G} [(\tilde{f} * g)(z)h(z)] \\
&= \mathbb{E}[f] \mathbb{E}[g] \mathbb{E}[h] \pm \|\tilde{f} * g\|_2 \|h\|_2 && \text{(Cauchy-Schwarz)} \\
&= \mathbb{E}[f] \mathbb{E}[g] \mathbb{E}[h] \pm Q^{-1/2} \|\tilde{f}\|_2 \|g\|_2 \|h\|_2. && \text{(Theorem A.4)}
\end{aligned}$$

Since $\tilde{f}$ and $\mathbb{E}[f]$ are orthogonal and f is 1-bounded, we have $\|\tilde{f}\|_2 = (\mathbb{E}[f^2] - \mathbb{E}[f]^2)^{1/2} \leq \|f\|_2$. $\square$

Note that we can also use [Corollary A.5](#) to count other simple patterns, such as tuples $(x, y, x^{-1}y) \in X \times Y \times Z$ or $(x, y, xy^{-1}) \in X \times Y \times Z$.

At this point, we have the necessary tools to prove [Theorem 1.6](#).

A.2. General setup. Throughout, we will work with a Q -quasirandom group G . We will consider subsets $X, Y, D \subseteq G$ of densities $\delta_X, \delta_Y, \delta_D$, and work within their *container set*

$$S(X, Y, D) := \{(x, y) \in G \times G : x \in X, y \in Y, xy \in D\} \subseteq G \times G.$$

Notice that for dense sets X, Y, D , [Corollary A.5](#) guarantees that $|S(X, Y, D)| \approx \delta_X \delta_Y \delta_D |G|^2$, or more precisely:

$$\left| |S(X, Y, D)| - \delta_X \delta_Y \delta_D |G|^2 \right| \leq Q^{-1/2} |G|^2. \quad (\text{A.1})$$

This will be one of two main ways that quasirandomness is utilized in the proof. The other way will be to construct sets which have small *cut norms*.

A.3. Cut norms. Our primary measure of pseudorandomness throughout this section will be the *cut norm*, sometimes referred to as discrepancy (see [\[16\]](#), for example).

Definition A.6 (Cut norm & pseudorandom). Let Ω_1, Ω_2 be finite sets and $T : \Omega_1 \times \Omega_2 \rightarrow \mathbb{R}$ be a function. We define the *cut norm* of T as

$$\|T\|_{\square} = \sup_{\substack{g_1 : \Omega_1 \rightarrow [-1, 1] \\ g_2 : \Omega_2 \rightarrow [-1, 1]}} \mathbb{E}_{\substack{x \in \Omega_1 \\ y \in \Omega_2}} [g_1(x)g_2(y)T(x, y)].$$

We say $T : \Omega_1 \times \Omega_2 \rightarrow [0, 1]$ is γ -pseudorandom if

$$\|T - \mathbb{E}_{x \in \Omega_1, y \in \Omega_2} [T(x, y)]\|_{\square} \leq \gamma.$$

Occasionally, we will refer to a set as pseudorandom if its indicator function is pseudorandom. Note that γ -pseudorandomness can be viewed as a *two-sided* assumption on how T correlates with rectangles, while *combinatorial spreadness* (see [Definition 3.1](#)) only gives upper bounds on how T correlates with rectangles. It was critical for us to work with this weaker, one-sided assumption in order to obtain quasipolynomial bounds in the abelian group setting (see the discussion in [Section 2](#)). On the other hand, we can obtain stronger quantitative bounds in the quasirandom setting because we have easy access to this stronger, two-sided assumption of γ -pseudorandomness. This ends up coming into play when we prove [Lemma A.8](#).

We now come to the second way in which we will utilize the quasirandomness of G . We will argue that for large enough sets $X, Y, D \subseteq G$, the set $T := \{(x, y) : xy \in D\} \subseteq X \times Y$ is γ -pseudorandom.

Lemma A.7. *Let G be a Q -quasirandom group, and let $X, Y, D \subseteq G$ be subsets with size $|X|, |Y| \geq |G|/Q^{1/8}$. Then, the set $T := \{(x, y) : xy \in D\} \subseteq X \times Y$ is $2Q^{-1/4}$ -pseudorandom.*

Proof. Let $\tau = |T|/|X||Y|$. We want to argue that for any functions $g_1 : X \rightarrow [-1, 1]$ and $g_2 : Y \rightarrow [-1, 1]$, we have

$$\left| \mathbb{E}_{\substack{x \in X \\ y \in Y}} [g_1(x)g_2(y)(\mathbb{1}_T - \tau)(x, y)] \right| = \left| \mathbb{E}_{\substack{x \in X \\ y \in Y}} [g_1(x)g_2(y)\mathbb{1}_T(x, y)] - \tau \mathbb{E}_{x \in X} [g_1(x)] \mathbb{E}_{y \in Y} [g_2(y)] \right| \leq 2Q^{-1/4}.$$

First, we will argue that τ is approximately δ_D . We have

$$|\tau - \delta_D| = \left| \mathbb{E}_{\substack{x \in X \\ y \in Y}} [\mathbb{1}_T(x, y) - \delta_D] \right| = (\delta_X \delta_Y)^{-1} \left| \mathbb{E}_{x, y \in G} [\mathbb{1}_X(x)\mathbb{1}_Y(y)(\mathbb{1}_D - \delta_D)(xy)] \right| \leq (\delta_X \delta_Y)^{-1} Q^{-1/2}$$

by [Corollary A.5](#). Recalling that $\mathbb{1}_T(x, y) = \mathbb{1}_D(xy)$, we have

$$\begin{aligned} \left| \mathbb{E}_{\substack{x \in X \\ y \in Y}} [g_1(x)g_2(y)\mathbb{1}_T(x, y)] - \tau \mathbb{E}_{x \in X} [g_1(x)] \mathbb{E}_{y \in Y} [g_2(y)] \right| &\leq |\tau - \delta_D| + \left| \mathbb{E}_{\substack{x \in X \\ y \in Y}} [g_1(x)g_2(y)(\mathbb{1}_T - \delta_D)(x, y)] \right| \\ &\leq |\tau - \delta_D| \\ &\quad + (\delta_X \delta_Y)^{-1} \left| \mathbb{E}_{x, y \in G} [g_1(x)g_2(y)(\mathbb{1}_D - \delta_D)(xy)] \right| \\ &\leq 2(\delta_X \delta_Y)^{-1} Q^{-1/2} \\ &\leq 2Q^{-1/4}, \end{aligned}$$

where the third inequality follows from [Corollary A.5](#), and in the last inequality we used the fact that $\delta_X, \delta_Y \geq Q^{-1/8}$. $\square$

We note that the same conclusion holds for the sets $\{(x, y) : x^{-1}y \in D\}$ and $\{(x, y) : xy^{-1} \in D\}$.

A.4. Von Neumann lemma. In this subsection, we give suitable conditions under which a set $A \subseteq S(X, Y, Z)$ contains roughly as many corners as a random set of the same density. More specifically, for a set $X \subseteq G$, let $\delta_X = |X|/|G|$. Suppose $A \subseteq S(X, Y, D)$ has size $|A| = \alpha \delta_X \delta_Y \delta_D |G|^2$. For sets X, Y, D which are dense enough, the size of $S(X, Y, D)$ will be roughly $\delta_X \delta_Y \delta_D |G|^2$ (see [\(A.1\)](#)), so one should morally view α as the density of A within its container $S(X, Y, D)$.

Define the trilinear form $\Phi(f_1, f_2, f_3) := \mathbb{E}_{x, y, z \in G} [f_1(x, y)f_2(zy^{-1}, y)f_3(x, x^{-1}z)]$. Observe that by the change of variable $z \rightarrow xgy$, $\Phi(f_1, f_2, f_3) = \mathbb{E}_{x, y, g \in G} [f_1(x, y)f_2(xg, y)f_3(x, gy)]$, so that $\Phi(\mathbb{1}_A, \mathbb{1}_A, \mathbb{1}_A)$ counts the number of corners in A (up to normalization). We will show that if A is sufficiently pseudorandom, then $\Phi(\mathbb{1}_A, \mathbb{1}_A, \mathbb{1}_A)$ is at least a constant factor times $\alpha^3 \delta_X^2 \delta_Y^2 \delta_D^2$, which is roughly the number of corners one would expect had A been chosen randomly from $S(X, Y, D)$. Concretely, for sufficiently quasirandom G , we will require two grid norms related to $\mathbb{1}_A$ to be bounded, as well as require that A has no columns which are too sparse.

We stress that the main difference between [Lemma A.8](#) and [Lemma 5.11](#) is the choice of p . Here, the choice of p depends only on α since we have easy access to γ -pseudorandomness. In [Lemma 5.11](#), the choice of p depends on α as well as δ_D , which is one reason we can obtain stronger quantitative bounds in the quasirandom setting. This is also why we required stronger lower bounds on δ_D when working in the abelian setting.

