ISRAEL JOURNAL OF MATHEMATICS **125** (2001), 347–367

# SUM-FREE SETS IN ABELIAN GROUPS

BY

VSEVOLOD F. LEV*

*Institute of Mathematics, The Hebrew University of Jerusalem*
*Givat Ram, Jerusalem 91904, Israel*
*e-mail: seva@math.huji.ac.il*

AND

TOMASZ ŁUCZAK**

*Department of Mathematics, Adam Mickiewicz University*
*Poznań, Poland*
*e-mail: tomasz@amu.edu.pl*

AND

TOMASZ SCHOEN

*Department of Mathematics, Christian Albrecht University*
*Kiel, Germany*
*and*
*Department of Discrete Mathematics, Adam Mickiewicz University*
*Poznań, Poland*
*e-mail: tos@numerik.uni-kiel.de*

ABSTRACT

We show that there is an absolute constant $\delta > 0$ such that the number of sum-free subsets of any finite abelian group $G$ is

$$\left(2^{\nu(G)} - 1\right) 2^{|G|/2} + O\left(2^{(1/2-\delta)|G|}\right),$$

where $\nu(G)$ is the number of even order components in the canonical decomposition of $G$ into a direct sum of its cyclic subgroups, and the implicit constant in the $O$-sign is absolute.

* This author was partially supported by the Edmund Landau Center for Research in Mathematical Analysis and Related Areas, sponsored by the Minerva Foundation (Germany).

** This author was partially supported by KBN grant 2 P03A 021 17.

Received March 13, 2000 and in revised form September 28, 2000

## 1. Introduction

A subset $A \subseteq G$ of an (additively written) group $G$ is said to be **sum-free** if no $a_1, a_2, a_3 \in A$ satisfy the equation $a_1 + a_2 = a_3$. In [A91], Alon proved that the number of sum-free subsets of any group $G$ of cardinality $n = |G|$ is at most $2^{(1/2+o(1))n}$ (as $n \to \infty$) and asked about the sharp form of this result. In the present paper we answer Alon's question for $G$ abelian.

THEOREM 1: *There is an absolute constant $\delta > 0$ such that the number of sum-free subsets of any abelian group $G$ of cardinality $n = |G|$ is*

$$\left(2^{\nu(G)} - 1\right)2^{n/2} + O\left(2^{(1/2-\delta)n}\right),$$

*where $\nu(G)$ is the number of even order components in the canonical decomposition of $G$ into a direct sum of its cyclic subgroups, and the implicit constant in the $O$-sign is absolute.*

Throughout the rest of the paper, $G$ is a finite abelian group of cardinality $n$; whenever appropriate, we tacitly assume $n$ to be sufficiently large. We write $\mathrm{SF}[G]$ for the set of all sum-free subsets of $G$. Let $H \subseteq G$ be a subgroup, and let $\bar{A} \in \mathrm{SF}[G/H]$. Suppose that a subset $A \subseteq G$ satisfies $\varphi_H(A) = \bar{A}$, where $\varphi_H\colon G \to G/H$ is the canonical homomorphism. Then plainly $A \in \mathrm{SF}[G]$, and we say that $A$ is induced by $\bar{A}$. For $H = G$, no non-empty sum-free subset of $G$ is induced by a sum-free subset of $G/H$. If $[G : H] = 2$ (that is, $H$ is an index two subgroup of $G$), then the sum-free subsets of $G$ induced by a sum-free subset of $G/H$ are all sets $A \subseteq G \setminus H$; it will be seen that it is these sets that contribute the main term to the asymptotic formula of Theorem 1. On the other hand, let $\mathrm{SF}^*[G]$ be the family of all $A \in \mathrm{SF}[G]$ *not* induced by any $\bar{A} \in \mathrm{SF}[G/H]$, where $H$ is an index two subgroup. We prove that this set is small; more precisely, we have

MAIN LEMMA: *There is an absolute constant $\delta > 0$ such that*

$$|\mathrm{SF}^*[G]| = O(2^{(1/2-\delta)n}),$$

*where the implicit constant in the $O$-sign is absolute.*

In particular, the number of all "primitive" sum-free subsets of $G$ (not induced by any $\bar{A} \in \mathrm{SF}[G/H]$ for a non-zero subgroup $H$) is $O(2^{(1/2-\delta)n})$. A result due to two of the present authors ([LS, Theorem 1]) shows that for $p$ prime, $|\mathrm{SF}[\mathbb{Z}_p]| \gg 2^{p/3}p$; therefore, $1/2 - \delta$ in the exponent cannot be replaced by $1/3$.

Theorem 1 will be deduced from the Main Lemma in Section 2, and the Main Lemma will be proved in Sections 3–7.

We now make several historical comments.

The first to consider sum-free sets was probably Schur with his celebrated theorem which states that it is impossible to partition the interval $[1, n]$ into a fixed number of sum-free subsets, provided that $n$ is large enough compared to the number of subsets. In [CE90], Cameron and Erdős investigated some properties of sum-free subsets of $[1, n]$ and conjectured that the number of such subsets is $O(2^{n/2})$. (The motivation for this conjecture is that the vast majority of sum-free sets $A \subseteq [1, n]$ are actually believed to be subsets of either the set $\{1, 3, \ldots, 2\lfloor (n-1)/2 \rfloor + 1\}$, or the set $[\lceil (n+1)/2 \rceil, n]$.) Having attracted much attention, the conjecture of Cameron and Erdős is, nevertheless, still open. One can expect that the problem gets easier if the condition $a_1 + a_2 \neq a_3$ ($a_i \in [1, n]$) is replaced by the stronger and more restrictive condition $a_1 + a_2 \not\equiv a_3 \pmod{n}$; that is, if sum-free subsets of $[1, n]$ are replaced by sum-free subsets of $\mathbb{Z}_n$. Indeed, Theorem 1 implies at once that $|\operatorname{SF}[\mathbb{Z}_n]| = O(2^{n/2})$, which can be viewed as a weak form of the conjecture.

For a survey of other results concerning sum-free sets and their generalizations we refer the reader to [B98, LS]. The rest of the paper is devoted to the proofs of Theorem 1 and the Main Lemma.

