RANDOM SUM-FREE SUBSETS OF ABELIAN GROUPS

JÓZSEF BALOGH, ROBERT MORRIS, AND WOJCIECH SAMOTIJ

Abstract. We characterize the structure of maximum-size sum-free subsets of a random subset of an abelian group G. In particular, we determine the threshold $p_c \approx \sqrt{\log n / n}$ above which, with high probability as $|G| \to \infty$, each such subset is contained in a maximum-size sum-free subset of G, whenever q divides $|G|$ for some (fixed) prime q with $q \equiv 2 \pmod 3$. Moreover, in the special case $G = \mathbb{Z}_{2n}$, we determine a sharp threshold for the above property. The proof uses recent 'transference' theorems of Conlon and Gowers, together with stability theorems for sum-free sets of abelian groups.

1. Introduction

One of the most important developments in Combinatorics over the past twenty years has been the introduction and proof of various 'random analogues' of well-known theorems in Extremal Graph Theory, Ramsey Theory, and Additive Combinatorics. These questions were first introduced for graphs by Babai, Simonovits, and Spencer [5], and for additive structures by Kohayakawa, Łuczak, and Rödl [28], and there has since been a tremendous interest in such problems (see for example [21, 23, 32, 33, 34]). This extensive study has recently culminated in the remarkable results of Conlon and Gowers [14] and Schacht [36] (see also Friedgut, Rödl, and Schacht [22]) in which a general theory was developed to attack such questions.

The main theorems in [14] and [36] resolved many long-standing open questions; however, they provide only asymptotic results. For example, they prove that, with high probability as $n \to \infty$, the largest triangle-free subgraph of the random graph $G_{n,p}$ has

$$\left(\frac{1}{2} + o(1)\right) p \binom{n}{2}$$

edges if $p \gg 1/\sqrt{n}$, which is best possible up to a constant. A much more precise question asks the following: for which functions $p = p(n)$ is, with high probability, the largest triangle-free subgraph of $G_{n,p}$ bipartite? This was answered (in the affirmative) in the case $p = 1/2$ by Erdős, Kleitman, and Rothschild [16], for $p \geqslant 1/2 - \delta$ by Babai, Simonovits, and Spencer [5], and for $p \geqslant n^{-\varepsilon}$ by Brightwell, Panagiotou, and Steger [7], where δ and ε are small positive constants. As pointed out in [7], the statement is false if $p \leqslant \frac{1}{10}\sqrt{\log n / n}$.

In the setting of additive number theory, the first results on such problems were obtained by Kohayakawa, Łuczak, and Rödl [28], who proved the following random version of

Research supported in part by: (JB) NSF CAREER Grant DMS-0745185, UIUC Campus Research Board Grants 09072 and 11067, and OTKA Grant K76099; (RM) ERC Advanced grant DMMCA, and a Research Fellowship from Murray Edwards College, Cambridge; (WS) Parker Fellowship and Schark Fellowship (from UIUC Mathematics Department) and ERC Advanced Grant DMMCA.

Roth's Theorem [35]. Given an abelian group G, a set $B \subseteq G$ is *p-random* if each element is chosen independently with probability p. We say that B is *δ-Roth* if every subset $A \subseteq B$ with $|A| \geqslant \delta|B|$ contains a 3-term arithmetic progression. Then, for every $\delta > 0$, if B is a p-random subset of $\mathbb{Z}_n$, and $p \geqslant C(\delta)/\sqrt{n}$, then B is δ-Roth with high probability. This result is again best possible up to the constant $C(\delta)$.

Given a sequence of abelian groups $(G_n)_{n\in\mathbb{N}}$ with $|G_n| = n$, we say that $p_c = p_c(n)$ is a *threshold function* for a property $\mathcal{P}$ if a p-random subset $B \subseteq G_n$ has $\mathcal{P}$ with high probability if $p = p(n) \gg p_c(n)$, and B does not have $\mathcal{P}$ with high probability if $p \ll p_c$. Bollobás and Thomason [6] proved that every monotone property has a threshold function. A threshold function is *sharp* if moreover the above holds with $p > (1+\varepsilon)p_c$ and $p < (1-\varepsilon)p_c$ for any fixed $\varepsilon > 0$. Friedgut and Kalai [20] proved that every monotone graph property has a sharp threshold, and further results were proved by Friedgut and Bourgain [19]. Non-monotone properties, such as the one which we shall be studying, are more complicated and such thresholds do not necessarily exist in general. We shall prove that such a threshold *does* exist in the setting described below, and moreover we shall determine it. For $\mathbb{Z}_{2n}$ we shall prove much more: that there exists a sharp threshold.

A set $A \subseteq G$ is said to be *sum-free* if there is no solution to the equation $x + y = z$ with $x, y, z \in A$, and such forbidden triples (x, y, z) are called *Schur triples*. (Note that we forbid some triples with $x = y$, and also some with $x = z$; the results and proofs in the case that such triples are allowed are identical.) In 1916, Schur [37] proved that, given any r-colouring of the integers $\mathbb{Z}$, there exists a monochromatic Schur triple. Graham, Rödl, and Ruciński [23] studied the random version of Schur's Theorem, and proved that the threshold function for the existence of a 2-colouring of a p-random set $B \subseteq \mathbb{Z}_n$ without a monochromatic Schur triple is $1/\sqrt{n}$. The extremal version of this question was open for fifteen years, until it was recently resolved by Conlon and Gowers [14] and Schacht [36].

Sum-free sets have been extensively studied over the past 40 years (see for example [9, 15, 24, 27, 30, 38]), mostly in the extremal setting. For example, in 1969 Diananda and Yap [15] determined the extremal density for a sum-free set in every abelian group G such that $|G|$ has a prime divisor q with $q \not\equiv 1 \pmod 3$. However, it was more than 30 years until the classification was completed by Green and Ruzsa [27] (see Theorem 2.1, below). Another well-studied problem was the Cameron-Erdős Conjecture (see [2, 8, 11, 12, 26, 31]), which asked for the number of sum-free sets of $\mathbb{Z}_n$, and was finally solved by Green [24] in 2004.

In this paper we shall study the analogue of the Babai-Simonovits-Spencer problem for sum-free sets. Indeed, let $G = \mathbb{Z}_{2n}$; it was proved in [15] that the maximum-size sum-free set in G is the set O_{2n} of odd numbers. We shall prove a probabilistic version of this statement, and determine the threshold p_c for the property that the unique maximum-size sum-free subset of a p-random subset B of G is the set $B \cap O_{2n}$. In fact, we shall do better: we shall determine a *sharp* threshold for this property.

For a set $B \subseteq G$, let $\mathrm{SF}_0(B)$ denote the collection of maximum-size sum-free subsets of B. The following theorem is our main result.

Theorem 1.1. *Let $n \in \mathbb{N}$, let $C = C(n) \gg \log n/n$, and let $p^2 n = C \log n$. Let $G_p \subseteq \mathbb{Z}_{2n}$ be a random set, with each element chosen independently with probability p. Then*

$$\mathbb{P}\Big(\mathrm{SF}_0(G_p) = \{G_p \cap O_{2n}\}\Big) \to \begin{cases} 0 & \text{if} \quad \limsup_n C(n) < 1/3 \\ 1 & \text{if} \quad \liminf_n C(n) > 1/3 \end{cases}$$

as $n \to \infty$. Moreover, if $C = 1/3$ then $\mathbb{P}\big(\mathrm{SF}_0(G_p) = \{G_p \cap O_{2n}\}\big) \to 0$ as $n \to \infty$.

We remark that the threshold for the asymptotic version of this statement (i.e., that the largest sum-free subset of G_p has $(1 + o(1))pn$ elements), determined by Conlon and Gowers [14] and Schacht [36], is $1/\sqrt{n}$, and so it differs from our threshold by a factor of $\sqrt{\log n}$.

We shall also determine the threshold $p_c(n)$ for all abelian groups of Type I(q) (see below), i.e., those for which $|G| = n$ has a (fixed) prime divisor q with $q \equiv 2 \pmod 3$. Say that a set $B \subseteq G$ is *sum-free good* if every maximum-size sum-free subset of B is of the form $B \cap A$ for some $A \in \mathrm{SF}_0(G)$.

Theorem 1.2. *Let q be a prime number such that $q \equiv 2 \pmod 3$. There exist positive constants c_q and C_q such that the following holds.*

Let $\mathcal{G} = (G_n)_{n \in q\mathbb{N}}$ be a sequence of abelian groups, with $|G_n| = n$, such that q divides $|G|$ for every $G \in \mathcal{G}$. Let $C = C(n) \gg \log n/n$, let $p^2 n = C \log n$, and let G_p be a p-random subset of $G = G_n$. Then

$$\mathbb{P}\Big(G_p \text{ is sum-free good}\Big) \to \begin{cases} 0 & \text{if} \quad C < c_q \\ 1 & \text{if} \quad C > C_q \end{cases}$$

as $n \to \infty$.

We remark that the conclusion of the theorem fails to hold when we do not assume that $|G|$ has a prime divisor q with $q \equiv 2 \pmod 3$. Indeed, we shall show (see Proposition 5.1) that if $G = \mathbb{Z}_{3q}$, where q is prime and $q \equiv 1 \pmod 3$, then the probability that G_p is sum-free good goes to zero (as $n = |G| \to \infty$) for all $p \ll (n \log n)^{-1/3}$. We shall also prove the same bound for the group $G = \mathbb{Z}_q$, where $q \equiv 2 \pmod 3$ (see Proposition 5.2), which shows that the condition $n \gg q$ in Theorem 1.2 is also necessary.

Finally, we note that, perhaps not surprisingly, the constant $C = 1/3$ in Theorem 1.1 is not the same for every abelian group G with $|G|$ even. Indeed, we shall show (see Proposition 5.3) that for the hypercube $G = \{0,1\}^k$, where $|G| = 2n$, the threshold p_c is at least $\sqrt{\log n/(2n)}$.

The remainder of the paper is organized as follows. In Section 2, we collect some extremal results on abelian groups, and the probabilistic tools that will be needed later. In Section 3 we state our main tool, a theorem of Conlon and Gowers [14] (see also Schacht [36]), which provides an asymptotic version of Theorems 1.1 and 1.2. Theorem 1.2 is proved in Section 4, and in Section 5 we prove the lower bounds for other abelian groups described above. Finally, in Section 6, we prove Theorem 1.1.

2. Preliminaries and tools

2.1. **Extremal results on abelian groups.** Let G be a finite abelian group. Given two subsets $A, B \subseteq G$, we let

$$A \pm B = \{a \pm b \colon a \in A \text{ and } b \in B\}.$$

Note that a subset A of G is *sum-free* if $(A + A) \cap A = \emptyset$. We begin with an important definition in the study of sum-free subsets of finite abelian groups.

Definition. Let G be a finite abelian group. We say that

(1) G is of type I if $|G|$ has at least one prime divisor q with $q \equiv 2 \pmod 3$.
(2) G is of type II if $|G|$ has no prime divisors q with $q \equiv 2 \pmod 3$, but $|G|$ is divisible by 3.
(3) G is of type III if every prime divisor q of $|G|$ satisfies $q \equiv 1 \pmod 3$.

Moreover, we say that G is of type I(q) if G is of type I and q is the smallest prime divisor of $|G|$ with $q \equiv 2 \pmod 3$.

Given an abelian group G, let $\mu(G)$ be the density of the largest sum-free subset of G (so that this subset has $\mu(G)|G|$ elements). As noted in the Introduction, the problem of determining $\mu(G)$ for an arbitrary G has been studied for more than 40 years, but only recently was it solved completely by Green and Ruzsa [27].

Theorem 2.1 (Diananda and Yap [15], Green and Ruzsa [27])**.** *Let G be an arbitrary finite abelian group. Then*

$$\mu(G) = \begin{cases} \frac{1}{3} + \frac{1}{3q} & \text{if } G \text{ is of type I}(q), \\ \frac{1}{3} & \text{if } G \text{ is of type II}, \\ \frac{1}{3} - \frac{1}{3m} & \text{if } G \text{ is of type III}, \end{cases}$$

where m is the exponent (largest order of any element) of G.

Note that it follows immediately from Theorem 2.1 that

$$2/7 \leqslant \mu(G) \leqslant 1/2$$

for every finite abelian group G.

Recall that $\mathrm{SF}_0(G)$ denotes the collection of all maximum-size sum-free subsets of G, i.e., those that have $\mu(G)|G|$ elements. As well as determining $\mu(G)$, Diananda and Yap [15] described $\mathrm{SF}_0(G)$ for all groups of type I (see also [27, Lemma 5.6]).

Theorem 2.2 (Diananda and Yap [15])**.** *Let G be a group of type $I(q)$, for some prime $q = 3k + 2$, and let $A \in \mathrm{SF}_0(G)$. There exists a homomorphism $\varphi \colon G \to \mathbb{Z}_q$ such that $A = \varphi^{-1}(\{k + 1, \ldots, 2k + 1\})$.*

In other words, every $A \in \mathrm{SF}_0(G)$ is a union of cosets of some subgroup H of G of index q, A/H is an arithmetic progression in G/H, and $A \cup (A + A) = G$.

We shall also need the following well-known bound on the number of homomorphisms from an arbitrary finite abelian group to a cyclic group of prime order, which follows easily from Kronecker's Decomposition Theorem.

Proposition 2.3. *For every prime q, the number of homomorphisms from a finite abelian group G to the cyclic group $\mathbb{Z}_q$ is at most $|G|$.*

Theorem 2.2 and Proposition 2.3 immediately imply the following.

Corollary 2.4. *Let G be an arbitrary group of type I. Then $|\operatorname{SF}_0(G)| \leqslant |G|$.*

We will also need the following corollary from the classification of maximum-size sum-free subsets of $\mathbb{Z}_{3q}$, where q is a prime with $q \equiv 1 \pmod 3$, due to Yap [39].