Lemma A.8. *Let G be a Q -quasirandom group, and let $X, Y, D \subseteq G$ be subsets of size $|X| = \delta_X |G|, |Y| = \delta_Y |G|, |D| = \delta_D |G|$. Additionally, let $A \subseteq S(X, Y, D)$ be a subset of size $|A| = \alpha \delta_X \delta_Y \delta_D |G|^2$. Let $\varepsilon \in (0, 1/10)$ and $p = \Omega(\log(1/\alpha)/\varepsilon^4)$ be a positive integer. Define the functions $F_1 : Y \times D \rightarrow \{0, 1\}$ where $F_1(y, z) = \mathbb{1}_A(zy^{-1}, y)$ and $F_2 : X \times D \rightarrow \{0, 1\}$ where $F_2(x, z) = \mathbb{1}_A(x, x^{-1}z)$. Suppose the following conditions hold:*

- (1) $\|F_1\|_{G(2,p)} < (1 + \varepsilon^2/36)\alpha\delta_X$,
- (2) $\|F_2\|_{G(2,p)} < 2\alpha\delta_Y$,
- (3) For all $y \in Y$, we have $\mathbb{E}_{x \in G}[\mathbb{1}_A(x, y)] \geq (1 - \varepsilon^2/36)\alpha\delta_X\delta_D$,
- (4) $Q^{-1} \leq (\varepsilon\alpha\delta_X\delta_Y\delta_D)^{O(p)}$.

Then,

$$\Phi(\mathbb{1}_A, \mathbb{1}_A, \mathbb{1}_A) \geq (1 - 12\varepsilon)\alpha^3\delta_X^2\delta_Y^2\delta_D^2.$$

Proof. For clarity, let $f := \mathbb{1}_A$ denote the indicator function for A . Additionally, let $S := S(G, Y, D) = \{(zy^{-1}, y) \in G \times G : y \in Y, z \in D\}$, and let $h = f - \alpha\delta_X\mathbb{1}_S$. Then, we have

$$\Phi(f, f, f) = \Phi(f, h, f) + \alpha\delta_X \cdot \Phi(f, \mathbb{1}_S, f).$$

We will proceed by lower bounding the second term. Afterwards, we will upper bound the magnitude of the first term, showing it is ultimately dominated by the second.

Using the observation that whenever $f(x, y) = 1$ and $f(x, x^{-1}z) = 1$ we must have $x \in X, y \in Y, z \in D$, we can lower bound

$$\begin{aligned} \Phi(f, \mathbb{1}_S, f) &= \mathbb{E}_{x, y, z \in G} [f(x, y)\mathbb{1}_S(zy^{-1}, y)f(x, x^{-1}z)] \\ &= \mathbb{E}_{x, y, z \in G} [f(x, y)\mathbb{1}_Y(y)\mathbb{1}_D(z)f(x, x^{-1}z)] \\ &= \mathbb{E}_{x, y, z \in G} [f(x, y)f(x, x^{-1}z)] \\ &= \mathbb{E}_{x \in G} \left(\mathbb{E}_{y \in G} f(x, y) \right)^2 \geq \delta_X^{-1} \left(\mathbb{E}_{x, y \in G} f(x, y) \right)^2 = \alpha^2\delta_X\delta_Y^2\delta_D^2, \end{aligned}$$

where we used the Cauchy-Schwarz inequality in the last line. In particular,

$$\alpha\delta_X \cdot \Phi(f, \mathbb{1}_S, f) \geq \alpha^3\delta_X^2\delta_Y^2\delta_D^2.$$

Thus, it remains to bound the magnitude of $\Phi(f, h, f)$. Since $f(x, y)$ is supported on $T := S(X, Y, D)$, it suffices to take the expectation over $(x, y) \in T$. Let $|T| = \tau|X||Y|$ and set $k = 2\lceil \log(1/\alpha) \rceil$. Then, we have

$$\begin{aligned} |\Phi(f, h, f)| &= \left| \mathbb{E}_{x, y, z \in G} f(x, y)h(zy^{-1}, y)f(x, x^{-1}z) \right| \\ &= \delta_X\delta_Y\delta_D \left| \mathbb{E}_{x \in X, y \in Y, z \in D} f(x, y)h(zy^{-1}, y)f(x, x^{-1}z) \right| \\ &= \tau\delta_X\delta_Y\delta_D \left| \mathbb{E}_{(x, y) \in T} f(x, y) \left(\mathbb{E}_{z \in D} h(zy^{-1}, y)f(x, x^{-1}z) \right) \right| \\ &\leq \tau\delta_X\delta_Y\delta_D \left(\frac{|A|}{|T|} \right)^{1-\frac{1}{k}} \left(\mathbb{E}_{(x, y) \in T} \left| \mathbb{E}_{z \in D} h(zy^{-1}, y)f(x, x^{-1}z) \right|^k \right)^{1/k}, \quad (\text{A.2}) \end{aligned}$$

where the last inequality follows by Hölder's inequality. Using (A.1) and the assumption $Q^{-1} \leq (\varepsilon\alpha\delta_X\delta_D\delta_Y)^{O(k)}$, we can show that τ is roughly δ_D :

$$\delta_D/2 \leq \delta_D - (\delta_X\delta_Y)^{-1}Q^{-1/2} \leq \tau \leq \delta_D + (\delta_X\delta_Y)^{-1}Q^{-1/2} \leq 2\delta_D.$$

Combining this bound with our choice of k and our upper bound on τ , we find

$$\tau\delta_X\delta_Y\delta_D \left(\frac{|A|}{|T|} \right)^{1-\frac{1}{k}} = \tau\delta_X\delta_Y\delta_D \left(\frac{\alpha\delta_X\delta_Y\delta_D}{\tau\delta_X\delta_Y} \right)^{1-\frac{1}{k}} \leq \tau\delta_X\delta_Y\delta_D \left(\frac{\alpha\delta_D}{\tau} \right) \left(\frac{\alpha}{2} \right)^{-\frac{1}{k}} \leq 4\alpha\delta_X\delta_Y\delta_D^2.$$

Thus, it suffices to bound the final factor of (A.2) by $3\varepsilon\alpha^2\delta_X\delta_Y$.

Define the function $H : Y \times D \rightarrow [-1, 1]$ as $H(y, z) = h(zy^{-1}, y)$, and recall that $F_2(x, z) = f(x, x^{-1}z)$. Rewriting, we have

$$\begin{aligned}
\left(\mathbb{E}_{(x,y) \in T} \left| \mathbb{E}_{z \in D} h(zy^{-1}, y) f(x, x^{-1}z) \right|^k \right)^{1/k} &= \left(\mathbb{E}_{(x,y) \in T} \left| \mathbb{E}_{z \in D} H(y, z) F_2(x, z) \right|^k \right)^{1/k} \\
&= \left(\mathbb{E}_{\substack{(x,y) \in T \\ z_1, \dots, z_k \in D}} \prod_{i=1}^k H(y, z_i) F_2(x, z_i) \right)^{1/k} \\
&= \left(\tau^{-1} \mathbb{E}_{\substack{x \in X, y \in Y \\ z_1, \dots, z_k \in D}} \left[\mathbb{1}_T(x, y) \left(\prod_{i=1}^k H(y, z_i) \right) \left(\prod_{i=1}^k F_2(x, z_i) \right) \right] \right)^{1/k} \\
&\leq \left(\tau^{-1} \left(2Q^{-1/4} + \tau \mathbb{E}_{\substack{x \in X, y \in Y \\ z_1, \dots, z_k \in D}} \prod_{i=1}^k H(y, z_i) F_2(x, z_i) \right) \right)^{1/k}
\end{aligned}$$

where the last inequality follows from the fact that T is $2Q^{-1/4}$ -pseudorandom by [Lemma A.7](#). Now, using our lower bound on τ and the assumption $Q^{-1} \leq (\varepsilon \alpha \delta_X \delta_Y \delta_D)^{O(k)}$, we know

$$(2\tau^{-1}Q^{-1/4})^{1/k} \leq \left(\frac{4Q^{-1/4}}{\delta_D} \right)^{1/k} \leq \varepsilon \alpha^2 \delta_X \delta_Y.$$

Hence by concavity it suffices to show that

$$\left(\mathbb{E}_{\substack{x \in X, y \in Y \\ z_1, \dots, z_k \in D}} \prod_{i=1}^k H(y, z_i) F_2(x, z_i) \right)^{1/k} \leq 2\varepsilon \alpha^2 \delta_X \delta_Y.$$

By the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
\left(\mathbb{E}_{\substack{x \in X, y \in Y \\ z_1, \dots, z_k \in D}} \prod_{i=1}^k H(y, z_i) F_2(x, z_i) \right)^{1/k} &\leq \left(\mathbb{E}_{\substack{y_1, y_2 \in Y \\ z_1, \dots, z_k \in D}} \prod_{i=1}^k H(y_1, z_i) H(y_2, z_i) \right)^{1/(2k)} \\
&\quad \cdot \left(\mathbb{E}_{\substack{x_1, x_2 \in X \\ z_1, \dots, z_k \in D}} \prod_{i=1}^k F_2(x_1, z_i) F_2(x_2, z_i) \right)^{1/(2k)} \\
&= \|H\|_{G(2,k)} \cdot \|F_2\|_{G(2,k)}.
\end{aligned}$$

The second hypothesis, along with the fact that grid norms are monotonic, bounds the second factor by $2\alpha\delta_Y$.

