A REFINEMENT OF THE CAMERON-ERDŐS CONJECTURE

NOGA ALON, JÓZSEF BALOGH, ROBERT MORRIS, AND WOJCIECH SAMOTIJ

Abstract. In this paper we study sum-free subsets of the set $\{1, \ldots, n\}$, that is, subsets of the first n positive integers which contain no solution to the equation $x + y = z$. Cameron and Erdős conjectured in 1990 that the number of such sets is $O(2^{n/2})$. This conjecture was confirmed by Green and, independently, by Sapozhenko. Here we prove a refined version of their theorem, by showing that the number of sum-free subsets of $[n]$ of size m is $2^{O(n/m)}\binom{\lceil n/2 \rceil}{m}$, for every $1 \leqslant m \leqslant \lceil n/2 \rceil$. For $m \geqslant \sqrt{n}$, this result is sharp up to the constant implicit in the $O(\cdot)$. Our proof uses a general bound on the number of independent sets of size m in 3-uniform hypergraphs, proved recently by the authors, and new bounds on the number of integer partitions with small sumset.

1. Introduction

What is the structure of a typical set of integers, of a given density, which avoids a certain arithmetic sub-structure? This fundamental question underlies much of Additive Combinatorics, and has been most extensively studied when the forbidden structure is a k-term arithmetic progression, see for example [22, 23, 29, 40, 48]. General systems of linear equations have also been studied, beginning with Rado [37] in 1933, and culminating in the recent advances of Green, Tao and Ziegler [30, 31]. The subject is extremely rich, and questions of this type have been attacked with tools from a wide variety of areas of mathematics, from Graph Theory to Number Theory, and from Ergodic Theory to Harmonic Analysis. See [49] for an excellent introduction to the area.

In this paper we shall consider *sum-free* sets of integers, that is, sets of integers which contain no solution of the equation $x + y = z$. It is easy to see that the odd numbers and the set $\{\lfloor n/2 \rfloor + 1, \ldots, n\}$ are the largest such subsets of $[n] = \{1, \ldots, n\}$. Both of these sets have $\lceil n/2 \rceil$ elements, and therefore there are at least $2^{\lceil n/2 \rceil}$ sum-free sets in $[n]$. In 1990, Cameron and Erdős [12] conjectured that this trivial lower bound is within a constant factor of the truth, that is, that the set $[n]$ contains only $O(2^{n/2})$ sum-free sets. Despite various attempts [2, 10, 19], their conjecture remained open for over ten years, until it was confirmed by Green [25] and, independently, by Sapozhenko [43]. We shall prove a natural generalization of the Cameron-Erdős Conjecture, by bounding the number of sum-free subsets of $[n]$ of size m, for all $1 \leqslant m \leqslant \lceil n/2 \rceil$. Moreover, we shall also give a quite

---

Date: March 4, 2013.

Research supported in part by: (NA) an ERC advanced grant, a USA-Israeli BSF grant, and the Israeli I-Core program; (JB) NSF CAREER Grant DMS-0745185, UIUC Campus Research Board Grant 11067, and OTKA Grant K76099; (RM) a CNPq bolsa de Produtividade em Pesquisa; (WS) ERC Advanced Grant DMMCA, and a Trinity College JRF.

precise structural description of almost all sum-free subsets of $[n]$ of size $m \geqslant C\sqrt{n \log n}$. Our proof uses a general bound on the number of independent sets of size m in 3-uniform hypergraphs, proved in [3], which allows one to deduce asymptotic structural results in the sparse setting (in fact, for all $m \gg \sqrt{n}$) from stability results in the dense setting (see Theorem 2.1). The dense stability result we shall use (see Proposition 2.2) was proved by Green [25]. The second main ingredient in the proofs of our main theorems will be some new bounds on the number of sets of integers with small sumset (see Theorems 1.3 and 1.4). Finally, we shall use Freiman's $3k - 4$ Theorem (see below) to count sets with an extremely small sumset.

The study of sum-free sets of integers dates back to 1916, when Schur [47] proved that if n is sufficiently large, then every r-colouring of $[n]$ contains a monochromatic triple (x, y, z) with $x + y = z$. (Such triples are thus often referred to as *Schur triples*.) Sum-free subsets of general Abelian groups have also been studied for many years, see e.g. [1, 4, 11, 50]. Diananda and Yap [16] and Green and Ruzsa [28] determined the maximum density $\mu(G)$ of a sum-free set in any finite Abelian group G, and in [28] it was moreover shown that any such group G has $2^{(1+o(1))\mu(G)|G|}$ sum-free subsets. For the group $\mathbb{Z}_p$, Sapozhenko [44] determined the number of sum-free subsets up to a constant factor, and for finite Abelian groups of Type I (those for which $|G|$ has a prime divisor $q \equiv 2 \pmod 3$) Green and Ruzsa [28] were able to determine the asymptotic number of sum-free subsets of G.

One of the most significant recent developments in Combinatorics has been the formulation and proof of various 'sparse analogues' of classical extremal, structural and Ramsey-type results. Beginning over 20 years ago (see, e.g., [6, 34, 38, 39]), and culminating in the recent breakthroughs of Conlon and Gowers [14] and Schacht [46] (see also [8, 21, 42, 45]), enormous progress has been made in understanding extremal structures in sparse random objects. For example, it is now known (see [14, 46]) that the theorem of Szemerédi [48] on k-term arithmetic progressions extends to sparse random sets of density $p \gg n^{-1/(k-1)}$, but not to those of density $p \ll n^{-1/(k-1)}$. A sparse analogue of Schur's Theorem was proved by Graham, Rödl and Ruciński [24], who showed that if $p \gg 1/\sqrt{n}$ and B is a p-random subset[1] of $\mathbb{Z}_n$, then with high probability every 2-colouring of B contains a monochromatic solution of $x + y = z$. A sharp version of this theorem was proved by Friedgut, Rödl, Ruciński and Tetali [20], but the extremal version, that is, the problem of determining the (asymptotic) size of the largest sum-free subset of a p-random subset of $\mathbb{Z}_n$, was open for 15 years before being resolved by Conlon and Gowers [14] and Schacht [46]. Even more recently, Balogh, Morris and Samotij [7] sharpened this result by proving that, for any finite Abelian group G of Type I(q)[2], if $|G| = n$ and $pn \geqslant C(q)\sqrt{n \log n}$, then with high probability every maximum-size sum-free subset of a p-random subset of G is contained in some sum-free subset of G of maximum size. In the case $G = \mathbb{Z}_{2n}$, they determined the sharp threshold.

For structural and enumerative problems, such as that addressed by our main theorem, results are known in only a few special cases. For example, Osthus, Prömel and Taraz [36]

[1]A p-random subset of a set X is a random subset of X, where each element is included with probability p, independently of all other elements.

[2]An Abelian group is of Type I(q) if $q \equiv 2 \pmod 3$ is the smallest such prime divisor of $|G|$.

proved that if $m \geqslant \big(\frac{\sqrt{3}}{4}+\varepsilon\big)n^{3/2}\sqrt{\log n}$, then almost all triangle-free graphs with m edges are bipartite, and that the constant $\sqrt{3}/4$ is best possible. This can be seen as a sparse version of the classical theorem of Erdős, Kleitman and Rothschild [17], which states that almost all triangle-free graphs are bipartite. In [3], the authors proved a sparse analogue of the theorem of Green and Ruzsa [28] mentioned above, by showing that if $m \geqslant C(q)\sqrt{n\log n}$, then almost every sum-free m-subset[3] of G is contained in some maximum-size sum-free set. We remark that there are only at most $|G|$ maximum-size sum-free subsets of such a group G, and that moreover they admit an elegant description.

In this paper we shall be interested in the corresponding question for the set $[n]$. As noted above, Cameron and Erdős [12] conjectured, and Green [25] and Sapozhenko [43] proved, that there are only $O(2^{n/2})$ sum-free subsets of $[n]$. Our main result is the following 'sparse analogue' of this theorem.

Theorem 1.1. *There exists a constant $C > 0$ such that, for every $n \in \mathbb{N}$ and every $1 \leqslant m \leqslant \lceil n/2 \rceil$, the set $[n]$ contains at most $2^{Cn/m}\binom{\lceil n/2\rceil}{m}$ sum-free sets of size m.*

If $m \geqslant \sqrt{n}$, then Theorem 1.1 is sharp up to the value of C, since in this case there is a constant $c > 0$ such that there are at least $2^{cn/m}\binom{n/2}{m}$ sum-free m-subsets of $[n]$ (see Proposition 3.1). Note that if $m \leqslant \sqrt{n}$ then the result is trivial, since in this case our upper bound is greater than $\binom{n}{m}$; however, if $m \ll \sqrt{n}$ then the problem is less interesting, since in this case it is relatively straightforward[4] to show that there are roughly $e^{-\mu}\binom{n}{m}$ sum-free m-subsets of $[n]$, where μ denotes the expected number of Schur triples in a random m-set. Since there are fewer than $2^{n/3}$ subsets of $[n]$ with at most $n/100$ elements, Theorem 1.1 easily implies the Cameron-Erdős Conjecture. However, Theorem 1.1 only implies that there are $O(2^{n/2})$ sum-free subsets of $[n]$, whereas Green [25] and Sapozhenko [43] proved that there are asymptotically $c(n)2^{n/2}$ such sets, where $c(n)$ takes two different constant values according to whether n is even or odd. Since for us the parity of n will not matter, we shall assume for simplicity throughout the paper that n is even; the proof in the case n is odd is identical.

We shall also prove the following structural description of a typical sum-free m-subset of $[n]$. Let O_n denote the set of odd numbers in $[n]$.

Theorem 1.2. *There exists $C > 0$ such that if $n \in \mathbb{N}$ and $m \geqslant C\sqrt{n\log n}$, then almost every sum-free subset $I \subset [n]$ of size m satisfies either $I \subset O_n$, or*

$$|S(I)| \leqslant \frac{Cn}{m} + \omega(n) \qquad \text{and} \qquad \sum_{a\in S(I)}\left(\frac{n}{2}-a\right) \leqslant \frac{Cn^3}{m^3} + \omega(n),$$

where $S(I) = \{x \in I : x \leqslant n/2\}$, and $\omega(n) \to \infty$ arbitrarily slowly as $n \to \infty$.

We remark that the upper bounds on $|S(I)|$ and $k(I) := \sum_{a\in S(I)}(n/2-a)$ in Theorem 1.2 are sharp up to a constant factor (see Section 6). Indeed, we shall show that if $m = o(n)$, then almost all sum-free m-sets $I \subset [n]$ have $|S(I)| = \Omega(n/m)$ and $k(I) = \Omega(n^3/m^3)$.

[3] An m-subset of a set X is simply a subset of X of size m.

[4] The proof is a standard application of the FKG and Janson inequalities, see Section 4, or, e.g., [5, 33].

Our proof of Theorems 1.1 and 1.2 has two main components. The first is a bound on the number of independent m-sets in 3-uniform hypergraphs (see Theorem 2.1), which was proved in [3], and used there to determine the asymptotic number of sum-free m-subsets of a finite Abelian group G such that $|G|$ has a prime factor $q \equiv 2 \pmod 3$, for every $m \geqslant C(q)\sqrt{n \log n}$. Using this theorem, together with a stability result from [25] (which follows from a result of Lev, Łuczak and Schoen [35]) it will be straightforward to bound the number of sum-free m-sets which contain at least δm even numbers, and at least δm elements less than $n/2$.

The second component involves counting *restricted integer partitions with small sumset.* Let $p(k)$ denote the number of integer partitions of k, so, for example, $p(3) = 3$ since $3 = 2 + 1 = 1 + 1 + 1$. In 1918, Hardy and Ramanujan [32] obtained an asymptotic formula for $p(k)$, proving that

$$p(k) \;=\; \frac{1 + o(1)}{4k\sqrt{3}} e^{\pi\sqrt{2k/3}}.$$

We shall study the following type of 'restricted' partition. Let $p^*_\ell(k)$ denote the number of integer partitions of k into ℓ distinct parts, i.e., the number of sets $S \subset \mathbb{N}$ such that $|S| = \ell$ and $\sum_{a\in S} a = k$. Thus, for example, $p^*_3(8) = 2$, since $8 = 5 + 2 + 1 = 4 + 3 + 1$. It is straightforward to show that $p^*_\ell(k) \leqslant \left(\frac{e^2 k}{\ell^2}\right)^\ell$, see Lemma 5.1.