## 2. Deduction of Theorem 1 from the Main Lemma

LEMMA 1: *The number of subgroups of $G$ is at most $2^{(\log_2 n)^2}$.*

*Proof:* Plainly, any subgroup of $G$ can be generated by at most $\lfloor \log_2 n \rfloor$ elements. If we do not require the elements to be distinct, then *exactly* $\lfloor \log_2 n \rfloor$ generators can be used. Thus, the number of subgroups does not exceed the number of ways to choose $\lfloor \log_2 n \rfloor$ elements of $G$, which is at most

$$n^{\lfloor \log_2 n \rfloor} \leq 2^{(\log_2 n)^2}. \qquad \blacksquare$$

LEMMA 2: *The number of index two subgroups of $G$ is $2^{\nu(G)} - 1$, where $\nu(G)$ is the number of even order components in the canonical decomposition of $G$ into a direct sum of its cyclic subgroups.*

*Proof:* Let $N(G)$ be the sought number of index two subgroups, so that for instance, $N(G) = 0$ if $G$ is cyclic of odd order and $N(G) = 1$ if $G$ is cyclic of even order. It suffices to prove that $N(G_1 \oplus G_2) + 1 = (N(G_1) + 1)(N(G_2) + 1)$

for any two finite abelian groups $G_1$ and $G_2$. For this, notice that if $H_1$ and $H_2$ are index two subgroups of $G_1$ and $G_2$, respectively, then each of

$$H_1 \oplus G_2,\ G_1 \oplus H_2, \text{ and } (H_1 \oplus H_2) \cup ((G_1 \setminus H_1) \oplus (G_2 \setminus H_2))$$

is, evidently, an index two subgroup of $G_1 \oplus G_2$. The number of all these subgroups is $N(G_1)+N(G_2)+N(G_1)N(G_2)$, and we leave it to the reader to verify that any index two subgroup of $G_1 \oplus G_2$ has this form. ∎

By the Main Lemma, to prove Theorem 1 we only need to count all sum-free $A \subseteq G$ induced by some $\bar{A} \in \mathrm{SF}[G/H]$, where $H$ is an index two subgroup. For $H$ fixed, such $A$ are those subsets of $G$ contained in the complement of $H$, and their number is $2^{n/2}$. Furthermore, if $H_1$ and $H_2$ are distinct index two subgroups and $A$ is contained in both complements $G \setminus H_i$, then $A \subseteq G \setminus (H_1 \cup H_2)$; for $H_1$ and $H_2$ fixed, the number of such $A$ is at most $2^{n/4}$. Now by the inclusion-exclusion argument and Lemmas 1 and 2, the number of $A$ in question is

$$\sum_{\substack{H \subseteq G\\ [G:H]=2}} 2^{n/2} - O\Bigg(\sum_{\substack{H_1,H_2 \subseteq G\\ [G:H_i]=2}} 2^{n/4}\Bigg) = (2^{\nu(G)}-1)2^{n/2} - O(2^{n/4+2(\log_2 n)^2}).$$

Theorem 1 follows.

## 3. Auxiliary results

We collect here some facts that will be used in the proof of the Main Lemma.

Let $A$ be a subset of $G$. The **period**, or **stabilizer** of $A$ is the subgroup of $G$ defined by

$$H(A) := \{g \in G\colon A+g=A\}.$$

In other words, $H(A)$ is the maximal subgroup of $G$ such that $A$ is a union of $H(A)$-cosets. The following theorem is essentially due to Kneser [Kn53, Kn55]; the version presented below can be found, for instance, in [Ke60, Theorem 3.1]. (In fact, it can be derived easily from Kneser's original result.)

THEOREM 2: *Let $A$ and $B$ be finite, non-empty subsets of an abelian group $G$. Suppose that $|A+B| \le |A|+|B|-1$. Then*

$$|A+B| = |A+H|+|B+H|-|H|,$$

*where $H = H(A+B)$.*

LEMMA 3: *Let $B \subseteq G$ be a finite, non-empty subset of $G$, not contained in a coset of a proper subgroup. Suppose, moreover, that $|B| \le \frac{2}{3}|G|$. Then there exists an element $b \in B$ such that $|(B+b) \cap B| \le \frac{5}{6}|B|$.*

*Proof:* For $g \in G$, define $f_B(g) := |(B+g) \setminus B| = |B| - |(B+g) \cap B|$. This function is often referred to as the function of Erdős–Heilbronn–Olson; its basic properties include

$$f_B(-g) = f_B(g), \tag{1}$$

$$f_B(g_1+g_2) \le f_B(g_1) + f_B(g_2), \tag{2}$$

and

$$\sum_{g \in B-B} f_B(g) = |B||B-B| - |B|^2. \tag{3}$$

(All this is easy to verify.) By averaging, we derive from (3) that there exist $b_1, b_2 \in B$ such that

$$f_B(b_1 - b_2) \ge |B| \left(1 - \frac{|B|}{|B-B|}\right),$$

and since $f_B(b_1 - b_2) \le f_B(b_1) + f_B(b_2)$ by (1) and (2), we have

$$f_B(b) \ge \frac{1}{2}|B| \left(1 - \frac{|B|}{|B-B|}\right)$$

either for $b = b_1$ or for $b = b_2$. Now

$$|(B+b) \cap B| = |B| - f_B(b) \le \frac{1}{2}|B| \left(1 + \frac{|B|}{|B-B|}\right),$$

and it remains to observe that $|B-B| \ge \frac{3}{2}|B|$ by Theorem 2: otherwise, letting $H := H(B-B)$, we get

$$\frac{3}{2}|B+H| \ge \frac{3}{2}|B| > |B-B| = 2|B+H| - |H|,$$
$$|B+H| < 2|H|,$$

whence $|B+H| = |H|$ and $B$ is contained in a coset of $H$. Thus $H = G$ or equivalently $B - B = G$, and therefore $\frac{3}{2}|B| > |G|$, a contradiction. ∎

LEMMA 4: *For any positive integer $y$ and real $x \ge y-1$ we have*

$$\binom{x}{y} < \left(\frac{x}{y}e\right)^y.$$

*Proof:* Using induction by $y$ (or by a quantitative version of the Stirling formula) one obtains $y! > (y/e)^y$, hence

$$\binom{x}{y} \le \frac{x^y}{y!} < \left(\frac{x}{y}e\right)^y. \quad \blacksquare$$

We will need an estimate for the tails of binomial and hypergeometric distributions. Recall that $X$ is distributed binomially with parameters $k$ and $p$ if it attains values from the interval $[0,k]$ with probabilities $\mathrm{Prob}\{X=i\} = \binom{k}{i}p^i(1-p)^{k-i}$, and in this case its expectation is $\mathrm{E}X = kp$. Furthermore, $X$ has a hypergeometric distribution with parameters $N, k$, and $m$, if it attains values from the interval $[0,N]$ with probabilities $\mathrm{Prob}\{X=i\} = \binom{k}{i}\binom{m}{N-i}/\binom{k+m}{N}$; the expectation of such a random variable is $\mathrm{E}X = mk/N$.

LEMMA 5 ([JLR00, Theorems 2.1 and 2.10]): *Suppose that $X$ has either a binomial or a hypergeometric distribution. Then for any $0 \le \varepsilon \le 1$ we have*

$$\mathrm{Prob}\{X \le (1-\varepsilon)\mathrm{E}X\} \le \exp\Big(-\frac{\varepsilon^2}{2}\,\mathrm{E}X\Big),$$

*and*

$$\mathrm{Prob}\{X \ge (1+\varepsilon)\mathrm{E}X\} \le \exp\Big(-\frac{\varepsilon^2}{3}\,\mathrm{E}X\Big).$$

## 4. Popular differences

We continue our preparations for the proof of the Main Lemma. Kneser's theorem shows that, "normally", the sumset $A+B$ contains at least $|A|+|B|-1$ elements. In this section we consider the case $B=-A$ so that $A+B$ becomes the difference set $A-A$ and show that, "normally", this set contains at least $2|A|(1-o(1))$ elements with large number of representations as a difference of two elements of $A$.