We will finish the proof by bounding the first factor by $\varepsilon\alpha\delta_X$. To do this, we will use the third item of the hypothesis as well as [Lemma 4.2](#) to reduce this quantity to the first hypothesis. First, note that

$$\|F_1\|_{G(1,1)} = (\delta_Y \delta_D)^{-1} \mathbb{E}_{y, z \in G} [f(zy^{-1}, y)] = \alpha \delta_X$$

by the definition of α . Thus $H = F_1 - \|F_1\|_{G(1,1)}$ by the definition of h . Observe the third item of the hypothesis guarantees lower-boundedness on the rows of H ; namely for all $y \in Y$, we have that

$$\mathbb{E}_{z \in D} [H(y, z)] = \delta_D^{-1} \mathbb{E}_{z \in G} [h(zy^{-1}, y)] \geq (1 - \varepsilon^2/36)\alpha\delta_X - \alpha\delta_X = -\varepsilon^2\alpha\delta_X/36.$$

Combining with the first item of the hypothesis, the contrapositive of [Lemma 4.2](#) gives that

$$\|H\|_{G(2,k)} < \varepsilon \alpha \delta_X,$$

as desired. $\square$

A.5. Density increment. Now we prove that if the grid norm assumptions of [Lemma A.8](#) are not satisfied, we can pass to a subset $X' \times Y' \times Z'$ where (essentially) the density of A in $S(X', Y', Z')$ increases by a constant factor. Our main tools will be [Theorem 3.5](#) and [Lemma 3.3](#). We remark that while the asymmetric conclusion present in [Theorem 3.5](#) is essential for our analysis over abelian groups, it is not necessary for the current setting. This is because the choice of p in [Lemma A.8](#) depends only on α , whereas the choice of p in [Lemma 5.11](#) depends on α as well as δ_D . Thus, it was crucial that the density of D' in the conclusion of [Lemma 5.13](#) did not depend on k (see [Section 2.4](#) for further discussion).

For the sake of brevity, we will simply use combinatorial spreadness as our pseudorandomness notion in this subsection to apply [Theorem 3.5](#) directly.

Lemma A.9. *Let G be a Q -quasirandom group, and let $X, Y, D \subseteq G$ be subsets of size $|X| = \delta_X |G|, |Y| = \delta_Y |G|, |D| = \delta_D |G|$. Additionally, let $A \subseteq S(X, Y, D)$ be a subset of size $|A| = \alpha \delta_X \delta_Y \delta_D |G|^2$. Let $\varepsilon \in (0, 1/10)$ and k be a positive integer. Define the functions $F_1 : Y \times D \rightarrow \{0, 1\}$ where $F_1(y, z) = \mathbb{1}_A(zy^{-1}, y)$ and $F_2 : X \times D \rightarrow \{0, 1\}$ where $F_2(x, z) = \mathbb{1}_A(x, x^{-1}z)$. Suppose X, Y, D satisfy $\delta_X, \delta_Y, \delta_D \geq Q^{-1/8}$, and*

$$Q^{-1} \leq (\varepsilon \alpha \delta_X \delta_Y \delta_D)^{O(\varepsilon^{-2} k \log(1/\alpha)^2 + \varepsilon^{-1} k \log(1/(\delta_X \delta_Y)))}$$

Then if $\|F_1\|_{G(2,k)} \geq (1+\varepsilon)\alpha\delta_X$, then there are $Y' \subseteq Y$ and $D' \subseteq D$ with $|Y'| \geq (\varepsilon\alpha/2)^{O(\varepsilon^{-1}k^2 \log(1/\alpha))} |Y|$ and $|D'| \geq (\varepsilon\alpha/2)^{O(\varepsilon^{-1} \log(1/\alpha))} |D|$, and

$$|A \cap S(X, Y', D')| \geq (1 + \Omega(\varepsilon)) \alpha \delta_X |Y'| |D'|.$$

Similarly, if $\|F_2\|_{G(2,k)} \geq (1+\varepsilon)\alpha\delta_Y$, then there are $X' \subseteq X$ and $D' \subseteq D$ with $|X'| \geq (\varepsilon\alpha/2)^{O(\varepsilon^{-1}k^2 \log(1/\alpha))} |X|$ and $|D'| \geq (\varepsilon\alpha/2)^{O(\varepsilon^{-1} \log(1/\alpha))} |D|$, and

$$|A \cap S(X', Y, D')| \geq (1 + \Omega(\varepsilon)) \alpha \delta_Y |X'| |D'|.$$

Proof. We only prove the former assertion, as the latter one has an identical proof. Let $\varepsilon_s = \Theta(\varepsilon)$ sufficiently small, and consider the set $T := \{(y, z) \in Y \times D : zy^{-1} \in X\}$ of size $|T| = |S(X, Y, Z)| = \tau |Y| |D|$. We will use [\(A.1\)](#) as well as our assumption $Q^{-1/2} \leq \varepsilon_s \delta_X \delta_Y \delta_D$ to argue that τ is roughly δ_X :

$$\tau \leq \delta_X + (\delta_Y \delta_D)^{-1} Q^{-1/2} \leq (1 + \varepsilon_s) \delta_X.$$

The same argument also shows that $(1 - \varepsilon_s) \delta_X \leq \tau$. We can then use this to write our grid norm assumption in terms of τ :

$$\|F_1\|_{G(2,k)} \geq (1 + \varepsilon) \alpha \delta_X \geq (1 + \varepsilon) (1 + \varepsilon_s)^{-1} \alpha \tau \geq \left(1 + \frac{\varepsilon}{2}\right) \alpha \tau.$$

Since Y and D are large enough, we may apply [Lemma A.7](#) to deduce that T is $2Q^{-1/4}$ -pseudorandom, and so it follows from the definition that T is $(\tau, 2Q^{-1/2})$ -combinatorially spread. We want to apply [Theorem 3.5](#) using T as a container, so it remains to verify that T is sufficiently spread. In particular, we need

$$2Q^{-1/2} \leq (\alpha \tau)^{O(\varepsilon^{-2} k \log(1/\alpha)^2 + \varepsilon^{-1} k \log(1/\tau))}.$$

Since we already showed that $\tau \leq 2\delta_X$, our assumption on Q satisfies the above inequality. Thus applying [Theorem 3.5](#) and [Lemma 3.3](#) with our choice of $\varepsilon_s = \Theta(\varepsilon)$ sufficiently small gives that there are functions $g_1 : Y \rightarrow \{0, 1\}$ and $g_2 : D \rightarrow \{0, 1\}$ satisfying:

$$\mathbb{E}_{y \in Y} [g_1(y)] \geq (\varepsilon\alpha/2)^{O(\varepsilon^{-1}k^2 \log(1/\alpha))} \quad \text{and} \quad \mathbb{E}_{z \in D} [g_2(z)] \geq (\varepsilon\alpha/2)^{O(\varepsilon^{-1} \log(1/\alpha))}$$

and

$$\begin{aligned}
\mathbb{E}_{y \in Y, z \in D} [F_1(y, z) g_1(y) g_2(z)] &\geq (1 - \varepsilon_s)(1 + \varepsilon/2) \alpha \tau \mathbb{E}_{y \in Y} [g_1(y)] \mathbb{E}_{z \in D} [g_2(z)] \\
&\geq (1 + \varepsilon/4) \alpha \tau \mathbb{E}_{y \in Y} [g_1(y)] \mathbb{E}_{z \in D} [g_2(z)] \\
&\geq (1 + \varepsilon/4)(1 - \varepsilon_s) \alpha \delta_X \mathbb{E}_{y \in Y} [g_1(y)] \mathbb{E}_{z \in D} [g_2(z)] \\
&\geq (1 + \varepsilon/8) \alpha \delta_X \mathbb{E}_{y \in Y} [g_1(y)] \mathbb{E}_{z \in D} [g_2(z)].
\end{aligned}$$

Letting Y' and D' be the indicator functions of g_1 and g_2 respectively completes the proof. $\square$