We shall bound the number of such partitions under the following more restrictive condition. Recall that, given sets $A, B \subset \mathbb{N}$, the sumset $A + B$ is defined to be the set $\{a + b : a \in A, b \in B\}$. The following theorem bounds the number of partitions of k into ℓ distinct parts, such that the resulting set S has 'small' sumset $S + S$.

Theorem 1.3. *For every $c_0 > 0$ and $\delta > 0$, there exists a $C = C(\delta, c_0) > 0$ such that the following holds. If $\ell^3 \geqslant Ck$ and $c \geqslant c_0$, then there are at most*

$$2^{\delta\ell}\left(\frac{2cek}{3\ell^2}\right)^\ell$$

sets $S \subset \mathbb{N}$ with $|S| = \ell$, $\sum_{a\in S} a = k$ and $|S + S| \leqslant ck/\ell$.

Sets with small sumset are a central object of interest in Combinatorial Number Theory, and have been extensively studied in recent years (see, e.g., [49]). It is easy to see that if $A, B \subset \mathbb{Z}$, then $|A + B| \geqslant |A| + |B| - 1$, with equality if and only if A and B are arithmetic progressions with the same common difference. The Cauchy-Davenport Theorem, proved by Cauchy [13] in 1813 and rediscovered by Davenport [15] in 1935, says that this result extends to the group $\mathbb{Z}_p$; more precisely, that

$$|A + B| \;\geqslant\; \min\big\{|A| + |B| - 1,\, p\big\}.$$

Many extensions of these results are now known; for example, the Freiman-Ruzsa Theorem (see [18, 41]) states that if $A \subset \mathbb{Z}$ and $|A + A| \leqslant C|A|$, then A is contained in a $O(1)$-dimensional generalized arithmetic progression of size $O(|A|)$. This result itself has many generalizations, culminating in the very recent theorem of Breuillard, Green and Tao [9], which is stated in the language of approximate groups.

Despite the enormous interest in such problems, very little seems to be known about the *number* of different sets with small sumset (see [26], for example). The following classical result, proved by Freiman [18] in 1959, implies a bound for sets with so-called 'doubling constant' less than 3.

Freiman's $3k-4$ Theorem. *If $A \subset \mathbb{Z}$ satisfies $|A+A| \leqslant 3|A|-4$, then A is contained in an arithmetic progression of size at most $|A+A|-|A|+1$.*

Observe that this implies that, for all $\lambda < 3$, there are at most $2^{o(\ell)}\binom{(\lambda-1)\ell}{\ell}$ sets $S \subset \mathbb{Z}$ such that $|S| = \ell$ and $|S+S| \leqslant \lambda\ell$, up to equivalence under translation and dilation. (That is, if we assume that $\min(S) = 0$ and S has no common divisor greater than one.) Our final theorem, which also follows from the proof of Theorem 1.3, provides an upper bound of this type whenever $|S+S| = O(|S|)$.[5] The following result will be crucial in the proof of Theorems 1.1 and 1.2 in the case $m = \Theta(n)$.

Theorem 1.4. *Let $\delta > 0$, and suppose that $\ell \in \mathbb{N}$ is sufficiently large and that $k \leqslant \ell^2/\delta$. Then for each $\lambda \geqslant 2$, there are at most*

$$2^{\delta\ell}\left(\frac{(4\lambda-3)e}{6}\right)^{\ell}$$

sets $S \subset \mathbb{N}$ with $|S| = \ell$, $\sum_{a \in S} a = k$, and $|S+S| \leqslant \lambda\ell$.

Theorems 1.3 and 1.4 are sufficient for our purposes; however, we believe the following stronger bound to be true.

Conjecture 1.5. *For every $\delta > 0$, there exists $C > 0$ such that the following holds. If $m \geqslant C\sqrt{N}$ and $m \geqslant C\log n$, then there are at most*

$$2^{\delta m}\binom{N/2}{m}$$

sets $S \subset [n]$ with $|S| = m$ and $|S+S| \leqslant N$.

Since $|S+S| \leqslant N$ for every m-subset $S \subset [N/2]$, the conjecture (if true) is close to optimal. Note that the condition $m \geqslant C\log n$ implies that $n \leqslant 2^{m/C}$, and thus guarantees that the number of translates of a given set S is negligible.

The rest of the paper is organised as follows. In Section 2, we shall recall the general structural theorem from [3] and deduce from it a bound on the number of sum-free m-sets which contain at least δm even numbers, and at least δm elements less than $n/2$. In Section 3 we shall prove a lower bound on the number of sum-free m-subsets of $[n]$, and in Section 4 we shall use Janson's inequality to bound the number of sum-free sets which contain at most δm even numbers. In Section 5 we shall prove Theorems 1.3 and 1.4. Finally, in Sections 6 and 7, we shall prove Theorems 1.1 and 1.2.

[5] Throughout the paper, $f(n) = O(g(n))$ means there exists an absolute constant, independent of all other variables, such that $f(n) \leqslant Cg(n)$.

2. Preliminaries

In this section we shall recall some of the main tools we shall use in the proofs of Theorems 1.1 and 1.2, and deduce that almost all sum-free m-sets $I \subset [n]$ either contain at most δm even elements, or satisfy $|I \setminus B| \leqslant \delta m$ for some interval B of length $n/2$. We remark that it will be fairly easy to bound from above the number of such sets if B contains more than δn elements less than $n/2$ (see Sections 4 and 6); the main difficulty will be counting sum-free sets that are almost contained in the interval $\{n/2+1, \ldots, n\}$.

2.1. A structural theorem for 3-uniform hypergraphs.

We begin by recalling from [3] our main tool: a theorem which allows one to deduce asymptotic structural results for sparse sum-free sets from stability results for dense sum-free sets. It is stated in the language of (sequences of) general 3-uniform hypergraphs $\mathcal{H} = (\mathcal{H}_n)_{n \in \mathbb{N}}$, where $|V(\mathcal{H}_n)| = n$. Throughout this section, the reader should think of the hypergraph $\mathcal{H}_n$ as encoding the Schur triples (that is, triples (x, y, z) with $x + y = z$) in $[n]$.

We next recall the type of dense stability property which we shall consider. Let $\alpha \in (0,1)$ and let $\mathcal{B} = (\mathcal{B}_n)_{n \in \mathbb{N}}$, where $\mathcal{B}_n$ is a family of subsets of $V(\mathcal{H}_n)$.

Definition. A sequence of hypergraphs $\mathcal{H} = (\mathcal{H}_n)_{n \in \mathbb{N}}$ is said to be $(\alpha, \mathcal{B})$-*stable* if for every $\gamma > 0$ there exists $\beta > 0$ such that the following holds for every $n \in \mathbb{N}$. If $A \subset V(\mathcal{H}_n)$ satisfies $|A| \geqslant (\alpha - \beta)n$, then either $e(\mathcal{H}_n[A]) \geqslant \beta e(\mathcal{H}_n)$, or $|A \setminus B| \leqslant \gamma n$ for some $B \in \mathcal{B}_n$.

Roughly speaking, a sequence of hypergraphs $(\mathcal{H}_n)$ is $(\alpha, \mathcal{B})$-stable if for every $A \subseteq V(\mathcal{H}_n)$ such that $|A|$ is almost as large as the independence number for $\mathcal{H}_n$, the set A is either very close to some 'extremal' set $B \in \mathcal{B}_n$, or it contains many (i.e., a positive fraction of all of the) edges of $\mathcal{H}_n$.

If $\mathcal{H}_n$ is a hypergraph and $m \in \mathbb{N}$, then let $\mathrm{SF}(\mathcal{H}_n, m)$ denote the collection of independent sets of size m in $\mathcal{H}_n$. Given a family of sets $\mathcal{B}_n$ and $\delta > 0$, we define

$$\mathrm{SF}_{\geqslant}^{(\delta)}(\mathcal{H}_n, \mathcal{B}_n, m) \,=\, \Big\{ I \in \mathrm{SF}(\mathcal{H}_n, m) : \, |I \setminus B| \geqslant \delta m \text{ for every } B \in \mathcal{B}_n \Big\}.$$

Roughly speaking, the collection $\mathrm{SF}_{\geqslant}^{(\delta)}$ contains all independent sets that are 'far' from all 'extremal' sets. Finally, for each $T \subset V(\mathcal{H}_n)$, let $d_{\mathcal{H}_n}(T) = \big|\big\{ e \in \mathcal{H}_n \,:\, T \subset e \big\}\big|$ and define

$$\Delta_2(\mathcal{H}_n) \,=\, \max\big\{ d_{\mathcal{H}_n}(T) \,:\, T \subset V(\mathcal{H}_n),\, |T| = 2 \big\}.$$

Note that if $\mathcal{H}_n$ encodes Schur triples in $[n]$, then $\Delta_2(\mathcal{H}_n) \leqslant 2$.

The following theorem, which was proved in [3], shows that if $\mathcal{H}$ is $(\alpha, \mathcal{B})$-stable and $m \gg \sqrt{n}$, then there are very few independent sets (i.e., sum-free sets) of size m in $\mathcal{H}_n$ which are far from every set $B \in \mathcal{B}_n$. We write $\|\mathcal{B}_n\| = \max\big\{ |B| : B \in \mathcal{B}_n \big\}$.

Theorem 2.1 (Theorem 4.1 of [3]). *Let $\alpha > 0$ and let $\mathcal{H} = (\mathcal{H}_n)_{n \in \mathbb{N}}$ be a sequence of 3-uniform hypergraphs which is $(\alpha, \mathcal{B})$-stable, has $e(\mathcal{H}_n) = \Theta(n^2)$ and $\Delta_2(\mathcal{H}_n) = O(1)$. If $\|\mathcal{B}_n\| \geqslant \alpha n$, then for every $\delta > 0$, there exists a $C > 0$ such that the following holds. If*

$m \geqslant C\sqrt{n}$ and n is sufficiently large, then

$$\big|\mathrm{SF}^{(\delta)}_{\geqslant}(\mathcal{H}_n, \mathcal{B}_n, m)\big| \;\leqslant\; \big(1-\varepsilon\big)^m \binom{\|\mathcal{B}_n\|}{m}$$

for some $\varepsilon = \varepsilon(\mathcal{H}, \delta) > 0$.

In the next subsection, we shall use this theorem, together with a result of Green [25], to deduce an approximate version of Theorem 1.2.

2.2. **Green's stability theorem.** Let $\mathcal{H} = (\mathcal{H}_n)_{n\in\mathbb{N}}$ be the sequence of hypergraphs which encodes Schur triples in $[n]$; that is, $V(\mathcal{H}_n) = [n]$ and $\{x,y,z\} \in E(\mathcal{H}_n)$ whenever $x+y=z$. The following stability result, due to Green [25], implies that $\mathcal{H}$ is $(\alpha, \mathcal{B})$-stable, where $\alpha = 1/2$ and

$$\mathcal{B}_n \;=\; \Big\{ \big\{a+1, \ldots, a+n/2\big\} \,:\; 0 \leqslant a \leqslant n/2 \Big\} \cup \big\{ O_n \big\}, \qquad (1)$$

where, as before, O_n denotes the odd numbers in $[n]$.

Proposition 2.2 ([25, Proposition 7])**.** *For any $\gamma > 0$, if $\beta = \beta(\gamma) > 0$ is sufficiently small, then the following holds. If $A \subset [n]$ with $|A| \geqslant (1/2 - \beta)n$, then either A contains at least βn^2 Schur triples, or $|A \setminus B| \leqslant \gamma n$ for some $B \in \mathcal{B}_n$.*

Using Theorem 2.1 and Proposition 2.2, we easily obtain the following corollary.

Proposition 2.3. *For every $\delta > 0$, there exist constants $C > 0$ and $\varepsilon > 0$ such that the following holds for every $n \in \mathbb{N}$ and $m \geqslant C\sqrt{n}$. There are at most*

$$2^{-\varepsilon m}\binom{n/2}{m}$$

sum-free subsets $I \subset [n]$ of size m such that $|I \setminus B| > \delta m$ for every $B \in \mathcal{B}_n$.

In particular, for almost every sum-free set $I \subset [n]$ of size m, either $|I \setminus O_n| \leqslant \delta m$, or $|I \setminus B| \leqslant \delta m$ for some interval B of length $n/2$.