For a subset $A \subseteq G$ of any (not necessarily finite) abelian group $G$ and a non-negative integer $K$, we denote by $D_K(A)$ the set of all those elements $g \in G$ which have at least $K$ distinct representations as $g = a_1 - a_2$ (with $a_1, a_2 \in A$).

PROPOSITION 1: *Let $A \subseteq G, K \in \mathbb{Z}^+$, and $D_K(A)$ be as above. Suppose that*

$$|D_K(A)| \le 2|A| - 5\sqrt{K|A-A|}.$$

*Then there is a subset $A' \subseteq A$ such that*

$$|A \setminus A'| \le \sqrt{K|A-A|} \quad \textit{and} \quad A' - A' \subseteq D_K(A).$$

The proof of Proposition 1 relies upon the following graph-theoretic lemma.

LEMMA 6: *For any graph $\Gamma = (V, E)$ of average degree $\bar{d} \ge (1-\lambda)|V|$, there exists an induced subgraph $\Gamma' = (V', E')$ such that*
(i) $|V'| \ge (1-\sqrt{\lambda})|V|$;
(ii) $\delta(\Gamma') > (1-2\sqrt{\lambda})|V|$ *(where $\delta(\Gamma')$ is the minimal degree of $\Gamma'$).*

*Proof:* We define $\Gamma'$ to be the subgraph of $\Gamma$, induced by all vertices $v \in V$ of degree $d(v) > (1-\sqrt{\lambda})|V|$. We have

$$\begin{aligned} \bar{d}|V| &= \sum_{d(v) \le (1-\sqrt{\lambda})|V|} d(v) + \sum_{d(v) > (1-\sqrt{\lambda})|V|} d(v) \\ &\le (1-\sqrt{\lambda})|V|(|V|-|V'|) + |V||V'|, \\ (1-\lambda)|V| &\le (1-\sqrt{\lambda})(|V|-|V'|) + |V'| \\ &= (1-\sqrt{\lambda})|V| + \sqrt{\lambda}|V'|, \\ |V'| &\ge (1-\sqrt{\lambda})|V|, \end{aligned}$$

which proves the first assertion. To prove the second assertion, notice that the degree in $\Gamma'$ of any vertex $v' \in V'$ is greater than

$$(1-\sqrt{\lambda})|V| - (|V|-|V'|) = |V'| - \sqrt{\lambda}|V| \ge (1-2\sqrt{\lambda})|V|. \quad \blacksquare$$

*Proof of Proposition 1:* We can assume that $K \le |A|$, as otherwise $\sqrt{K|A-A|} > |A|$ and the assertion is trivial.

Consider the graph $\Gamma = (A, E)$ on the system of vertices $A$, where $(a_1, a_2) \in E$ if and only if $a_1 - a_2 \in D_K(A)$. The edges of the complement of $\Gamma$ correspond to elements $c \in (A-A) \setminus D_K(A)$. Any such element yields at most $K-1$ edges, and elements $c$ and $-c$ yield the same edge. Therefore, the number of edges of the complement is at most

$$\frac{1}{2}(K-1)|(A-A) \setminus D_K(A)| \le \frac{1}{2}(K-1)|A-A|,$$

the number of edges of $\Gamma$ is at least

$$\binom{|A|}{2} - \frac{1}{2}(K-1)|A-A|,$$

the average degree of $\Gamma$ is at least

$$|A|-1-(K-1)|A-A|/|A|\geq|A|-K|A-A|/|A|=|A|(1-K|A-A|/|A|^2),$$

and by Lemma 6 there is a subgraph $\Gamma'=(A',E')$ such that

$$|A'|\geq|A|(1-\sqrt{K|A-A|/|A|^2})=|A|-\sqrt{K|A-A|},$$

and for any $a'\in A'$ the neighborhood $N(a')$ of $a'$ in $\Gamma'$ is "large":

$$|N(a')|>|A|-2\sqrt{K|A-A|}.$$

Assume that $A'-A'\not\subseteq D_K(A)$ (otherwise we are done). Then there exist two elements $a_1'$ and $a_2'$ of $A'$ such that $a_2'-a_1'\notin D_K(A)$, and hence the number of representations of $a_2'-a_1'$ as a difference of two elements of $A$ does not exceed $K-1$. It follows that

$$|(a_1'-N(a_1'))\cap(a_2'-N(a_2'))|\leq K-1,$$

and since $a_j'-N(a_j')\subseteq D_K(A)$ $(j=1,2)$, we conclude that

$$\begin{aligned}|D_K(A)|&\geq|N(a_1')|+|N(a_2')|-(K-1)\\&>2|A|-4\sqrt{K|A-A|}-K>2|A|-5\sqrt{K|A-A|}.\qquad\blacksquare\end{aligned}$$

Why are we interested in the elements of $A-A$ with a large number of representations? Suppose that any $A\in\mathrm{SF}[G]$ contains a "small" subset $R$, such that its difference set $R-R$ is "large". Since the number of possible sets $R$ is small (as $|R|$ is small), and since the number of sets $A$ corresponding to a given $R$ is small also (as $A\subseteq G\setminus(A-A)\subseteq G\setminus(R-R)$), this would help us to bound the total number of $A$ possible. Indeed, it is easy to show that there exists a "small" $R$ such that $R-R$ contains the set $D_K(A)$ with a suitably chosen $K$.

Lemma 7: *For any $A\subseteq G$, any $p\in(0,1)$, and any integer $K\geq 0$ such that $p^2K\geq 6\ln n$, there exists a subset $R\subseteq A$ with the following properties:*
(i) $|R|\leq 2p|A|$;
(ii) $D_K(A)\subseteq R-R$.

*Proof:* Let $R\subseteq A$ be a random subset of $A$ for which the elements of $A$ are chosen randomly and independently with probability $p$ each. Plainly, the expected cardinality of $R$ is $p|A|$, and by Markov's inequality, (i) holds with probability at least $1/2$.

Fix $d \in D_K(A)$. For any representation $d = a_1 - a_2$ $(a_1, a_2 \in A)$ there are at most two other representations of $d$ of this form in which $a_1$ or $a_2$ are used (there can be one representation $d = a_3 - a_1$ and one $d = a_2 - a_4$). As the total number of representations is at least $K$, we can select at least $K/3$ representations disjoint in the sense that no $a \in A$ is used in two distinct representations. The probability that a given representation "survives" in $R$ is $p^2$, the probability that it is destroyed is $1 - p^2$, the probability that *all* selected representations are destroyed is less then or equal to $(1 - p^2)^{K/3}$; thus,

$$\mathrm{Prob}\{d \notin R - R\} \leq (1 - p^2)^{K/3} < e^{-p^2 K/3} \leq 1/n^2,$$

whence

$$\mathrm{Prob}\{D_K(A) \not\subseteq R - R\} \leq \sum_{d \in D_K(A)} \mathrm{Prob}\{d \notin R - R\} \leq 1/n.$$

Therefore, (ii) holds with probability at least $1 - 1/n > 1/2$, and the result follows. ∎

Below we choose $K = \lfloor n^{2/3} \rfloor$, $p = n^{-1/7}/2$, and think of $R$ as being associated with $A$ uniquely; in other words, for each $A$ we select and fix one particular set $R$ of all those, the existence of which is guaranteed by Lemma 7. We abbreviate $D_K(A)$ by $D$; thus, we have

$$|R| < n^{6/7}, \quad D \subseteq R - R. \tag{4}$$

## 5. Proof of the Main Lemma, I. Small sum-free sets

To prove the Main Lemma we split the family of all sum-free subsets of $G$ into several sub-families and show that each of them contains not more than $2^{(1/2-\delta)n}$ sets for some constant $\delta > 0$. In this section we estimate the number of sum-free subsets of cardinality less than $(1 - \varepsilon)n/4$, where $\varepsilon$ is a positive constant. We follow closely Alon's argument from [A91] and use some of his intermediate results.