A.6. Obtaining spreadness. In this subsection, we will use a density increment algorithm to reach a state $A \subseteq S(X, Y, D)$ where the conditions of the von Neumann Lemma (Lemma A.8) are satisfied. We recall the useful definition of asymmetric combinatorial spreadness (Definitions 5.14 and 5.15): a function $f : \Omega_1 \times \Omega_2 \rightarrow [0, 1]$ is (s, t, ε) -combinatorially spread if for all functions $g_1 : \Omega_1 \rightarrow \{0, 1\}$ and $g_2 : \Omega_2 \rightarrow \{0, 1\}$ with

$$\mathbb{E}[g_1(x)] \geq 2^{-s} \quad \text{and} \quad \mathbb{E}[g_2(y)] \geq 2^{-t},$$

it holds that

$$\mathbb{E}[f(x, y) g_1(x) g_2(y)] \leq (1 + \varepsilon) \mathbb{E}[f] \mathbb{E}[g_1] \mathbb{E}[g_2].$$

Note that unlike Definition 3.1, the g_i are simply subsets of Ω_i , rather than $[0, 1]$ -valued functions. Moreover, we say A is (s, t, ε) -combinatorially spread in a container $S(X, Y, D)$ if the functions $F_1 : Y \times D \rightarrow \{0, 1\}$, $F_2 : X \times D \rightarrow \{0, 1\}$ defined as

$$F_1(y, z) = \mathbb{1}_{A \cap S(X, Y, D)}(zy^{-1}, y) \quad \text{and} \quad F_2(x, z) = \mathbb{1}_{A \cap S(X, Y, D)}(x, x^{-1}z)$$

are (s, t, ε) -combinatorially spread.

Lemma A.10. *Let G be a Q -quasirandom group. Let $t \geq 1$ and $\varepsilon \in (0, 1/64)$. Suppose that $A \subseteq G \times G$ has size $|A| = \alpha|G|^2$. As long as $s, t \leq O\left(\frac{\varepsilon \log Q}{\log(1/\alpha)}\right)$, then there exist sets $X, Y, D \subseteq G$ such that*

- (1) *The set A is $(s, t, 5\varepsilon^{1/2})$ -combinatorially spread in the container $S(X, Y, D)$.*
- (2) *$|D| \geq 2^{-O(t\varepsilon^{-1} \log(1/\alpha))} |G|$.*
- (3) *$|X|, |Y| \geq 2^{-O(s\varepsilon^{-1} \log(1/\alpha))} |G|$.*
- (4) *The columns of $A \cap S(X, Y, D)$ are lower bounded; namely for all $y \in Y$,*

$$\mathbb{E}_{x \in G} [\mathbb{1}_{A \cap S(X, Y, D)}(x, y)] \geq (1 - 2\varepsilon^{1/2}) \alpha^* \delta_X \delta_D,$$

$$\text{where } \alpha^* = \frac{|A \cap S(X, Y, D)|}{\delta_X \delta_Y \delta_D |G|^2} \geq \alpha.$$

Proof. Perform the following algorithm, beginning with $X, Y, D = G$. As long as A is not $(s+1, t, \varepsilon)$ -combinatorially spread in its container $S(X, Y, D)$, iteratively restrict to containers $S(X_1, Y, D_1)$ or $S(X, Y_1, D_1)$ to obtain a $(1 + \varepsilon)$ density increment as is guaranteed by Definition 5.15. At the end, remove all columns from A that violate (4).

We start by analyzing how X, Y, D change in one iteration of the algorithm, before the column removal phase. For brevity, let f be the indicator of $A \cap S(X, Y, D)$. If A is not $(s+1, t, \varepsilon)$ -combinatorially spread, without loss of generality we can find subsets $X_1 \subseteq X$ and $D_1 \subseteq D$ with $|X_1| \geq 2^{-s-1}|X|$ and $|D_1| \geq 2^{-t}|D|$ where

$$\mathbb{E}_{x \in X, z \in D} [\mathbb{1}_{X_1}(x) \mathbb{1}_{D_1}(z) f(x, x^{-1}z)] \geq (1 + \varepsilon) \mathbb{E}_{x \in X, z \in D} [f(x, x^{-1}z)] \cdot \mathbb{E}_{x \in X} [\mathbb{1}_{X_1}] \cdot \mathbb{E}_{z \in D} [\mathbb{1}_{D_1}],$$

or equivalently after normalizing,

$$\mathbb{E}_{x \in X_1, z \in D_1} [f(x, x^{-1}z)] \geq (1 + \varepsilon) \mathbb{E}_{x \in X, z \in D} [f(x, x^{-1}z)].$$

In other words,

$$\frac{|A \cap S(X_1, Y, D_1)|}{|X_1||D_1|} \geq (1 + \varepsilon) \frac{|A \cap S(X, Y, D)|}{|X||D|}.$$

We then recurse on $A \cap S(X_1, Y, D_1)$. The claim is that in at most $O(\varepsilon^{-1} \log(1/\alpha))$ iterations, we obtain sets X^*, Y^*, D^* of size at least $|X^*|, |Y^*| \geq 2^{-O(s\varepsilon^{-1} \log(1/\alpha))}|G|$ and $|D^*| \geq 2^{-O(t\varepsilon^{-1} \log(1/\alpha))}|G|$ which satisfy (2) and (3). Otherwise, every iteration has failed to find a container where A is spread, at which point we have the guarantee that

$$\frac{|A \cap S(X^*, Y^*, D^*)|}{\delta_{X^*} \delta_{Y^*} \delta_{D^*} |G|^2} \geq (1 + \varepsilon)^{O(\varepsilon^{-1} \log(1/\alpha))} \cdot \alpha > 2.$$

Our choice of s, t implies that $|X^*|, |Y^*| \geq 2^{-O(s\varepsilon^{-1} \log(1/\alpha))}|G| \geq |G|/Q^{1/12}$, and similarly $|D^*| \geq |G|/Q^{1/12}$, so it follows from (A.1) that $|S(X^*, Y^*, D^*)| \leq 2\delta_{X^*} \delta_{Y^*} \delta_{D^*} |G|^2$. Combining with the above inequality gives

$$\frac{|A \cap S(X^*, Y^*, D^*)|}{|S(X^*, Y^*, D^*)|} > \frac{|A \cap S(X^*, Y^*, D^*)|}{2\delta_{X^*} \delta_{Y^*} \delta_{D^*} |G|^2} > 1,$$

which is a contradiction.

To complete the proof of the lemma, we must guarantee that $A \cap S(X^*, Y^*, D^*)$ has lower-bounded columns. To achieve this, we will simply remove all of the columns which violate (4). More formally, now let f be the indicator of $A \cap S(X^*, Y^*, D^*)$ and $\alpha^* = |A \cap S(X^*, Y^*, D^*)|/(\delta_{X^*} \delta_{Y^*} \delta_{D^*} |G|^2)$, and let $L = \{y \in Y^* : \mathbb{E}_{x \in G} [f(x, y)] < (1 - \varepsilon^{1/2})\alpha^* \delta_{X^*} \delta_{D^*}\}$. We may assume $|L| \leq 4\varepsilon^{1/2}|Y^*|$, as otherwise Lemma 5.12² implies there exists a subset $Y^+ \subseteq Y^*$ with $\delta_{Y^+} \geq \delta_{Y^*}/2$ such that

$$\frac{|A \cap S(X^*, Y^+, D^*)|}{\delta_{X^*} \delta_{Y^+} \delta_{D^*} |G|^2} > (1 + \varepsilon)\alpha^*,$$

which contradicts the (s, t, ε) -combinatorial spreadness of A in $S(X^*, Y^*, D^*)$. Thus, we define $Y^+ := Y^* \setminus L$ so that $|Y^+| \geq (1 - 4\varepsilon^{1/2})|Y^*|$. Define

$$\alpha^+ := \frac{|A \cap S(X^*, Y^+, D^*)|}{\delta_{X^*} \delta_{Y^+} \delta_{D^*} |G|^2}.$$

Note that $\alpha^+ \geq \alpha^*$ since we only deleted sparse columns. Additionally, we have $\alpha^+ \leq (1 + \varepsilon)\alpha^*$ by $(s + 1, t, \varepsilon)$ -combinatorial spreadness. Thus, in terms of the new density α^+ , all of the columns with $y \in Y^+$ satisfy

$$\mathbb{E}_{x \in G} [\mathbb{1}_{S(X^*, Y^+, D^*)}(x, y)] \geq (1 - \varepsilon^{1/2})\alpha^* \delta_{X^*} \delta_{D^*} \geq (1 - 2\varepsilon^{1/2})\alpha^+ \delta_{X^*} \delta_{D^*}.$$