Proof. Let $\mathcal{H} = (\mathcal{H}_n)_{n\in\mathbb{N}}$ be the sequence of hypergraphs which encodes Schur triples in $[n]$, as above, and let $\mathcal{B} = (\mathcal{B}_n)_{n\in\mathbb{N}}$ be the collection of intervals of length $n/2$, plus the odds, as in (1). Set $\alpha = 1/2$, and observe that $\mathcal{H}$ is $(\alpha, \mathcal{B})$-stable, by Proposition 2.2.

Now, by Theorem 2.1, if $C = C(\delta) > 0$ is sufficiently large, and $m \geqslant C\sqrt{n}$, then

$$\big|\mathrm{SF}^{(\delta)}_{\geqslant}(\mathcal{H}_n, \mathcal{B}_n, m)\big| \;\leqslant\; 2^{-\varepsilon m}\binom{\|\mathcal{B}_n\|}{m} \;=\; 2^{-\varepsilon m}\binom{n/2}{m}$$

for some $\varepsilon = \varepsilon(\delta) > 0$. Since there are at least $\binom{n/2}{m}$ sum-free m-subsets of $[n]$, it follows that for almost every such set I we have $|I \setminus B| \leqslant \delta n$ for some $B \in \mathcal{B}_n$, as required. ☐

As we remarked above, it will be relatively straightforward to count the sets I that contain fewer than δm even elements, using Janson's inequality (see Section 4), and those that contain more than δm elements less than $n/2$, using induction on n (see Section 6). Thus, Proposition 2.3 essentially reduces the problem of counting sum-free m-sets in $[n]$ to counting the sum-free sets that are almost contained in the interval $\{n/2+1, \ldots, n\}$.

2.3. **Binomial coefficient inequalities.** We shall make frequent use of some simple inequalities involving binomial coefficients; for convenience, we collect them here. Note first that $\binom{a}{b} \leqslant \left(\frac{ea}{b}\right)^b$ and that $\binom{a}{b}$ is increasing in a. Next, observe that if $a > b > c \geqslant 0$, then

$$\binom{a}{b-c} \leqslant \left(\frac{b}{a-b}\right)^c \binom{a}{b} \qquad \text{and} \qquad \binom{a-c}{b} \leqslant \left(\frac{a-c}{a}\right)^b \binom{a}{b}, \tag{2}$$

and hence

$$\binom{a-c}{b-d} \leqslant \left(\frac{a-c}{a}\right)^{b-d} \left(\frac{b}{a-b}\right)^d \binom{a}{b}. \tag{3}$$

We shall also use several times the observation that, for every $a \geqslant 1$ and $b > 0$,

$$\sum_{k=1}^{\infty} k^a e^{-bk} \leqslant c \cdot \frac{\Gamma(a+1)}{b^{a+1}} \leqslant C \left(\frac{a}{b}\right)^{a+1} e^{-(a+1)}, \tag{4}$$

where $\Gamma(\cdot)$ is Euler's Gamma function, for some absolute constants $C > c > 0$.

For other standard probabilistic bounds, such as the FKG inequality and Chernoff's inequality, we refer the reader to [5].

3. A lower bound on the number of sum-free sets

In this section we shall prove the following simple proposition, which shows that the bound in Theorem 1.1 is tight.

Proposition 3.1. *If $m \geqslant \sqrt{n}$, then there are $2^{\Omega(n/m)}\binom{n/2}{m}$ sum-free subsets of $[n]$ of size m.*

Proof. Let $c > 0$ be a sufficiently small absolute constant and set $a = cn^2/m^2$. We claim that if S is a uniformly chosen random m-subset of $U = \{n/2 - a, \ldots, n\}$, then

$$\mathbb{P}\big(S \text{ is sum-free}\big) \geqslant \exp\left(-\frac{cn}{2m}\right). \tag{5}$$

In order to prove (5), we shall in fact choose the elements of S independently at random with probability $p = 4m/n$, and bound $\mathbb{P}_p\big(S \text{ is sum-free} \,|\, |S| = m\big)$, which is clearly equivalent. (Note that the proposition is trivial if $m = \Omega(n)$, so we may assume that p is sufficiently small.)

First observe that there are at most $a^2 + a$ triples $\{x, y, z\}$ in U with $x + y = z$, and at most $a + 1$ pairs $\{x, y\}$ in U with $2x = y$. Thus, by the FKG inequality,

$$\mathbb{P}_p\big(S \text{ is sum-free}\big) \geqslant \big(1 - p^3\big)^{a^2+a}\big(1 - p^2\big)^{a+1} \geqslant \exp\left(-\frac{cn}{3m}\right)$$

since $c > 0$ is sufficiently small, by our choices of a and p. Next, note that, by Chernoff's inequality,

$$\mathbb{P}_p\big(|S| < m\big) \leqslant e^{-cm} \leqslant e^{-cn/m},$$

since $m \geqslant \sqrt{n}$. Finally, observe that $g(t) = \mathbb{P}_p\big(S \text{ is sum-free} \,|\, |S| = t\big)$ is decreasing in t.

It follows immediately that

$$\mathbb{P}_p\big(S \text{ is sum-free} \,|\, |S| = m\big) \geqslant \exp\left(-\frac{cn}{2m}\right),$$

which proves (5). Hence the number of sum-free m-sets in $\{n/2-a,\dots,n\}$ is at least

$$\binom{n/2+a}{m}\exp\left(-\frac{cn}{2m}\right) \geqslant \binom{n/2}{m}\exp\left(\frac{am}{n}-\frac{cn}{2m}\right) = \exp\left(\frac{cn}{2m}\right)\binom{n/2}{m},$$

where the inequality follows from (2) and the fact that $\left(1+\frac{2a}{n}\right) \geqslant e^{a/n}$. □

4. Janson argument

In this section, we shall count the sum-free sets that have few even elements. Recall that O_n denotes the odd numbers in $[n]$.

Proposition 4.1. *If $\delta > 0$ is sufficiently small, then there are at most $2^{O(n/m)}\binom{n/2}{m}$ sum-free subsets $I \subset [n]$ with $|I| = m$ and $|I \setminus O_n| \leqslant \delta m$, for every $m, n \in \mathbb{N}$.*

We remark that an argument similar to the one presented in this section was used in [3] in a somewhat more general context, see also [7]. Indeed, the following result was proved in [3].

Proposition 4.2 ([3, Proposition 5.1])**.** *There exists constants $\delta > 0$ and $C > 0$ such that the following holds for every $m \geqslant C\sqrt{n \log n}$. There are at most*

$$\left(1+\frac{1}{n^3}\right)\binom{n/2}{m}$$

sum-free subsets $I \subset [n]$ with $|I| = m$ and $|I \setminus O_n| \leqslant \delta m$.

Proposition 4.2 clearly implies Proposition 4.1 in the case $m \geqslant C\sqrt{n\log n}$. Furthermore, Proposition 4.1 is trivial if $m \leqslant O\big(\sqrt{n}\big)$, since then the claimed upper bound is greater than $\binom{n}{m}$. Thus, we need only consider the case $C\sqrt{n} \leqslant m \leqslant C\sqrt{n\log n}$.

Recall the following well-known result, which is an easy corollary of Janson's inequality (see [5, 33]), combined with Pittel's inequality (see [33]). We refer the reader to [3, Section 5] for a proof.

Lemma 4.3 (Hypergeometric Janson Inequality)**.** *Suppose that $\{U_i\}_{i\in J}$ is a family of subsets of an n-element set X and let $m \in \{0,\dots,n\}$. Let*

$$\mu = \sum_{i\in J}(m/n)^{|U_i|} \quad \text{and} \quad \Delta = \sum_{i\sim j}(m/n)^{|U_i\cup U_j|},$$

where the second sum is over ordered pairs (i,j) such that $i \neq j$ and $U_i \cap U_j \neq \emptyset$. Let R be a uniformly chosen random m-subset of X. Then

$$\mathbb{P}\big(U_i \not\subseteq R \text{ for all } i \in J\big) \leqslant C\cdot\max\left\{e^{-\mu/2}, e^{-\mu^2/2\Delta}\right\},$$

for some absolute constant $C > 0$.

We now turn to the proof of Proposition 4.1.

Let $C > 0$ be a sufficiently large constant, and recall that we may assume that $C\sqrt{n} \leqslant m \leqslant C\sqrt{n\log n}$. We begin by proving the following claim.

Claim. *For some constant $c > 0$, there are at most*

$$C \cdot \binom{n/2}{k} \max\left\{e^{-ckm^2/n}, e^{-cm}\right\} \binom{n/2}{m-k} \tag{6}$$

sum-free m-sets $I \subset [n]$ with $|I \setminus O_n| = k \leqslant \delta m$.

Proof of claim. Let $k \leqslant \delta m$ and let S be an arbitrary k-subset of $[n] \setminus O_n$. Let $\{U_i\}_{i \in J}$ be the collection of pairs $\{x, y\} \subset O_n$ such that either $x + y = z$ or $x - y = z$ for some $z \in S$. In order to bound the number of sum-free m-sets I with $I \setminus O_n = S$, we shall apply the Hypergeometric Janson Inequality to the collection $\{U_i\}_{i \in J}$ and the set $X = O_n$, with R a uniformly chosen random $(m - k)$-subset of X. Note that if $S \cup R$ is sum-free, then $U_i \nsubseteq R$ for all $i \in J$.

Let μ and Δ be the quantities defined in the statement of Lemma 4.3, and observe that for every even number z, there are either at least $n/10$ pairs $\{x, y\} \subset O_n$ with $x + y = z$ (if $z \geqslant n/2$), or at least $n/5$ such pairs with $x - y = z$ (if $z \leqslant n/2$). Thus $nk/20 \leqslant |J| \leqslant nk$, since each pair can be counted at most twice. Observe that each vertex $x \in O_n$ lies in at most $2k$ of the U_i. Hence

$$\mu \geqslant \frac{nk}{20} \cdot \frac{(m-k)^2}{n^2} \geqslant \frac{km^2}{30n} \quad \text{and} \quad \Delta \leqslant (2k)^2 \binom{|J|}{2k} \left(\frac{m}{n}\right)^3 \leqslant \frac{2k^2 m^3}{n^2}.$$

By the Hypergeometric Janson Inequality, if $c = 10^{-4}$ then there are at most

$$C \cdot \max\left\{e^{-ckm^2/n}, e^{-cm}\right\} \binom{n/2}{m-k}$$

sets $R \subset O_n$ of size $m - k$ such that $S \cup R$ is sum-free. Summing over choices of S, we obtain the claimed bound. □

Now, by (2) and since $m \leqslant C\sqrt{n \log n} \leqslant n/6$, if $k \geqslant n/m$ then (6) is at most

$$C \cdot \binom{n/2}{k} \left(\frac{m}{n/2 - m}\right)^k e^{-cm} \binom{n/2}{m} \leqslant C \cdot \left(\frac{3em}{2k}\right)^k e^{-cm} \binom{n/2}{m} \ll \binom{n/2}{m},$$

assuming $\delta > 0$ is sufficiently small. However, if $k \leqslant n/m$ then (6) is at most

$$C \cdot \left(\frac{3em}{2k} e^{-cm^2/n}\right)^k \binom{n/2}{m} \leqslant 2^{O(n/m)} \binom{n/2}{m}.$$

To see the final inequality, observe that (since $xe^{-x/n} \leqslant n$) we have $me^{-cm^2/n} \leqslant n/cm$, and use the fact that $k \mapsto (a/k)^k$ is maximized when $k = a/e$. This completes the proof of Proposition 4.1.

5. Partitions and sumsets

In this section we shall prove Theorems 1.3 and 1.4. Recall that

$$p_\ell^*(k) = \#\big\{\text{partitions of } k \text{ into } \ell \text{ distinct parts}\big\}.$$

We shall use the following easy upper bound on $p_\ell^*(k)$ in the proofs of Theorems 1.1 and 1.2.

