# Quasipolynomial bounds for the corners theorem

Talk by Mehtaab Sawhney

Notes by Sanjana Das

April 17, 2025

This is joint work with Michael Jaber, Yang P. Liu, Shachar Lovett, and Anthony Ostuni.

## §1 Introduction

### §1.1 Context

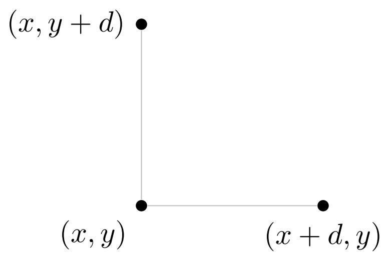
This work broadly falls in the context of theorems in additive combinatorics around Szemerédi's theorem.

**Theorem 1.1** (Szemerédi 1975)
Let $A \subseteq [N]$ with $|A| \geq \delta N$. If $N$ is sufficiently large as a function of $\delta$ and $k$, then $A$ contains a $k$-AP — i.e., it contains $x$, $x + y$, $x + 2y$, $\ldots$, $x + (k-1)y$ where $y \neq 0$.

Szemerédi's theorem is about finding some kind of regular structure in a 1-dimensional set which is dense. Since Szemerédi's theorem, there's been lots of work towards multidimensional versions of this. One of the first was the following:

**Theorem 1.2** (Ajtai–Szemerédi 1974)
Let $A \subseteq [N]$ with $|A| \geq \delta N^2$. If $N$ is sufficiently large as a function of $\delta$, then $A$ contains three points $(x, y)$, $(x + d, y)$, and $(x, y + d)$ with $d > 0$.

So if you have a dense subset of the 2-dimensional integer lattice, then it must contain a *corner* — three points forming an isosceles right triangle.




This is a very particular 3-point configuration in 2 dimensions, but we actually know much more. Historically, after Szemerédi's theorem, Furstenberg proved it by ergodic methods in 1977, and soon afterwards there was a generalization which completely generalized this result.

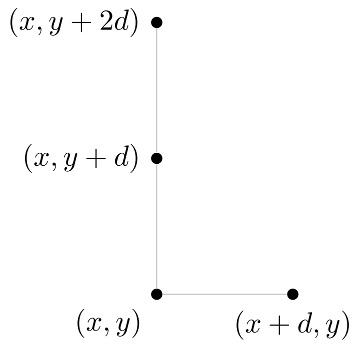
**Theorem 1.3** (Furstenberg–Katznelson 1977)
Let $P$ be any fixed subset in $d$ dimensions. Let $A \subseteq [N]$ with $|A| \geq \delta N^d$. If $N$ is sufficiently large as a function of $\delta$, $P$, and $d$, then $A$ contains $x + t \cdot P$ with $t \in \mathbb{Z}$.

This notation means you take your fixed configuration $P$, and you're allowed to dilate it by an integer $t$ and to shift it by $x$; so you're finding a copy of $P$ inside your set.

The proof of this is by ergodic theory. To this day, at this level of generality we have very poor understanding of the function $N(\delta, P, d)$ (how large $N$ has to be for this to hold). The Furstenberg–Katznelson proof gives no bounds because it uses ergodic theory. Now there are proofs using the hypergraph regularity lemma, but you can't write down good effective bounds — the current bound is maybe a tower of 2's of height $1/\delta$, or worse (e.g., for four points in a square).

But in some cases, we do have a decent understanding. In 1 dimensions, we have reasonable bounds. This is essentially contained in the work of Gowers 1998 — he gave an alternate proof of Szemerédi's theorem which gave effective bounds, and that can be used to handle all 1-dimensional patterns.

There are exactly two other cases where we have effective bounds. One is corners (solving a question of Gowers); this was handled by Shkredov 2005. The other pattern we have is four points forming an L, i.e., $(x, y)$, $(x + d, y)$, $(x, y + d)$, $(x, y + 2d)$. This was handled by Peluse 2024.




## §1.2 Corners

Today we'll focus on corners. Shkredov gave a reasonable quantitative rate:

**Theorem 1.4** (Shkredov 2005)
Let $A \subseteq [N]^2$ with $|A| \geq N^2 \cdot (\log\log N)^{-\Omega(1)}$. Then $A$ contains a corner.

(The constant in the exponent is something explicit, like $\frac{1}{73}$.)

Our main theorem improves this quantitative version.