Recall that a graph $\Gamma$ is $r$**-regular** if each vertex of $\Gamma$ has degree $r$, and that a subset $A \subseteq V(\Gamma)$ is **independent** if it induces an empty subgraph of $\Gamma$.

LEMMA 8: *Let $\varepsilon > 0$ be fixed. Assuming $r$ to be large enough, for any $r$-regular graph $\Gamma$ on $n$ vertices the number of independent sets of at most $(1 - \varepsilon)n/4$ vertices of $\Gamma$ is smaller than $2^{n/2 - \varepsilon^2 n/6}$.*

*Proof:* Alon [A91, Corollary 3.2] showed that there is a spanning bipartite subgraph $\Gamma' \subseteq \Gamma$ such that the degree of any vertex of $\Gamma'$ is between $r/2 - r^{5/8}/2$

and $r/2+r^{5/8}/2$. Let $E$ be the edge set, and let $U$ and $V$ be the partite sets of $\Gamma'$, labeled so that $|U|\le|V|$. Then evidently

$$\begin{aligned}|V|(r/2-r^{5/8}/2)\le|E|\le|U|(r/2+r^{5/8}/2),\\ |V|(r-r^{5/8})\le(n-|V|)(r+r^{5/8}),\end{aligned}$$

whence

$$|V|\le m:=\lfloor n(1+r^{-3/8})/2\rfloor.$$

Let $I(s,t)$ denote the number of all $t$ element subsets of $U$ which have exactly $s$ neighbors in $V$. By [A91, Corollary 2.5], there exists an absolute constant $C$ such that for every $t\ge 2m/\sqrt{r}$ we have

$$I(s,t)\le\binom{s+Cmr^{-1/7}}{t}.$$

On the one hand, the number of independent sets $A$ with at most $\ell:=\lfloor(1-\varepsilon)n/4\rfloor$ elements satisfying $t:=|A\cap U|<\lceil 2m/\sqrt{r}\rceil$ is bounded from above by

$$\sum_{0\le t<\lceil 2m/\sqrt{r}\rceil}\binom{m}{t}\binom{m}{\ell}<n\binom{m}{\lceil 2m/\sqrt{r}\rceil}\binom{m}{\ell}\le nr^{2m/\sqrt{r}}\binom{m}{\ell}.$$

On the other hand, the number of independent sets $A$ of cardinality $i:=|A|\le\ell$ with $t:=|A\cap U|\ge 2m/\sqrt{r}$ does not exceed

$$\begin{aligned}&\sum_{i=\lceil 2m/\sqrt{r}\rceil}^{\ell}\ \sum_{t=\lceil 2m/\sqrt{r}\rceil}^{i}\ \sum_{s=0}^{|V|+t-i} I(s,t)\binom{|V|-s}{i-t}\\
&\le\sum_{i=\lceil 2m/\sqrt{r}\rceil}^{\ell}\ \sum_{t=\lceil 2m/\sqrt{r}\rceil}^{i}\ \sum_{s=0}^{|V|+t-i}\binom{s+Cmr^{-1/7}}{t}\binom{|V|-s}{i-t}\\
&\le\sum_{i=\lceil 2m/\sqrt{r}\rceil}^{\ell}\ \sum_{t=\lceil 2m/\sqrt{r}\rceil}^{i}\ \sum_{s=0}^{|V|+t-i}\binom{|V|+Cmr^{-1/7}}{i}\\
&\le n^3\binom{m+Cmr^{-1/7}}{\ell}\\
&\le n^3e^{C\ell r^{-1/7}}\binom{m}{\ell}.\end{aligned}$$

Thus, the total number of independent sets with at most $\ell$ vertices is bounded from above by

$$nr^{2m/\sqrt{r}}\binom{m}{\ell} + n^3 e^{Clr^{-1/7}}\binom{m}{\ell} \le n^3\, 2^{Cnr^{-1/7}}\binom{m}{\ell}$$
$$\le n^3\, 2^{Cnr^{-1/7}+m}\sum_{i=0}^{\ell}\binom{m}{i}2^{-m}.$$

The latter sum is the probability that a random variable, distributed binomially with parameters $m$ and $1/2$, attains value not larger than $\ell \le (1-\varepsilon)n/4 \le (1-\varepsilon)m/2$; by Lemma 5, this sum does not exceed $e^{-0.25\varepsilon^2 m}$, and it remains to observe that

$$n^3 2^{Cnr^{-1/7}+m} e^{-0.25\varepsilon^2 m} \le 2^{n/2-\varepsilon^2 n/8\ln 2+2Cnr^{-1/7}} \le 2^{n/2-\varepsilon^2 n/6}. \qquad \blacksquare$$

LEMMA 9: *Let $\varepsilon > 0$ be fixed. Assuming $n$ to be large enough, we have*

$$\#\{A \in \mathrm{SF}[G]\colon |A| \le (1-\varepsilon)n/4\} = O(2^{(1/2-\varepsilon^2/7)n})$$

*(where the implicit constant in the $O$-sign depends on $\varepsilon$ only).*

*Proof:* Let $r = \lfloor \log n \rfloor$. Given a sum-free set $A \subseteq G$ of cardinality $r \le |A| \le (1-\varepsilon)n/4$, select an $r$-element subset $B \subseteq A$ and define $r_0 := |B \cup (-B)|$, so that $r \le r_0 \le 2r$. Consider the graph on the vertex set $G$, in which two vertices $u$ and $v$ are adjacent if and only if $u - v \in \pm B$. This graph is $(r_0)$-regular, and each $A \in \mathrm{SF}[G]$ containing $B$ is its independent set. (Otherwise, we would have $a' - a'' \in B \subseteq A$ for some $a', a'' \in A$.) Thus, by Lemma 8 the number of $A \in \mathrm{SF}[G]$ with at most $(1-\varepsilon)n/4$ elements does not exceed

$$\sum_{i=0}^{r-1}\binom{n}{i} + \binom{n}{r}2^{(1/2-\varepsilon^2/6)n} = O_\varepsilon(2^{(1/2-\varepsilon^2/7)n}). \qquad \blacksquare$$

It is worth pointing out that our Lemma 9 is "parallel" to a result of Bilu [B98, Theorem 1.1], where a similar estimate for the number of sum-free subsets of the interval $[1, n]$ is established. Bilu's result implies at once the desired estimate for the group $\mathbb{Z}_n$; however, his approach, based on Szemerédi's theorem, is not applicable for a generic abelian group $G$.

## 6. Proof of the Main Lemma, II. Small (popular) difference sets

First, we estimate the number of $A \in \mathrm{SF}^*[G]$ for which $D$ is "small".