Corollary 2.5. *If q is a prime with $q \equiv 1 \pmod 3$, then $|\operatorname{SF}_0(\mathbb{Z}_{3q})| \leqslant 2q$.*

In the proof of Theorem 1.2, we will use the following two lemmas, which were proved in [27] and [30] respectively. The first lemma establishes a strong stability property for large sum-free subsets of groups of type I.

Lemma 2.6 (Green and Ruzsa [27])**.** *Let G be an abelian group of type $I(q)$. If A is a sum-free subset of G, and*

$$|A| \geqslant \left(\mu(G) - \frac{1}{3q^2+3q}\right)|G|,$$

then A is contained in some $A' \in \operatorname{SF}_0(G)$.

The second lemma is a rather straightforward corollary of a much stronger result of Lev, Łuczak, and Schoen [30].

Lemma 2.7 (Lev, Łuczak, and Schoen [30])**.** *Let $\varepsilon > 0$, let G be a finite abelian group, and let $A \subseteq G$. If*

$$|A| \geqslant \left(\frac{1}{3} + \varepsilon\right)|G|$$

then one of the following holds:

(*a*) *$|A \setminus A'| \leqslant \varepsilon|G|$ for some sum-free $A' \subseteq A$.*
(*b*) *A contains at least $\varepsilon^3|G|^2/27$ Schur triples.*

We remark that a more general statement, the so-called Removal Lemma for groups, was proved by Green [25] (for abelian groups) and Král, Serra, and Vena [29] (for arbitrary groups). Lemmas 2.6 and 2.7 immediately imply the following.

Corollary 2.8. *Let G be a group of type $I(q)$, where $q \equiv 2 \pmod 3$, and let $0 < \varepsilon < \frac{1}{9q^2+9q}$. Let $A \subseteq G$, and suppose that*

$$|A| \geqslant \big(\mu(G) - \varepsilon\big)|G|.$$

Then one of the following holds:

(*a*) *$|A \setminus A'| \leqslant \varepsilon|G|$ for some $A' \in \operatorname{SF}_0(G)$.*
(*b*) *A contains at least $\varepsilon^3|G|^2/27$ Schur triples.*

Proof. Suppose first that $|A \setminus A''| > \varepsilon|G|$ for every sum-free $A'' \subseteq A$. By Theorem 2.1 and our choice of ε, we have $|A| \geqslant (1/3 + \varepsilon)|G|$. Hence, by Lemma 2.7, A contains at least $\varepsilon^3|G|^2/27$ Schur triples, as required.

So assume that there exists a sum-free set $A'' \subseteq A$ with $|A \setminus A''| \leqslant \varepsilon|G|$. Then

$$|A''| \,\geqslant\, |A| - \varepsilon|G| \,\geqslant\, \big(\mu(G) - 2\varepsilon\big)|G| \,>\, \left(\mu(G) - \frac{1}{3q^2+3q}\right)|G|,$$

and so, by Lemma 2.6, A'' is contained in some $A' \in \mathrm{SF}_0(G)$. But then

$$|A \setminus A'| \,\leqslant\, |A \setminus A''| \,\leqslant\, \varepsilon|G|,$$

and so we are done in this case as well. □

2.2. **Probabilistic tools.** To finish this section, we shall recall three well-known probabilistic inequalities: the FKG inequality, Janson's inequality, and Chernoff's inequality.

Given an arbitrary set X and a real number $p \in [0,1]$, we denote by X_p the random subset of X, where each element is included with probability p independently of all other elements. In the proof of our main theorem, we shall need several bounds on the probabilities of events of the form

$$\bigwedge_{i \in I} (B_i \nsubseteq X_p),$$

where B_i are subsets of X. The first such estimate can be easily derived from the FKG Inequality; see, e.g., [4, Section 6.3].

The FKG inequality. *Suppose that $\{B_i\}_{i\in I}$ is a family of subsets of a finite set X and let $p \in [0,1]$. Then*

$$\mathbb{P}\big(B_i \nsubseteq X_p \text{ for all } i \in I\big) \,\geqslant\, \prod_{i\in I} \mathbb{P}\big(B_i \nsubseteq X_p\big) \,=\, \prod_{i \in I}\big(1 - p^{|B_i|}\big). \tag{1}$$

The second result, due to Janson (see, e.g., [4, Section 8.1]), gives an upper bound on the probability in the left-hand side of (1) expressed in terms of the intersection pattern of the sets B_i.

Janson's inequality. *Suppose that $\{B_i\}_{i\in I}$ is a family of subsets of a finite set X and let $p \in [0,1]$. Let*

$$M = \prod_{i \in I}\big(1 - p^{|B_i|}\big), \quad \mu = \sum_{i\in I} p^{|B_i|}, \quad \text{and} \quad \Delta = \sum_{i \sim j} p^{|B_i \cup B_j|},$$

where $i \sim j$ denotes the fact that $i \neq j$ and $B_i \cap B_j \neq \emptyset$. Then,

$$\mathbb{P}(B_i \nsubseteq X_p \text{ for all } i \in I) \leqslant \min\left\{Me^{\Delta/(2-2p)}, e^{-\mu+\Delta/2}\right\}.$$

Furthermore, if $\Delta \geqslant \mu$, then

$$\mathbb{P}(B_i \nsubseteq X_p \text{ for all } i \in I) \leqslant e^{-\mu^2/(2\Delta)}.$$

Finally, we will need the following well-known concentration result for binomial random variables; see, e.g., [4, Appendix A].

Chernoff's inequality. *Let X be the binomial random variable with parameters n and p. Then for every positive a,*

$$\mathbb{P}\big(X - pn > a\big) \, < \, \exp\left(-\frac{a^2}{2pn} + \frac{a^3}{2(pn)^2}\right) \quad \textit{and} \quad \mathbb{P}\big(X - pn < -a\big) \, < \, \exp\left(-\frac{a^2}{2pn}\right).$$

2.3. **Notation.** For the sake of brevity, we shall write y for the set $\{y\}$. We shall also use Δ to denote the "correlation measure" as in Janson's inequality, above, and $\Delta(G)$ for the maximum degree in a graph G. We trust that neither of these will confuse the reader.

3. The Conlon-Gowers Method

In order to prove the 1-statement in Theorem 1.2, i.e., that every maximum-size sum-free subset of G_p is of the form $A \cap G_p$ for some $A \in \mathrm{SF}_0(G)$, provided that p is sufficiently large, we need the following approximate version of this statement, which, as we will show below, can be derived from Corollary 2.8 using the *transference* theorems of Conlon and Gowers [14].

Theorem 3.1. *For every $\varepsilon > 0$, and every prime q with $q \equiv 2 \pmod 3$, there exists a constant $C = C(q, \varepsilon) > 0$ such that the following holds. Let G be an arbitrary n-element group of type $I(q)$. If*

$$p \, \geqslant \, \frac{C}{\sqrt{n}},$$

then a.a.s. for every sum-free subset of G_p with

$$|B| \, \geqslant \, \left(\mu(G) - \frac{\varepsilon}{40q^2 + 40q}\right) p|G|,$$

there is an $A \in \mathrm{SF}_0(G)$ such that $|B \setminus A| \leqslant \varepsilon pn$.

Unfortunately, for technical reasons, the methods of Conlon and Gowers [14] can only be applied under the additional assumption that $p \geqslant C(q,\varepsilon)^{-1}$. Therefore, we will show how to derive Theorem 3.1 from the following statement, which in turn can be proved using the aforementioned transference theorems.

Theorem 3.2. *For every $\varepsilon > 0$, and every prime q with $q \equiv 2 \pmod 3$, there exists a constant $C = C(q, \varepsilon) > 0$ such that the following holds. Let G be an arbitrary n-element group of type $I(q)$. If*

$$\frac{C}{\sqrt{n}} \, \leqslant \, p \, \leqslant \, \frac{1}{C},$$

then a.a.s. for every sum-free subset of G_p with

$$|B| \, \geqslant \, \left(\mu(G) - \frac{1}{10q^2 + 10q}\right) p|G|,$$

there is an $A \in \mathrm{SF}_0(G)$ such that $|B \setminus A| \leqslant \varepsilon pn$.

In order to deduce Theorem 3.1 from Theorem 3.2, we simply chop p into sufficiently small pieces, apply Theorem 3.2 to each piece, and then show that we obtain the same set $A \in \mathrm{SF}_0(G)$ for (almost) each of the pieces.

Proof of Theorem 3.1. Clearly, it suffices to consider the case $p \geqslant C(q,\varepsilon)^{-1}$. To this end, fix some $p \in (0,1]$, let $c_q = 1/(10q^2 + 10q)$, let $\varepsilon' = \varepsilon c_q/4$, and let M be the least positive integer satisfying $2p/M \leqslant 1/C'$, where $C' = C(q,\varepsilon')$ is the constant defined in the statement of Theorem 3.2. Assign to each $x \in G$ a number $t(x) \in [0,1]$ uniformly at random and for all $i \in [M]$, let

$$G^i = \{x \in G \colon (i-1)p/M \leqslant t(x) \leqslant ip/M\}.$$

It is not hard to see that $G^1 \cup \ldots \cup G^M$ has the same distribution as G_p, that G^i has the same distribution as $G_{p/M}$ for every $i \in [M]$, and that $G^i \cup G^j$ has the same distribution as $G_{2p/M}$ for all distinct $i,j \in [M]$. Since $M = O(1)$, a.a.s. the following statements hold simultaneously:

(*i*) For all $i \in [M]$, G^i satisfies the assertion of Theorem 3.2 with $\varepsilon = \varepsilon'$.

(*ii*) For all distinct $i,j \in [M]$, $G^i \cup G^j$ satisfies the assertion of Theorem 3.2 with $\varepsilon = \varepsilon'$.

(*iii*) For all $i \in [M]$ and distinct $A, A' \in \mathrm{SF}_0(G)$, $|G^i \cap A| \leqslant (\mu(G) + \varepsilon')pn/M$ and

$$|G^i \cap A \cap A'| \;\leqslant\; \left(\mu(G) - \frac{1}{2q^2}\right)\frac{pn}{M}.$$

To see that (*iii*) holds, observe that $|A \cap A'| \leqslant (\mu(G) - 1/q^2)n$ since, by Theorem 2.2, $A \setminus A'$ is a union of cosets of some subgroup $H \cap H'$, where H and H' are subgroups of index q (and hence $H \cap H'$ has index q or q^2). Now, since $p = \Theta(1)$ and $|\mathrm{SF}_0(G)| \leqslant n$, by Corollary 2.4, the result follows by Chernoff's inequality.

Now, let B be a sum-free subset of $G^1 \cup \ldots \cup G^M$ with $|B| \geqslant (\mu(G) - \varepsilon')pn$. For each $i \in [M]$, let $B^i = G^i \cap B$ and let

$$I \;=\; \left\{ i \in [M] : \, |B^i| \geqslant \big(\mu(G) - c_q\big)\frac{pn}{M} \right\}.$$

Suppose that $i \in I$. Since $B^i \subset B$ is sum-free, (*i*) implies that $|B^i \setminus A^i| \leqslant \varepsilon' pn/M$ for some $A^i \in \mathrm{SF}_0(G)$. Moreover, for every distinct $i,j \in I$, the set $B^i \cup B^j \subset B$ is also sum-free and $|B^i \cup B^j| = |B^i| + |B^j| \geqslant (\mu(G) - c_q)2pn/M$. Thus, by (*ii*), we have $|(B^i \cup B^j) \setminus A| \leqslant 2\varepsilon' pn/M$ for some $A \in \mathrm{SF}_0(G)$. In particular, it follows from (*iii*) that $A^i = A^j = A$, since otherwise

$$|G^i \cap A \cap A^i| \;\geqslant\; |B^i \cap A \cap A^i| \;\geqslant\; |B^i| - |B^i \setminus A^i| - |B^i \setminus A| \;\geqslant\; \Big(\mu(G) - c_q - 3\varepsilon'\Big)\frac{pn}{M},$$

which contradicts (*iii*), since $c_q + 3\varepsilon' \leqslant 1/(2q^2)$.

Let A be the unique maximum-size sum-free subset of G satisfying $A^i = A$ for all $i \in I$; we claim that $|B \setminus A| \leqslant \varepsilon pn$. Indeed, by the definition of A^i and (*iii*),

$$|B^i| \leqslant |G^i \cap A| + |B^i \setminus A| \leqslant (\mu(G) + 2\varepsilon')pn/M$$

and hence

$$\big(\mu(G) - \varepsilon'\big)pn \;\leqslant\; |B| \;\leqslant\; |I|\Big(\mu(G) + 2\varepsilon'\Big)\frac{pn}{M} + \big(M - |I|\big)\Big(\mu(G) - c_q\Big)\frac{pn}{M},$$

which implies that $|I| \geqslant (1-\varepsilon)M$. We conclude that

$$|B \setminus A| \;\leqslant\; \sum_{i \in I} |B^i \setminus A| + \sum_{i \notin I} |B^i| \;\leqslant\; |I| \cdot \frac{\varepsilon' pn}{M} + \big(M - |I|\big)\mu(G) \cdot \frac{pn}{M} \;\leqslant\; \varepsilon pn,$$

as required. □

In the remainder of this section, we shall sketch the proof of Theorem 3.2. Let G_p be a p-random subset of G. Following [14], we define the associated measure of G_p, denoted μ, by $\mu = p^{-1} \cdot \chi_{G_p}$, i.e., $\mu(x) = p^{-1}$ if $x \in G_p$ and $\mu(x) = 0$ otherwise. Let S be the collection of all Schur triples in G and note that $|S| = \Theta(n^2)$. Moreover, let $\mathcal{V} = \mathrm{SF}_0(G) \cup \{G\}$ and note that (by Corollary 2.4) $|\mathcal{V}| \leqslant n+1$. The transference theorem proved in [14] asserts that a.a.s. for every function $f\colon G \to \mathbb{R}$ with $0 \leqslant f \leqslant \mu$, there is a function $g\colon G \to [0,1]$ such that

$$\mathbb{E}_{s\in S} f(s_1)f(s_2)f(s_3) \geqslant \mathbb{E}_{s\in S} g(s_1)g(s_2)g(s_3) - \varepsilon \tag{2}$$

and

$$\left|\sum_{x\in V} f(x) - \sum_{x\in V} g(x)\right| \leqslant \varepsilon|G| \quad \text{for all } V \in \mathcal{V}. \tag{3}$$

Moreover, we may assume that g takes values only in $\{0,1\}$, i.e., g is the characteristic function of some subset of G; see, e.g., [14, Corollary 9.7]. In particular, if we let f to be p^{-1} times the characteristic function of some subset of G_p, then we will see that for every $B \subseteq G_p$, there exists a $B' \subseteq G$ such that $|B' \cap V| \approx p|B \cap V|$ for every $V \in \mathcal{V}$ and the number of Schur triples in B' is by at most εn^2 larger than p^{-3} times the number of Schur triples in B. Hence, if $B \subseteq G_p$ is sum-free and $|B| \geqslant (\mu(G) - c_q)pn$ for some small positive constant c_q, then the corresponding set B' has at least $(\mu(G) - c_q - \varepsilon)n$ elements and at most εn^2 Schur triples. By Corollary 2.8, there is an $A \in \mathrm{SF}_0(G)$ such that $|B' \setminus A| \leqslant 3\varepsilon^{1/3}n$. Finally, since $A \in \mathcal{V}$, we can conclude that $|B \setminus A| = O(\varepsilon^{1/3}pn)$.