**Theorem 1.5** (JLLOS 2025+)
Let $A \subseteq [N]^2$ with $|A| \geq N^2 \cdot e^{-c(\log N)^{1/600}}$. Then $A$ contains a corner.

In particular, this bound is two exponentials better in terms of the density savings.

Before diving into discussions of the proof, we'll make a few remarks. First, the shape of this bound is optimal — there's a lower bound of the form $N^2 \cdot \mathsf{quasipoly}(N)$ (meaning there exist sets of that size with no corners). This comes from examples of large sets which avoid 3-APs, due to Behrend. So up to changing $\frac{1}{600}$ to maybe $\frac{1}{2}$, this bound is correct — at least its shape is correct.

**Remark 1.6.** By a well-known projection argument, upper bounds for corners also give upper bounds for sets avoiding 3-APs. The first people to prove quasipolynomial bounds for 3-APs were Kelley–Meka 2023. (This is not an alternate proof of their result; we use their ideas.)

### §1.3 Number-on-forehead complexity

This has a nice application to computational complexity. In the number on forehead model, imagine you have three people, and they're trying to compute some function of three numbers $x$, $y$, and $z$. Player 1 has $x$ taped to their forehead, Player 2 has $y$, and Player 3 has $z$. So as the player, you can see the other two numbers, but not your own.

**Question 1.7.** What's the minimum number of bits that the players need to communicate to compute a certain function?

The corners problem is closely related to computing the function $\mathbf{1}[x + y + z = N]$, called the Exactly-$N$ function. There's a randomized protocol taking $O(1)$ rounds of communication; but the question is, can you get good bounds on the non-randomized setting?

By a reduction of Chandra–Furst–Lipton 1983, Theorem 1.5 implies that Exactly-$N$ needs $\Omega((\log N)^{1/600})$ rounds of communication (for deterministic protocols). The trivial algorithm for computing this function would be for Player 2 to tell Player 1 what his number is; this takes $\log N$ rounds. There's a better algorithm that takes $(\log N)^c$ rounds for some $c < 1$, so this bound is optimal up to the power of log.

(By 'rounds of communication,' we mean the total number of bits they communicate.)

This is for 3-player number on forehead, and essentially solves it up to the power of log. You can also look at 4-player number on forehead, where you have $x$, $y$, $z$, $w$ and four players. For 4-player Exactly-$N$, Theorem 1.5 gives a lower bound of $\Omega(\log\log\log\log N)$ rounds. Maybe this is not so interesting, but before, you couldn't even really write down a function that tended to $\infty$ (though we knew one existed). (It's expected the right answer should be a power of log, but we're very far away from proving anything like that.)

### §1.4 Setup

As one last thing, all the setup we've stated makes sense in the context of general abelian groups. So for the rest of the talk, we'll consider $A \subseteq \mathbb{F}_2^n \times \mathbb{F}_2^n$ instead. So $N$ from the earlier statements becomes $2^n$, and we still want to find configurations $(x, y)$, $(x, y + d)$, $(x + d, y)$ in $A$. (Essentially, statements on $[N]$ are the same as ones on $\mathbb{Z}/N\mathbb{Z}$, but working with finite fields instead makes many of the technical aspects easier to think about.)

Also, we've been parametrizing corners as $(x, y)$, $(x, y + d)$, $(x + d, y)$ so far. But for the proof, it's useful to parametrize corners differently — we substitute $d = x + y + z$, so the parametrization becomes $(x, y)$, $(x, x + z)$, $(y + z, y)$. (We're in characteristic 2, so for example, $x + 2y + z = x + z$.) So we get this simpler parametrization, and we'll be operating with it for the rest of the talk.

## §2 Shkredov's argument

To discuss the proof, it's important to understand the work of Shkredov — it's the starting point. We'll win two logs over Shkredov's bound, so it's important to see where you lose each log in that proof.

For concreteness, let $A \subseteq \mathbb{F}_2^n \times \mathbb{F}_2^n$ with $|A| = \alpha \cdot 4^n$. For the rest of this computation, we write $\mathbf{1}_A(x, y)$ for the function dictating whether $(x, y) \in A$, and $f_A(x, y) = \mathbf{1}_A(x, y) - \alpha$ for the balanced indicator function.