Lemma 5.1. *For every $k, \ell \in \mathbb{N}$,*

$$p_\ell^*(k) \leqslant \left(\frac{e^2 k}{\ell^2}\right)^\ell.$$

Proof. Note that $p_\ell^*(k) = 0$ if $k < \binom{\ell+1}{2}$, and that the result is trivial if $\ell \leqslant 3$. So assume that $\ell \geqslant 4$, and consider putting k identical balls into ℓ labelled boxes. There are $\binom{k+\ell-1}{\ell-1} = \frac{\ell}{k+\ell}\binom{k+\ell}{\ell}$ ways to do so, and each partition of k into ℓ distinct parts is counted exactly $\ell!$ times. Using the bound $\ell! \geqslant \sqrt{2\pi\ell}\left(\frac{\ell}{e}\right)^\ell$, it follows that, if $k \geqslant \binom{\ell+1}{2}$ and $\ell \geqslant 4$, then

$$p_\ell^*(k) \leqslant \frac{1}{\ell!} \cdot \frac{\ell}{k+\ell}\binom{k+\ell}{\ell} \leqslant \frac{2}{\ell+2} \cdot \frac{1}{\sqrt{2\pi\ell}}\left(\frac{e}{\ell}\right)^\ell \left(\frac{e(k+\ell)}{\ell}\right)^\ell \leqslant \left(\frac{e^2 k}{\ell^2}\right)^\ell,$$

as required, since $\left(\frac{k+\ell}{k}\right)^\ell < e^2 < 6\sqrt{2\pi}$. □

In order to motivate the proofs of Theorems 1.3 and 1.4, we shall first sketch an easy proof of a weaker bound and an incorrect proof of a sharper one; we will use ideas from both in the actual proof. We begin with the weaker bound: given $c, \delta > 0$, let $C = C(\delta, c) > 0$ be sufficiently large, and suppose that $\ell^3 \geqslant Ck$. We claim that there are at most

$$2^{\delta\ell}\left(\frac{cek}{\ell^2}\right)^\ell \tag{7}$$

sets $S \subset \mathbb{N}$ with $|S| = \ell$, $\sum_{a \in S} a = k$, and $|S+S| \leqslant ck/\ell$. Note that this is weaker than Theorem 1.3 by a factor of $(3/2)^\ell$.

We shall count 'good' sequences $(a_1, \ldots, a_\ell)$ of length ℓ, that is, sequences such that the underlying set $S = \{a_1, \ldots, a_\ell\}$ satisfies $|S| = \ell$, $\sum_{a \in S} a = k$ and $|S+S| \leqslant ck/\ell$. Note that each such set S will appear as a sequence exactly $\ell!$ times. Set $S_j = \{a_1, \ldots, a_j\}$ and observe that $\big|(S_j + a_{j+1}) \setminus (S_j + S_j)\big| \geqslant \delta|S_j|$ for at most $\delta\ell$ indices $j \in [\ell]$, since otherwise $|S+S| \geqslant \delta(\delta\ell/2)^2 > ck/\ell$, where the last inequality follows because $\ell^3 \geqslant Ck$.

We now make a simple but key observation: that, for every set $S \subset \mathbb{N}$, there are at most $(1-\delta)^{-1}|S+S|$ elements $y \in \mathbb{N}$ such that

$$\big|(S+y) \setminus (S+S)\big| \leqslant \delta|S|. \tag{8}$$

To prove this, observe that there are $|S| \cdot |S+S|$ pairs (a,b) with $a \in S$ and $b \in S+S$, and that if (8) holds then $a + y = b$ for at least $(1-\delta)|S|$ pairs $(a,b) \in S \times (S+S)$. For each pair (a,b) there is at most one such y, and so there are at most $\frac{|S|\cdot|S+S|}{(1-\delta)|S|}$ elements y which satisfy (8), as claimed.

Thus, as in the proof of Lemma 5.1, the number of good sequences $(a_1, \ldots, a_\ell)$ is at most

$$\begin{aligned}\binom{\ell}{\delta\ell}\binom{k+\delta\ell}{\delta\ell}\left(\frac{|S+S|}{1-\delta}\right)^{(1-\delta)\ell} &\leqslant \left(\frac{e}{\delta}\right)^{\delta\ell}\left(\frac{2ek}{\delta\ell}\right)^{\delta\ell}\left(\frac{1}{|S+S|}\right)^{\delta\ell} e^{O(\delta\ell)}|S+S|^\ell \\ &\leqslant \left(\frac{2e^2}{\delta^2 c}\right)^{\delta\ell} e^{O(\delta\ell)}\left(\frac{ck}{\ell}\right)^\ell \leqslant 2^{O(\sqrt{\delta}\ell)}\left(\frac{ck}{\ell}\right)^\ell,\end{aligned} \tag{9}$$

assuming $\delta > 0$ is sufficiently small. Dividing by $\ell!$, we obtain (7), as claimed.

Next, define the *span* of a set $S \subset \mathbb{N}$ to be $\max(S) - \min(S)$ and observe that, for any set S, we have $2 \cdot \mathrm{span}(S) = \mathrm{span}(S+S)$. Let B_j denote the set of elements y (as above) for which $|(S_j + y) \setminus (S_j + S_j)| \leqslant \delta |S_j|$. Intuitively, one would expect that $B_j + S_j \approx S_j + S_j$, which implies that

$$\mathrm{span}(B_j) + \mathrm{span}(S_j) \, = \, \mathrm{span}(B_j + S_j) \, \approx \, \mathrm{span}(S_j + S_j) \, = \, 2 \cdot \mathrm{span}(S_j), \tag{10}$$

and hence $\mathrm{span}(B_j) \approx \mathrm{span}(S_j)$. This would imply that $\min(S_j) + B_j$ and $\max(S_j) + B_j$ are (almost) disjoint, since

$$\max\big(\min(S_j) + B_j\big) \, = \, \min(S_j) + \max(B_j) \, \approx \, \max(S_j) + \min(B_j) \, = \, \min\big(\max(S_j) + B_j\big).$$

Hence, if $B_j + S_j \approx S_j + S_j$, then it follows from the argument above that

$$|B_j| \, \lesssim \, \frac{|S_j + S_j|}{2},$$

which would win us a factor of roughly 2^ℓ in (9). Unfortunately, (10) does not hold in general; for example, if $S_j = [k] \cup X$, where X is a random subset of $\{k+1, \ldots, 2k\}$ of size δk, then $S_j + S_j \approx [4k]$ and $B_j \approx [3k]$, and so $\mathrm{span}(S_j) \approx 2k$ but $\mathrm{span}(B_j) \approx 3k$. Nevertheless, we shall be able to prove an approximate version of (10) (see (14), below) by considering an interval J on which S_j is sufficiently dense close to the extremal values, $\max(J)$ and $\min(J)$.

Proof of Theorem 1.3. Let $c \geqslant c_0 > 0$ and $\delta > 0$, and note that without loss of generality we may assume that $\delta = \delta(c_0)$ is sufficiently small. Let $C = C(\delta, c_0) > 0$ be sufficiently large; with foresight, we remark that $C = 1/\delta^{13}$ will suffice. Note also that if $c \geqslant 3e/2$, then the theorem follows immediately from Lemma 5.1; we shall therefore assume that $c < 3e/2$.

Given $S \subset \mathbb{N}$ with $|S| = \ell$, $\sum_{a \in S} a = k$ and $|S+S| \leqslant ck/\ell$, let $S_* = \max\big\{\delta k/\ell, \, \min(S)\big\}$ and $S^* = \min\big\{k/\delta\ell, \, \max(S)\big\}$, so $\delta k/\ell \leqslant S_* \leqslant S^* \leqslant k/\delta\ell$. Moreover, define

$$t_* \, = \, \min\Big\{ t : \big| S \cap \big[(1+\delta)^t S_*, (1+\delta)^{t+1} S_* \big] \big| \geqslant \delta^3 \ell \Big\}$$

and $t^* = \min\Big\{ t : \big| S \cap \big[(1-\delta)^{t+1} S^*, (1-\delta)^t S^* \big] \big| \geqslant \delta^3 \ell \Big\}$ (or ∞ if no such t exists), and set

$$J \, = \, [J_*, J^*] \, = \, \big[(1+\delta)^{t_*} S_*, (1-\delta)^{t^*} S^* \big].$$

Note that the interval J has 'dense ends', that is, the sets $[(1-\delta)J^*, J^*]$ and $[J_*, (1+\delta)J_*]$ each contain more than $\delta^3\ell$ elements of S. We first deal with an easy special case.

Case 1: $\max\big\{t^*, t_*\big\} \geqslant 1/\delta^2$.

Suppose first that $t^* \geqslant 1/\delta^2$. Note that S contains at most $\delta\ell$ elements greater than S^*, since $\sum_{a \in S} a = k$, and hence at most $2\delta\ell$ elements of S greater than

$$(1-\delta)^{1/\delta^2} S^* \, \leqslant \, e^{-1/\delta} \frac{k}{\delta \ell} \, \leqslant \, \frac{\delta k}{\ell}.$$

Let $s \leqslant 2\delta\ell$ denote the number of elements of S greater than $\delta k/\ell$, and note that $\binom{a}{\ell-s} \leqslant \left(\frac{e^2 a}{\ell}\right)^{\ell-s}$, since $e(\ell - s) \geqslant \ell$. By Lemma 5.1, it follows that there are at most

$$\left(\frac{e^2 k}{s^2}\right)^s \binom{\delta k/\ell}{\ell - s} \leqslant \left(\frac{e^2 k}{s^2}\right)^s \left(\frac{e^2 \delta k}{\ell^2}\right)^{\ell - s} = \left(\frac{\ell^2}{\delta s^2}\right)^s \left(\frac{e^2 \delta k}{\ell^2}\right)^{\ell} \leqslant \left(\frac{\sqrt{\delta} k}{\ell^2}\right)^{\ell}, \tag{11}$$

such sets S, where the last inequality follows since $\delta > 0$ is sufficiently small and $s \leqslant \ell/3$. Summing over s, it follows that there are at most $\ell\left(\frac{\sqrt{\delta}k}{\ell^2}\right)^{\ell}$ sets S with $t^* \geqslant 1/\delta^2$. Since $\delta = \delta(c_0)$ was chosen sufficiently small, the required bound follows.

Similarly, if $t_* \geqslant 1/\delta^2$, then S contains at most $2\delta\ell$ elements greater than $\delta k/\ell$ (at most $\delta\ell$ in the range $[\delta k/\ell, k/\delta\ell] \subset [\delta k/\ell, (1+\delta)^{1/\delta^2}\delta k/\ell]$, and at most $\delta\ell$ greater than $k/\delta\ell$). Thus $\ell\left(\frac{\sqrt{\delta}k}{\ell^2}\right)^{\ell}$ is also an upper bound on the number of sets S with $t_* \geqslant 1/\delta^2$.

We shall now apply the argument which failed to work above to the set $S_J := S \cap J$.

Case 2: $\max\{t^*, t_*\} \leqslant 1/\delta^2$.

Since $t^*, t_* \leqslant 1/\delta^2$, there are at most $3\delta\ell$ elements[6] of S outside the set $[0, \delta k/\ell] \cup J$. Moreover, we may assume that S has at least $\ell/3$ elements larger than $\delta k/\ell$, since the inequalities in (11) hold for every $s \leqslant \ell/3$. It follows that $|S_J| \geqslant \ell/4$, and therefore, using the trivial bound $k = \sum_{a \in S} a \geqslant |S_J| \cdot J_*$, we deduce that $J_* \leqslant 4k/\ell$.

Set $J_0 = \{a \in S : a \leqslant \delta k/\ell\}$, let $b = |J_0| \leqslant 2\ell/3$ and set $r = |S \setminus (J \cup J_0)| \leqslant 3\delta\ell$. Suppose first that $\text{span}(J) < ck/8\ell$. Then, by Lemma 5.1, the number of choices for S is at most[7]

$$\sum_{b=0}^{2\ell/3} \sum_{r=0}^{3\delta\ell} \sum_{J_*, J^*} \binom{\delta k/\ell}{b} \left(\frac{e^2 k}{r^2}\right)^r \binom{ck/8\ell}{\ell - b - r} \leqslant k^2 \ell^2 \max_{b \leqslant 2\ell/3,\, r \leqslant 3\delta\ell} \left(\frac{e\delta k}{\ell b}\right)^b \left(\frac{e^2 k}{r^2}\right)^r \left(\frac{cek}{2\ell^2}\right)^{\ell - b - r}$$

$$= k^2 \ell^2 \max_{b \leqslant 2\ell/3,\, r \leqslant 3\delta\ell} \left(\frac{2\delta\ell}{bc}\right)^b \left(\frac{2e\ell^2}{cr^2}\right)^r \left(\frac{cek}{2\ell^2}\right)^{\ell} \leqslant 2^{\sqrt{\delta}\ell} \left(\frac{cek}{2\ell^2}\right)^{\ell}$$

if $\delta = \delta(c_0) > 0$ is sufficiently small, since the maximum occurs at $r = 3\delta\ell$ and $b = 2\delta\ell/(ec)$.