To complete the proof of Theorem 3.2, we still need to argue how the results of Conlon and Gowers [14] imply that a.a.s. every function f with $0 \leqslant f \leqslant \mu$ can be approximated in the above sense by some $g\colon G \to \{0,1\}$. This implication would be a direct consequence of [14, Corollary 9.7] if S, the collection of Schur triples in G, was a so called *good system*, i.e., if S satisfied the following conditions:

- (i) No sequence in S has repeated elements.
- (ii) S is *homogeneous*, i.e., for every $j \in [3]$ and every $x \in G$, the set $\{s \in S\colon s_j = x\}$, denoted $S_j(x)$, has the same size.
- (iii) S has *two degrees of freedom*, i.e., whenever some $s, t \in S$ satisfy $s_i = t_i$ for two distinct indices i, then $s = t$.

Indeed, if S satisfied (i)–(iii), then we could simply apply [14, Corollary 9.7], since for families with two degrees of freedom, the threshold for p required for the transference theorem is $n^{-\gamma/2}$, where γ is defined by $|S_j(x)| = n^\gamma$. Since in our case $\gamma = 1$, the threshold is at $n^{-1/2}$.

Unfortunately, in our setting some sequences in S have repeated elements (Schur triples of the form $(x, x, 2x)$, $(x, 0, x)$, or $(0, x, x)$) and when we remove them, the new set S is no longer homogeneous as, e.g., some $y \in G$ satisfy $y = x + x$ for many different x. To overcome this problem, let

$$D = \big\{y \in G\colon y = x + x \text{ for more than } \sqrt{n} \text{ different } x\big\}.$$

Since there are only n sums of the form $x + x$, it follows that $|D| \leqslant \sqrt{n}$. Next, we let $X = G \setminus (D \cup \{0\})$ and, instead of working with the collection of all Schur triples in G, we define the new set S as follows:

$$S = \big\{(x, y, x+y) \in X^3 \colon x \neq y\big\}.$$

Since $0 \notin X$, it follows that no triple in S has repeated elements. Moreover, for all $j \in [3]$ and every $x \in X$,

$$n - 2\sqrt{n} - 1 \leqslant |S_j(x)| \leqslant n,$$

i.e., the new set S satisfies $|S_j(x)| = (1+o(1))n$ for all $j \in [3]$ and $x \in X$.

Before being able to state the version of the transference theorem that we are actually going to use, we need to do some preparation. Recall that for every $j \in [3]$ and $x \in X$, we defined $S_j(x) = \{s \in S \colon s_j = x\}$. Following [14], given functions $h_1, h_2, h_3 \colon X \to \mathbb{R}$ and $j \in [3]$, we define their *jth convolution* $*_j(h_1, h_2, h_3)$ by

$$\begin{aligned} *_j(h_1,h_2,h_3)(x) &= \frac{|X|}{|S|} \sum_{s \in S_j(x)} h_1(s_1)\ldots h_{j-1}(s_{j-1})h_{j+1}(s_{j+1})h_3(s_3) \\ &= \frac{|X||S_j(x)|}{|S|} \mathbb{E}_{s \in S_j(x)} h_1(s_1)\ldots h_{j-1}(s_{j-1})h_{j+1}(s_{j+1})h_3(s_3). \end{aligned}$$

Moreover, we define an inner product of real-valued functions on X by

$$\langle f, g\rangle = \frac{1}{|X|}\sum_{x \in X} f(x)g(x).$$

A crucial observation is that our (slightly modified in comparison with [14, Definition 3.2]) definition of the convolutions $*_j$ guarantees that for each j,

$$\langle h_j, *_j(h_1,h_2,h_3)\rangle = \frac{1}{|S|}\sum_{s\in S} h_1(s_1)h_2(s_2)h_3(s_3) = \mathbb{E}_{s \in S} h_1(s_1)h_2(s_2)h_3(s_3).$$

This allows us to write

$$\mathbb{E}_{s\in S} f(s_1)f(s_2)f(s_3) - \mathbb{E}_{s \in S} g(s_1)g(s_2)g(s_3) = \sum_{j=1}^{3} \langle f-g, *_j(g,\ldots,g,f,\ldots,f)\rangle \qquad (4)$$

and

$$\sum_{x\in V} f(x) - \sum_{x \in V} g(x) = \langle f-g, \chi_V\rangle \cdot |X| \quad \text{for every } V \subseteq X. \qquad (5)$$

One of the main ideas in [14] is to use (4) and (5) to show that, if the inner products of two functions f and g with elements of some large class of functions on X do not differ much from one another, then f and g will satisfy (2) and (3). Due to various technical complications that would arise in the most straightforward approach, i.e., letting this large class of functions to contain all the convolutions $*_j$, Conlon and Gowers work with so called *capped convolutions* $\circ_j$, defined by $\circ_j(h_1,h_2,h_3) = \min\{*_j(h_1,h_2,h_3), 2\}$. They then define the class of *basic anti-uniform functions*, which are functions of the form $\circ_j(g_1,\ldots,g_j,f_{j+1},\ldots,f_3)$, where $0 \leqslant g_i \leqslant 1$, $0 \leqslant f_i \leqslant \mu$, and μ is the associated measure of a p-random subset coming from some finite sequence, or of the form χ_V for some $V \in \mathcal{V}$ (the characteristic functions are included in order to guarantee that (3) will also hold).

After all these preparations, we can finally state the transference theorem of Conlon and Gowers [14] in the version which is best suited for our needs.

Main assumption. *Let $d, m \in \mathbb{N}$ and $\eta, \lambda > 0$ be given. Let $U_1, \ldots, U_m$ be independent p-random subsets of X, and let $\mu_1, \ldots, \mu_m$ be their associated measures. If $p \geqslant p_0$, then the following properties hold with probability $1 - n^{-C'}$, where C' is a large enough constant:*

0. $\|\mu_i\|_1 = 1 + o(1)$ *for each* $1 \leqslant i \leqslant m$.
1. *If* $1 \leqslant i_1 < i_2 < i_3 \leqslant m$, *and* $j \in [3]$, *then*
$$\| *_j (\mu_{i_1}, \mu_{i_2}, \mu_{i_3}) - \circ_j(\mu_{i_1}, \mu_{i_2}, \mu_{i_3})\|_1 \leqslant \eta.$$
2. *If* $j \in \{2, 3\}$ *and* $1 \leqslant i_{j+1} < \ldots < i_3 \leqslant m$, *then*
$$\| *_j (1, \ldots, 1, \mu_{i_{j+1}}, \ldots, \mu_{i_3})\|_\infty \leqslant 2.$$
3. $|\langle \mu - 1, \xi \rangle| < \lambda$ *if* ξ *is a product of at most* d *basic anti-uniform functions.*

Theorem 3.3 ([14, Theorems 4.10 and 9.3])**.** *Let $\varepsilon > 0$. Let X be a finite set, let S be a collection of ordered subsets of X of size 3, and let $\mathcal{V}$ be a collection of subsets of X. Then there exist constants $C, \eta, \lambda > 0$ and $d, m \in \mathbb{N}$ such that the following holds.*

Let p_0 be such that the main assumption holds for the triple $(S, p_0, \mathcal{V})$ and the constants η, λ, d, and m. Let U be a p-random subset of X, where $Cp_0 \leqslant p \leqslant 1/C$, and let μ be the associated measure of U.

Then, with probability $1 - o(1)$, for every function $f \colon X \to \mathbb{R}$ with $0 \leqslant f \leqslant \mu$, there exists a function $g \colon X \to \{0, 1\}$ such that (2) and (3) hold.

Thus, in order to deduce that the conclusion we require, it suffices to check that the main assumption holds for our choice of X, S, $\mathcal{V}$ and $p_0 = C/\sqrt{n}$, for a large enough constant C. Indeed, Property 0 easily follows from Chernoff's inequality, as it simply says that, with probability at least $1 - n^{-C'}$, a p-random subset of X has $(1 + o(1))p|X|$ elements. Moreover, it is shown in [14] that Property 3 is implied by Properties 0–2, together with the fact that $|\mathcal{V}| \leqslant n + 1 = 2^{o(p|X|)}$, so we only need to argue that our system satisfies Properties 1 and 2.

This is done in [14] in the case when S is a homogeneous system with two degrees of freedom. In our case S does have two degrees of freedom but we only know that it is 'almost homogeneous', i.e., that $|S_j(x)| = (1 + o(1))n$ for every j and x. Fortunately, it is not hard (though somewhat tedious) to check that this is a sufficiently strong assumption to keep the arguments of [14] valid; see [14, Lemma 7.2] and [14, Lemma 8.4] plus the discussion below it.

Finally, let us mention briefly where the lower bound $p \geqslant C/\sqrt{n}$ in Theorem 3.2 comes from. Following [14], for each $x \in X$ let
$$K(x) \,=\, \big\{ y \in X \,:\, S_1(x) \cap S_3(y) \neq \emptyset \big\},$$
and for each $y \in K(x)$, let
$$W(x, y) \,=\, \mathbb{E}_{s \in S_1(x) \cap S_3(y)} \mu(s_2),$$
where μ is the associated measure of some p-random subset of X. A crucial assumption in the proof of Property 2 (see [14, Lemma 8.4]) is that $W(x, y) \ll p|K(x)|$ for every $x, y \in X$.

Now, since our set S has two degrees of freedom, we have $|S_1(x) \cap S_3(y)| \leqslant 1$ for every $x, y \in X$, and therefore $|K(x)| = |S_1(x)| = (1 + o(1))n$. Furthermore,

$$W(x, y) \leqslant \max_{s \in S_1(x) \cap S_3(y)} \mu(s_2) \leqslant \frac{1}{p},$$

so it is enough to require that $p^{-1} \ll p(1 + o(1))n$, which is equivalent to $p \gg n^{-1/2}$.

4. Abelian groups of Type I

In this section we shall prove Theorem 1.2. For the sake of clarity of the argument, from now on, we will assume not only that q divides $|G|$, but also that G is of type $\mathrm{I}(q)$, i.e., that q is the smallest prime q' that divides $|G|$ and satisfies $q' \equiv 2 \pmod 3$. Since for each q, there are at most finitely many primes q' smaller than q that satisfy $q' \equiv 2 \pmod 3$, this assumption clearly does not affect the validity of our argument.

We begin with the 0-statement, i.e., that if

$$\frac{\log n}{n} \ll p(n) \leqslant c_q \sqrt{\frac{\log n}{n}},$$

then with high probability not all maximum-size sum-free subsets of G_p are of the form $A \cap G_p$, with $A \in \mathrm{SF}_0(G)$; in fact we shall prove that *none* of them have this form. The proof uses Janson's inequality and the second moment method.

Remark 4.1. If $\log n/n \ll p(n) \ll n^{-2/3}$, then the 0-statement in Theorem 1.2 becomes almost trivial. Indeed, since G contains at most n^2 Schur triples, then with high probability the set G_p itself is sum-free, and by Chernoff's inequality, $|A \cap G_p| < |G_p|$ for every $A \in \mathrm{SF}_0(G)$.

Proof of the 0-statement in Theorem 1.2. We wish to prove that, for each prime $q \equiv 2 \pmod 3$, if c_q is sufficiently small then the following holds with probability tending to 1 as $n \to \infty$. Let G be an abelian group of type $\mathrm{I}(q)$ with $|G| = n$, let

$$\frac{\log n}{n} \ll p \leqslant c_q \sqrt{\frac{\log n}{n}},$$

and let G_p be a random p-subset of G. Then, for any maximum-size sum-free subset $B \subseteq G_p$, we have $|B| > |A \cap G_p|$ for every $A \in \mathrm{SF}_0(G)$.

The proof will be by the second moment method. To be precise, we shall show that, given $A \in \mathrm{SF}_0(G)$, with high probability there exist at least $10\sqrt{pn \log n}$ elements $x \in G_p$, each chosen from a sum-free subset of a subgroup of G disjoint from A, such that $(A \cap G_p) \cup \{x\}$ is sum-free. It will be easy to bound the expected number of such elements using the FKG inequality; to bound the variance we shall need to calculate more carefully, using Janson's inequality. The result then follows by Chernoff's inequality, since the size of the sets $\{A \cap G_p : A \in \mathrm{SF}_0(G)\}$ is highly concentrated.