Let's try to interpret the condition that $A$ has no corners analytically. It's relatively straightforward to show this means that

$$|\mathbb{E}_{x,y,z} f_A(x,y)\mathbf{1}_A(x,x+z)\mathbf{1}_A(y,y+z)| \gtrsim \alpha^3$$

(throughout this talk, $x$, $y$, $z$ always range over $\mathbb{F}_2^n$) — what's happening is that if we expand $f_A(x,y) = \mathbf{1}_A(x,y) - \alpha$, the first term $\mathbf{1}_A(x,y)\mathbf{1}_A(x,x+z)\mathbf{1}_A(y,y+z)$ has very small contribution because there are no corners in $A$, while the second term $\alpha\mathbf{1}_A(x,x+z)\mathbf{1}_A(y,y+z)$ results in the $\alpha^3$.

This is additive combinatorics, so we'll do lots of Cauchy–Schwarz; our goal is to use this statement to derive some structure on $f_A$. First, by the triangle inequality, we can bound the left-hand side by

$$\mathbb{E}_{x,z} \left|\mathbb{E}_y f_A(x,y)\mathbf{1}_A(y,y+z)\right|.$$

Then by Cauchy–Schwarz, this is at most

$$\left(\mathbb{E}_{x,z} \left(\mathbb{E}_y f_A(x,y)\mathbf{1}_A(y,y+z)\right)^2\right)^{1/2}.$$

If we expand the square by creating two copies $y_1$ and $y_2$ and switch the order of summation, this becomes

$$\left(\mathbb{E}_{y_1,y_2} \left(\mathbb{E}_x f_A(x,y_1)f_A(x,y_2)\right)\left(\mathbb{E}_z \mathbf{1}_A(y_1,y_1+z)\mathbf{1}_A(y_2,y_2+z)\right)\right)^{1/2}.$$

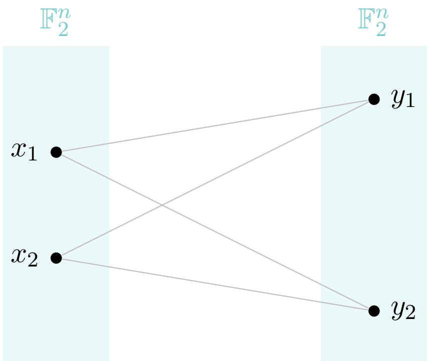
All of these things are 1-bounded, so we can use the triangle inequality and drop the second term; so this is at most

$$\left(\mathbb{E}_{y_1,y_2} \left|\mathbb{E}_x f_A(x,y_1)f_A(x,y_2)\right|\right)^{1/2}.$$

Now we do Cauchy–Schwarz again to bound this by

$$\left(\mathbb{E}_{x_1,x_2,y_1,y_2} f_A(x_1,y_1)f_A(x_1,y_2)f_A(x_2,y_1)f_A(x_2,y_2)\right)^{1/4}. \quad (2.1)$$

Now let's examine what we've proven. We can think of the above expression as some sort of 4-cycle count: We can imagine a bipartite graph with $\mathbb{F}_2^n$ on each side, and with edges $f_A(x,y) = \mathbf{1}_A(x,y) - \alpha$. (If we didn't have the $-\alpha$, then this would be precisely a graphical representation of $A$; the $-\alpha$ makes it a weighted graph with mean 0.) Then this expression is the density of 4-cycles in this weighted graph.

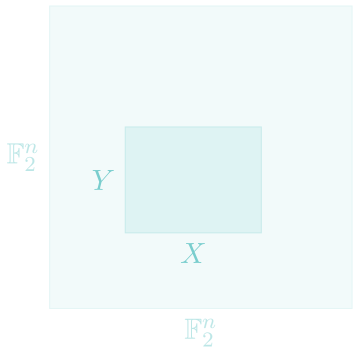



Due to Chung–Graham–Wilson, if this 4-cycle count is large (what we've shown is that it's bigger than $\alpha^{12}$), this implies there is a subset of this graph which has higher density than you expect. (This is by some further applications of Cauchy–Schwarz.) So this implies there exist sets $X$ and $Y$ such that

$$\mathbb{E}\mathbf{1}_X(x)\mathbf{1}_Y(y)\mathbf{1}_A(x,y) \geq (\alpha + \alpha^{O(1)})\mathbb{E}\mathbf{1}_X(x)\mathbf{1}_Y(y).$$

In words, what this says is that on the rectangle $X \times Y$, the density of $A$ went up by $\alpha^{O(1)}$ — its original density was $\alpha$, but its density on $X \times Y$ is $\alpha + \alpha^{O(1)}$. Right now this is not so interesting because $X$ and $Y$ could just be singletons, but actually we can also ensure that $\mathbb{E}\mathbf{1}_X \geq \alpha^{O(1)}$ and $\mathbb{E}\mathbf{1}_Y \geq \alpha^{O(1)}$.