Thus we may assume that $\text{span}(J) \geqslant ck/8\ell$, from which it follows that

$$\text{span}(J) = J^* - J_* \geqslant \sqrt{\delta}\big(J^* + J_*\big), \tag{12}$$

since $J_* \leqslant 4k/\ell$, and so $J^* \geqslant J_* + ck/8\ell \geqslant \big(1 + c/32\big) J_*$.

Let us count sequences $\mathbf{a} = (a_1, \dots, a_\ell)$ of distinct elements such that $S = \{a_1, \dots, a_\ell\}$ satisfies $\sum_{a \in S} a = k$ and $|S + S| \leqslant ck/\ell$. Given such a sequence $\mathbf{a}$, for each $j \in [\ell]$, set

$$S_j = \{a_1, \dots, a_j\} \cap J$$

and define

$$B_j = \Big\{ y \in \mathbb{N} : \big|(S_j + y) \setminus (S_j + S_j)\big| \leqslant \delta^6 |S_j| \Big\}. \tag{13}$$

[6] At most $(t_* + t^*)\delta^3\ell$ elements in the range $[\delta k/\ell, k/\delta\ell]$, and at most $\delta\ell$ elements greater than $k/\delta\ell$.

[7] Here, as usual, $0^0 = 1$.

We make the following key claim.

Claim. *Suppose that the intervals* $[J_*, (1+\delta)J_*]$ *and* $[(1-\delta)J^*, J^*]$ *contain more than* $\delta^5 \ell$ *elements of* $\big\{a_1, \ldots, a_{\delta\ell}\big\}$ *each. Then*

$$|B_j| \,\leqslant\, \left(\frac{2}{3} + \delta\right) |S + S|$$

for every $\delta\ell \leqslant j \leqslant \ell$.

Proof of claim. Fix j with $\delta\ell \leqslant j \leqslant \ell$. Recall from (13) that $\big|\big(S_j + y\big) \setminus \big(S_j + S_j\big)\big| \leqslant \delta^6 \ell$ for every $y \in B_j$, and that $2 \cdot \mathrm{span}(S_j) = \mathrm{span}(S_j + S_j)$. We claim that

$$\big(1 - \sqrt{\delta}\big)\mathrm{span}(J) + \mathrm{span}(B_j) \,\leqslant\, \mathrm{span}(S_j + S_j). \tag{14}$$

Indeed, since $[(1-\delta)J^*, J^*]$ contains more than $\delta^5\ell$ elements of $\big\{a_1, \ldots, a_{\delta\ell}\big\}$, and hence of S_j, it follows that $x^* + \max(B_j) \in S_j + S_j$ for some $x^* \in S_j \cap [(1-\delta)J^*, J^*]$. Similarly, our assumption on $[J_*, (1+\delta)J_*] \cap \big\{a_1, \ldots, a_{\delta\ell}\big\}$ implies that $x_* + \min(B_j) \in S_j + S_j$ for some $x_* \in S_j \cap [J_*, (1+\delta)J_*]$, and therefore

$$\mathrm{span}(S_j + S_j) \,\geqslant\, \mathrm{span}(B_j) + x^* - x_* \,\geqslant\, \mathrm{span}(B_j) + \mathrm{span}(J) - \delta(J^* + J_*).$$

But by (12) we have $\mathrm{span}(J) \geqslant \sqrt{\delta}\big(J^* + J_*\big)$, so (14) follows.

Now $S_j \subset J$, and so $2 \cdot \mathrm{span}(J) \geqslant \mathrm{span}(S_j + S_j)$, which, together with (14), implies that

$$\mathrm{span}(B_j) \,\leqslant\, \frac{1 + \sqrt{\delta}}{2}\mathrm{span}(S_j + S_j) \,=\, \big(1 + \sqrt{\delta}\big)\mathrm{span}(S_j). \tag{15}$$

Next, define the set

$$A \,=\, \Big\{a \in S_j \,:\, \big|\big(a + B_j\big) \cap \big(S_j + S_j\big)\big| \geqslant \big(1 - \delta\big)|B_j|\Big\}.$$

Note that

$$|S_j \setminus A| \cdot \delta |B_j| \,\leqslant\, \Big|\big\{(a, y) \in S_j \times B_j \,:\, a + y \notin S_j + S_j\big\}\Big| \,\leqslant\, \delta^6 |S_j| \cdot |B_j|,$$

and hence $|A| \geqslant (1 - \delta^5)|S_j|$. Let $A_* = \min A$ and $A^* = \max A$, and consider the set

$$D \,=\, \big(A_* + B_j\big) \cup \big(A^* + B_j\big).$$

Subclaim: $|D| \geqslant 3|B_j|/2$.

Proof of subclaim. Since $[J_*, (1+\delta)J_*]$ and $[(1-\delta)J^*, J^*]$ each contain more than $\delta^5\ell$ elements of $\big\{a_1, \ldots, a_{\delta\ell}\big\}$ and $|A| \geqslant (1 - \delta^5)|S_j|$, we have

$$\mathrm{span}(A) \,\geqslant\, \mathrm{span}(S_j) - \delta\big(J_* + J^*\big) \,\geqslant\, \big(1 - 2\sqrt{\delta}\big)\mathrm{span}(S_j),$$

where the second inequality follows by (12), and the fact that $2 \cdot \mathrm{span}(S_j) \geqslant \mathrm{span}(J)$. Thus, by (15),

$$\mathrm{span}(B_j) \,\leqslant\, \big(1 + \sqrt{\delta}\big)\mathrm{span}(S_j) \,\leqslant\, \big(1 + 4\sqrt{\delta}\big)\mathrm{span}(A) \,<\, 2 \cdot \mathrm{span}(A),$$

and so $\big|\big(A_* + B_j\big) \cap \big(A^* + B_j\big)\big| \leqslant |B_j|/2$, which easily implies the subclaim. □

Finally, observe that $|D \setminus (S_j + S_j)| \leqslant 2\delta |B_j|$ by the definition of A. Hence

$$\left(\frac{3}{2} - 2\delta\right) |B_j| \,\leqslant\, |S_j + S_j| \,\leqslant\, |S + S|,$$

as required. □

Now, recall that (by definition), the intervals $[J_*, (1+\delta)J_*]$ and $[(1-\delta)J^*, J^*]$ each contain at least $\delta^3 \ell$ elements of S. Let $\mathcal{I}(S)$ denote the collection of orderings of the elements of S such that the intervals $[J_*, (1+\delta)J_*]$ and $[(1-\delta)J^*, J^*]$ each contain more than $\delta^5 \ell$ elements of $\big\{a_1, \ldots, a_{\delta\ell}\big\}$, and write $\mathbf{a} \in \mathcal{I}$ if $\mathbf{a} \in \mathcal{I}\big(\{a_1, \ldots, a_\ell\}\big)$. Observe that, given a random ordering $\mathbf{a} = (a_1, \ldots, a_\ell)$ of the elements of S, the probability that $\mathbf{a} \in \mathcal{I}(S)$ is at least $1/2$. Thus there are at least $\ell!/2$ orderings $\mathbf{a} \in \mathcal{I}$.

In order to count sequences $\mathbf{a} \in \mathcal{I}$, recall that $J_0 = \big\{ a \in S : a \leqslant \delta k/\ell \big\}$ and $b = |J_0|$, let $\hat{J}_0 = \big\{ j \in [\ell] : a_j \in J_0 \big\}$, and set

$$Q \,=\, \big\{1, \ldots, \delta\ell\big\} \cup \Big\{ j \in [\ell] \,:\, a_j \not\in B_{j-1} \text{ and } j \not\in \hat{J}_0 \Big\}.$$

We claim that if $\mathbf{a} \in \mathcal{I}$, then $|Q| \leqslant 5\delta\ell$. To see this, recall that $r = |S \setminus (J \cup J_0)| \leqslant 3\delta\ell$, and suppose that there are at least $\delta\ell$ values of $j \geqslant \delta\ell$ with $a_{j+1} \in J \setminus B_j$. Then each such j adds at least $\delta^6 |S_j| \geqslant \delta^{11}\ell$ new elements to $S_J + S_J$, since $\mathbf{a} \in \mathcal{I}$ implies that $|S_{\delta\ell}| \geqslant \delta^5 \ell$. But then

$$|S + S| \,\geqslant\, |S_J + S_J| \,\geqslant\, \delta\ell \cdot \delta^{11}\ell \,>\, ck/\ell,$$

since $\ell^3 \geqslant Ck = k/\delta^{13} > ck/\delta^{12}$, which contradicts our assumption.

Thus, by the Claim, and setting $q = |Q|$, the number of choices for S is at most

$$\sum_{b=0}^{2\ell/3} \sum_{q=0}^{5\delta\ell} \frac{2}{\ell!} \binom{\ell}{q} \binom{k+q-1}{q-1} \binom{\ell}{b} \left(\frac{\delta k}{\ell}\right)^b \left(\left(\frac{2}{3} + \delta\right) \frac{ck}{\ell} \right)^{\ell - q - b}. \tag{16}$$

Indeed, for each choice of the sets $\hat{J}_0$ and Q, and the values of a_j for each $j \in \hat{J}_0 \cup Q$, there are at most $|B_j| \leqslant \left(\frac{2}{3} + \delta\right) |S + S|$ choices for each remaining element a_{j+1}. Recall that $q \leqslant 5\delta\ell$, by the observations above, and that each set S is counted at least $\ell!/2$ times as a sequence $\mathbf{a} \in \mathcal{I}$. Since $|S + S| \leqslant ck/\ell$, (16) follows.

Finally, note that the summand in (16) is bounded above by

$$\frac{2}{\ell!} \left(\frac{e\ell}{q} \cdot \frac{2k}{q} \cdot \frac{3\ell}{2ck} \right)^q \left(\frac{e\ell}{b} \cdot \frac{\delta k}{\ell} \cdot \frac{3\ell}{2ck} \right)^b \left(\left(\frac{2}{3} + \delta\right) \frac{ck}{\ell} \right)^\ell \,=\, \frac{2^{O(\delta\ell)}}{\ell!} \left(\frac{3e\ell^2}{cq^2} \right)^q \left(\frac{3e\delta\ell}{2cb} \right)^b \left(\frac{2ck}{3\ell} \right)^\ell$$

which is maximized with $b = 3\delta\ell/2c$ and $q = 5\delta\ell$, and so is at most

$$2^{\gamma\ell} \left(\frac{2cek}{3\ell^2} \right)^\ell,$$

where $\gamma = \gamma(\delta, c_0) \to 0$ as $\delta \to 0$ for any fixed $c_0 > 0$. Since we chose $\delta = \delta(c_0)$ to be sufficiently small, the theorem follows. □

The proof of Theorem 1.4 is almost identical to the proof of Theorem 1.3 given above; we need only to add the following observations: that $S_j \subset B_j$, and that $a_{j+1} \not\in S_j$.

Proof of Theorem 1.4. Let $\delta > 0$, and note that without loss of generality we may assume that δ is sufficiently small. Suppose that $\ell \in \mathbb{N}$ is sufficiently large and that $\binom{\ell}{2} \leqslant k \leqslant \ell^2/\delta$, and set $c_0 = 2\ell^2/k$. Let $C = C(\delta^3, c_0) > 0$ be given by Theorem 1.3, and note that since $\delta \leqslant \ell^2/k \leqslant 2$, C depends only on δ. Let $\lambda \geqslant 2$ and set $c = \lambda\ell^2/k$, and note that $\lambda\ell = ck/\ell$ and $\ell^3 \geqslant (\delta\ell)k \geqslant Ck$.

We are therefore in the setting of Theorem 1.3, and hence we can repeat the proof above up to (16), except replacing δ everywhere by δ^3. Using the observations that $S_j \subset B_j$ and $a_{j+1} \not\in S_j$, we deduce that the number of choices for S is at most

$$\sum_{b=0}^{\ell} \sum_{q=0}^{5\delta^3\ell} \frac{2}{\ell!} \binom{\ell}{q} \binom{k+q-1}{q-1} \binom{\ell}{b} \left(\frac{\delta^3 k}{\ell} \right)^b \prod_{j \in Z} \left(\left(\frac{2}{3} + \delta^3 \right) |S+S| - |S_j| \right), \tag{17}$$

where $Z = \big\{ j > \delta^3 \ell \,:\, a_{j+1} \in B_j \big\}$, so in particular $|Z| = \ell - q - b$.