To begin, observe that for any $A \in \mathrm{SF}_0(G)$, we have $|A| \geqslant n/3$ (by Theorem 2.1) and

$$\mathbb{P}\Big(\big| |A \cap G_p| - p|A| \big| > 4\sqrt{pn \log n} \Big) \leqslant n^{-4},$$

by Chernoff's inequality (with $a = 4\sqrt{pn \log n}$) and by our lower bound on p. By Corollary 2.4, it follows that, with high probability,

$$\big||A \cap G_p| - p|A|\big| \leqslant 4\sqrt{pn \log n} \quad \text{for every } A \in \mathrm{SF}_0(G). \tag{6}$$

Throughout the rest of the proof, let $A \in \mathrm{SF}_0(G)$ be fixed. By Theorem 2.2, there exists a subgroup H of G of index q such that A is a union of cosets of H and $A \cup (A+A) = G$. Since H is not sum-free (it is a subgroup), it follows that $A \cap H = \emptyset$. Recall that $\mu(H) \geqslant 2/7$, by Theorem 2.1, and let $E \in \mathrm{SF}_0(H)$ be an arbitrary maximum-size sum-free subset of H, so

$$|E| \,=\, \mu(H)|H| \,\geqslant\, \frac{2|G|}{7q}.$$

We shall find in E our elements x such that $(A \cap G_p) \cup \{x\}$ is sum-free. To this end, for each $x \in E$ we define

$$\begin{aligned} C_1(x) &= \Big\{ y \in A \colon x = y + y \Big\}, \\ C_2(x) &= \left\{ \{y, z\} \in \binom{A}{2} \colon x = y + z \right\}, \\ C_3(x) &= \left\{ \{y, z\} \in \binom{A}{2} \colon x = y - z \right\}, \end{aligned}$$

and let $C(x) = C_1(x) \cup C_2(x) \cup C_3(x)$. Note that $|C_2(x)| \leqslant n/2$ and $|C_3(x)| \leqslant n$ for every $x \in E$.

We shall say that an element $x \in E$ is *safe* if no set in $C(x)$ is fully contained in G_p. Thus x is safe if and only if the set $(A \cap G_p) \cup \{x\}$ is sum-free. We shall show below that, with high probability, $E \cap G_p$ contains more than $10\sqrt{pn \log n}$ safe elements. Since E is sum-free, and H is a subgroup, we have $E \pm E \subseteq H \setminus E$. Since $A \cap H = \emptyset$, it follows that the set

$$B \;:=\; \big(A \cap G_p\big) \cup \big\{ x \in E \cap G_p : \; x \text{ is safe} \big\}$$

is sum-free. By (6), it will follow that B is larger than $A' \cap G_p$ for every $A' \in \mathrm{SF}_0(G)$.

We begin by giving a lower bound on the expected number of safe elements in E. In fact, in order to simplify the calculation of the variance, below, we shall focus on a subset $E'' \subseteq E$, defined as follows. First let $E' = \{x \in E \colon |C_1(x)| < 7q\}$, and note that, since the $C_1(x)$ are disjoint subsets of A, we have

$$\big|E \setminus E'\big| \cdot 7q \,\leqslant\, |A| \,\leqslant\, n,$$

and so $|E'| \geqslant |E| - n/(7q) \geqslant n/(7q)$. Now let E'' be an arbitrary subset of E' of cardinality at least $|E'|/2$ such that there are no distinct $x, y \in E''$ with $x = -y$. Finally, let S be the number of safe elements in E''.

For each $x \in E$, denote by S_x the event that x is safe. By the FKG inequality, we have

$$\mathbb{P}(S_x) \geqslant (1-p)^{|C_1(x)|}(1-p^2)^{|C_2(x) \cup C_3(x)|}. \tag{7}$$

Thus, since $p \leqslant c_q\sqrt{\frac{\log n}{n}}$, and using the bounds $|E''| \geqslant n/(14q)$, $|C_1(x)| \leqslant 7q$, $|C_2(x)| \leqslant n/2$, and $|C_3(x)| \leqslant n$, we have

$$\mathbb{E}[S] \,=\, \sum_{x \in E''} \mathbb{P}(S_x) \,\geqslant\, |E''|(1-p)^{7q}(1-p^2)^{3n/2} \,\geqslant\, \frac{n}{20q} \cdot e^{-2p^2 n} \,\geqslant\, \sqrt{n}, \tag{8}$$

where the last inequality holds if c_q is sufficiently small.

The following bound on $\mathrm{Var}(S)$ will allow us to apply Chebychev's inequality.

Claim: $\mathrm{Var}[S] = o(\mathbb{E}[S]^2)$.

Proof of claim. Given distinct elements $x, y \in E$, define a graph $J(x,y)$ on the elements of $C(x) \cup C(y)$ as follows: let $B_1 \sim B_2$ if $B_1 \cap B_2 \neq \emptyset$ and $B_1 \neq B_2$. Now define

$$\Delta_{x,y} = \sum_{B_1 \sim B_2} p^{|B_1 \cup B_2|},$$

where the sum is taken over all edges of $J(x,y)$. For the sake of brevity, let $C_1(x,y) = C_1(x) \cup C_1(y)$ and $C'(x,y) = C_2(x) \cup C_2(y) \cup C_3(x) \cup C_3(y)$. By Janson's inequality,

$$\mathbb{P}(S_x \wedge S_y) \leqslant (1-p)^{|C_1(x,y)|}(1-p^2)^{|C'(x,y)|}e^{\Delta_{x,y}}. \tag{9}$$

We claim that, for each $x, y \in E''$, with $x \neq y$,

$$\Delta_{x,y} \leqslant |C_1(x,y)| \cdot 6p^2 + |C'(x,y)| \cdot 12p^3 = O\big(qp^2 + np^3\big). \tag{10}$$

Indeed, if $B_1 \in C_1(x,y) \cup C'(x,y)$ and $B_2 \in C'(x,y)$ are such that $B_1 \cap B_2 = \{z\}$, then $B_2 = \{z, z'\}$, where

$$z' \in \big\{x - z, z - x, z + x, y - z, z - y, z + y\big\}.$$

Thus, given $B_1 \ni z$, there are at most six sets $B_2 \in C'(x,y)$ such that $B_1 \cap B_2 = \{z\}$. It follows that for every $B_1 \in C_1(x,y)$, there are at most six sets $B_2 \in C'(x,y)$ with $B_1 \cap B_2 \neq \emptyset$ and for every $B_1 \in C'(x,y)$, there are at most twelve sets $B_2 \in C'(x,y)$ with $B_1 \cap B_2 \neq \emptyset$. The second inequality follows since $|C_1(x,y)| = |C_1(x)| + |C_1(y)| < 14q$ and $|C'(x,y)| \leqslant 3n$.

Combining (7), (9), and (10), we obtain

$$\mathbb{P}(S_x \wedge S_y) \,\leqslant\, \mathbb{P}(S_x)\mathbb{P}(S_y)(1-p^2)^{C^*(x,y)}e^{O(qp^2+np^3)}, \tag{11}$$

where $C^*(x,y) = |C'(x,y)| - |C_2(x) \cup C_3(x)| - |C_2(y) \cup C_3(y)|$. Hence, it only remains to bound $C^*(x,y)$ from below.

Recall that E'' does not contain any pairs $\{x,y\}$ with $x = -y$. Hence $C_3(x) \cap C_3(y) = \emptyset$, and trivially $C_2(x) \cap C_2(y) = \emptyset$. It follows by elementary manipulation that

$$|C'(x,y)| \geqslant |C_2(x) \cup C_3(x)| + |C_2(y) \cup C_3(y)| - |C_2(x) \cap C_3(y)| - |C_2(y) \cap C_3(x)|, \tag{12}$$

i.e., $C^*(x,y) \geqslant -|C_2(x) \cap C_3(y)| - |C_2(y) \cap C_3(x)|$.

If $C_2(x) \cap C_3(y) \neq \emptyset$, then there exist $a, b \in A$ such that $x = a + b$ and $y = a - b$, and thus $2a = x + y$ and $2b = x - y$. We split into two cases.

Case 1: $|G|$ is odd.

In this case the equations $2a = x + y$ and $2b = x - y$ have at most one solution (a, b), and so $|C_2(x) \cap C_3(y)| \leqslant 1$. Hence $C^*(x, y) \geqslant -2$, and so, by (11),

$$\mathbb{P}(S_x \wedge S_y) \leqslant \mathbb{P}(S_x)\mathbb{P}(S_y)(1 - p^2)^{-2} e^{O(qp^2 + np^3)}.$$

Thus, using the bounds $p \ll n^{-1/3}$ and $\mathbb{E}[S] \gg 1$,

$$\begin{aligned} \mathrm{Var}[S] &= \sum_{x,y \in E''} \Big(\mathbb{P}(S_x \wedge S_y) - \mathbb{P}(S_x)\mathbb{P}(S_y) \Big) \leqslant \mathbb{E}[S] + \mathbb{E}[S]^2 \Big((1 - p^2)^{-2} e^{O(qp^2 + np^3)} - 1 \Big) \\ &\leqslant \mathbb{E}[S] + O\big(qp^2 + np^3\big)\mathbb{E}[S]^2 = \mathbb{E}[S] + o(\mathbb{E}[S]^2) = o(\mathbb{E}[S]^2), \end{aligned}$$

as required.

Case 2: $|G|$ is even.

Since $|G|$ is even, it follows (by Theorem 2.2) that H is of index 2, and so $A = G \setminus H$. Let

$$I := \big\{ h \in H : h + h = 0 \big\},$$

note that I is a subgroup of H, and let $i = |I|$. We claim that

$$|C_2(x) \cap C_3(y)| = 0 \quad \text{or} \quad i/2 \leqslant |C_2(x) \cap C_3(y)| \leqslant i.$$

To see this, simply observe that if $\{a, b\} \in C_2(x) \cap C_3(y)$ and $h \in I$, then $\{a + h, b + h\} \in C_2(x) \cap C_3(y)$. Indeed, if $x = a+b$ and $y = a-b$ for some $a, b \in A$, then $x = (a+h)+(b+h)$ and $y = (a+h)-(b+h)$, and $a+h, b+h \in A$ since $A = G \setminus H$. Conversely, if $x = a+b = a'+b'$ and $y = a - b = a' - b'$ for some $a, a', b, b' \in A$, then $a' - a = b' - b \in I$. Hence, there are precisely i ordered pairs (a, b) with $\{a, b\} \in C_2(x) \cap C_3(y)$.

Next, note that there are at most $n^2/(2i)$ pairs $\{x, y\} \in \binom{E''}{2}$ with $|C_2(x) \cap C_3(y)| + |C_2(y) \cap C_3(x)| > 0$. Indeed, there are at most $n^2/4$ quadruples (a, b, x, y) in $A^2 \times (E'')^2$ with $x = a + b$ and $y = a - b$, and there are at least $i/2$ such quadruples for each pair $\{x, y\}$ as above. Moreover, for every $x \in E''$, either $|C_3(x)| = 0$ or $|C_3(x)| \geqslant n/4$, since any solution $(a, b) \in A^2$ of the equation $x = a - b$ may be shifted by an arbitrary element of H. Hence, for such an x,

$$\mathbb{P}(S_x \wedge S_y) - \mathbb{P}(S_x)\mathbb{P}(S_y) \leqslant \mathbb{P}(S_x) \leqslant (1 - p^2)^{|C_3(x)|/3} = (1 - p^2)^{n/12} \leqslant e^{-p^2 n/12}, \quad (13)$$

where the second inequality follows because there exists a matching in $C_3(x)$ of size $|C_3(x)|/3$. (This follows because the graph $C_3(x)$ has maximum degree 2; the worst case is a disjoint union of triangles.)

Finally, we divide once again into two cases: $i \leqslant \sqrt{n}$ and $i > \sqrt{n}$. In the former case we have

$$C^*(x, y) \geqslant -|C_2(x) \cap C_3(y)| - |C_2(y) \cap C_3(x)| \geqslant -2\sqrt{n},$$

and so, by (11),

$$\mathbb{P}(S_x \wedge S_y) \leqslant \mathbb{P}(S_x)\mathbb{P}(S_y)(1 - p^2)^{-2\sqrt{n}} e^{O(qp^2 + np^3)} = (1 + o(1))\mathbb{P}(S_x)\mathbb{P}(S_y),$$

since $p^2\sqrt{n} + p^3 n = o(1)$. Hence $\mathrm{Var}[S] = o(\mathbb{E}[S]^2)$, as required.

If $i > \sqrt{n}$, we partition the sum $\sum_{x,y \in E''} \big(\mathbb{P}(S_x \wedge S_y) - \mathbb{P}(S_x)\mathbb{P}(S_y) \big)$ into two parts, according to whether or not $|C_2(x) \cap C_3(y)| + |C_2(y) \cap C_3(x)| > 0$. By (13), we obtain

$$\sum_{x,y \in E''} \Big(\mathbb{P}(S_x \wedge S_y) - \mathbb{P}(S_x)\mathbb{P}(S_y) \Big) \leqslant \mathbb{E}[S] + \mathbb{E}[S]^2 \left(e^{O(qp^2 + np^3)} - 1 \right) + \frac{n^2}{2i} \cdot e^{-p^2 n/12}.$$

Now, recalling from (8) that $\mathbb{E}[S] \geqslant \frac{n}{20q} e^{-2p^2 n}$, and noting that $\frac{1}{\sqrt{n}} e^{-p^2 n/12} \leqslant e^{-4p^2 n}$ if c_q is sufficiently small, we deduce that

$$\mathrm{Var}[S] \,=\, o\Big(\mathbb{E}[S]^2 + n^2 e^{-4p^2 n} \Big) \,=\, o(\mathbb{E}[S]^2),$$

as required. □

Since $\mathrm{Var}[S] = o(\mathbb{E}[S]^2)$, the number of safe elements in E'' satisfies

$$S \,\geqslant\, \frac{\mathbb{E}[S]}{2} \,\geqslant\, \frac{n}{40q} \cdot e^{-2p^2 n},$$

with high probability, by Chebyshev's inequality. Finally, note that the events $\{S_x\}_{x \in E''}$ depend only on the set $A \cap G_p$, and recall that $A \cap E'' = \emptyset$. Hence the events $\{x \in G_p\}_{x \in E''}$ are independent of the events $\{S_x\}_{x \in E''}$, and so, by Chernoff's inequality, the number of safe elements in $E'' \cap G_p$ is at least $pS/2$ with high probability.