What have we accomplished? We originally had some set, and we passed to an arbitrary subset $X$ in one dimension and $Y$ in the other such that on the subgrid $X \times Y$, we've increased the density.




That seems promising, and the idea of the *density increment strategy* is to sort of iterate this.

The first log loss already comes from here — when we do a density increment, we go

$$\alpha \mapsto \alpha + \alpha^{O(1)}.$$

We won't work out the density increment calculation, but if you do, it leads to logarithmic-type bounds. So that's at least one log lost.

For where the second log loss comes from, we've sort of cheated here — after we apply this result, $A$ no longer lives on $\mathbb{F}_2^n \times \mathbb{F}_2^n$, but instead on some arbitrary product space $X \times Y$. We want to run this argument again to get another density increment, but we can't, because we're no longer in $\mathbb{F}_2^n \times \mathbb{F}_2^n$. (This part was Shkredov's main innovation — it was known before that you could get to a density increment on some $X \times Y$, but it wasn't clear what to do with this.)

What Shkredov does is show that one can *pseudorandomize* $X$ and $Y$ to have no large Fourier coefficients. Roughly, you take your sets $X$ and $Y$, and if either of them has a large Fourier coefficient, then this means you have a density increment along some subspace, so you can split your sets along that subspace. And you can run an iterative argument to find $X$ and $Y$ which live on some smaller subspace and are pseudorandom on that subspace. It turns out that if they're pseudorandom, then you can run an argument like this (and we'll use that in the proof).

But there's some argument that uses Fourier coefficients here, and this loses an additional log.

So each of these costs you one log, and we need to remove both to get quasipolynomial bounds. To preserve everyone's sanity, we won't discuss the second (the pseudorandomization step); it's complicated, and the solution to the first (the density increment step) informs the solution to the second. We'll really discuss what you replace the first step with (the log loss from the density increment) and how; that'll be the focus of the rest of the talk.

## §3 A better density increment

The idea (which comes from work of Kelley–Meka and Kelley–Lovett–Meka) is to replace the density increment by something better. We'd like to get a density increment of the form $\alpha \mapsto \alpha(1 + O(1))$ instead of $\alpha \mapsto \alpha + \alpha^{O(1)}$ — instead of going up by a power of $\alpha$, you want to replace $\alpha$ with $\alpha \cdot 1.01$. If you can get this, it'll save you one log.

But the Cauchy–Schwarz argument will always lose a power of $\alpha$ — at some point we're replacing functions like $\mathbf{1}_A$ with the fact that they're 1-bounded, and it turns out that will always lose things like this. So the idea is some kind of Höldering magic which gets around this issue; it's sort of a surprising thing that this is possible.

## §3.1 Setup

For some setup, when we pass to smaller rectangles, $A$ is going to live in some container set. So for the rest of the talk, we'll have three sets $X, Y, D \subseteq \mathbb{F}_2^n$; we'll write $\mathbb{E}\mathbf{1}_X(x) = \delta_X$, $\mathbb{E}\mathbf{1}_Y(y) = \delta_Y$, and $\mathbb{E}\mathbf{1}_D(z) = \delta_D$ for their densities. We'll always have

$$\mathbf{1}_A(x,y) \leq \mathbf{1}_X(x)\mathbf{1}_Y(y)\mathbf{1}_D(x+y)$$

(in words, this says that for all $(x, y) \in A$, you need $x \in X$, $y \in Y$, and $x + y \in D$). We'll additionally have that $X$, $Y$, and $D$ are 'pseudorandom.'

Also, previously we used $\alpha$ to measure the density of $A$ relative to the entire space, but now it's better to measure it relative to the container sets; so we'll have

$$\mathbb{E}_A\mathbf{1}_A(x,y) = \alpha\delta_X\delta_Y\delta_Z.$$

We'll also have a parameter $k \asymp \log(1/\alpha)$. Finally, we also need an analog of the function $f_A$ from before; here that'll be

$$f_A(x,y) = \mathbf{1}_A(x,y) - \alpha\delta_X\delta_Y\delta_D.$$

## §3.2 Computations

First, by a quick computation (similarly to before), the fact that $A$ is corner-free means

$$c\alpha^3\delta_X^2\delta_Y^2\delta_D^2 \leq |\mathbb{E}_{x,y,z}\mathbf{1}_A(x,y)\mathbf{1}_A(x,x+z)f_A(y+z,y)|$$

(where $c$ is some absolute constant that may change line to line). There will be lots of factors of $\delta_X$ and so on floating around, so here's the way to think about them: First, $A$ is an $\alpha$-fraction of the container; there are three $A$'s on the right-hand side, so we get an $\alpha^3$. If we look at factors of $\delta_X$, the first term is telling us that $x \in X$, which contributes one factor of $\delta_X$; the second term doesn't give any additional constraint involving $X$; and the third term contributes another $\delta_X$. So that's why you get two factors of $\delta_X$; the other normalizations work out similarly.