We conclude by showing that A is $(s, t, 5\varepsilon^{1/2})$ -combinatorially spread in the container $S(X^*, Y^+, D^*)$. For clarity, define $F_1 : Y^* \times D^* \rightarrow \{0, 1\}$ and $F_2 : X^* \times D^* \rightarrow \{0, 1\}$ with $F_1(y, z) = f(zy^{-1}, y)$ and $F_2(x, z) = f(x, x^{-1}z)$. The $(s, t, 5\varepsilon^{1/2})$ -combinatorial spreadness of F_2 follows easily, since for any $X' \subseteq X^*$ with size $|X'| \geq 2^{-(s+1)}|X^*|$ and $D' \subseteq D^*$ with size $|D'| \geq 2^{-t}|D^*|$, we have

$$|A \cap S(X', Y^+, D')| \leq |A \cap S(X', Y^*, D')| \leq (1 + \varepsilon)\alpha^* \delta_{X'} \delta_{Y^*} \delta_{D'} |G|^2.$$

This combined with the fact that $\delta_{Y^+} \geq (1 - 4\varepsilon^{1/2})\delta_{Y^*}$ and $\alpha^+ \geq \alpha^*$ gives

$$|A \cap S(X', Y^+, D')| \leq (1 + 5\varepsilon^{1/2})\alpha^+ \delta_{X'} \delta_{Y^+} \delta_{D'} |G|^2,$$

²While Lemma 5.12 is stated for X, Y, D contained in a subspace W , the statement and proof migrate identically to the case of containment in an arbitrary group G .

which is equivalent to F_2 being $(s+1, t, 5\varepsilon^{1/2})$ -combinatorially spread. Showing spreadness of F_1 is only slightly more subtle. For $Y' \subseteq Y^+$ with size $|Y'| \geq 2^{-(s+1)}|Y^*|$ and $D' \subseteq D^*$ of size at least $2^{-t}|D^*|$, we again have

$$|A \cap S(X^*, Y', D')| \leq (1 + \varepsilon)\alpha^* \delta_{X^*} \delta_{Y'} \delta_{D'} |G|^2.$$

Since $|Y^+| \geq (1 - 4\varepsilon^{1/2})|Y^*| \geq |Y^*|/2$, this means that Y' must have density at least $2 \cdot 2^{-(s+1)} = 2^{-s}$ in Y^+ , which means F_1 is $(s, t, 5\varepsilon^{1/2})$ -combinatorially spread. $\square$

A.7. Completing the proof. In this short section, we combine the previous pieces we have developed to establish [Theorem A.1](#).

Proof of Theorem A.1. We prove the contrapositive. Let $A \subseteq G \times G$ be a set of size $|A| = \alpha|G|^2$ for $\alpha \geq \exp(-c(\log Q)^{1/11})$ for some small enough $c > 0$. The proof proceeds by restricting A to a large container $S(X, Y, D)$ where A is combinatorially spread using [Lemma A.10](#). Then, we will argue that combinatorial spreadness is enough to ensure bounded grid norms using [Lemma A.9](#). At that point, we can apply [Lemma A.8](#) to show that A contains many corners.

Let ε be a sufficiently small constant, and let

$$s = O(\log(1/\alpha)^4) \quad \text{and} \quad t = O(\log(1/\alpha)^2)$$

for large enough implicit constants. Our choice of s, t guarantees that $s, t \leq O(\log Q / \log(1/\alpha))$. By applying [Lemma A.10](#) with $s, t, \varepsilon_s = \Theta(\varepsilon^4)$ small enough, there exist sets $X, Y, D \subseteq G$ satisfying the following properties:

- (1) The set A is $(s, t, O(\varepsilon^2))$ -combinatorially spread in the container $S(X, Y, D)$.
- (2) $|X|, |Y|, |D| \geq 2^{-O(\log(1/\alpha)^5)}|G|$.
- (3) The columns of $A \cap S(X, Y, D)$ are lower bounded; namely for all $y \in Y$,

$$\mathbb{E}_{x \in G} [\mathbb{1}_{A \cap S(X, Y, D)}(x, y)] \geq (1 - O(\varepsilon^2))\alpha^* \delta_X \delta_D$$

$$\text{where } \alpha^* = \frac{|A \cap S(X, Y, D)|}{\delta_X \delta_Y \delta_D |G|^2} \geq \alpha.$$

Let f denote the indicator of $A \cap S(X, Y, D)$, and define the functions $F_1 : Y \times D \rightarrow \{0, 1\}$ with $F_1(y, z) = f(zy^{-1}, y)$ and $F_2 : X \times D$ with $F_2(x, z) = f(x, x^{-1}z)$. We will argue that the combinatorial spreadness of A in the container $S(X, Y, D)$ implies certain grid norms are bounded. In particular, we will show that

$$\|F_1\|_{G(2, k)} \leq (1 + \varepsilon^2/36)\alpha^* \delta_X \quad \text{and} \quad \|F_2\|_{G(2, k)} \leq 2\alpha^* \delta_Y$$

for $k = O(\log(1/\alpha^*)/\varepsilon^4) = O(\log(1/\alpha))$ a large enough integer. Without loss of generality, assume the latter does not hold. We want to apply [Lemma A.9](#), so it suffices to show that X, Y, D are sufficiently dense, and that Q^{-1} is sufficiently small. First, we have $|X|, |Y|, |D| \geq |G|/2^{O(\log(1/\alpha)^5)} \geq |G|/Q^{1/8}$ by the choice of α . Secondly, we have

$$Q^{-1} \leq 2^{-O(\log(1/\alpha)^{11})} \leq (\varepsilon \alpha \delta_X \delta_Y \delta_D)^{O(\varepsilon^{-2} \log(1/\alpha)^3 + \varepsilon^{-1} \log(1/\alpha) \log(1/(\delta_X \delta_Y)))}.$$

[Lemma A.9](#) implies that there are $Y' \subseteq Y$ and $D' \subseteq D$ with $|Y'| \geq (\varepsilon \alpha/2)^{O(k^2 \log(1/\alpha))}|Y|$ and $|D'| \geq (\varepsilon \alpha/2)^{O(\log(1/\alpha))}|D|$, and

$$|A \cap S(X, Y', D')| \geq (1 + \Omega(\varepsilon^2))\alpha \delta_X |Y'| |D'|.$$

This, however contradicts the $(O(\log(1/\alpha)^4), O(\log(1/\alpha)^2), O(\varepsilon^2))$ -combinatorial spreadness of A in the container $S(X, Y, D)$. By our choice of Q , we have $Q^{-1} \leq 2^{-O(\log(1/\alpha)^6)} \leq (\varepsilon \alpha \delta_X \delta_Y \delta_D)^{O(k)}$ so [Lemma A.8](#) implies that

$$\Phi(f, f, f) \geq (1 - O(\varepsilon))\alpha^3 \delta_X^2 \delta_Y^2 \delta_D^2.$$

In other words, A contains at least

$$(1 - O(\varepsilon))\alpha^3\delta_X^2\delta_Y^2\delta_D^2|G|^2 \geq 2^{-O(\log(1/\alpha)^5)}|G|^3 > |G|^2$$

many corners. Since there are at most $|G|^2$ trivial corners (i.e., triples $(x, y), (xg, y), (x, gy)$ with $g = \mathbf{1}_G$), the result follows. $\square$

APPENDIX B. ALMOST PERIODICITY

We first state the key almost periodicity result which we require. This is proven as [59, Theorem 5.4]; we quote the statement from [49, Theorem 8] (which while stated for $\mathbb{Z}/N\mathbb{Z}$ follows for general G by changing each occurrence of $\mathbb{Z}/N\mathbb{Z}$ in the half page deduction from [59, Theorem 5.4] to G).

Theorem B.1. *Let $\varepsilon \in (0, 1/2)$. Let $B_1, B_2, \dots$ be a (d, η) -small sequence of Bohr sets (see Definition 6.6).*

Let $Y \subseteq B_1$ and $Z \subseteq B_2$ with $\beta = \mathbb{E}_{x \sim B_1}[\mathbb{1}_Y(x)]$ and $\gamma = \mathbb{E}_{x \sim B_2}[\mathbb{1}_Z(x)]$. Let $D \subseteq G$ be such that $|Y| \leq |D| \leq 2|B_1|$.

Then there exists a regular Bohr set $B' \subseteq B_2$ of dimension at most $d + d'$ where

$$d' = O(\varepsilon^{-4} \log(2\beta^{-1})^3 \log(2\gamma^{-1}))$$

and radius $r_2 \cdot (\varepsilon\beta)/(24d^3d')$ and such that

$$\left| \mathbb{E}_{\substack{b \sim B' \\ y \sim B_1 \\ z \sim B_2}} \mathbb{1}_D(z - y + b) \mathbb{1}_Y(y) \mathbb{1}_Z(z) - \mathbb{E}_{\substack{y \sim B_1 \\ z \sim B_2}} \mathbb{1}_D(z - y) \mathbb{1}_Y(y) \mathbb{1}_Z(z) \right| \leq \varepsilon\beta\gamma.$$

We now give the proof of Theorem 6.16.