Recalling that $|S_j| \geqslant j - q - b$ and $|S+S| \leqslant \lambda\ell$, and applying the AM-GM inequality, we see that

$$\prod_{j \in Z} \left(\left(\frac{2}{3} + \delta^3 \right) |S+S| - |S_j| \right) \leqslant \left(\left(\frac{2}{3} + \delta^3 \right) \lambda\ell + q + b - \frac{1}{|Z|} \sum_{j \in Z} j \right)^{\ell - q - b},$$

and hence (17) is at most

$$\sum_{b=0}^{\ell} \sum_{q=0}^{5\delta^3\ell} \frac{2^{O(\delta^2\ell)}}{\ell!} \left(\frac{e^3 k\ell}{q^2} \right)^q \left(\frac{\delta^3 e k}{b} \right)^b \left(\frac{2\lambda\ell}{3} - \frac{\ell}{2} + \frac{3(q+b)}{2} \right)^{\ell - q - b}. \tag{18}$$

Since $\binom{\ell}{2} \leqslant k \leqslant \ell^2/\delta$ and $\lambda \geqslant 2$, the summand in (18) is maximized with $q = 5\delta^3\ell$ and $b \leqslant 3\delta^2\ell/c$, and hence (18) is at most

$$2^{\delta\ell} \left(\frac{(4\lambda - 3)e}{6} \right)^{\ell},$$

as required. □

We end this section by briefly discussing the factor $(4/3)^\ell$ which separates the bound in Theorem 1.3 from that in Conjecture 1.5. Although it may seem obvious that we lost this factor in the subclaim, where one might hope that $|D| \geqslant (2-\delta)|B_j|$, we remark that this is in fact not the case. Indeed, let $x \ll y \ll z$, and consider the set $S_j = (T+z) \cup (T+2z)$, where $T = U \cup W$ is composed of $U = [y, y+x]$ and W, a random subset of $[0, 2y]$ of density ε. Note that if $\delta z > 2y$, then this set is dense near its extremes, and moreover $B_j \approx [z, 2y+z] \cup [2z, 2y+2z]$ and $A_j \approx (U+z) \cup (U+2z)$. Thus

$$A_j + B_j \approx \big[y+z, x+3y+z\big] \cup \big[y+2z, x+3y+2z\big] \cup \big[y+3z, x+3y+3z\big]$$

has size roughly $3|B_j|/2$, and so in this case the subclaim is sharp.

It therefore seems that in order to prove Conjecture 1.5, one would have to use some structural properties of S_j. Indeed, since $B_j \approx B_{j+1}$ for each $j \in [\ell]$ and a typical S_j is a 'random-like' subset of $B_1 \cup \ldots \cup B_{j-1}$, one might hope to bound the probability that (i.e., the number of S_j for which) in the subclaim we have $|D| < (2-\delta)|B_j|$. However, since the bounds in Theorems 1.3 and 1.4 will suffice to prove Theorems 1.1 and 1.2, we shall not pursue this matter here.

6. The proof of Theorem 1.1

We are now ready to prove Theorem 1.1, which generalizes the Cameron-Erdős Conjecture to sum-free sets of size m, and its structural analogue, Theorem 1.2. Both theorems will follow from essentially the same proof; we shall first prove Theorem 1.1, and then (in Section 7) point out how the proof can be adapted to deduce Theorem 1.2. As noted earlier, we shall for simplicity assume throughout that n is even.

The proof is fairly long and technical so, in order to aid the reader, we shall start by giving a brief sketch. The argument is broken into a series of six claims, each relying on the earlier ones; the first five being relatively straightforward, and the last being somewhat more involved.

We begin, in Claim 1, by using the Hypergeometric Janson Inequality to give a general upper bound on the number of sum-free m-sets $I \subset [n]$ with $S = I \cap [n/2]$ fixed. In Claim 2, we use this bound, together with Propositions 2.3 and 4.2, to prove the theorem in the case $m \geqslant (1/2-\delta)n$. Then, in Claim 3, we use Claim 2, Propositions 2.3 and 4.2, and induction on n, to deal with the case $|S| \geqslant \delta m$. Writing $\ell = |S|$ and $k = \sum_{a \in S}(n/2 - a)$, in Claims 4 and 5 we use Claims 1, 2 and 3 and Lemma 5.1 to deal with the (easy) cases $\ell = O(n/m)$ and $k \gg \ell^2 n/m$. Finally, in Claim 6, we deal with the remaining (harder) cases; however, since we now have $\ell^3 \geqslant Ck$, we may apply Theorem 1.3 in place of Lemma 5.1. In fact, it turns out that when $m = \Theta(n)$ the bound in Theorem 1.3 is not quite strong enough for our purposes, but in this case we have $|S+S| = O(|S|)$, and so may instead use Theorem 1.4. Each of the claims is stated in such a way as to facilitate the deduction of Theorem 1.2, which follows by exactly the same argument, with a couple of minor tweaks.

Proof of Theorem 1.1. Fix $\delta > 0$ sufficiently small, let $C = C(\delta) > 0$ be a sufficiently large constant, and let $n \in \mathbb{N}$ and $1 \leqslant m \leqslant n/2$. We shall show that there are at most $2^{Cn/m}\binom{n/2}{m}$ sum-free m-subsets of $[n]$. We shall use induction on n, and so we will assume that the result holds for all smaller even values of n. Note that the result is trivial if $m \leqslant C^{1/3}\sqrt{n}$, since in that case $2^{Cn/m}\binom{n/2}{m} \geqslant \binom{n}{m}$. We shall therefore assume that $m \geqslant C^{1/3}\sqrt{n}$, and hence (via our choice of C) that n is sufficiently large.

For any set $I \subset [n]$, let
$$S(I) = \big\{ x \in I \, : \, x \leqslant n/2 \big\}$$
denote the collection of elements of I which are at most $n/2$, as in the statement of Theorem 1.2. Moreover, given a set $S \subset [n/2]$, let $S' = \big\{ x \in S : x > n/4 \big\}$.

We begin by giving a general bound on the number of sum-free m-sets $I \subset [n]$ with $S(I) = S$, for each $S \subset [n/2]$. For each $\ell \in [m]$ and $k \in \mathbb{N}$, let $\mathcal{S}(k,\ell)$ denote the collection of sets $S \subset [n/2]$ such that $|S| = \ell$ and

$$\sum_{a \in S} \left(\frac{n}{2} - a \right) \, = \, k.$$

The following claim follows easily from the Hypergeometric Janson Inequality.

Claim 1. *For every $k, \ell \in \mathbb{N}$ with $\ell \leqslant m/2$, and every $S \in \mathcal{S}(k,\ell)$, there are at most*

$$\min \left\{ \binom{n/2 - |S' + S'|}{m - \ell}, \, C \cdot \max \left\{ e^{-km^2/2n^2}, e^{-km/8n\ell} \right\} \binom{n/2}{m - \ell} \right\}$$

sum-free m-sets $I \subset [n]$ such that $S(I) = S$.

Proof of Claim 1. Since I is sum-free and $S' \subset I$, it follows that I contains no element of $S' + S' \subset \{n/2+1, \ldots, n\}$. Since $S(I) = S$ and $|S| = \ell$, the first bound follows. For the second bound, we use the Hypergeometric Janson Inequality. Define the graph G of 'forbidden pairs' by setting

$$V(G) = \big\{ n/2 + 1, \ldots, n \big\} \quad \text{and} \quad E(G) = \big\{ \{x, x+s\} \colon s \in S \big\},$$

and observe that $I \setminus S$ is an independent set in G, and that G has k edges and maximum degree at most 2ℓ, since $S(I) = S \in \mathcal{S}(k,\ell)$.

Let μ and Δ be the quantities defined in the statement of Lemma 4.3 and note that we are applying the lemma with $|X| = n/2$ and $|R| = m - \ell$. Recalling that $\ell \leqslant m/2$, we have

$$\mu \, = \, k \cdot \frac{(m-\ell)^2}{(n/2)^2} \, \geqslant \, \frac{km^2}{n^2} \quad \text{and} \quad \Delta \, \leqslant \, (2\ell)^2 \binom{k}{\ell} \left(\frac{m - \ell}{(n/2)} \right)^3 \, \leqslant \, \frac{32k\ell(m-\ell)^3}{n^3}.$$

Thus $\mu/2 \geqslant km^2/2n^2$ and $\mu^2/2\Delta \geqslant km/8n\ell$, and so the claimed inequality follows. $\square$

In the calculation below, we shall on several occasions wish to make the assumption that $n - 2m \geqslant \delta n$. The next claim deals with the complementary case.

Claim 2. *If $m \geqslant \left(\frac{1}{2} - \delta \right) n$, then there are at most $\left(\frac{1}{2^\ell} + \frac{1}{n^2} \right) \binom{n/2}{m}$ sum-free m-sets $I \subset [n]$ such that $|S(I)| = \ell$ and $I \not\subset O_n$.*

Proof of Claim 2. The result is trivial if $\ell = 0$, so let us assume that $\ell \geqslant 1$. Note that the conditions of Propositions 2.3 and 4.2 are satisfied, since $\delta > 0$ is sufficiently small and n is sufficiently large. Thus, by Proposition 2.3, there exists $\varepsilon = \varepsilon(\delta) > 0$ such that all but $2^{-\varepsilon m} \binom{n/2}{m}$ sum-free m-sets $I \subset [n]$ satisfy either $|I \setminus O_n| \leqslant \delta m$, or $|I \setminus B| \leqslant \delta m$ for some interval B of length $n/2$. Moreover, by Proposition 4.2 there are at most $\frac{1}{n^3} \binom{n/2}{m}$ sum-free m-sets $I \subset [n]$ such that $1 \leqslant |I \setminus O_n| \leqslant \delta m$. It thus suffices to count sum-free m-sets $I \subset [n]$ such that $|I \setminus B| \leqslant \delta m$ for some interval B of length $n/2$.

We shall divide into three cases, depending on the size of $\min(I)$. Set

$$a(I) \, := \, \frac{n}{2} - \min(I),$$

and fix an interval B of length $n/2$ such that $|I \setminus B| \leqslant \delta m$.

Case 1: $a(I) \geqslant 4\delta m$.

We claim that no such sum-free set I exists. Indeed, note that $n/2 - a(I) \in I$ and that $\max(B) - \min(B) = n/2 - 1$. Since I is sum-free, it follows that I contains at most half of the elements of the set[8]

$$\big\{ \min(B), \dots, \min(B) + 4\delta m - 1 \big\} \cup \big\{ \max(B) - a(I) + 1, \dots, \max(B) - a(I) + 4\delta m \big\}.$$

Thus $|I| \leqslant |B| - 4\delta m + \delta m < \left(\frac{1}{2} - \delta\right) n$, which contradicts our assumption on $|I|$.

Case 2: $2\ell \leqslant a(I) \leqslant 4\delta m$.

Set $a = a(I)$ and note that, as in Case 1, since $n/2 - a \in I$, it follows that I contains at most a elements of the set $\{n/2 + 1, \dots, n/2 + a\} \cup \{n - a + 1, \dots, n\}$. Thus I contains at least $m - a - \ell \geqslant m - 3a/2$ elements of $\{n/2 + a + 1, \dots, n - a\}$. The remaining elements are contained in a set of size $3a$, and thus, by (3), there are at most

$$2^{3a} \binom{n/2 - 2a}{m - 3a/2} = 2^{3a} \binom{n/2 - 2a}{t - a/2} \leqslant 2^{3a} \left(\frac{t}{n/2 - t} \right)^{a/2} \binom{n/2}{t} \leqslant \frac{1}{2^{\ell+1}} \binom{n/2}{m}$$

such sum-free m-sets, where $t = n/2 - m \leqslant \delta n$.

Case 3: $a(I) \leqslant \min\{2\ell, 4\delta m\}$.

Fix some set $S \subset [n/2 - 4\delta m, n/2]$ with $|S| = \ell$, and note that $|S' + S'| \geqslant 3\ell/2$, where $S' = S \setminus [n/4]$, since $S' = S$. Hence, by Claim 1 and using (3), the number of sum-free m-sets $I \subset [n]$ such that $S(I) = S$ is at most

$$\binom{n/2 - |S' + S'|}{m - \ell} \leqslant \binom{n/2 - 3\ell/2}{m - \ell} = \binom{n/2 - 3\ell/2}{t - \ell/2} \leqslant \left(\frac{t}{n/2 - t} \right)^{\ell/2} \binom{n/2}{t}, \tag{19}$$

where again $t = n/2 - m \leqslant \delta n$. Now, since $a(I) \leqslant 2\ell$, there are at most $2^{2\ell}$ choices for the set $S(I)$. Therefore, by (19), there are at most

$$2^{2\ell} \binom{n/2 - |S' + S'|}{m - \ell} \leqslant \left(\frac{16t}{n/2 - t} \right)^{\ell/2} \binom{n/2}{t} \leqslant \frac{1}{2^{\ell+1}} \binom{n/2}{m}$$

such sum-free m-sets $I \subset [n]$ with $|S(I)| = \ell$.