LEMMA 10: *The number of $A \in \mathrm{SF}^*[G]$ satisfying*

$$|D| \le 2|A| - 5n^{5/6}$$

*is $O(2^{0.46n})$.*

*Proof:* As $2|A| - 5n^{5/6} \le 2|A| - 5\sqrt{K|A-A|}$, by Proposition 1 for any $A$ under consideration there exists $A' \subseteq A$ such that

$$|A'| \ge |A| - \sqrt{K|A-A|} \ge |A| - n^{5/6}, \quad A' - A' \subseteq D, \tag{5}$$

and then

$$|A' - A'| \le |D| \le 2|A| - 5n^{5/6} \le 2|A'| - 3n^{5/6}.$$

Thus, letting $H := H(A' - A')$, by Theorem 2 we get $|A' - A'| = 2|A' + H| - |H|$, whence

$$2(|A' + H| - |A'|) \le |H| - 3n^{5/6}$$

and we conclude that

$$|H| \ge 3n^{5/6} \quad \text{and} \quad |A' + H| - |A'| < |H|/2. \tag{6}$$

We now observe that $[G : H] \ne 1$, since otherwise $A' - A' = G$ (contradicting the fact that $A'$ is sum-free), and similarly $[G : H] \ne 2$, since otherwise $A' - A' = H$, $A \subseteq G \setminus (A' - A') = G \setminus H$ and $A$ is contained in the complement of an index two subgroup. Therefore, we have

$$|H| \le n/3. \tag{7}$$

Furthermore, we note that

$$|A' + H| \le n/2 \tag{8}$$

(else for any $g \in G$ by the pigeonhole principle holds $(A' + H) \cap (g + (A' + H)) \ne \emptyset$, hence $g \in (A' + H) - (A' + H)$ implying that $A' - A' = (A' + H) - (A' + H) = G$), and that by (6) and (7)

$$|A' + H| - |A'| < n/6. \tag{9}$$

We now make the counting. To each $A$ there correspond a set $A'$ and a subgroup $H$. The number of subgroups $H$ possible is, by Lemma 1, less than

$2^{(\log_2 n)^2}$, and for any $H$ given the number of sets $A'+H$ possible is at most $2^{n/|H|}$; thus, by (6) the number of choices for $A'+H$ is at most $2^{(\log_2 n)^2}2^{(n^{1/6}/3)} = 2^{o(n)}$. Next, by (8) and (9) for any $A'+H$ given, the number of sets $A'$ possible is at most

$$\sum_{0\le i<n/6}\binom{n/2}{i} = O(2^{0.4592n}).$$

Finally, by (5) for any $A'$ there are at most

$$\sum_{0\le i\le n^{5/6}}\binom{n}{i} = 2^{o(n)}$$

corresponding sets $A$. Putting everything together, we see that the total number of $A$ is at most

$$2^{0.4592n+o(n)} = O(2^{0.46n}). \quad \blacksquare$$

Having established Lemma 10, we can concentrate on sets $A$ such that

$$|D| > 2|A| - 5n^{5/6}. \tag{10}$$

Moreover, by Lemma 9 we can restrict ourselves to studying the sets $A$ of cardinality

$$|A| > n/4 - 10^{-8}n. \tag{11}$$

In our next lemma we count $A$ which, in addition to these two properties, have "small" difference set.

LEMMA 11: *The number of $A \in \mathrm{SF}^*[G]$ satisfying (10), (11), and*

$$|A - A| \le n/2 + 10^{-7}n$$

*is $O(2^{0.42n})$.*

*Proof:* Consider the set $R \subseteq A$ with properties (4). By (10), (11) and the assumptions of the lemma we have

$$|R-R| \ge |D| > 2|A| - 5n^{5/6} > n/2 - 3\cdot 10^{-8}n \ge |A-A| - 2\cdot 10^{-7}n,$$

hence one can add to $R$ at most $4\cdot 10^{-7}n$ elements of $A$ to obtain a set $A''\subseteq A$ of cardinality $|A''| \le |R| + 4\cdot 10^{-7}n \le 5\cdot 10^{-7}n$ such that $A''-A'' = A-A$. Clearly, such an $A''$ can be chosen from $G$ at no more than $n\binom{n}{\lfloor 5\cdot 10^{-7}n\rfloor}$ ways.

We put $B := A'' - A'' = A - A$ and note that $A \subseteq a - B$ for every $a \in A$.

If $B$ is contained in a coset of a subgroup $H \subset G$, then so is $A$, and in this case $k := [G : H] \geq 3$: otherwise, $A$ and $A - A$ are disjoint subsets of $H$, whence

$$\frac{1}{2}n = |H| \geq |A - A| + |A| \geq |D| + |A| > 3|A| + o(n) > \frac{3}{4}n - 4 \cdot 10^{-8}n,$$

a contradiction. For $H$ given, the number of $A$ contained in an $H$-coset is at most $k2^{|H|} = k2^{n/k} < n2^{n/3}$, hence by Lemma 1 the total number of $A \in \mathrm{SF}^*[G]$ for which $B$ is contained in a coset of a proper subgroup is $O(2^{0.34n})$.

Suppose now that $B$ is *not* contained in a coset of a proper subgroup. Applying then Lemma 3 to the set $B$, we find an element $b = a_1 - a_2$ ($a_1, a_2 \in A$) with the property that

$$|(a_1 - B) \cap (a_2 - B)| = |(-B) \cap (-(a_1 - a_2) - B)| = |B \cap (b + B)| \leq \frac{5}{6}|B|.$$

Notice that $B \supseteq a_i - A$, whence $A \subseteq a_i - B$ for $i = 1, 2$. Thus, once $a_1, a_2$, and $A''$ are selected, the remaining elements of $A$ are to be chosen from the set $(a_1 - B) \cap (a_2 - B)$ of cardinality at most $\frac{5}{6}(n/2 + 10^{-7}n)$. Consequently, the number of possible sets $A$ satisfying the assumptions of the lemma is at most

$$n^3 \binom{n}{\lfloor 5 \cdot 10^{-7}n \rfloor} 2^{\frac{5}{6}(1/2 + 10^{-7})n} = O(2^{0.42n}). \qquad \blacksquare$$

## 7. Proof of the Main Lemma, III. Conclusion

We now take care of the remaining and most complicated case, that of $A, A - A$, and $D$ all "large". More precisely, by Lemmas 9, 10, and 11, to conclude the proof of the Main Lemma it suffices to count $A \in \mathrm{SF}[G]$ such that (10), (11), and

$$|A - A| > n/2 + 10^{-7}n \tag{12}$$

hold.