Hence

$$\big| \big\{ x \in E'' \cap G_p \colon x \text{ is safe} \big\} \big| \,\geqslant\, \frac{pn}{80q} \cdot e^{-2p^2 n} \,>\, 10\sqrt{pn \log n},$$

where the last inequality follows from the bounds $\frac{\log n}{n} \ll p < c_q \sqrt{\frac{\log n}{n}}$, provided that c_q is sufficiently small and n is sufficiently large. Thus, by (6),

$$B \,:=\, \big(A \cap G_p\big) \cup \big\{ x \in E'' \cap G_p : \, x \text{ is safe} \big\}$$

is larger than $A' \cap G_p$ for every $A' \in \mathrm{SF}_0(G)$, and is sum-free, as required. □

We now turn to the 1-statement in Theorem 1.2. The proof uses Theorem 3.1, together with Janson's inequality.

Proof of the 1*-statement in Theorem 1.2.* Let q be a prime with $q \equiv 2 \pmod 3$. We shall prove that if C_q is sufficiently large, then the following holds with high probability as $n \to \infty$. Let G be an abelian group of type $\mathrm{I}(q)$ with $|G| = n$, let

$$p \,\geqslant\, C_q \sqrt{\frac{\log n}{n}},$$

and let G_p be a random p-subset of G. Then every maximum-size sum-free subset $B \subseteq G_p$ is of the form $A \cap G_p$ for some $A \in \mathrm{SF}_0(G)$.

The proof will be roughly as follows. If a maximum-size sum-free subset $B \subseteq G_p$ is not of the form $A \cap G_p$, where $A \in \mathrm{SF}_0(G)$, then there must exist sets $S \subseteq G_p \setminus A$ and $T \subseteq A \cap G_p$, with $|S| \geqslant |T|$, such that $B = S \cup (A \cap G_p) \setminus T$. We shall show that the expected number of such pairs (S,T) is small when $|S| \leqslant \varepsilon pn$, using Janson's inequality. The case $|B \setminus A| > \varepsilon pn$ for every $A \in \mathrm{SF}_0(G)$ is dealt with using Theorem 3.1.

Let $\varepsilon > 0$ be sufficiently small, and let B be a maximum-size sum-free subset of G_p. Note that $|B| \geqslant |A \cap G_p|$ for any $A \in \mathrm{SF}_0(G)$, and so, by Chernoff's inequality,

$$|B| \;\geqslant\; \left(\mu(G) - \frac{\varepsilon}{40q^2 + 40q}\right) p|G|$$

with high probability as $n \to \infty$. Hence, by Theorem 3.1, a.a.s. we have $|B \setminus A| \leqslant \varepsilon pn$ for some $A \in \mathrm{SF}_0(G)$. We shall prove that in fact, with high probability, $B = A \cap G_p$.

We shall say that a pair of sets (S,T) is *bad* for a set $A \in \mathrm{SF}_0(G)$ if the following conditions hold:

(*a*) $S \subseteq G_p \setminus A$ and $T \subseteq A \cap G_p$,

(*b*) $0 < |S| = |T| \leqslant \varepsilon pn$,

(*c*) $S \cup (A \cap G_p) \setminus T$ is sum-free.

We shall prove that, for every $A \in \mathrm{SF}_0(G)$, the expected number of bad pairs (S,T) is $o(1/n)$ as $n \to \infty$. It will follow (by Corollary 2.4) that with high probability no such pair exists for any $A \in \mathrm{SF}_0(G)$. We remark that a bound of the form $o(1)$ does not suffice here, since the events "$|B \setminus A| \leqslant \varepsilon pn$" and "there exists a pair (S,T) which is bad for A" are not independent of one another.

As in the proof of the lower bound, for every $x \in G \setminus A$, define

$$C_1(x) = \Big\{ y \in A \colon x = y + y \Big\} \quad \text{and} \quad C_2(x) = \left\{ \{y,z\} \in \binom{A}{2} \colon x = y + z \right\}.$$

We begin by proving two easy properties of the sets $C_1(x)$ and $C_2(x)$, which will be useful in what follows.

Claim 1: For every $x \in G \setminus A$, $\max\{|C_1(x)|, |C_2(x)|\} \geqslant n/(3q)$.

Proof. By Theorem 2.2, there exists a subgroup H of G of index q such that A is a union of cosets of H and $A \cup (A + A) = G$. It follows that $x = y + z$ for some $y, z \in A$, and that $y + h, z - h \in A$ for every $h \in H$. Thus $\{y + h, z - h\} \in C_1(x) \cup C_2(x)$ for every $h \in H$, and hence $|C_1(x)| + 2|C_2(x)| \geqslant |H| = n/q$. □

Let $A^* = \{x \in G \setminus A \colon |C_1(x)| \geqslant n/(3q)\}$.

Claim 2: With probability $1 - o(1/n)$, there is no bad pair (S,T) with $S \cap A^* \neq \emptyset$.

Proof. Let $x \in A^*$, so $x \in G \setminus A$ and $|C_1(x)| \geqslant n/(3q)$. If (S,T) is a bad pair for A with $x \in S$, then $C_1(x) \cap G_p \subseteq T$, since $S \cup (A \cap G_p) \setminus T$ is sum-free, and $|T| \leqslant \varepsilon pn$. But by Chernoff's inequality, we have

$$|C_1(x) \cap G_p| \;\geqslant\; \frac{pn}{6q} \;>\; \varepsilon pn \;\geqslant\; |T|$$

with probability at least $1 - \exp\left(-\frac{pn}{24q}\right) \geqslant 1 - n^{-3}$. By the union bound, with probability $1 - o(1/n)$ this holds for every $x \in A^*$, and hence with probability $1 - o(1/n)$ there is no bad pair (S,T) with $S \cap A^* \neq \emptyset$. □

We now arrive at an important definition. Given a set $S \subseteq G \setminus (A \cup A^*)$ of size k, let $\mathcal{G}_S$ denote the graph with vertex set A and edge set $\bigcup_{x \in S} C_2(x)$. The key observation about $\mathcal{G}_S$ is as follows: if (S,T) is a bad pair for A, then $A \cap G_p \setminus T$ is an independent set in $\mathcal{G}_S$. This follows because the endpoints of an edge of $\mathcal{G}_S$ sum to an element of S.

Recall that, by Claims 1 and 2 and the definition of A^*, we may assume that $|C_2(x)| \geqslant n/(3q)$ for every $x \in S$. Since, by definition, $C_2(x) \cap C_2(x') = \emptyset$ if $x \neq x'$, we have $e(\mathcal{G}_S) \geqslant kn/(3q)$. Moreover, each $C_2(x)$ is a matching (i.e., no two edges share an endpoint) and hence $\Delta(\mathcal{G}_S) \leqslant |S| = k$.

Given a subset $T \subseteq A$ of size k, let $\mathcal{G}_{S,T} = \mathcal{G}_S[A \setminus T]$. Note that $e(\mathcal{G}_{S,T}) \geqslant kn/(3q) - k^2$ and $\Delta(\mathcal{G}_{S,T}) \leqslant k$ for every such T. Also if (S,T) is a bad pair for A, then $e(\mathcal{G}_{S,T}[G_p]) = 0$. Since the sets S, T, and $A \setminus T$ are pairwise disjoint, the events $S \subseteq G_p$, $T \subseteq G_p$, and $e(\mathcal{G}_{S,T}[G_p]) = 0$ are independent. Hence, the expected number of bad pairs for A is at most

$$\sum_{S,T} \mathbb{P}\big(S \subseteq G_p\big)\, \mathbb{P}\big(T \subseteq G_p\big)\, \mathbb{P}\big(e(\mathcal{G}_{S,T}[G_p]) = 0\big), \tag{14}$$

where the summation ranges over all $S \subseteq G \setminus (A \cup A^*)$ and $T \subseteq A$ with $1 \leqslant |S| = |T| \leqslant \varepsilon pn$.

Fix a k with $1 \leqslant k \leqslant \varepsilon pn$ and let $S \subseteq G \setminus (A \cup A^*)$ and $T \subseteq A$ be sets of size k. Let $\mu = \mathbb{E}[e(\mathcal{G}_{S,T}[G_p])]$ and observe that

$$\mu \;=\; p^2 e\big(\mathcal{G}_{S,T}\big) \;\geqslant\; p^2 \left(\frac{kn}{3q} - k^2 \right) \;\geqslant\; \frac{p^2 kn}{6q}. \tag{15}$$

since $k \leqslant \varepsilon n$ and ε is sufficiently small. Furthermore, let $\Delta = \sum_{B_1 \sim B_2} p^{|B_1 \cup B_2|}$, where the summation ranges over all $B_1, B_2 \in E(\mathcal{G}_{S,T})$ such that $B_1 \neq B_2$ and $B_1 \cap B_2 \neq \emptyset$ (in other words, B_1 and B_2 are two edges of $\mathcal{G}_{S,T}$ sharing an endpoint), and note that

$$\Delta \;\leqslant\; n \binom{\Delta(\mathcal{G}_{S,T})}{2} p^3 \;\leqslant\; p^3 k^2 n. \tag{16}$$

By Janson's inequality, if $\Delta \leqslant \mu$, then (15) and our assumption on the growth rate of p imply that

$$\mathbb{P}\big(e(\mathcal{G}_{S,T}[G_p]) = 0\big) \;\leqslant\; e^{-\mu/2} \;\leqslant\; \Big(e^{-p^2 n/(12q)}\Big)^k \;\leqslant\; n^{-5k},$$

and if $\Delta \geqslant \mu$, then (16) implies that

$$\mathbb{P}\big(e(\mathcal{G}_{S,T}[G_p]) = 0\big) \;\leqslant\; e^{-\mu^2/(2\Delta)} \;\leqslant\; e^{-pn/(72q^2)}.$$

Hence, using (14), for those k such that $\Delta \leqslant \mu$, the expected number of bad pairs (S,T) with $|S| = |T| = k$ is at most

$$\binom{n}{k}^2 p^{2k} n^{-5k} \;\leqslant\; n^{-3k} \;\leqslant\; n^{-3}.$$

Also, recalling that $k \leqslant \varepsilon pn$, for those k such that $\Delta > \mu$, the expected number of bad pairs (S,T) with $|S| = |T| = k$ is at most

$$\binom{n}{k}^2 p^{2k} e^{-\frac{pn}{72q^2}} \;\leqslant\; \left(\frac{epn}{k}\right)^{2k} e^{-\frac{pn}{72q^2}} \;\leqslant\; \left[\left(\frac{e}{\varepsilon}\right)^{2\varepsilon} e^{-\frac{1}{72q^2}} \right]^{pn} \;\leqslant\; e^{-\sqrt{n}}$$

whenever ε is sufficiently small and n is sufficiently large (depending only on q). It follows that (14) can be bounded by $\varepsilon pn \cdot \max\{n^{-3}, e^{-\sqrt{n}}\}$, and so this is an upper bound on the probability that there exists a bad pair (S,T) for A. Since $|\,\mathrm{SF}_0(G)| \leqslant n$, the expected number of pairs (S,T) which are bad for some $A \in \mathrm{SF}_0(G)$ tends to 0 as $n \to \infty$, and so this completes the proof. □

5. Lower bounds for other abelian groups

In this section we shall prove the following three propositions, which show that the threshold for some groups can be much larger than that determined in Theorems 1.1 and 1.2. The first of these results shows that for certain groups of Type II, the threshold is at least $(n \log n)^{-1/3}$. Recall that G_p denotes a p-random subset of G.

Proposition 5.1. *Let q be a prime with $q \equiv 1 \pmod 3$, let $n = 3q$, and let $G = \mathbb{Z}_n$. If*

$$\frac{\log n}{n} \ll p(n) \ll \left(\frac{1}{n \log n}\right)^{1/3},$$

then, with high probability as $n \to \infty$, there exists a sum-free subset of G_p which is larger than $A \cap G_p$ for every $A \in \mathrm{SF}_0(G)$.

The second result gives the same bounds for groups of Type I of prime order.

Proposition 5.2. *Let q be a prime with $q \equiv 2 \pmod 3$, let $n = q$, and let $G = \mathbb{Z}_n$. If*

$$\frac{\log n}{n} \ll p(n) \ll \left(\frac{1}{n \log n}\right)^{1/3},$$

then, with high probability as $n \to \infty$, there exists a sum-free subset of G_p which is larger than $A \cap G_p$ for every $A \in \mathrm{SF}_0(G)$.

The third proposition shows that for the hypercube $\{0,1\}^k$ on $2n$ vertices, the threshold is different from the threshold for $\mathbb{Z}_{2n}$.

Proposition 5.3. *Let $C < 1/2$ and $k \in \mathbb{N}$, let $n = 2^{k-1}$, and let $G = \mathbb{Z}_2^k$. If*

$$\frac{\log n}{n} \ll p \leqslant \sqrt{\frac{C \log n}{n}},$$

then, with high probability as $n \to \infty$, there exists a sum-free subset of G_p which is larger than $A \cap G_p$ for every $A \in \mathrm{SF}_0(G)$.

The proofs of Propositions 5.1 and 5.2 are almost identical, and are based on the following general statement providing a lower bound for the size of a largest sum-free subset of a p-random subset of $\mathbb{Z}_n$.