Finally, summing over the various cases, the claim follows. $\square$

From now on we shall assume that $n - 2m \geqslant 2\delta n$. Recall that Claim 1 allows us to count sum-free sets with at most δm elements less than $n/2$. We shall use the induction hypothesis to count the sets I that have more than δm elements in $[n/2]$.

Claim 3. *There are at most $\delta \cdot 2^{Cn/m} \binom{n/2}{m}$ sum-free m-sets $I \subset [n]$ with at least δm elements less than $n/2$.*

[8] For simplicity, we assumed here that $a(I) \leqslant n/2 - 4\delta m$. However, if I contains an element $b < 4\delta m$ then we can easily find a matching of pairs $\{x, x + b\}$ of size $4\delta m$ in B.

Proof of Claim 3. Recall that $m \geqslant C^{1/3}\sqrt{n}$ and that $C = C(\delta) > 0$ is sufficiently large. Thus, by Proposition 2.3, it follows that there exists $\varepsilon = \varepsilon(\delta) > 0$ such that all but $2^{-\varepsilon m}\binom{n/2}{m}$ sum-free m-sets $I \subset [n]$ satisfy either $|I \setminus O_n| \leqslant \delta^3 m$, or $|I \setminus B| \leqslant \delta^3 m$ for some interval B of length $n/2$. Moreover, since $\delta > 0$ is sufficiently small, by Proposition 4.1 there are at most $2^{Cn/2m}\binom{n/2}{m}$ sum-free m-sets $I \subset [n]$ with $|I \setminus O_n| \leqslant \delta^3 m$. We may therefore restrict our attention to the collection $\mathcal{X}$ of sum-free m-sets $I \subset [n]$ that satisfy $|I \setminus B| \leqslant \delta^3 m$ for some interval B of length $n/2$.

First, we shall show that there are only few sets in $\mathcal{X}$ which contain more than $\delta^3 m$ elements less than $n/2 - 2\delta^2 n$. Indeed, such a set contains at most $\delta^3 m$ elements of the interval $\{n - 2\delta^2 n + 1, \ldots, n\}$ and hence, by the induction hypothesis and (3), there are at most

$$2^{C(n-\delta^2 n)/(m-s)}\binom{n/2 - \delta^2 n}{m-s}\binom{2\delta^2 n}{s} \leqslant 2^{Cn/m}\big(1 - 2\delta^2\big)^{m-s}\left(\frac{2m}{n-2m}\right)^s\left(\frac{2e\delta^2 n}{s}\right)^s\binom{n/2}{m}$$

such sets with $s \leqslant \delta^3 m$ elements greater than $n - 2\delta^2 n$. Summing over s, and recalling that $n - 2m \geqslant \delta n$, it follows that there are at most

$$\sum_{s=0}^{\delta^3 m} 2^{Cn/m} e^{-2\delta^2(m-s)}\left(\frac{2m}{\delta n}\cdot\frac{2e\delta^2 n}{s}\right)^s\binom{n/2}{m} \leqslant \delta^3 m \cdot e^{-\delta^2 m}\left(\frac{4e}{\delta^2}\right)^{\delta^3 m} 2^{Cn/m}\binom{n/2}{m} \qquad (20)$$

such sets, which is at most $\delta^2 \cdot 2^{Cn/m}\binom{n/2}{m}$, since $\delta > 0$ was chosen sufficiently small and $m \geqslant C^{1/3}$ is sufficiently large.

It only remains to count the sets in $\mathcal{X}$ which contain at least $\delta m - \delta^3 m > \delta m/2$ elements of the interval $\{n/2 - 2\delta^2 n, \ldots, n/2\}$, and at most $\delta^3 m$ elements less than $n/2 - 2\delta^2 n$. Note that $|S + S| \geqslant 2|S| - 1$ for every $S \subset \mathbb{Z}$ and so, by Claim 1, there are at most

$$\sum_{a=0}^{\delta^3 m}\sum_{b=\delta m/2}^{m-a}\binom{n/2}{a}\binom{2\delta^2 n + 1}{b}\binom{n/2 - \delta m}{m-a-b} \qquad (21)$$

such sum-free sets. Note that $m \leqslant 5\delta n$ (else there are no such sets), and so by (3) we have

$$\binom{n/2 - \delta m}{m-a-b} \leqslant \left(\frac{n/2 - \delta m}{n/2}\right)^{m-a-b}\left(\frac{m}{n/2 - m}\right)^{a+b}\binom{n/2}{m} \leqslant \left(\frac{3m}{n}\right)^{a+b}\binom{n/2}{m}.$$

Hence (21) is at most

$$\sum_{a=0}^{\delta^3 m}\sum_{b=\delta m/2}^{m-a}\left(\frac{3em}{2a}\right)^a\left(\frac{7e\delta^2 m}{b}\right)^b\binom{n/2}{m} \leqslant m^2\left(\frac{3e}{2\delta^3}\right)^{\delta^3 m}\big(14e\delta\big)^{\delta m/2}\binom{n/2}{m} \leqslant e^{-\delta m}\binom{n/2}{m},$$

so this completes the proof of the claim. □

Since $\delta > 0$ was arbitrary, we may from now on restrict our attention to those sum-free subsets $I \subset [n]$ for which $|S(I)| \leqslant \delta^3 m$. The remainder of the proof involves some careful counting using Theorems 1.3 and 1.4 and Lemma 5.1. We shall break up the calculation into three claims. In the first two, which are fairly straightforward, we count the sets I for which $|S(I)|$ is small (Claim 4) or $|S(I)|$ is not small but $\sum_{a \in S(I)}(n/2-a)$ is large (Claim 5). Finally in Claim 6, which is much more delicate, we count the remaining sets.

Claim 4. *If $\ell \leqslant (2C\delta)n/m$, then there are at most $2^{Cn/2m}\binom{n/2}{m}$ sum-free m-sets $I \subset [n]$ with $|S(I)| = \ell$.*

Proof of Claim 4. Let $k \in \mathbb{N}$ and fix a set $S \in \mathcal{S}(k,\ell)$. By Claim 1, there are at most

$$C \cdot \max\left\{e^{-km^2/2n^2}, e^{-km/8n\ell}\right\}\binom{n/2}{m-\ell} \tag{22}$$

sum-free m-sets $I \subset [n]$ with $S(I) = S$. Recall that $\ell \leqslant \delta^3 m$ and $n - 2m \geqslant \delta n$, and suppose first that $e^{-km^2/2n^2} \geqslant e^{-km/8n\ell}$, i.e., that $\ell \leqslant n/4m$. By Lemma 5.1, there are at most $\left(\frac{e^2k}{\ell^2}\right)^\ell$ choices for S, and hence, using (2), we can bound the number of sets I as follows:

$$\sum_k \sum_{S \in \mathcal{S}(k,\ell)} C \cdot e^{-km^2/2n^2}\binom{n/2}{m-\ell} \leqslant \sum_k C\left(\frac{e^2k}{\ell^2}\right)^\ell e^{-km^2/2n^2}\left(\frac{2m}{n-2m}\right)^\ell\binom{n/2}{m}. \tag{23}$$

Now, using (4) to bound the sum over k, this is at most

$$C^2\left(\frac{e^2}{\ell^2}\cdot\frac{2m}{\delta n}\right)^\ell\left(\frac{2n^2\ell}{em^2}\right)^{\ell+1}\binom{n/2}{m} \leqslant \left(\frac{Cn}{m}\right)^3\left(\frac{4en}{\delta m\ell}\right)^\ell\binom{n/2}{m} \leqslant 2^{Cn/3m}\binom{n/2}{m},$$

where in the last two steps we used the bound $\ell \leqslant n/4m$.

Suppose next that $e^{-km^2/2n^2} \leqslant e^{-km/8n\ell}$, i.e., that $n/4m \leqslant \ell \leqslant (2C\delta)n/m$. The calculation is almost the same:

$$\begin{aligned}\sum_k \sum_{S \in \mathcal{S}(k,\ell)} C \cdot e^{-km/8n\ell}\binom{n/2}{m-\ell} &\leqslant \sum_k C\left(\frac{e^2k}{\ell^2}\right)^\ell e^{-km/8n\ell}\left(\frac{2m}{n-2m}\right)^\ell\binom{n/2}{m}\\ &\leqslant \left(\frac{Cn}{m}\right)^3\left(\frac{4e}{\delta}\right)^\ell\binom{n/2}{m} \leqslant \left(\left(4e/\delta\right)^{3\delta}\right)^{Cn/m}\binom{n/2}{m} \leqslant 2^{Cn/3m}\binom{n/2}{m},\end{aligned} \tag{24}$$

where we again used Lemma 5.1, (2), (4) and the bound $\ell \leqslant (2C\delta)n/m$. □

Claim 5. *If $\ell \geqslant n/4m$, then there are at most $e^{-\ell}\binom{n/2}{m}$ sum-free m-sets $I \subset [n]$ such that $S(I) \in \mathcal{S}(k,\ell)$ for some $k \geqslant \ell^2 n/\delta m$.*

Proof of Claim 5. The calculation is similar to that in the previous claim. Indeed, note that (22) still holds, and that $e^{-km^2/2n^2} \leqslant e^{-km/8n\ell}$, since $\ell \geqslant n/4m$. Thus, by Lemma 5.1 and (2), and since $n - 2m \geqslant \delta n$,

$$\sum_{k \geqslant \ell^2 n/\delta m}\sum_{S \in \mathcal{S}(k,\ell)}\binom{n/2}{m-\ell}e^{-km/8n\ell} \leqslant \sum_{k \geqslant \ell^2 n/\delta m}\left(\frac{e^2k}{\ell^2}\cdot\frac{2m}{\delta n}\right)^\ell e^{-km/8n\ell}\binom{n/2}{m} \tag{25}$$

is an upper bound on the number of sum-free m-sets $I \subset [n]$ such that $S(I) \in \mathcal{S}(k,\ell)$ for some $k \geqslant \ell^2 n/\delta m$.

Now, note that $k \geqslant \ell^2 n/\delta m > 3\ell \cdot 8n\ell/m$ since $\delta > 0$ is sufficiently small. Therefore

$$\sum_{k=\ell^2 n/\delta m}^{\infty} \left(\frac{e^2 k}{\ell^2} \cdot \frac{2m}{\delta n} \right)^{\ell} e^{-km/8n\ell} \;\leqslant\; \frac{16 n\ell}{m} \left(\frac{2e^2}{\delta^2} \right)^{\ell} e^{-\ell/8\delta} \;\leqslant\; e^{-\ell},$$

since $g(x) = x^a e^{-bx}$ is decreasing on $[a/b, \infty)$ and $g(x + 1/b) < g(x)/2$ if $x > 3a/b$. It follows that the right-hand side of (25) is bounded above by $e^{-\ell}\binom{n/2}{m}$, as claimed. □

Since $3 < 2\sqrt{e}$, the following claim now completes the proof of Theorem 1.1.

Claim 6. *If $k \leqslant \ell^2 n/\delta m$ and $\ell \geqslant (2C\delta)n/m$, then there are at most*

$$\left(2^{O(\delta\ell)} \left(\frac{3}{2\sqrt{e}} \right)^{\ell} + e^{-\delta m} \right) \binom{n/2}{m} \tag{26}$$

sum-free m-sets $I \subset [n]$ with $S(I) \in \mathcal{S}(k, \ell)$.

This is the most difficult case, and we shall have to count more carefully, using Theorem 1.3 (in the case $m = o(n)$) and Theorem 1.4 (in the case $m = \Theta(n)$). Given a set $S \subset [n/2]$, we write $S'' = S \cap [n/4]$ and $S' = S \setminus S''$. For simplicity, we shall fix integers $k', \ell' \in \mathbb{N}$ and consider only sets in

$$\mathcal{S} \;=\; \big\{ S \in \mathcal{S}(k, \ell) : \, S' \in \mathcal{S}(k', \ell') \big\}.$$

Note that summing over choices of k' and ℓ' only costs us a factor of $k\ell = O(\ell^4)$, which is absorbed by the error term $2^{O(\delta\ell)}$. Set $\ell'' = \ell - \ell'$ and $k'' = k - k'$, and observe that

$$\ell'' \;=\; \ell - \ell' \;=\; \big| S \cap [n/4] \big| \;\leqslant\; \frac{4k}{n} \;\leqslant\; \frac{4\ell^2}{\delta m} \;\leqslant\; \delta\ell, \tag{27}$$

since $k \leqslant \ell^2 n/\delta m$ and $\ell \leqslant \delta^3 m$. We shall use this fact (that $\ell' \approx \ell$) on numerous occasions in the calculation below.