Since the proof is somewhat technical, we first describe briefly its main idea. To construct $A$ we first choose the small subset $R \subseteq A$. The remaining elements of $A$ must then be selected from the set $G \setminus (R - R)$, the cardinality of which is $n - |R - R| \leq n - |D| < n/2 + 3 \cdot 10^{-8}n$ (only slightly exceeding $n/2$ in the worst case). We select $A \setminus R$ in two rounds, first choosing a set $Z \subseteq A$ of $\lfloor |A|/2 \rfloor$ elements, and then finding $A \setminus (R \cup Z)$. If $Z$ is chosen "at random", then each element $d \in A - A$ with probability at least $1/4$ belongs to $Z - Z$. Hence, we can expect that

$$\begin{aligned}
|(Z-Z)\setminus(R-R)| &\geq \frac{1}{4}|(A-A)\setminus(R-R)| \\
&= \frac{1}{4}|A-A| - \frac{1}{4}|R-R|, \\
|(R\cup Z)-(R\cup Z)| &\geq |(Z-Z)\cup(R-R)| \\
&\geq \frac{3}{4}|R-R| + \frac{1}{4}|A-A| \\
&\geq \frac{3}{4}|D| + \frac{1}{4}(n/2+10^{-7}n) \\
&\geq \frac{3}{2}|A| + \frac{1}{8}n + \frac{1}{4}10^{-7}n - 4n^{5/6} \\
&\geq \frac{1}{2}n + \left(\frac{1}{4}10^{-7} - \frac{3}{2}10^{-8}\right)n - 4n^{5/6} \\
&> \frac{1}{2}n + \delta n
\end{aligned}$$

(with some $\delta > 0$). As $A \subseteq G\setminus((R\cup Z)-(R\cup Z))$, we expect that after choosing $Z$, the set $A\setminus(R\cup Z)$ is to be chosen from at most $n-|(Z\cup R)-(Z\cup R)| < n/2-\delta n$ elements of $G$; hence, the number of choices for $A\setminus(R\cup Z)$ is bounded from above by $2^{n/2-\delta n}$. This is small enough to compensate for the choices of $R$ and $Z$. Unfortunately, a fair amount of work is needed to make the above argument rigorous. The main difficulty is that if $d_1, d_2 \in A-A$ and $Z$ is a random subset of $A\setminus R$, then the events that $d_i \in (Z\cup R)-(Z\cup R)$ for $i=1,2$ are not independent. Hence our main task will be, roughly speaking, to approximate $|(Z\cup R)-(Z\cup R)|$ by a sum of independent random variables.

For $A$ (and therefore, $R=R(A)$) given, let $X=X(A)$ be a set of pairs $(b_i', b_i'')$ ($b_i', b_i'' \in A$) which satisfies the following conditions and is maximal subject to these conditions:

(i) all differences $b_i' - b_i''$ are pairwise distinct and do not belong to $R-R$;

(ii) $\{b_i', b_i''\}\cap\{b_j', b_j''\} \subseteq R$ (for any $i \neq j$).

We put $X^b = \cup_i\{b_i', b_i''\}$ so that by the maximality of $X$, for any $d \in A-A$ there is a representation $d = a' - a''$ such that either $a' \in R\cup X^b$, or $a'' \in R\cup X^b$. (To see this, consider separately the cases $d \in R-R$; $d = b_i' - b_i''$ for some $i$; and $d \notin R-R, d \neq b_i' - b_i''$.)

Next, we introduce yet another set of pairs associated with $A$: specifically, let $Y = Y(A)$ be a set of pairs $(c_i', c_i'')$ ($c_i' \in R\cup X^b, c_i'' \in A\setminus(R\cup X^b)$) which satisfies and is maximal subject to the following conditions:

(i) all differences $c_i' - c_i''$ are pairwise distinct and do not belong to $(R\cup X^b) - (R\cup X^b)$;

(ii) $c_i'' \neq c_j''$ (for any $i \neq j$).

We put $Y^c = \cup_i \{c_i''\}$ and note that $R \cup X^b \cup Y^c \subseteq A$, and moreover,

$$(R \cup X^b \cup Y^c) - (R \cup X^b \cup Y^c) = A - A. \tag{13}$$

(To verify, assume that $d \in (A-A) \setminus ((R \cup X^b) - (R \cup X^b))$ and write $d = a' - a''$, where exactly one of $a', a''$ belongs to $R \cup X^b$. Now if $d = \pm(c_i' - c_i'')$ for some $i$, then $d \in \pm((R \cup X^b) - Y^c)$; otherwise $(a', a''), (a'', a') \notin Y$ and the maximality of $Y$ shows that of the elements $a'$ and $a''$ one which *does not* belong to $R \cup X^b$, belongs to $Y^c$ — whence, again, $d \in \pm((R \cup X^b) - Y^c)$.)

We split the proof into three cases, depending on the cardinalities of $X$ and $Y$. We set $m = |A - A|$ and $m_0 = \lfloor n/2 + 10^{-7} n \rfloor$; thus, $m \geq m_0$ by (12).

Case 1: $|X| < (m - n/2)/10^6$ and $|Y| < (m - n/2)/100$.

To construct $A$, we first choose the set $R \cup X^b \cup Y^c$ of cardinality $i := |R \cup X^b \cup Y^c| \leq (m - n/2)/99$ and then select other elements of $A$. By (13) and in view of $A \cap (A - A) = \emptyset$, the number of sets $A$ satisfying all of the assumptions is at most

$$\begin{aligned}
\sum_{m \geq m_0} \sum_{i=1}^{\lfloor (m-n/2)/99 \rfloor} \binom{n}{i} 2^{n-m} &\leq n^2 \max_{m \geq m_0} \binom{n}{\lceil (m-n/2)/99 \rceil} 2^{n-m} \\
&\leq n^2 \max_{m \geq m_0} \left( \frac{99ne}{m - n/2} \right)^{(m-n/2)/99+1} 2^{n-m} \\
&\leq n^2 \max_{m \geq m_0} (2.7 \cdot 10^9)^{(m-n/2)/99+1} 2^{n-m} \\
&\leq n^2 \max_{m \geq m_0} 2^{((m-n/2)/99+1) \ln(2.7 \cdot 10^9)/\ln 2 + (n-m)} \\
&\leq \max_{m \geq m_0} 2^{0.32(m-n/2)+(n-m)+32} \\
&= 2^{0.32(m_0-n/2)+(n-m_0)+32} \\
&= O(2^{n/2 - 0.68 \cdot 10^{-7} n}).
\end{aligned}$$

Case 2: $|X| < (m - n/2)/10^6$ and $|Y| \geq (m - n/2)/100$.

Again, we build a set $A$ in several steps. First we choose $R$ and $X^b$. Then, we select $\lfloor |A|/2 \rfloor$ elements of $A \setminus (R \cup X^b)$ and, finally, the remaining $|A| - |R \cup X^b| - \lfloor |A|/2 \rfloor$ elements of $A$. More formally, for a given $R$ and $X$, instead of counting sum-free sets $A$ with $A \supseteq (R \cup X^b)$, we shall estimate the number $w$ of pairs of sets $(Z, A \setminus (R \cup X^b \cup Z))$, where $Z \subseteq A \setminus (R \cup X^b)$ and $|Z| = \lfloor |A|/2 \rfloor$. Our hope is that, since $Z$ contains half of the elements of $A$, it will contain a considerable fraction of elements of $Y^c$ and thus substantially decrease the number of choices for the elements from $A \setminus (R \cup X^b \cup Z)$.