Then by Hölder's inequality, we get

$$(c\alpha^3)^k(\delta_X^2\delta_Y^2\delta_D^2)^k \leq (\mathbb{E}_{x,y}\mathbf{1}_A(x,y))^{k-1}\left(\mathbb{E}_{x,y}\mathbf{1}_A(x,y)\left(\mathbb{E}_z\mathbf{1}_A(x,x+z)f_A(y+z,y)\right)^k\right).$$

(Recall that $k$ is quite large — it's a power of $\log(1/\alpha)$.) We know what the first term on the right-hand side is — we have $\mathbb{E}_{x,y}\mathbf{1}_A(x,y) = \alpha\delta_X\delta_Y\delta_D$ — so this becomes

$$(c\alpha)^{2k}(\delta_X\delta_Y\delta_D)^{k+1} \leq \mathbb{E}_{x,y}\mathbf{1}_A(x,y)\left(\mathbb{E}_z\mathbf{1}_A(x,x+z)f_A(y+z,y)\right)^k.$$

Now we'll do something that at first seems extremely lossy — we'll take the term $\mathbf{1}_A(x,y)$ and replace it by the simpler function $\mathbf{1}_X(x)\mathbf{1}_Y(y)\mathbf{1}_D(x+y)$. This seems like we've made a huge loss — we've replaced $A$ by this upper bound coming from the container, which seems like it should lose a factor of $\alpha$. But the point is that we've chosen $k$ to be so large that this just corresponds to changing the constant $c$ to some slightly smaller constant. So as long as we don't need to track this constant carefully, this loss isn't too bad.

Now we've gotten that

$$(c\alpha)^{2k}(\delta_X\delta_Y\delta_D)^{k+1} \leq \mathbb{E}_{x,y}\mathbf{1}_X(x)\mathbf{1}_Y(y)\mathbf{1}_D(x+y)\left(\mathbb{E}_z\mathbf{1}_A(x,x+z)f_A(y+z,y)\right)^k.$$

The second point is that we don't actually need to keep around the $\mathbf{1}_X$ and $\mathbf{1}_Y$ terms — if the second factor $\mathbb{E}_z(\cdots)$ is nonzero, then we automatically know $x \in X$ and $y \in Y$. So we can erase those terms and get

$$(c\alpha)^{2k}(\delta_X\delta_Y\delta_D)^{k+1} \leq \mathbb{E}_{x,y}\mathbf{1}_D(x+y)\left(\mathbb{E}_z\mathbf{1}_A(x,x+z)f_A(y+z,y)\right)^k.$$

Next, since we're only trying to win one log factor right now (this becomes a key difficulty when trying to win the second), we can assume $D$ is Fourier-pseudorandom. If one is careful, this means we can replace $\mathbf{1}_D$ by its average; so we essentially get

$$(c\alpha)^{2k}(\delta_X\delta_Y\delta_D)^{k+1} \leq \mathbb{E}_{x,y}\delta_D\left(\mathbb{E}_z\mathbf{1}_A(x,x+z)f_A(y+z,y)\right)^k.$$

> **Remark 3.1.** That's why taking this high power $k$ and Höldering is so useful — it means you can afford to lose some powers of $\alpha$ in this way. The tradeoff is that we are going to eventually lose something in $k$ — we're going to get a multiplicative density increment, but it'll be on a rectangle of size $2^{-\mathsf{poly}(k)}$. It turns out that the density increment argument can tolerate really bad losses in the size if you can get a multiplicative density increment, so this works out to be better.