Proof of Theorem 6.16. We first begin with essentially the dependent random choice argument of [42]; we follow the proof as in [22, Theorem 3.7]. We let B_3 be a regular Bohr set with the same frequencies as B_i and such that $r_3/r_2 \in [\eta/2, \eta]$. Observe that

$$\mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2 \\ z \sim B_3}} f_1(x - z) f_2(y + z) g(x + y) \geq (1 + 9\varepsilon/10) \cdot \mathbb{E}[f_1] \cdot \mathbb{E}[f_2] \cdot \mathbb{E}[g].$$

Let $p = 1000 \lceil k \log(1/\varepsilon)/\varepsilon \rceil$ and observe that

$$\mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2 \\ z \sim B_3}} f_1(x - z) f_2(y + z) g(x + y) \leq \left(\mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} f_1(x - z) \right)^{p-1} \cdot \left(\mathbb{E}_{x \sim B_1} \left(\mathbb{E}_{y \sim B_2} f_2(y + z) g(x + y) \right)^p \right).$$

Via the definition of p , this implies that

$$(1 + 7\varepsilon/8)^p \cdot (\mathbb{E}[f_2] \mathbb{E}[g])^p \leq \mathbb{E}_{\substack{x \sim B_1 \\ y_1, \dots, y_p \sim B_2 \\ z \sim B_3}} \prod_{j=1}^p f_2(y_j + z) g(x + y_j).$$

Define

$$\mathcal{E}_1(y_1, \dots, y_p) = \mathbb{1} \left[\mathbb{E}_{x \sim B_1} \prod_{j=1}^p g(x + y_j) \leq (\mathbb{E}[f_2] \cdot \mathbb{E}[g])^p \right] \cdot \mathbb{1} \left[\mathbb{E}_{z \sim B_3} \prod_{j=1}^p f_2(y_j + z) \leq (\mathbb{E}[f_2] \cdot \mathbb{E}[g])^p \right]$$

and observe that

$$(1 + 3\varepsilon/4)^p \cdot (\mathbb{E}[f_2] \mathbb{E}[g])^p \leq \mathbb{E}_{\substack{x \sim B_1 \\ y_1, \dots, y_p \sim B_2 \\ z \sim B_3}} (1 - \mathcal{E}_1(y_1, \dots, y_p)) \prod_{j=1}^p f_2(y_j + z) g(x + y_j).$$

We define

$$\mathcal{E}_2(x, z) = \mathbb{1} \left[\mathbb{E}_{y \sim B_2} f_2(y + z)g(x + y) \leq (1 + 5\varepsilon/8) \cdot (\mathbb{E}[f_2] \cdot \mathbb{E}[g]) \right].$$

Observe that

$$\begin{aligned} & \mathbb{E}_{\substack{x \sim B_1 \\ y_1, \dots, y_p \sim B_2 \\ z \sim B_3}} \mathcal{E}_2(x, z) \prod_{j=1}^p f_2(y_j + z)g(x + y_j) \\ & \leq 10^{-5} \cdot \varepsilon \cdot \mathbb{E}_{\substack{x \sim B_1 \\ y_1, \dots, y_p \sim B_2 \\ z \sim B_3}} (1 - \mathcal{E}_1(y_1, \dots, y_p)) \prod_{j=1}^p f_2(y_j + z)g(x + y_j). \end{aligned}$$

Then there exists $(y_1, \dots, y_p) \in B_2^{\otimes p}$ such that $\mathcal{E}_1(y_1, \dots, y_p) = 0$ and

$$\mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} \mathcal{E}_2(x, z) \prod_{j=1}^p f_2(y_j + z)g(x + y_j) \leq 10^{-5} \cdot \varepsilon \cdot \mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} \prod_{j=1}^p f_2(y_j + z)g(x + y_j).$$

Let $H_1(x) = \prod_{j=1}^p g(x + y_j)$ and $H_2(z) = \prod_{j=1}^p f_2(y_j + z)$. By the definition of $\mathcal{E}_1$ we have that $\mathbb{E}[H_1] \geq (\mathbb{E}[f_1] \cdot \mathbb{E}[g])^p$ and $\mathbb{E}[H_2] \geq (\mathbb{E}[f_1] \cdot \mathbb{E}[g])^p$. The above conclusion is then equivalent to

$$\mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} H_1(x) \mathcal{E}_2(x, z) H_2(z) \leq 10^{-5} \cdot \varepsilon \cdot \mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} H_1(x) H_2(z).$$

If we now forget the precise details of the construction, we may assume that $H_1 : B_1 \rightarrow \{0, 1\}$ and $H_2 : B_3 \rightarrow \{0, 1\}$ with $\mathbb{E}[H_1] \geq (\mathbb{E}[f_1] \cdot \mathbb{E}[g])^p/2$ and

$$\mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} H_1(x) \mathcal{E}_2(x, z) H_2(z) \leq 10^{-5} \cdot \varepsilon \cdot \mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} H_1(x) H_2(z)$$

via [Lemma 3.3](#).

We now observe that

$$\mathbb{E}_{y \sim B_2} f_2(y + z)g(x + y) = \mathbb{E}_{y \sim B_2} f_2(y)g(x + y - z) + O(\eta d).$$

Let $\mathcal{E}_3(t) = \mathbb{1}[\mathbb{E}_{y \sim B_2} f_2(y)g(t + y) \leq (1 + 9\varepsilon/16) \cdot \mathbb{E}[f_2] \mathbb{E}[g]]$ and note that we have

$$\mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} H_1(x) \mathcal{E}_3(x - z) H_2(z) \leq 10^{-5} \cdot \varepsilon \cdot \mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} H_1(x) H_2(z).$$

This is equivalent to

$$\mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} H_1(x) (1 - \mathcal{E}_3(x - z)) H_2(z) \geq (1 - 10^{-5} \cdot \varepsilon) \cdot \mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} H_1(x) H_2(z).$$

Observe that if $\mathbb{E}_{x \sim B_1} (1 - \mathcal{E}_3(x)) \leq \mathbb{E}[H_1(x)]/2$ then

$$\mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} H_1(x) (1 - \mathcal{E}_3(x - z)) H_2(z) \leq \mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} (1 - \mathcal{E}_3(x - z)) H_2(z) \leq 3/4 \cdot \mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} H_1(x) H_2(z)$$

which is a contradiction.

We now apply [Theorem B.1](#). We first suppose that $\mathbb{E}_{x \sim B_1} (1 - \mathcal{E}_3(x)) \geq \mathbb{E}[H_1(x)]$. Then $(1 - \mathcal{E}_3)$ will be $\mathbb{1}_D$, H_1 will be $\mathbb{1}_Y$, and H_2 will be $\mathbb{1}_Z$. There exists a Bohr set B' of dimension bounded by $d + d'$ with

$$d' \ll \varepsilon^{-4} p^4 \cdot \log(1/(\mathbb{E}[f_2] \mathbb{E}[g]))^4 \ll \varepsilon^{-8} (\log(1/\varepsilon))^4 \cdot k^8$$

and radius $r' \gg r_3 \cdot \varepsilon \cdot (\mathbb{E}[f_2] \mathbb{E}[g])^{-O(k \log(1/\varepsilon)/\varepsilon)} / d^4$ such that

$$\mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3 \\ t \sim B'}} H_1(x)(1 - \mathcal{E}_3(x - z + t))H_2(z) \geq (1 - 10^{-4} \cdot \varepsilon) \cdot \mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} H_1(x)H_2(z).$$

Else we have that $\mathbb{E}_{x \sim B_1}(1 - \mathcal{E}_3(x)) \in [\mathbb{E}[H_1(x)]/2, \mathbb{E}[H_1(x)]]$ and observe that

$$\mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} H_1(x + z)(1 - \mathcal{E}_3(x))H_2(z) \geq (1 - 2 \cdot 10^{-5} \cdot \varepsilon) \cdot \mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} H_1(x)H_2(z).$$

Thus (taking $\mathbb{1}_Z = H_2$, $\mathbb{1}_D = H_1$ and $\mathbb{1}_Y(y) = (1 - \mathcal{E}_3)(-y)$) there exists B' with the same properties as earlier such that

$$\mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3 \\ t \sim B'}} H_1(x + z + t)(1 - \mathcal{E}_3(x))H_2(z) \geq (1 - 5 \cdot 10^{-5} \cdot \varepsilon) \cdot \mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} H_1(x)H_2(z).$$

Via a change of variable we get

$$\begin{aligned} \mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3 \\ t \sim B'}} H_1(x)(1 - \mathcal{E}_3(x - z - t))H_2(z) &= \mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3 \\ t \sim B'}} H_1(x)(1 - \mathcal{E}_3(x - z + t))H_2(z) \\ &\geq (1 - 10^{-4} \cdot \varepsilon) \cdot \mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} H_1(x)H_2(z); \end{aligned}$$

this is the same conclusion as earlier.