For given positive integers $k$ and $\ell$, let us count such pairs with $|A| = k$ and $|R \cup X^b| = \ell$. We first estimate the number $w'_{k,\ell}$ of the pairs $(Z, A \setminus (R \cup X^b \cup Z))$ for which

$$|Z \cap Y^c| \le |Y|/3. \tag{14}$$

Thus, we fix set $A$ and count the number of all subsets $Z$ of $A$ with $Z \subseteq A \setminus (R \cup X^b)$, $|Z| = \lfloor k/2 \rfloor$ for which (14) holds. Equivalently, we may estimate the probability that for the random subset $\mathcal{Z}$, chosen uniformly at random from all subsets of $A \setminus (R \cup X^b)$ with $\lfloor |A|/2 \rfloor$ elements, we have $|\mathcal{Z} \cap Y^c| \le |Y|/3$. Note that the random variable $\mathcal{Y} = |\mathcal{Z} \cap Y^c|$ has the hypergeometric distribution with parameters $|A| - |R \cup X^b|$, $|Y|$, and $\lfloor |A|/2 \rfloor$. In particular, for the expectation of $\mathcal{Y}$ we get

$$\frac{|Y| \cdot \lfloor |A|/2 \rfloor}{|A| - |R \cup X|} > \frac{|Y|}{2}.$$

Hence, Lemma 5 gives

$$\begin{aligned}\operatorname{Prob}\{|\mathcal{Z} \cap Y^c| \le |Y|/3\} &= \operatorname{Prob}\{|\mathcal{Y}| \le |Y|/3\}\\ &\le \exp(-|Y|/100) \le \exp\Big(-\frac{m_0 - n/2}{10^4}\Big).\end{aligned}$$

Thus, to estimate $w'_{k,\ell}$, it is enough to bound the number of choices for $A$ and multiply the result by

$$\binom{k-\ell}{\lfloor k/2 \rfloor} \exp\Big( - \frac{m_0 - n/2}{10^4}\Big).$$

Consequently, from the assumptions we have $|R - R| \ge n/2 - 3 \cdot 10^{-17} n$, and $\ell \le 3(m - n/2)/10^6$, so that

$$\begin{aligned}\frac{w'_{k,\ell}}{\binom{k-\ell}{\lfloor k/2 \rfloor}} &\le \sum_{m \ge m_0} \binom{n}{\ell} 2^{n/2 + 3\cdot 10^{-17} n} \exp\Big(-\frac{m_0 - n/2}{10^4}\Big)\\ &\le n \binom{n}{\lceil 3(m_0 - n/2)/10^6 \rceil} 2^{n/2 + 3\cdot 10^{-17} n} \exp\Big(-\frac{m_0 - n/2}{10^4}\Big)\\ &\le n \Big(\frac{10^6 n}{m_0 - n/2} \cdot e^{-32}\Big)^{3(m_0 - n/2)/10^6 + 1} 2^{n/2 + 3\cdot 10^{-17} n}\\ &= O(n (10^{13} e^{-32})^{3\cdot 10^{-13} n} 2^{n/2 + 3 \cdot 10^{-17} n}) = O(2^{n/2 - 2\cdot 10^{-13} n}).\end{aligned} \tag{15}$$

Now we estimate the number of pairs $w''_{k,\ell}$ ($|A|=k$ and $|R\cup X^b|=\ell$) for which (14) does not hold. Note that in this case

$$\begin{aligned}\big|(R\cup X^b\cup Z)-(R\cup X^b\cup Z)\big| &\geq |R-R|+|Y|/3\\ &\geq n/2-3n/10^{17}+(m-n/2)/300\\ &\geq n/2+(m-n/2)/400.\end{aligned}$$

Hence, choosing first $R\cup X^b$, then $Z$ from at most $n-|(R\cup X^b)-(R\cup X^b)|$ elements and, finally, selecting $A\setminus(R\cup X^b\cup Z)$ from the available set of not more than $n/2-(m-n/2)/400-\lfloor |A|/2\rfloor$ elements, we arrive at

$$w''_{k,\ell}\leq \sum_{m\geq m_0}\binom{n}{\ell}\binom{n/2+3\cdot 10^{-17}n}{\lfloor k/2\rfloor}\binom{n/2-(m-n/2)/400-\lfloor k/2\rfloor}{k-\ell-\lfloor k/2\rfloor}.$$

Because of the combinatorial identity

$$\binom{n/2+3\cdot 10^{-17}n}{\lfloor k/2\rfloor}\binom{n/2+3\cdot 10^{-17}n-\lfloor k/2\rfloor}{k-\ell-\lfloor k/2\rfloor}=\binom{n/2+3\cdot 10^{-17}n}{k-\ell}\binom{k-\ell}{\lfloor k/2\rfloor},$$

we can bound $\frac{w''_{k,\ell}}{\binom{k-\ell}{\lfloor k/2\rfloor}}$ from above by

$$\begin{aligned}&2^{n/2+3\cdot 10^{-17}n}\sum_{m\geq m_0}\binom{n}{\ell}\frac{\binom{n/2-(m-n/2)/400-\lfloor k/2\rfloor}{k-\ell-\lfloor k/2\rfloor}}{\binom{n/2+3\cdot 10^{-17}n-\lfloor k/2\rfloor}{k-\ell-\lfloor k/2\rfloor}}\\ &\leq 2^{n/2+3\cdot 10^{-17}n}\sum_{m\geq m_0}\left(\frac{en}{\lceil 3(m-n/2)/10^6\rceil}\right)^{3(m-n/2)/10^6+1}\\ &\qquad\qquad\times\left(\frac{n/2-(m-n/2)/400-\lfloor k/2\rfloor}{n/2+3\cdot 10^{-17}n-\lfloor k/2\rfloor}\right)^{0.1n}\\ &\leq n2^{n/2+3\cdot 10^{-17}n}\left(\frac{en}{3(m_0-n/2)/10^6}\right)^{3(m_0-n/2)/10^6+1}\\ &\qquad\qquad\times\left(1-\frac{m_0-n/2}{200n}\right)^{0.1n}\\ &=O(n2^{n/2+3\cdot 10^{-17}n}10^{4(m_0-n/2)/10^5}2^{-(m_0-n/2)/2\cdot 10^3})\\ &=O(2^{n/2-10^{-12}n}).\end{aligned}\tag{16}$$

Note that if by $\sigma_{k,\ell}$ we denote the number of all sum-free sets $A$ with $|A|=k$ and $|R\cup X^b|=\ell$, then

$$\begin{aligned}\sigma_{k,\ell}\binom{k-\ell}{\lfloor k/2\rfloor}&=\big|\{(Z,A\setminus(R\cup X^b\cup Z)\colon |A|=k, |R\cup X^b|=\ell, A\in \mathrm{SF}[G]\}\big|\\ &=w'_{k,\ell}+w''_{k,\ell}.\end{aligned}$$

Hence, from (15) and (16) we infer that the number of subsets $A \in \mathrm{SF}[G]$, for which $|X| < (m-n/2)/10^6$ but $|Y| \ge (m-n/2)/100$, is bounded from above by

$$\sum_k \sum_\ell \frac{w'_{k,\ell} + w''_{k,\ell}}{\binom{k-\ell}{\lfloor k/2 \rfloor}} = O\big(n^2\big(2^{n/2-2\cdot 10^{-13}n} + 2^{n/2-10^{-12}n}\big)\big) = O(2^{n/2-10^{-13}n}).$$

CASE 3: $|X| > (m-n/2)/10^6$.