Lemma 5.4. *Let $n \in \mathbb{N}$, and let $G = \mathbb{Z}_n$. Let $m = \min\{n, p^{-2}\}/100$. If*

$$\frac{\log n}{n} \ll p(n) = o(1),$$

then, with high probability, the largest sum-free subset of G_p has at least

$$\frac{pn}{3} + \frac{pm}{4} - 4\sqrt{pn \log n}$$

elements.

Proof. The idea is as follows: first we construct a sum-free set $A \subseteq G$ of size about $n/3 - 2m$; then we consider $A \cap G_p$, and observe that a.a.s. it has at least $pn/3 - 2pm - 4\sqrt{pn \log n}$ elements; finally we show that, with high probability, we can add $9pm/4$ elements to $A \cap G_p$ while still remaining sum-free.

We begin by constructing $A = \{\ell, \dots, r\}$, where $\ell = \lceil n/3 \rceil + \lceil 4m \rceil + 1$ and $r = \lfloor 2n/3 \rfloor + \lfloor 2m \rfloor$. Observe that A is sum-free, since $2\ell > r$ and $2r - n < \ell$, and that

$$\frac{n}{4} \leqslant \frac{n}{3} - 2m - 3 \leqslant |A| \leqslant \frac{n}{3} - 2m.$$

By Chernoff's inequality (applied with $a = 3\sqrt{pn \log n}$), and our lower bound on p, we have

$$\left| |A \cap G_p| - \left(\frac{pn}{3} - 2pm \right) \right| \leqslant 4\sqrt{pn \log n} \tag{17}$$

with high probability.

Next, let $A' = \{\ell', \dots, \ell - 1\}$, and $A'' = \{r', \dots, r\} \subseteq A$, where $\ell' = \lceil n/3 \rceil + \lceil m \rceil + 1$ and $r' = \lfloor 2n/3 \rfloor - \lceil m \rceil$. Since A is sum-free, $2\ell' > r$ and $(r' - 1) + r - n < \ell'$, it follows that every Schur triple (x, y, z) in $A \cup A'$ satisfies $x, y \in A''$ and $z \in A'$.

For every $x \in A'$, let

$$C_1(x) = \Big\{ y \in A'' \colon x = y + y \Big\}, \quad C_2(x) = \left\{ \{y, z\} \in \binom{A''}{2} \colon x = y + z \right\},$$

and $C(x) = C_1(x) \cup C_2(x)$. Moreover, note that $|C_1(x)| \leqslant 1$ and $|C_2(x)| \leqslant |A''|/2 \leqslant 2m$ for every $x \in A'$. Call an $x \in A'$ *safe* if no $B \in C(x)$ is fully contained in G_p. By the above observation about the Schur triples in $A \cup A'$, the set

$$\big(A \cap G_p \big) \cup \big\{ x \in A' \colon x \text{ is safe} \big\}$$

is sum-free.

In the remainder of the proof, we will show that a.a.s. $A' \cap G_p$ contains at least $9pm/4$ safe elements. Together with (17), this will imply that the size of a largest-sum free subset of G_p is at least $pn/3 + pm/4 - 4\sqrt{pn \log n}$.

For each $x \in A'$, denote by S_x the event that x is safe, and let S be the number of safe elements in A'. As in the proof of the lower bound in Theorem 1.2, by the FKG inequality we have

$$\mathbb{E}[S] = \sum_{x \in A'} \mathbb{P}(S_x) \geqslant \sum_{x \in A'} (1 - p)^{|C_1(x)|} (1 - p^2)^{|C_2(x)|} \geqslant |A'| e^{-3p^2 m} \geqslant \frac{11m}{4}.$$

Here we used the bounds $|C_1(x)| \leqslant 1$, $|C_2(x)| \leqslant 2m$, $|A'| \geqslant 3m-2$ and $p^2 m \leqslant 1/100$. Similarly, using Janson's inequality, we obtain

$$\begin{aligned}\mathrm{Var}[S] &= \sum_{x,y\in A'} \Big(\mathbb{P}(S_x \wedge S_y) - \mathbb{P}(S_x)\mathbb{P}(S_y)\Big) \leqslant \mathbb{E}[S] + \mathbb{E}[S]^2 \left(e^{O(p^2+p^3 m)} - 1\right) \\ &\leqslant \mathbb{E}[S] + O\big(p^2 + p^3 m\big)\cdot \mathbb{E}[S]^2 = o\big(\mathbb{E}[S]^2\big),\end{aligned}$$

To see this, observe that (9), (10), and (11) still hold (with n replaced by $2m$), and that $C^*(x,y) = 0$. The last inequality follows from the fact that $p^2 + p^3 m \ll 1$.

By Chebyshev's inequality,

$$\mathbb{P}\left(\big|S - \mathbb{E}[S]\big| > \frac{\mathbb{E}[S]}{11}\right) \leqslant \frac{121\,\mathrm{Var}[S]}{\mathbb{E}[S]^2} = o(1),$$

and it follows that, with high probability, the number of safe elements in A' satisfies

$$S \geqslant \frac{10\mathbb{E}[S]}{11} \geqslant \frac{5m}{2}.$$

Finally, note that for every $x \in A'$, the event S_x is independent of the events $\{y \in G_p\}_{y\in A'}$. Hence, by Chernoff's inequality, with high probability the number of safe elements in $A' \cap G_p$ satisfies

$$\big|\{x \in A' \cap G_p \colon x \text{ is safe}\}\big| \geqslant \frac{9pS}{10} \geqslant \frac{9pm}{4},$$

as required. □

Propositions 5.1 and 5.2 both follow easily from Lemma 5.4; since the proofs are almost identical, we shall prove only the former, and leave the details of the latter to the reader.

Proof of Proposition 5.1. Fix an $A \in \mathrm{SF}_0(G)$ and recall that $|A| = n/3$ by Theorem 2.1. Thus, by Chernoff's inequality (with $a = 4\sqrt{pn\log n}$) and our lower bound on p, we have

$$\mathbb{P}\left(\left||A\cap G_p| - \frac{pn}{3}\right| > 4\sqrt{pn\log n}\right) \leqslant \frac{1}{n^4}.$$

Hence, by Corollary 2.5 (Corollary 2.4 in the proof of Proposition 5.2), with high probability,

$$|A \cap G_p| \leqslant \frac{pn}{3} + 4\sqrt{pn\log n} \quad \text{for every } A \in \mathrm{SF}_0(G). \tag{18}$$

Now, by Theorem 5.4, with high probability G_p contains a sum-free subset with at least

$$\frac{pn}{3} + \frac{pm}{4} - 4\sqrt{pn\log n}$$

elements, where $m = \min\{n, p^{-2}\}/100$.

Finally, observe that $pm \gg \sqrt{pn\log n}$, since if $m = n/100$, then this is equivalent to $pn \gg \log n$ and if $m = 1/(100p^2)$, then it is equivalent to $p^3 n \log n \ll 1$. Hence, by (18), there exists a sum-free subset of G_p larger than $A \cap G_p$ for every $A \in \mathrm{SF}_0(G)$. □

We shall now prove Proposition 5.3. Note that for the hypercube, the conditions $x = y+z$ and $x = y-z$ are the same, and so we have fewer restrictions for a vertex to be safe. This allows us to show that the threshold is different in this case.

Proof of Proposition 5.3. For each non-zero element $a \in G = \mathbb{Z}_2^k$, we shall denote the even and odd cosets of the subgroup (subspace) orthogonal to a by $E(a)$ and $O(a)$, respectively. Note that $n = |E(a)| = |O(a)|$, that $O(a)$ is sum-free, and that, by Theorem 2.2, every element of $\mathrm{SF}_0(G)$ is of the form $O(a)$ for some non-zero $a \in G$. By Chernoff's inequality, for every $a \in G \setminus \{0\}$,

$$\mathbb{P}\Big(\big| |O(a) \cap G_p| - pn \big| > 4\sqrt{pn \log n} \Big) \leqslant \frac{1}{n^2},$$

and hence, with high probability,

$$\big| |O(a) \cap G_p| - pn \big| \leqslant 4\sqrt{pn \log n} \quad \text{for every } a \in G \setminus \{0\}.$$

Let $\mathbf{1} \in G$ be the all-ones vector. To simplify the notation, let $E = E(\mathbf{1})$ and $O = O(\mathbf{1})$. Since E is isomorphic to $\mathbb{Z}_2^{k-1}$, E contains a sum-free subset of size 2^{k-2}. Choose one such subset and denote it E'. Note that $0 \notin E'$.

For each $x \in E'$, let

$$C(x) = \left\{ \{y, z\} \in \binom{O}{2} : x = y + z \right\},$$

and note that $|C(x)| = n/2$. Call an $x \in E'$ *safe* if no $B \in C(x)$ is fully contained in G_p, and for every $x \in E'$, denote by S_x the event that x is safe. Let $\varepsilon = 1/4 - C/2 > 0$, let

$$\mu := \frac{np^2}{2} \leqslant \frac{C \log n}{2} = \left(\frac{1}{4} - \varepsilon \right) \log n,$$

and observe that, by the FKG inequality,

$$\mathbb{P}(S_x) \geqslant (1 - p^2)^{n/2} \geqslant \exp\Big(-\mu - O\big(np^4\big) \Big) \geqslant \frac{n^{-1/4+\varepsilon}}{2},$$

since $np^4 \ll 1$. Let S be the number of safe elements in E' and note that

$$\mathbb{E}[S] = \sum_{x \in E'} \mathbb{P}(S_x) \geqslant \frac{n^{3/4+\varepsilon}}{4}.$$

Since for two distinct $x, y \in E'$, the sets $C(x)$ and $C(y)$ are disjoint, it follows from Janson's inequality that

$$\mathbb{P}(S_x \wedge S_y) \leqslant (1 - p^2)^{|C(x) \cup C(y)|} e^{np^3} = (1 + o(1)) \mathbb{P}(S_x) \mathbb{P}(S_y)$$

and hence $\mathrm{Var}[S] = o(\mathbb{E}[S]^2)$. By the inequalities of Chebyshev and Chernoff (as in the proof of Lemma 5.4), with high probability the set

$$\big(O \cap G_p\big) \cup \big\{ x \in E' \cap G_p \colon x \text{ is safe} \big\}$$

is sum-free and larger than $O(a) \cap G_p$ for every non-zero $a \in G$. □

6. The group $\mathbb{Z}_{2n}$

Let E_{2n} and O_{2n} denote the even and odd cosets of $\mathbb{Z}_{2n}$, respectively, and recall that O_{2n} is the unique maximum-size sum-free subset of $\mathbb{Z}_{2n}$. Throughout this section, let $G = \mathbb{Z}_{2n}$, and recall that G_p denotes a p-random subset of G. We shall prove Theorem 1.1, which gives a sharp threshold for the property that $\mathrm{SF}_0(G_p) = \{G_p \cap O_{2n}\}$.

For each $x \in E_{2n} \cap [n-1]$, let

$$C_1(x) = \Big\{ y \in O_{2n} \colon x = y + y \Big\},$$

$$C_2(x) = \left\{ \{y, z\} \in \binom{O_{2n}}{2} \colon x = y + z \right\},$$

$$C_3(x) = \left\{ \{y, z\} \in \binom{O_{2n}}{2} \colon x = y - z \right\},$$

and let $C(x) = C_1(x) \cup C_2(x) \cup C_3(x)$. Call an element $x \in E_{2n}$ *safe* if no $B \in C(x)$ is fully contained in G_p, and observe that x is safe if and only if $(G_p \cap O_{2n}) \cup \{x\}$ is sum-free.

We begin by proving the 0-statement; even computing the precise value, it is somewhat simpler than in the general case.

Proof of the 0*-statement in Theorem 1.1.* Let

$$\frac{\log n}{n} \ll p \leqslant \sqrt{\frac{\log n}{3n}},$$

and let G_p be a p-random subset of $G = \mathbb{Z}_{2n}$. We shall prove that, with high probability, $G_p \cap E_{2n}$ contains a safe element, and hence there exists a sum-free subset $B \subseteq G_p$ larger than $G_p \cap O_{2n}$.

Indeed, for each $x \in E_{2n} \cap [n-1]$, denote by S_x the event that x is safe, and let S denote the number of safe elements in $E_{2n} \cap [n-1]$. Note that $|C_1(x)| \leqslant 2$, $\frac{n}{2} - 1 \leqslant |C_2(x)| \leqslant \frac{n}{2}$, and $|C_3(x)| = n$. By the FKG inequality,

$$\mathbb{P}(S_x) \geqslant (1-p)^{|C_1(x)|}(1-p^2)^{|C_2(x) \cup C_3(x)|} \geqslant \exp\left(-\frac{3p^2 n}{2} + O\big(p + p^4 n\big) \right), \tag{19}$$

and so, since $3p^2 n \leqslant \log n$,

$$\mathbb{E}[S] \geqslant \sum_{x \in E_{2n} \cap [n-1]} \mathbb{P}(S_x) \geqslant \frac{n}{3} \cdot e^{-3p^2 n/2} \geqslant \frac{1}{3}\sqrt{n}.$$

The calculation of the variance is similar to those in the proofs of Theorem 1.2 and Lemma 5.4, and so we omit the details, noting only that $C_2(x) \cap C_2(y) = C_3(x) \cap C_3(y) = \emptyset$ and $|C_2(x) \cap C_3(y)| \leqslant 4$ if $x \neq -y$. We obtain

$$\begin{aligned} \mathrm{Var}[S] &= \sum_{x,y \in E_{2n} \cap [n-1]} \Big(\mathbb{P}(S_x \wedge S_y) - \mathbb{P}(S_x)\mathbb{P}(S_y) \Big) \leqslant \mathbb{E}[S] + \mathbb{E}[S]^2 \left(e^{O(p^2 + np^3)} - 1 \right) \\ &\leqslant \mathbb{E}[S] + O\big(p^2 + np^3\big)\mathbb{E}[S]^2 = o\big(\mathbb{E}[S]^2\big), \end{aligned}$$

since $p^3 n = o(1)$. Hence, by Chebyshev's inequality, with high probability the number of safe elements satisfies

$$S \geqslant \frac{\mathbb{E}[S]}{2} \geqslant \frac{1}{6}\sqrt{n}.$$

Finally, for each $x \in E_{2n}$, the event $x \in G_p$ is independent of all the events $\{S_x\}_{x \in E_{2n}}$, so the probability that no safe element belongs to G_p (given $|S| \geqslant \sqrt{n}/6$) is at most

$$(1-p)^{\sqrt{n}/6} \leqslant e^{-(\log n)/100} = o(1)$$

as required. □

In order to obtain the 1-statement, we shall have to work quite a bit harder. To show that all sum-free sets B with at least εpn even numbers are smaller than $G_p \cap O_{2n}$, we shall use the Conlon-Gowers method (Theorem 3.1), as in the proof of Theorem 1.2. When $|B \cap E_{2n}| = o(pn)$, however, we shall need to study carefully the structure of the graph $\mathcal{G}_S$ for each $S \subseteq E_{2n}$, where $V(\mathcal{G}_S) = O_{2n}$ and $E(\mathcal{G}_S)$ consists of all pairs $\{a, b\}$ such that either $a + b \in S$, or $a - b \in S$.