Then if you do Cauchy–Schwarz (and a trick called 'spectral positivity' that we won't go into), at the end of time you get that

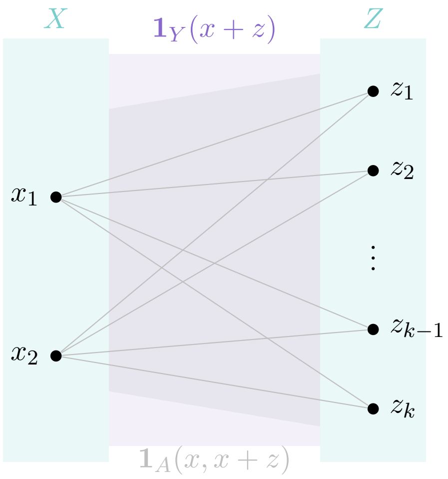
$$\mathbb{E}_{x_1,x_2,z_1,\ldots,z_k}\prod_{i\in[2]}\prod_{j\in[k]}\mathbf{1}_A(x_i,x_i+z_j) \geq ((1+c)\alpha)^{2k}\delta_X^2\delta_Y^{2k}\delta_D^k. \tag{3.1}$$

(In some sense, this is analogous to saying that if the $L^p$ norm of a vector is large, e.g., twice the $L^2$ norm, then there should be a set of size $2^{-p}$ on which it's double what you'd expect. Here $f_A$ is not purely positive; what spectral positivity does is let you derive a purely positive conclusion, replacing $f_A$ with $\mathbf{1}_A$.)

## §3.3 Graph counting

In Shkredov's argument, the expression (2.1) we got corresponded to cycles of length 4, or copies of $K_{2,2}$. In (3.1), we have something more complicated. Let's try to interpret this as a graph; the rest of the talk will be about graphs and graph counting lemmas.

We have a set $X$ on the left and $Z$ on the right. In (3.1), there are 2 vertices on the left and $k$ on the right, and the graph $G$ is given by $\mathbf{1}_A(x,x+z)$. So this quantity is computing the number of copies of $K_{2,k}$. And we also have an additional property: Recll that $\mathbf{1}_A(x,x+z) \leq \mathbf{1}_Y(x+z)$. So we can think of $G$ as living inside some bigger graph given by $Y$ — there's a big graph with edges given by $\mathbf{1}_Y(x+z)$, and our graph is an $\alpha$-subset of this big graph.




So we have a graph with many copies of $K_{2,k}$ in it, and we want to derive some conclusion from this. (You can think of $\delta_X^2$ and the other factors as normalizations.) In Shkredov's argument, we did this by quoting a quasirandomness result; but now we're in a slightly different setting.

There's an important condition that $Y$ is Fourier-pseudorandom. That implies the big underlying graph (defined by $Y$) is quasirandom — so we have a dense subset of a very quasirandom set.

**Theorem 3.2**

In this setup, there exist 1-bounded nonnegative functions $g_1$ and $g_2$ with $\mathbb{E}[g_1] \geq e^{-\log(1/\alpha)^{O(1)}k^{O(1)}}\delta_X$ and $\mathbb{E}[g_2] \geq e^{-\log(1/\alpha)^{O(1)}k^{O(1)}}\delta_D$, such that

$$\mathbb{E}[g_1(x)g_2(z)\mathbf{1}_A(x, x+z)] \geq (1+c)\alpha\mathbb{E}[g_1]\mathbb{E}[g_2].$$

This conclusion takes a second to parse, but what it means is that we've taken $A$ and found a multiplicative density increment (on $g_1$ and $g_2$). The size of the rectangles are now $\mathbb{E}[g_1]$ and $\mathbb{E}[g_2]$, and that's important to track. Originally the left side was supported on $X$, whose density was $\delta_X$; so the amount the density went down is the $e^{-\log(1/\alpha)^{O(1)}k^{O(1)}}$ factor.

It's important that this depends only on $\alpha$ and $k$, not the $\delta$'s. If you naively tried proving a statement like this, you'd pick up dependences on $\delta_Y$, because our graph $G$ is sparse (as a graph between $X$ and $Y$). But the quasirandomness of $Y$ means that this doesn't happen — this is some kind of 'dense model' statement.

To say a bit about the proof: Proving this statement is already quite hard. But the main idea is to use an idea called *densification* due to Conlon–Fox–Zhao. Very roughly, this is a technique saying that if you're trying to prove these kinds of quasirandomness statements in the setting of a 'pseudorandom majorant' (here that's $\mathbf{1}_Y(x+z)$), as long as it's sufficiently pseudorandom you can 'pretend it doesn't exist' (and here this means the densities of $g_1$ and $g_2$ don't depend on $\delta_Y$). The point is that because $Y$ is so pseudorandom, you can hope to win back this factor of $Y$.

So this gives the multiplicative density increment; and once you have this, you can run an argument which at least saves one log.