As in the previous case we first select all elements from $R$ and then count pairs $(Z, A \setminus (R \cup Z))$, where $Z \subseteq A \setminus R$ and $|Z| = \lfloor |A|/2 \rfloor$.

Thus, fix $k = |A|$ and $\ell = |R|$. Let $\tilde{w}'_{k,\ell}$ count pairs $(Z, A \setminus (R \cup Z))$ such that $Z \subseteq A \setminus R$, $|Z| = \lfloor |A|/2 \rfloor$, and the number $s(Z, A)$ of elements $(b', b'') \in X$ for which $b', b'' \in R \cup Z$ is at most $|X|/10$. As before, we estimate $s(\mathcal{Z}, A)$ for the random subset $\mathcal{Z}$ of $A \setminus R$ of $\lfloor |A|/2 \rfloor$ elements.

The distribution of $s(\mathcal{Z}, A)$ is neither hypergeometric nor binomial, so we cannot apply Lemma 5 directly. Thus, instead of $\mathcal{Z}$, we study the random set $\mathcal{X}$, obtained by putting an element $x \in A \setminus R$ into $\mathcal{X}$ with probability 2/5, independently for each $x \in A \setminus R$. Since the function $s(\cdot, A)$ is non-decreasing, we have

$$\begin{aligned}\mathrm{Prob}\{s(\mathcal{Z}, A) \le |X|/10\} \le\, & \mathrm{Prob}\{|\mathcal{X}| \ge |\mathcal{Z}|\} \\ & + \mathrm{Prob}\big\{\{s(\mathcal{X}, A) \le |X|/10\} \wedge \{|\mathcal{X}| \le |\mathcal{Z}|\}\big\} \\ \le\, & \mathrm{Prob}\{|\mathcal{X}| \ge \lfloor |A|/2 \rfloor\} + \mathrm{Prob}\{s(\mathcal{X}, A) \le |X|/10\}.\end{aligned}$$

Note that $\mathcal{X}$ is a binomially distributed random variable with expectation

$$\mathrm{E}\mathcal{X} = \frac{2}{5}|A \setminus R| \le \frac{2}{5}|A|.$$

Furthermore, the random variable $s(\mathcal{X}, A)$ is the sum of $|X|$ zero-one independent random variables $\{I_d\colon (b', b'') \in X\}$, where for each $(b', b'') \in X$,

$$\mathrm{Prob}\{I_d = 1\} = \big(2/5\big)^{2-|\{b', b''\} \cap (A \setminus R)|},$$

so that

$$\mathrm{E}s(\mathcal{X}, A) \ge \frac{4}{25}|X|.$$

Hence, Lemma 5 implies that

$$\begin{aligned}\mathrm{Prob}\{s(\mathcal{Z}, A) \le |X|/10\} & \le \exp(-|A|/50) + \exp(-|X|/100) \\ & \le 2\exp(-|X|/100) \le 2\exp\Big(-\frac{m_0 - n/2}{10^6}\Big).\end{aligned}$$

Thus, as in the Case 2, since $|R-R| \geq n/2 - 3\cdot 10^{-17}n$, one obtains

$$\begin{aligned}\frac{\tilde{w}'_{k,\ell}}{\binom{k-\ell}{\lfloor k/2\rfloor}} &\leq \sum_{m\geq m_0}\binom{n}{\ell}2^{n-|R-R|}2\exp\Big(-\frac{m_0-n/2}{10^6}\Big)\\ &\leq n^{n^{6/7}}2^{n/2+4\cdot 10^{-17}n+1}\exp(-10^{-16}n)\\ &= O(2^{n/2-10^{-17}n}).\end{aligned}$$

In order to estimate the number $\tilde{w}''_{k,\ell}$ of pairs $(Z, A\setminus(R\cup Z))$ such that $|A|=k$, $|R|=\ell$ and $s(Z,A) > |X|/10$, we remark that in this case

$$\big|(R\cup Z)-(R\cup Z)\big| \geq |R-R| + \frac{|X|}{10} \geq \frac{n}{2} + \frac{n}{10^{15}}.$$

Hence

$$\frac{\tilde{w}''_{k,\ell}}{\binom{k-\ell}{\lfloor k/2\rfloor}} \leq \sum_{m\geq m_0}\binom{n}{\ell}\binom{n/2+3\cdot 10^{-17}n}{\lfloor k/2\rfloor}\cdot\binom{n/2-(2m-n)/3\cdot 10^7-\lfloor k/2\rfloor}{k-\ell-\lfloor k/2\rfloor}.$$

Thus, arguing as in Case 2, one can bound $\frac{\tilde{w}''_{k,\ell}}{\binom{k-\ell}{\lfloor k/2\rfloor}}$ from above by

$$\begin{aligned}n^{n^{6/7}}\sum_{m\geq m_0} 2^{n/2+3\cdot 10^{-17}n}&\Big(\frac{n/2-n/10^{15}}{n/2+3\cdot 10^{-17}n}\Big)^{k-\ell-\lfloor k/2\rfloor}\\ &\leq n^{n^{6/7}}2^{n/2+4\cdot 10^{-17}n}\Big(1-\frac{1}{10^{15}}\Big)^{0.1n}\\ &\leq n^{n^{6/7}}2^{n/2+4\cdot 10^{-17}n}2^{-10^{-15}n} = O(2^{n/2-10^{-16}n}).\end{aligned}$$

Consequently, as in the previous case, one can bound the number of sum-free subsets $A$ of $G$ with $|X| > (m-n/2)/10^6$ by

$$\sum_k\sum_\ell \frac{\tilde{w}'_{k,\ell}+\tilde{w}''_{k,\ell}}{\binom{k-\ell}{\lfloor k/2\rfloor}} = O(n^2(2^{n/2-10^{-17}n}+2^{n/2-10^{-16}n})) = O(2^{n/2-10^{-18}n}).$$

This completes the proof of the Main Lemma.

## References

[A91] N. Alon, *Independent sets in regular graphs and sum-free subsets of finite groups,* Israel Journal of Mathematics **73** (1991), 247–256.

[B98] Y. Bilu, *Sum-free sets and related sets,* Combinatorica **18** (1998), 449–459.

[CE90] P. J. Cameron and P. Erdős, *On the number of sets of integers with various properties,* in *Number Theory* (R. A. Mollin, ed.), de Gruyter, Berlin, 1990, pp. 61–79.

[JLR00] S. Janson, T. Łuczak, and A. Ruciński, *Random Graphs,* Wiley, New York, 2000.

[Ke60] J. H. B. Kemperman, *On small sumsets in Abelian group,* Acta Mathematica **103** (1960), 63–88.

[Kn53] M. Kneser, *Abschtzung der asymptotischen Dichte von Summenmengen,* Mathematische Zeitschrift **58** (1953), 459–484.

[Kn55] M. Kneser, *Ein Satz über abelschen Gruppen mit Anwendungen auf die Geometrie der Zahlen,* Mathematische Zeitschrift **61** (1955), 429–434.

[LS] V. Lev and T. Schoen, *Cameron–Erdős modulo a prime,* submitted.