Proof of the 1-statement in Theorem 1.1. Let $\delta > 0$, let $C = C(n) \geqslant \dfrac{1}{3} + \delta$, let

$$p = p(n) = \sqrt{\frac{C \log n}{n}},$$

and let G_p be a p-random subset of $G = \mathbb{Z}_{2n}$. We shall prove that, with high probability, $G_p \cap O_{2n}$ is the unique maximum-size sum-free subset of G_p.

Let $\varepsilon > 0$ be sufficiently small, and let B be a maximum-size sum-free subset of G_p. Note that $|B| \geqslant |G_p \cap O_{2n}|$, and so, by Chernoff's inequality,

$$|B| \geqslant \left(\frac{1}{2} - \varepsilon^2\right) p|G|$$

with high probability as $n \to \infty$. Hence, by Theorem 3.1, we have $|B \setminus O_{2n}| \leqslant \varepsilon pn$ with high probability.

Let $S = B \cap E_{2n}$, and suppose that $|S| = k$ for some positive k. Since B is at least as large as $G_p \cap O_{2n}$, there must exist a set $T \subseteq G_p \cap O_{2n}$, with $|T| \leqslant k$, such that

$$B = S \cup \big(G_p \cap O_{2n}\big) \setminus T.$$

We shall bound the expected number of such pairs (S, T), with $|S| = k$ and T minimal.

For each set $S \subseteq E_{2n}$, define $\mathcal{G}_S$ to be the graph with vertex set O_{2n} whose edges are all pairs $\{a, b\} \in \binom{O_{2n}}{2}$, such that either $a + b \in S$, or $a - b \in S$. Note that $0 \notin B$, and that we ignore loops[1]. We say that a pair of sets (S, T) is *good* if the following conditions hold:

- (i) $S \subseteq G_p \cap E_{2n}$, with $|S| = k$,
- (ii) $T \subseteq G_p \cap O_{2n}$, with $|T| \leqslant k$,
- (iii) $G_p \setminus T$ is an independent set in $\mathcal{G}_S$,

[1]Hence we prove a slightly stronger result – even if we allow sums of the form $a + a = 2a$ to exist in a sum-free set, the odd numbers are still the best.

(iv) $G_p \setminus T'$ is not independent for every $T' \subsetneq T$.

It is a simple (but key) observation that if $B = S \cup (G_p \cap O_{2n}) \setminus T$ is a maximum-size sum-free subset of G_p with $|T| \leqslant |S| = k$, then (S, T') is a good pair for some $T' \subseteq T$. Indeed, every edge of $\mathcal{G}_S[G_p \cap O_{2n}]$ must have an endpoint in T, since $S \cup (G_p \cap O_{2n}) \setminus T$ is sum-free. Now take $T' \subseteq T$ to be minimal such that this holds.

Let m denote the number of pairs $\{x, 2n - x\} \subseteq S$, where $x \in E_{2n} \cap [n-1]$, and let $\mathbf{1}_S(x)$ denote the indicator function of the event $x \in S$.

Claim 1: $e(\mathcal{G}_S) \geqslant \left(\dfrac{3k - \mathbf{1}_S(n)}{2} - m \right) n - O\big(k^2\big)$, and $\Delta(\mathcal{G}_S) \leqslant 3k$.

Proof of claim. Each $x \in S$ (other than $x = n$, which contributes $n/2$) contributes n edges $\{a, b\}$ with $a - b = x$ to $\mathcal{G}_S$, with two edges incident to each vertex of O_{2n}. The edge sets corresponding to different members of S are disjoint, except for those corresponding to x and $2n - x$, which are identical. Thus the pairs $\{a, b\}$ with $a - b \in S$ contribute at least $\big(k - m - \frac{1}{2}\mathbf{1}_S(n)\big)n$ edges to $\mathcal{G}_S$, and there are at most $2(k - m) \leqslant 2k$ such edges incident to each vertex.

Now consider the pairs $\{a, b\}$ with $a + b \in S$. Each $x \in S$ contributes at least $(n-2)/2$ such edges (since there are at most two loops, at $x/2$ and $n + x/2$), and these sets of edges are disjoint for different members of S. Moreover, each vertex is incident to at most k such edges, and so $\Delta(\mathcal{G}_S) \leqslant 3k$.

Finally, if for some $\{a, b\}$, both $a + b$ and $a - b$ are in S, then $\{a, b\} = \{x + y, x - y\}$ or $\{a, b\} = \{n + x + y, n + x - y\}$, where $2x, 2y \in S$. There are at most k^2 such pairs $\{2x, 2y\}$, and so the number of pairs $\{a, b\}$ with $a + b \in S$ and $a - b \notin S$ is at least $nk/2 - 3k^2$. Hence the total number of edges is at least

$$\left(k - m - \frac{\mathbf{1}_S(n)}{2} \right) n + \frac{nk}{2} - 3k^2,$$

as required. □

The following idea is key:

Claim 2: If a pair (S, T) is good, then there exists a subset $U \subseteq (G_p \cap O_{2n}) \setminus T$ with $|U| \leqslant |T|$, such that $T \subseteq N_{\mathcal{G}_S}(U)$ and in $\mathcal{G}_S$ there is matching of size $|U|$ from U to T.

Proof of claim. To see this, we simply take a maximal matching M from T to $G_p \setminus T$ in $\mathcal{G}_S$ and let U be the set of vertices in $G_p \setminus T$ that are incident to M. By construction, $|U| = |M| \leqslant |T|$ and M is a matching of size $|U|$ from U to T.

It remains to prove that $T \subseteq N_{\mathcal{G}_S}(U)$. Suppose not, i.e., assume that there is a vertex $a \in T \setminus N_{\mathcal{G}_S}(U)$. Since $G_p \setminus T$ is an independent set in $\mathcal{G}_S$, and by the minimality of T, it follows that a has a neighbor $b \in G_p \setminus T$. But $a \notin N_{\mathcal{G}_S}(U)$, so $b \notin U$, and thus $M \cup \{a, b\}$ is a matching from T to $G_p \setminus T$ which is larger than M. This contradicts the choice of M, and so $T \subseteq N_{\mathcal{G}_S}(U)$ as claimed. □

The plan for the rest of the proof is now clear. Say that a triple (S, T, U) is good if (S, T) is good, $U \subseteq (G_p \cap O_{2n}) \setminus T$ and $T \subseteq N_{\mathcal{G}_S}(U)$. Let $Z' = Z'(k, \ell, m, j)$ denote the number of good triples (S,T,U) with $|S| = k$, $|T| = \ell$, $|U| = j$, and with m pairs $\{x, 2n-x\}$ in S, and let

$$Z \;:=\; \sum_{k=1}^{n} \sum_{m=0}^{k/2} \sum_{\ell=0}^{k} \sum_{j=0}^{\ell} Z'(k,\ell,m,j).$$

By Claim 2 and the observations above, if there exists a maximum-size sum-free set $B \neq G_p \cap O_{2n}$, then $Z \geqslant 1$, and hence

$$\mathbb{P}\Big(\mathrm{SF}_0(G_p) \neq \{G_p \cap O_{2n}\} \Big) \;\leqslant\; \mathbb{E}\big[Z\big] \;=\; \sum_{k,\ell,m,j} \mathbb{E}\big[Z'(k,\ell,m,j)\big]. \tag{20}$$

It thus will suffice to bound $\mathbb{E}[Z'(k,\ell,m,j)]$ for each k, ℓ, m and j. We shall prove that $\mathbb{E}[Z'(k,\ell,m,j)] \leqslant n^{-\varepsilon k}$ if $kp \leqslant 1$, and $\mathbb{E}[Z'(k,\ell,m,j)] \leqslant e^{-\sqrt{n}}$ otherwise.

Let us fix k, m, ℓ and j, and count the triples (S,T,U) which contribute to $Z'(k,\ell,m,j)$. There are at most $3^k\binom{n}{k-m}$ choices for the set S if $n \not\in S$ (since S intersects $k-m$ of the pairs $\{x, 2n-x\}$), and similarly there are at most $3^k\binom{n}{k-m-1}$ choices for S if $n \in S$. Regardless of whether $n \in S$ or $n \not\in S$, there are at most $\binom{n}{k}$ choices for S with $|S| = k$.

Now, for each $S \subseteq E_{2n}$ and $\ell, j \in \mathbb{N}$, let $W(S,\ell,j)$ denote the number of pairs (T,U) such that $T, U \subseteq G_p \cap O_{2n}$, $T \cap U = \emptyset$ and $T \subseteq N_{\mathcal{G}_S}(U)$, with $|T| = \ell$ and $|U| = j$.

Claim 3: If $|S| = k$ and $0 \leqslant j \leqslant \ell \leqslant k \leqslant \varepsilon p n$, then

$$\mathbb{E}\big[W(S,\ell,j)\big] \;\leqslant\; (3e^2p^2n)^k \;\ll\; (C\log n)^{2k}. \tag{21}$$

Proof of claim. We have at most $\binom{n}{j}$ choices for U, and, given S and U, there are at most $\binom{3kj}{\ell}$ choices for T, since $T \subseteq N_{\mathcal{G}_S}(U)$ and $\Delta(\mathcal{G}_S) \leqslant 3k$. Since T and U are disjoint, the probability that T and U are contained in G_p is $p^{\ell+j}$.

To simplify the computation, note that for fixed ℓ and k with $0 \leqslant \ell \leqslant k \leqslant \varepsilon p n$, the functions $\binom{n}{j}p^j$ and $\binom{3kj}{\ell}$ are increasing in j if $0 \leqslant j \leqslant \ell$. Therefore,

$$\mathbb{E}\big[W(S,\ell,j)\big] \;\leqslant\; p^{\ell+j}\binom{n}{j}\binom{3kj}{\ell} \;\leqslant\; p^{2\ell}\binom{n}{\ell}\binom{3k\ell}{\ell} \;\leqslant\; \left(\frac{3e^2p^2nk}{\ell}\right)^{\ell}. \tag{22}$$

Since, for any $a > 0$, the function $x \mapsto \left(\frac{a}{x}\right)^x$ is increasing for $0 \leqslant x \leqslant \frac{a}{e}$, and $3e^2p^2n \gg 1$, and $0 \leqslant \ell \leqslant k$, the quantity in the right-hand side of (22) is maximized when $\ell = k$. This yields (21). □

Finally, recall that if (S,T,U) is good, then no edge of the graph

$$\mathcal{G}_{S,T,U} \;:=\; \mathcal{G}_S\big[O_{2n} \setminus (T \cup U)\big]$$

has both its endpoints in G_p. Let $\mathcal{S}(k,m)$ denote the set of $S \subseteq E_{2n}$ with $|S| = k$ and with m pairs $\{x, 2n-x\}$ in S. We have

$$\mathbb{E}\big[Z'(k,\ell,m,j)\big] \;\leqslant\; \sum_{S \in \mathcal{S}(k,m)} \mathbb{P}(S \subseteq G_p)\cdot\mathbb{E}\big[W(S,\ell,j)\big]\cdot\max_{T,U}\Big\{\mathbb{P}\Big(e\big(\mathcal{G}_{S,T,U}[G_p]\big) = 0\Big)\Big\}, \tag{23}$$

where the maximum is taken over all pairs (T, U) as in Claim 3. Hence, to complete the proof, it only remains to give a uniform bound on the probability that $e(\mathcal{G}_{S,T,U}[G_p]) = 0$. We shall do so using Janson's inequality.

Note first that, by Claim 1, $\Delta(\mathcal{G}_{S,T,U}) \leqslant \Delta(\mathcal{G}_S) \leqslant 3k$ and

$$e(\mathcal{G}_{S,T,U}) \geqslant e(\mathcal{G}_S) - (|T| + |U|)\Delta(\mathcal{G}_S) = \left(\frac{3k - 2m - \mathbf{1}_S(n)}{2}\right) n - O\big(k^2\big).$$

Let $\mu = \frac{3k}{2} p^2 n$, let $\mu' = p^2 e(\mathcal{G}_{S,T,U})$ (the expected number of edges in $\mathcal{G}_{S,T,U}[G_p]$) and observe that

$$k^2 p^3 n \leqslant \Delta := \sum_{B_1 \sim B_2} p^{|B_1 \cup B_2|} \leqslant \binom{3k}{2} p^3 n.$$

where the summation is over pairs of edges of $\mathcal{G}_{S,T,U}$ sharing an endpoint. There are two cases.