This immediately gives the desired conclusion modulo unwinding definitions. Observe that we have that

$$\mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3 \\ t \sim B'}} H_1(x)(1 - \mathcal{E}_3(x - z + t))H_2(z) \geq (1 - 10^{-4} \cdot \varepsilon) \cdot \mathbb{E}_{\substack{x \sim B_1 \\ z \sim B_3}} H_1(x)H_2(z).$$

Via the definition of $\mathcal{E}_3$, we obtain that

$$\mathbb{E}_{\substack{x \sim B_1 \\ y \sim B_2 \\ z \sim B_3}} H_1(x)H_2(z)f_2(y) \left(\mathbb{E}_{t \sim B'} g(x - z + y + t) \right) \geq (1 + 17\varepsilon/32) \cdot \mathbb{E}[H_1] \cdot \mathbb{E}[H_2] \cdot \mathbb{E}[f_2] \cdot \mathbb{E}[g].$$

Therefore there exists $x \in B_1, y \in B_2, z \in B_3$ such that

$$\left(\mathbb{E}_{t \sim B'} g(x - z + y + t) \right) \geq (1 + 17\varepsilon/32) \cdot \mathbb{E}[g]$$

as desired. □

REFERENCES

- [1] M. Ajtai and E. Szemerédi, *Sets of lattice points that form no squares*, Studia Sci. Math. Hungar. **9** (1974), 9–11. [2](#)
- [2] Ryan Alweiss, Shachar Lovett, Kewen Wu, and Jiapeng Zhang, *Improved bounds for the sunflower lemma*, Ann. of Math. (2) **194** (2021), 795–815. [12](#)
- [3] Tim Austin, *Ajtai-Szemerédi theorems over quasirandom groups*, Recent trends in combinatorics, IMA Vol. Math. Appl., vol. 159, Springer, [Cham], 2016, pp. 453–484. [4](#)
- [4] László Babai, Nikolay Nikolov, and László Pyber, *Product growth and mixing in finite groups*, Proceedings of the Nineteenth Annual ACM-SIAM Symposium on Discrete Algorithms, ACM, New York, 2008, pp. 248–257. [71](#)
- [5] László Babai, Noam Nisan, and Márió Szegedy, *Multiparty protocols and logspace-hard pseudorandom sequences*, Proceedings of the Twenty-first Annual ACM Symposium on Theory of Computing, STOC 1989, Seattle, Washington, USA, May 15–17, 1989, Association for Computing Machinery (ACM), New York, 1989, p. 1–11. [1, 71](#)

- [6] Paul Beame, Matei David, Toniann Pitassi, and Philipp Woelfel, *Separating deterministic from randomized multiparty communication complexity*, Theory Comput. **6** (2010), 201–225. [2](#)
- [7] Paul Beame and Dang-Trinh Huynh-Ngoc, *Multiparty communication complexity and threshold circuit size of AC^0* , 2009 50th Annual IEEE Symposium on Foundations of Computer Science—FOCS 2009, IEEE Computer Soc., Los Alamitos, CA, 2009, pp. 53–62. [1](#)
- [8] F. A. Behrend, *On sets of integers which contain no three terms in arithmetical progression*, Proc. Nat. Acad. Sci. U.S.A. **32** (1946), 331–332. [1](#), [3](#)
- [9] Vitaly Bergelson, Randall McCutcheon, and Qing Zhang, *A Roth theorem for amenable groups*, Amer. J. Math. **119** (1997), 1173–1211. [3](#)
- [10] Amey Bhargale and Subhash Khot, *Optimal inapproximability of satisfiable k -LIN over non-abelian groups*, STOC '21—Proceedings of the 53rd Annual ACM SIGACT Symposium on Theory of Computing, ACM, New York, [2021] ©2021, pp. 1615–1628. [71](#)
- [11] Thomas F. Bloom and Olof Sisask, *The Kelley-Meka bounds for sets free of three-term arithmetic progressions*, Essent. Number Theory **2** (2023), 15–44. [6](#)
- [12] J. Bourgain, *On triples in arithmetic progression*, Geom. Funct. Anal. **9** (1999), 968–984. [36](#)
- [13] Jean Bourgain and Alex Gamburd, *Uniform expansion bounds for Cayley graphs of $SL_2(\mathbb{F}_p)$* , Ann. of Math. (2) **167** (2008), 625–642. [71](#)
- [14] Ashok K. Chandra, Merrick L. Furst, and Richard J. Lipton, *Multi-party protocols*, Proceedings of the 15th Annual ACM Symposium on Theory of Computing, 25–27 April, 1983, Boston, Massachusetts, USA (David S. Johnson, Ronald Fagin, Michael L. Fredman, David Harel, Richard M. Karp, Nancy A. Lynch, Christos H. Papadimitriou, Ronald L. Rivest, Walter L. Ruzzo, and Joel I. Seiferas, eds.), ACM, 1983, pp. 94–99. [1](#), [2](#), [3](#)
- [15] Matthias Christandl, Omar Fawzi, Hoang Ta, and Jeroen Zuiddam, *Larger corner-free sets from combinatorial degenerations*, 13th Innovations in Theoretical Computer Science Conference, LIPIcs. Leibniz Int. Proc. Inform., vol. 215, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2022, pp. Art. No. 48, 20. [3](#)
- [16] David Conlon, Jacob Fox, and Yufei Zhao, *Extremal results in sparse pseudorandom graphs*, Adv. Math. **256** (2014), 206–290. [9](#), [10](#), [72](#)
- [17] David Conlon, Jacob Fox, and Yufei Zhao, *A relative Szemerédi theorem*, Geom. Funct. Anal. **25** (2015), 733–762. [10](#)
- [18] Ernie Croot, Vsevolod F. Lev, and Péter Pál Pach, *Progression-free sets in $\mathbb{Z}_4^n$ are exponentially small*, Ann. of Math. (2) **185** (2017), 331–337. [3](#)
- [19] Harm Derksen, Chin Ho Lee, and Emanuele Viola, *Boosting uniformity in quasirandom groups: fast and simple*, 2024 IEEE 65th Annual Symposium on Foundations of Computer Science—FOCS 2024, IEEE Computer Soc., Los Alamitos, CA, [2024] ©2024, pp. 1425–1430. [71](#)
- [20] Jordan S. Ellenberg and Dion Gijswijt, *On large subsets of $\mathbb{F}_q^n$ with no three-term arithmetic progression*, Ann. of Math. (2) **185** (2017), 339–343. [3](#)
- [21] Christian Elsholtz, Zach Hunter, Laura Proske, and Lisa Sauermann, *Improving Behrend’s construction: Sets without arithmetic progressions in integers and over finite fields*. [3](#)
- [22] Yuval Filmus, Hamed Hatami, Kaave Hosseini, and Esty Kelman, *Sparse graph counting and Kelley-Meka bounds for binary systems*, 2024 IEEE 65th Annual Symposium on Foundations of Computer Science (FOCS), IEEE, 2024, pp. 1559–1578. [37](#), [45](#), [80](#)
- [23] H. Furstenberg and Y. Katznelson, *An ergodic Szemerédi theorem for commuting transformations*, J. Analyse Math. **34** (1978), 275–291. [2](#)
- [24] W. T. Gowers, *Arithmetic progressions in sparse sets*, Current developments in mathematics, 2000, Int. Press, Somerville, MA, 2001, pp. 149–196. [2](#)
- [25] W. T. Gowers, *Hypergraph regularity and the multidimensional Szemerédi theorem*, Ann. of Math. (2) **166** (2007), 897–946. [2](#)
- [26] W. T. Gowers, *Quasirandom groups*, Combin. Probab. Comput. **17** (2008), 363–387. [70](#), [71](#)
- [27] W. T. Gowers and Emanuele Viola, *Interleaved group products*, SIAM J. Comput. **48** (2019), 554–580. [71](#)
- [28] Ron Graham and József Solymosi, *Monochromatic equilateral right triangles on the integer grid*, Topics in discrete mathematics, Algorithms Combin., vol. 26, Springer, Berlin, 2006, pp. 129–132. [3](#)
- [29] B. Green, *100 open problems*, <https://people.maths.ox.ac.uk/greenbj/papers/open-problems.pdf>. [2](#)
- [30] Ben Green, *An argument of Shkredov in the finite field setting*, <https://people.maths.ox.ac.uk/greenbj/papers/corners.pdf>. [2](#)
- [31] Ben Green, *Finite field models in additive combinatorics*, Surveys in combinatorics 2005, London Math. Soc. Lecture Note Ser., vol. 327, Cambridge Univ. Press, Cambridge, 2005, pp. 1–27. [2](#), [3](#)
- [32] Ben Green, *Lower bounds for corner-free sets*, New Zealand J. Math. **51** (2021), 1–2. [3](#)
- [33] Ben Green and Terence Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. (2) **167** (2008), 481–547. [10](#)