Case 1. $\Delta \leqslant \mu'$**.**

In this case we shall show that $\mathbb{E}[Z'(k, \ell, m, j)] \leqslant n^{-\varepsilon k}$. By Janson's inequality, we have

$$\mathbb{P}\Big(e\big(\mathcal{G}_{S,T,U}[G_p]\big) = 0\Big) \leqslant e^{-\mu' + \Delta/2} \leqslant e^{-\mu + \Delta/2} \exp\left(\frac{2m + \mathbf{1}_S(n)}{2} p^2 n + O(p^2 k^2)\right). \quad (24)$$

Suppose first that $\Delta \leqslant \varepsilon\mu$. Then $kp \leqslant 1$, and so

$$-\mu + \frac{\Delta}{2} + O\big(p^2 k^2\big) \leqslant -\big(1 - \varepsilon\big)\mu = -\left(1 - \varepsilon\right) \frac{3C}{2} k \log n \leqslant -\left(\frac{1}{2} + 2\varepsilon\right) k \log n, \quad (25)$$

since $p^2 n = C \log n$ and $C > 1/3 + 2\varepsilon$, if $\varepsilon > 0$ is sufficiently small. By the comments before Claim 3, we have

$$\begin{aligned} \sum_{S \in \mathcal{S}(k,m)} \exp\left(\frac{2m + \mathbf{1}_S(n)}{2} p^2 n\right) &\leqslant 3^k \left[\binom{n}{k-m} e^{p^2 nm} + \binom{n}{k-m-1} e^{p^2 n(m+1)}\right] \\ &\leqslant 3^k e^{p^2 nm} \left[\left(\frac{en}{k-m}\right)^{k-m} + \left(\frac{en}{k-m-1}\right)^{k-m-1} e^{p^2 n}\right]. \end{aligned}$$

But $m \leqslant k/2$, so $\frac{en}{k-m-1} \leqslant \frac{3en}{k}$. Thus

$$\sum_{S \in \mathcal{S}(k,m)} \exp\left(\frac{2m + \mathbf{1}_S(n)}{2} p^2 n\right) \leqslant 2^{O(k)} n^{Cm} \left(\frac{n}{k}\right)^{k-m} \big(1 + kn^{C-1}\big). \quad (26)$$

Since the vertex set of $\mathcal{G}_{S,T,U}$ is disjoint from $S \cup T \cup U$, it follows that the events $e(\mathcal{G}_{S,T,U}[G_p]) = 0$ and $S, T, U \subseteq G_p$ are all independent. Therefore, by Claim 3 and (23)-(26), and some simple manipulation,

$$\begin{aligned} \mathbb{E}\big[Z'(k, \ell, m, j)\big] &\leqslant p^k (C \log n)^{2k} e^{-\mu + \Delta/2 + O(p^2 k^2)} \sum_{S \in \mathcal{S}(k,m)} \exp\left(\frac{2m + \mathbf{1}_S(n)}{2} p^2 n\right) \\ &\leqslant \left(O(1) \cdot p \cdot (C \log n)^2 \cdot n^{-1/2 - 2\varepsilon} \cdot \frac{n}{k}\right)^k \big(kn^{C-1}\big)^m \big(1 + kn^{C-1}\big) \ll n^{-\varepsilon k} \end{aligned}$$

as $n \to \infty$, assuming that $C < 1$. If $C \geqslant 1$ then the same calculation works, but we do not need to estimate so precisely; we leave the details to the reader.

The case $\Delta > \varepsilon\mu$ is similar, so we shall skip some of the details. Note that $k \gg \sqrt{n}$, that (26) still holds, and that

$$-\mu + \frac{\Delta}{2} + O\big(p^2k^2\big) \;\leqslant\; -\frac{2\mu}{5} \;=\; -\frac{3C}{5}k\log n.$$

Thus

$$\mathbb{E}\big[Z'(k,\ell,m,j)\big] \;\leqslant\; \left(O(1)\cdot p\cdot (C\log n)^2\cdot n^{-3C/5}\cdot\frac{n}{k}\right)^k \big(kn^{C-1}\big)^m\big(1+kn^{C-1}\big) \;\ll\; n^{-\varepsilon k},$$

using the fact that $k \geqslant 2m$. Thus, if $\Delta \leqslant \mu'$ then $\mathbb{E}[Z'(k,\ell,m,j)] \leqslant n^{-\varepsilon k}$, as claimed.

Case 2. $\Delta \geqslant \mu'$**.**

In this case, we shall show that $\mathbb{E}[Z'(k,\ell,m,j)] \leqslant e^{-\sqrt{n}}$. Indeed, by Janson's inequality,

$$\mathbb{P}\Big(e\big(\mathcal{G}_{S,T,U}[G_p]\big)=0\Big) \;\leqslant\; \exp\left(-\frac{(\mu')^2}{2\Delta}\right) \;\leqslant\; \exp\left(-\frac{\mu^2}{18\Delta}\right),$$

since $\mu' \geqslant \frac{\mu}{3}$. We shall also need an improved version of Claim 3 when k is large, since the bound $\binom{3kj}{\ell}$ on the number of choices for T becomes very bad. Fortunately the following bound is trivial:

$$\mathbb{E}[W(S,\ell,j)] \;\leqslant\; \binom{n}{\ell}\binom{n}{j}p^{\ell+j} \;\leqslant\; \binom{n}{k}^2 p^{2k}.$$

Recall that we have at most $\binom{n}{k}$ choices for S, and that $k \leqslant \varepsilon pn$ and $pn \gg \sqrt{n}$. By (23),

$$\mathbb{E}[Z'(k,\ell,m,j)] \;\leqslant\; \binom{n}{k}^3 p^{3k}e^{-\mu^2/(18\Delta)} \;\leqslant\; \left(\frac{epn}{k}\right)^{3k} e^{-pn/2} \;\leqslant\; e^{-\sqrt{n}},$$

as claimed.

Having bounded $\mathbb{E}[Z']$ in both cases, the result now follows easily by summing over k, ℓ, m and j. Indeed, by (20), together with the application of Theorem 3.1 noted at the start of the proof, we have

$$\begin{aligned}\mathbb{P}\Big(\mathrm{SF}_0(G_p) \neq \{G_p\cap O_{2n}\}\Big) \;&\leqslant\; \sum_{k=1}^{\varepsilon pn}\sum_{m=0}^{k/2}\sum_{\ell=0}^{k}\sum_{j=0}^{\ell}\mathbb{E}\big[Z'(k,\ell,m,j)\big] \;+\; o(1)\\ &\leqslant\; \sum_{k=1}^{n}(k+1)^3\max\Big\{n^{-\varepsilon k}, e^{-\sqrt{n}}\Big\} \;+\; o(1) \;\to\; 0\end{aligned}$$

as $n \to \infty$, as required. □

Acknowledgment. The bulk of this work was done during the IPAM Long Program, Combinatorics: Methods and Applications in Mathematics and Computer Science. We would like to thank the organizers of this program, the members of staff at IPAM, and all its participants for creating a fantastic research environment. We would also like to thank

David Conlon, Gonzalo Fiz Pontiveros, Simon Griffiths, Tomasz Łuczak, and Vera Sós for helpful and stimulating discussions.

References

[1] H.L. Abbott and E.T.H Wang, Sum-free sets of integers, *Proc. Amer. Math. Soc.*, **67** (1977), 11–16.

[2] N. Alon, Independent sets in regular graphs and sum-free subsets of abelian groups, *Israel J. Math.*, **78** (1991), 247–256.

[3] N. Alon and D.J. Kleitman, Sum-free subsets, in A tribute to Paul Erdős (A. Baker, B. Bollobás and A. Hajnal, eds), Cambridge University Press, Cambridge, 1990, 13–26.

[4] N. Alon and J. Spencer, The Probabilistic Method (3rd edition), Wiley Interscience, 2008.

[5] L. Babai, M. Simonovits and J. Spencer, Extremal subgraphs of random graphs, *J. Graph Theory*, **14** (1990), 599–622.

[6] B. Bollobás and A. Thomason, Threshold functions, *Combinatorica*, **7** (1986), 35–38.

[7] G. Brightwell, K. Panagiotou and A. Steger, On extremal subgraphs of random graphs, *Proc. 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, 477–485, ACM, New York, 2007.

[8] N.J. Calkin, On the number of sum-free sets, *Bull. London Math. Soc.*, **22** (1990), 141–144.

[9] P.J. Cameron, On the structure of a random sum-free set, *Prob. Theory Rel. Fields*, **76** (1987), 523–531.

[10] P.J. Cameron, Portrait of a typical sum-free set, in Surveys in Combinatorics (C. Whitehead, ed.), *London Math. Soc. Lecture Notes*, **123**, 13–42, Cambridge University Press, 1987.

[11] P.J. Cameron and P. Erdős, On the number of sets of integers with various properties, in Number Theory (R.A. Mollin, ed.), 61–79, Walter de Gruyter, Berlin, 1990.

[12] P.J. Cameron and P. Erdős, Notes on sum-free and related sets, *Combin. Probab. Computing*, **8** (1999), 95–107.

[13] S.L.G. Choi, J. Komlós and E. Szemerédi, On sum-free subsequences, *Trans. Amer. Math. Soc.*, **212** (1975), 307–313.

[14] D. Conlon and W.T. Gowers, Combinatorial Theorems in sparse random sets, submitted.

[15] P.H. Diananda and H.P. Yap, Maximal sum-free sets of elements of finite groups, *Proc. Japan Academy*, **45** (1969), 1–5.

[16] P. Erdős, D.J. Kleitman and B.L. Rothschild, Asymptotic enumeration of K_n-free graphs, in *Colloquio Internazionale sulle Teorie Combinatorie* (Rome, 1973), Vol. II, pp. 19–27. *Atti dei Convegni Lincei*, **17**, Accad. Naz. Lincei, Rome, 1976.

[17] P. Frankl, R. Graham and V. Rödl, Quantitative theorems for regular systems of equations, *J. Combin. Theory Ser. A*, **47** (1988), 246–261.

[18] G.A. Freiman, On the structure and the number of sum-free sets, *Journées Arithmétiques*, 1991 (Geneva), Ast'erisque, **209** (1992) 195–201.

[19] E. Friedgut, Sharp thresholds of graph properties, and the k-sat problem, with an appendix by Jean Bourgain, *J. Amer. Math. Soc.*, **12** (1999), 1017–1054.

[20] E. Friedgut and G. Kalai, Every monotone graph property has a sharp threshold, *Proc. Amer. Math. Soc.*, **124** (1996), 2993–3002.

[21] E. Friedgut, V. Rödl, A. Ruciński and P. Tetali, A sharp threshold for random graphs with a monochromatic triangle in every edge coloring, *Mem. Amer. Math. Soc.*, **179** (2006), 66pp.

[22] E. Friedgut, V. Rödl and M. Schacht, Ramsey properties of random discrete structures, *Random Structures Algorithms*, **37** (2010), 407–436

[23] R. Graham, V. Rödl and A. Ruciński, On Schur properties of random subsets of integers, *J. Number Theory*, **61** (1996), 388–408.

[24] B.J. Green, The Cameron-Erdős conjecture, *Bull. London Math. Soc.*, **36** (2004), 769-778.

[25] B.J. Green, A Szemerédi-type regularity lemma in abelian groups, *GAFA*, **15** (2005), 340–376.

[26] B.J. Green and I.Z. Ruzsa, Counting sumsets and sum-free sets modulo a prime, *Studia Sci. Math. Hungarica*, **41** (2004), 285–293.

[27] B.J. Green and I.Z. Ruzsa, Sum-free sets in abelian groups, *Israel J. Math.*, **147** (2005), 157–188.
[28] Y. Kohayakawa, T. Łuczak and V. Rödl, Arithmetic progressions of length three in subsets of a random set, *Acta Arith.*, **75** (1996), 133–163.
[29] D. Král, O. Serra and L. Vena, A combinatorial proof of the Removal Lemma for groups, *J. Combin. Theory Ser. A*, **116** (2009), 971–978.
[30] V. F. Lev, T. Łuczak and T. Schoen, Sum-free sets in abelian groups, *Israel J. Math.*, **125** (2001), 347–367.
[31] V.F. Lev and T. Schoen, Cameron-Erdős modulo a prime, *Finite Fields Applications*, **8** (2002), 108–119.
[32] T. Łuczak, Randomness and regularity, International Congress of Mathematicians, Vol. III, *Eur. Math. Soc.*, Zurich (2006), 899–909.
[33] V. Rödl and A. Ruciński, Threshold functions for Ramsey properties, *J. Amer. Math. Soc.*, **8** (1995), 917–942.
[34] V. Rödl and A. Ruciński, Rado partition theorem for random subsets of integers, *Proc. London Math. Soc.*, **74** (1997), 481–502.
[35] K.F. Roth, (1953), On certain sets of integers, *J. London Math. Soc.*, **28** (1952), 104–109.
[36] M. Schacht, Extremal results for random discrete structures, submitted.
[37] I. Schur, Uber die Kongruenz $x^m + y^m \equiv z^m \pmod{p}$, *ber. Deutsch. Mat. Verein.*, **25** (1916), 114–117.
[38] H.P. Yap, Maximal sum-free sets of group elements, *J. London Math. Soc.*, **44** (1969), 131–136.
[39] H.P. Yap, Structure of maximal sum-free sets in groups of order $3p$, *Proc. Japan Academy*, **46** (1970), 758–762.

Department of Mathematics, University of Illinois, 1409 W. Green Street, Urbana, IL 61801; and Department of Mathematics, University of California, San Diego, La Jolla, CA 92093
E-mail address: jobal@math.uiuc.edu

IMPA, Estrada Dona Castorina 110, Jardim Botânico, Rio de Janeiro, RJ, Brasil
E-mail address: rob@impa.br

School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel; and Trinity College, Cambridge CB2 1TQ, UK
E-mail address: samotij@post.tau.ac.il