- [34] Ben Green and Terence Tao, *New bounds for Szemerédi’s theorem. II. A new bound for $r_4(N)$* , Analytic number theory, Cambridge Univ. Press, Cambridge, 2009, pp. 180–204. [4](#)
- [35] Lianna Hambardzumyan, Toniann Pitassi, Suhail Sherif, Morgan Shirley, and Adi Shraibman, *An improved protocol for EXACTLYN with more than 3 players*, 15th Innovations in Theoretical Computer Science Conference, LIPIcs. Leibniz Int. Proc. Inform., vol. 287, Schloss Dagstuhl. Leibniz-Zent. Inform., Wadern, 2024, pp. Art. No. 58, 23. [3](#)
- [36] Hamed Hatami, *Graph norms and Sidorenko’s conjecture*, Israel J. Math. **175** (2010), 125–150. [7](#), [19](#)
- [37] Zach Hunter, *Corner-free sets via the torus*. [3](#)
- [38] Michael Jaber, Shachar Lovett, and Anthony Ostuni, *Strong bounds for skew corner-free sets*. [2](#)
- [39] Peter Keevash and Noam Lifshitz, *Sharp hypercontractivity for symmetric groups and its applications*, arXiv preprint arXiv:2307.15030 (2023), Available at <https://arxiv.org/pdf/2307.15030>. [5](#)
- [40] Zander Kelley, *Problems in extremal combinatorics and connections with multiparty communication complexity*, 2024, <https://www.youtube.com/watch?v=YkHabH4iPVo>. [2](#)
- [41] Zander Kelley, Shachar Lovett, and Raghu Meka, *Explicit separations between randomized and deterministic Number-on-Forehead communication*, STOC’24—Proceedings of the 56th Annual ACM Symposium on Theory of Computing, ACM, New York, [2024] ©2024, pp. 1299–1310. [2](#), [7](#), [8](#), [9](#), [14](#), [15](#), [20](#), [21](#)
- [42] Zander Kelley and Raghu Meka, *Strong bounds for 3-progressions*, 2023 IEEE 64th Annual Symposium on Foundations of Computer Science—FOCS 2023, IEEE Computer Soc., Los Alamitos, CA, [2023] ©2023, pp. 933–973. [2](#), [4](#), [6](#), [11](#), [12](#), [20](#), [23](#), [45](#), [80](#)
- [43] Andrey Kupavskii and Dmitrii Zakharov, *Spread approximations for forbidden intersections problems*, Adv. Math. **445** (2024), Paper No. 109653, 29. [12](#)
- [44] Michael T. Lacey and William McClain, *On an argument of Shkredov on two-dimensional corners*, Online J. Anal. Comb. (2007), Art. 2, 21. [3](#)
- [45] James Leng, Ashwin Sah, and Mehtaab Sawhney, *Improved Bounds for Szemerédi’s Theorem*, arXiv:2402.17995. [4](#)
- [46] Nati Linial and Adi Shraibman, *Larger corner-free sets from better NOF exactly-N protocols*, Discrete Anal. (2021), Paper No. 19, 9. [2](#), [3](#)
- [47] A. Lubotzky, R. Phillips, and P. Sarnak, *Ramanujan graphs*, Combinatorica **8** (1988), 261–277. [71](#)
- [48] Raghu Meka, *Strong bounds for 3-progressions*, 2023, Talk as part the Breakthroughs lecture series at the Simons Institute for the Theory of Computing. <https://www.youtube.com/watch?v=WN7rJPWy6z8>. [2](#)
- [49] Luka Milićević, *Good bounds for sets lacking skew corners*, arXiv:2404.07180. [2](#), [13](#), [37](#), [80](#)
- [50] Brendan Nagle, Vojtěch Rödl, and Mathias Schacht, *The counting lemma for regular k -uniform hypergraphs*, Random Structures Algorithms **28** (2006), 113–179. [2](#)
- [51] Noam Nisan and Avi Wigderson, *Rounds in communication complexity revisited*, SIAM J. Comput. **22** (1993), 211–219. [1](#)
- [52] Sarah Peluse, *Finite field models in arithmetic combinatorics—twenty years on*, Surveys in combinatorics 2024, London Math. Soc. Lecture Note Ser., vol. 493, Cambridge Univ. Press, Cambridge, 2024, pp. 159–199. [2](#)
- [53] Sarah Peluse, *Subsets of $\mathbb{F}_p^n \times \mathbb{F}_p^n$ without L-shaped configurations*, Compos. Math. **160** (2024), 176–236. [2](#)
- [54] D. H. J. Polymath, *A new proof of the density Hales-Jewett theorem*, Ann. of Math. (2) **175** (2012), 1283–1327. [2](#)
- [55] L. Pyber, *How abelian is a finite group?*, The mathematics of Paul Erdős, I, Algorithms Combin., vol. 13, Springer, Berlin, 1997, pp. 372–384. [4](#)
- [56] Ran Raz, *The BNS-Chung criterion for multi-party communication complexity*, Comput. Complexity **9** (2000), 113–122. [1](#)
- [57] Vojtěch Rödl and Jozef Skokan, *Regularity lemma for k -uniform hypergraphs*, Random Structures Algorithms **25** (2004), 1–42. [2](#)
- [58] Tom Sanders, *Solving $xz = y^2$ in certain subsets of finite groups*, Q. J. Math. **68** (2017), 243–273. [5](#)
- [59] Tomasz Schoen and Olof Sisask, *Roth’s theorem for four variables and additive structures in sums of sparse sets*, Forum Math. Sigma **4** (2016), Paper No. e5, 28. [80](#)
- [60] Jean-Pierre Serre, *Linear representations of finite groups*, Graduate Texts in Mathematics, vol. Vol. 42, Springer-Verlag, New York-Heidelberg, 1977. [71](#)
- [61] I. D. Shkredov, *On a problem of Gowers*, Dokl. Akad. Nauk **400** (2005), 169–172. [2](#)
- [62] I. D. Shkredov, *On a generalization of Szemerédi’s theorem*, Proc. London Math. Soc. (3) **93** (2006), 723–760. [2](#), [3](#), [6](#), [8](#), [11](#), [12](#)
- [63] Adi Shraibman, *A note on multiparty communication complexity and the Hales-Jewett theorem*, Inform. Process. Lett. **139** (2018), 44–48. [2](#)
- [64] József Solymosi, *Note on a generalization of Roth’s theorem*, Discrete and computational geometry, Algorithms Combin., vol. 25, Springer, Berlin, 2003, pp. 825–827. [2](#)

- [65] E. Szemerédi, *On sets of integers containing no four elements in arithmetic progression*, Number Theory (Colloq., János Bolyai Math. Soc., Debrecen, 1968), Colloq. Math. Soc. János Bolyai, vol. 2, North-Holland, Amsterdam-London, 1970, pp. 197–204. [2](#)
- [66] E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245. [2](#)
- [67] Terence Tao, *254B, Notes 3: Quasirandom groups, expansion, and Selberg’s $3/16$ theorem*, <https://terrytao.wordpress.com/2011/12/16/254b-notes-3-quasirandom-groups-expansion-and-selbergs-316-theorem/>. [71](#)
- [68] Terence Tao, *A variant of the hypergraph removal lemma*, J. Combin. Theory Ser. A **113** (2006), 1257–1280. [2](#)
- [69] Terence Tao and Van H. Vu, *Additive combinatorics*, paperback ed., Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge University Press, Cambridge, 2010. [36](#)
- [70] Emanuele Viola, *Guest column: Non-abelian combinatorics and communication complexity*, ACM SIGACT News **50** (2019), 52–74. [4](#)
- [71] Yufei Zhao, *Graph theory and additive combinatorics—exploring structure and randomness*, Cambridge University Press, Cambridge, 2023. [2](#)

DEPARTMENT OF COMPUTER SCIENCE, UNIVERSITY OF TEXAS, AUSTIN, TX 78712

Email address: mjjaber@cs.utexas.edu

COMPUTER SCIENCE DEPARTMENT, CARNEGIE MELLON UNIVERSITY, PITTSBURGH, PA 15213

Email address: yang17@andrew.cmu.edu

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, UNIVERSITY OF CALIFORNIA, SAN DIEGO, CA 92093

Email address: slovett@ucsd.edu

DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING, UNIVERSITY OF CALIFORNIA, SAN DIEGO, CA 92093

Email address: aostuni@ucsd.edu

DEPARTMENT OF MATHEMATICS, COLUMBIA UNIVERSITY, NEW YORK, NY 10027

Email address: m.sawhney@columbia.edu