Proof of Claim 6. We begin by slightly improving the bound in Claim 1; this will allow us to deal with the case in which S'' is reasonably large. To be precise, we shall show that if $n - 2m \geqslant \delta n$ and $\ell \leqslant \delta^3 m$, then there are at most

$$C \sum_{S \in \mathcal{S}} \max\left\{ e^{-\delta |S''| m^2/n}, e^{-\delta m} \right\} \binom{n/2 - |S' + S'|}{m - \ell} \;+\; e^{-\delta m} \binom{n/2}{m} \tag{28}$$

sum-free m-sets $I \subset [n]$ with $S(I) \in \mathcal{S}$.

To prove (28), we shall partition $\mathcal{S}$ into two sets by setting

$$\mathcal{S}_1 \;=\; \big\{ S \in \mathcal{S} : \, |S' + S'| \geqslant \delta n \big\},$$

and $\mathcal{S}_2 = \mathcal{S} \setminus \mathcal{S}_1$. By Claim 1, and using (3), there are at most

$$\binom{n/2}{\ell}\binom{n/2 - \delta n}{m - \ell} \;\leqslant\; \big(1 - 2\delta\big)^{m-\ell} \left(\frac{2m}{\delta n} \right)^{\ell} \left(\frac{en}{2\ell} \right)^{\ell} \binom{n/2}{m} \;\leqslant\; e^{-\delta m} \binom{n/2}{m},$$

choices for I with $S(I) \in \mathcal{S}_1$, since $n - 2m \geqslant \delta n$ and $\ell \leqslant \delta^3 m$.

Suppose now that $S \in \mathcal{S}_2$. Similarly as in the proof of Claim 1, in order to bound the number of sets I with $S(I) = S$ we apply the Hypergeometric Janson Inequality to the graph G with

$$V(G) = \big\{n/2+1,\dots,n\big\} \setminus (S'+S') \quad \text{and} \quad E(G) = \big\{\{x, x+s\}\colon s \in S''\big\}.$$

Observe that $k'' \geqslant \ell'' n/4$, and hence G has at least

$$k'' - (2\ell'')|S'+S'| \geqslant 2\ell''\left(\frac{n}{8} - |S'+S'|\right) \geqslant \frac{\ell'' n}{8}$$

edges, and maximum degree at most $2\ell''$. Hence, letting μ and Δ to be the quantities defined in the statement of Lemma 4.3, we have

$$\mu \geqslant \frac{\ell'' n}{8}\cdot\frac{(m-\ell)^2}{(n/2)^2} \geqslant \frac{\ell'' m^2}{3n} \quad \text{and} \quad \Delta \leqslant \frac{n}{2}\cdot\big(2\ell''\big)^2\left(\frac{m-\ell}{n/2-\delta n}\right)^3 \leqslant \frac{20(\ell'')^2 m^3}{n^2}.$$

Thus $\mu/2 \geqslant \delta \ell'' m^2/n$ and $\mu^2/2\Delta \geqslant \delta m$, and so (28) follows.

To see why the bound in (28) is so useful, observe that

$$\left(\frac{m}{|S''|}\right)^{|S''|} \max\Big\{e^{-\delta|S''|m^2/n}, e^{-\delta m}\Big\} \leqslant 2^{\delta\ell}. \tag{29}$$

Indeed, if $\ell'' \geqslant n/m$ then the left-hand side of (29) is at most 1, since $\ell'' \leqslant \delta^3 m$. On the other hand, if $\ell'' \leqslant n/m$ then

$$\left(\frac{m}{\ell''}\right)^{\ell''} \max\Big\{e^{-\delta\ell'' m^2/n}, e^{-\delta m}\Big\} = \left(\frac{m}{\ell''}\cdot e^{-\delta m^2/n}\right)^{\ell''} \leqslant \left(\frac{n/\delta m}{\ell''}\right)^{\ell''} \leqslant e^{n/\delta m} \leqslant 2^{\delta\ell},$$

since $me^{-\delta m^2/n} \leqslant n/\delta m$, as in the proof of Proposition 4.1, and the function $x \mapsto (c/x)^x$ is maximized when $x = c/e$; the last inequality holds since $\ell \geqslant (2C\delta)n/m$.

In order to complete the calculation and obtain (26), we break into cases according to the order of magnitude of m. We begin with the case $m \leqslant \delta n$.

Case 1: $C^{1/3}\sqrt{n} \leqslant m \leqslant \delta n$.

We wish to bound (28) from above; the key idea is to break up $\mathcal{S}$ according to the size of $|S'+S'|$. If it is very small (at most $\delta^3 k'/\ell'$), then we easily obtain a bound via Theorem 1.3, and if it is very large (at least $(n/\delta m)\cdot k'/\ell'$), then the trivial bound of Lemma 5.1 suffices. In the middle range, we shall partition into sufficiently small intervals (see (34), below); inside such an interval we have an upper bound on $|S'+S'|$, which allows us to apply Theorem 1.3, and also a lower bound, which is useful when we apply (28).

(*a*) $|S'+S'| \leqslant \delta^3 k'/\ell'$.

Let $\mathcal{S}^a = \big\{S \in \mathcal{S} : |S'+S'| \leqslant \delta^3 k'/\ell'\big\}$. We shall first apply Theorem 1.3 to bound the number of sets $S \in \mathcal{S}^a$; we may do so since

$$\frac{(\ell')^3}{k'} \geqslant \frac{\ell^3}{2k} \geqslant \ell\cdot\frac{\delta m}{2n} \geqslant C\delta^2, \tag{30}$$

which follows by (27), and since $k' \leqslant k \leqslant \ell^2 n/\delta m$ and $\ell \geqslant (2C\delta)n/m$. Thus, by Theorem 1.3, and using (27), we obtain

$$|\mathcal{S}^a| \leqslant 2^{\delta\ell}\left(\frac{2\delta^3 e k'}{3\ell'^2}\right)^{\ell'}\binom{n/4}{\ell''} \leqslant \left(\frac{\delta^2 k}{\ell^2}\right)^{\ell'}\left(\frac{n}{\ell''}\right)^{\ell''}, \tag{31}$$

Now, combining (31) with (3), and recalling that $\ell = \ell' + \ell''$ and $m \leqslant \delta n$, we obtain

$$\sum_{S \in \mathcal{S}^a}\binom{n/2}{m-\ell} \leqslant \left(\frac{\delta^2 k}{\ell^2}\right)^{\ell'}\left(\frac{n}{\ell''}\right)^{\ell''}\left(\frac{2m}{n-2m}\right)^{\ell}\binom{n/2}{m}$$
$$\leqslant \left(\frac{\delta^2 k}{\ell^2}\cdot\frac{4m}{n}\right)^{\ell'}\left(\frac{4m}{\ell''}\right)^{\ell''}\binom{n/2}{m} \leqslant (5\delta)^{\ell'}\left(\frac{4m}{\ell''}\right)^{\ell''}\binom{n/2}{m}.$$

To see the final inequality, recall that $k/\ell^2 \leqslant n/\delta m$, by assumption. Hence, using (27) and (29), it follows that

$$\sum_{S \in \mathcal{S}^a}\max\left\{e^{-\delta|S''|m^2/n}, e^{-\delta m}\right\}\binom{n/2}{m-\ell} \leqslant (6\delta)^{\ell}\binom{n/2}{m}. \tag{32}$$

(*b*) $|S' + S'| \geqslant (n/\delta m)\cdot k'/\ell'$.

Let $\mathcal{S}^b = \left\{S \in \mathcal{S} : |S'+S'| \geqslant (n/\delta m)\cdot k'/\ell'\right\}$. By Lemma 5.1, and using (27), we have

$$|\mathcal{S}^b| \leqslant \left(\frac{e^2 k'}{\ell'^2}\right)^{\ell'}\binom{n/4}{\ell''} \leqslant \left(\frac{e^3 k'}{\ell^2}\right)^{\ell'}\left(\frac{n}{\ell''}\right)^{\ell''},$$

Thus, by (3) and recalling that $\ell = \ell' + \ell''$ and $m \leqslant \delta n$,

$$\sum_{S \in \mathcal{S}^b}\binom{n/2 - |S'+S'|}{m-\ell} \leqslant \left(\frac{e^3 k'}{\ell^2}\right)^{\ell'}\left(\frac{n}{\ell''}\right)^{\ell''}\left(1 - \frac{2k'}{\delta m \ell'}\right)^{m-\ell}\left(\frac{2m}{n-2m}\right)^{\ell}\binom{n/2}{m}$$
$$\leqslant \left(\frac{e^3 k'}{\ell^2}\cdot\frac{4m}{n}\right)^{\ell'}\left(\frac{4m}{\ell''}\right)^{\ell''}\exp\left(-\frac{k'}{\delta\ell'}\right)\binom{n/2}{m} \leqslant \delta^{\ell'}\left(\frac{4m}{\ell''}\right)^{\ell''}\binom{n/2}{m}$$

where in the final step we used our assumption that $m \leqslant \delta n$, and the inequality (4) (with $a = \ell'$ and $b = 1/\delta\ell'$) to maximize[9] over k'. Using (27) and (29), it follows that

$$\sum_{S \in \mathcal{S}^b}\max\left\{e^{-\delta|S''|m^2/n}, e^{-\delta m}\right\}\binom{n/2 - |S'+S'|}{m-\ell} \leqslant (2\delta)^{\ell}\binom{n/2}{m}. \tag{33}$$

(*c*) $\delta^3 k'/\ell' \leqslant |S'+S'| \leqslant (n/\delta m)\cdot k'/\ell'$.

For each $c > 0$ let $\mathcal{S}^{(c)}$ denote the collection of sets $S \in \mathcal{S}$ with

$$\frac{ck'}{\ell'} \leqslant |S'+S'| \leqslant \frac{(1+\delta)ck'}{\ell'}. \tag{34}$$

[9]To be precise, it follows from (4) that $\max_{k'}(k')^{\ell'}e^{-k'/\delta\ell'} \leqslant \sum_{k'}(k')^{\ell'}e^{-k'/\delta\ell'} \leqslant C\big(\delta\ell'^2/e\big)^{\ell'+1} \leqslant (\delta\ell^2)^{\ell'}$.

We shall first bound the sum in (28) restricted to $\mathcal{S}^{(c)}$ for each $\delta^3 \leqslant c \leqslant n/\delta m$, and then sum over choices of c.

First, note that the conditions of Theorem 1.3 hold, by (30). It follows that

$$|\mathcal{S}^{(c)}| \leqslant 2^{3\delta\ell} \left(\frac{2cek'}{3\ell'^2}\right)^{\ell'} \binom{n/4}{\ell''},$$

since $|S' + S'| \leqslant (1+\delta)ck'/\ell'$ for every $S \in \mathcal{S}^{(c)}$. Moreover,

$$\binom{n/2 - |S'+S'|}{m-\ell} \leqslant \left(1 - \frac{2ck'}{\ell' n}\right)^{m-\ell} \left(\frac{2m}{n-2m}\right)^{\ell} \binom{n/2}{m}$$

by (3), and because $|S'+S'| \geqslant ck'/\ell'$ for every $S \in \mathcal{S}^{(c)}$. Since $m \leqslant \delta n$, it follows that

$$\sum_{S \in \mathcal{S}^{(c)}} \binom{n/2 - |S'+S'|}{m-\ell} \leqslant 2^{O(\delta\ell)} \left(\frac{2cek'}{3\ell'^2}\right)^{\ell'} \binom{n/4}{\ell''} \left(1 - \frac{2ck'}{\ell' n}\right)^{m-\ell} \left(\frac{2m}{n}\right)^{\ell} \binom{n/2}{m}$$

$$\leqslant 2^{O(\delta\ell)} \left(\frac{4cek'm}{3\ell'^2 n}\right)^{\ell'} \left(\frac{em}{2\ell''}\right)^{\ell''} \exp\left(-\frac{2ck'(m-\ell)}{\ell' n}\right) \binom{n/2}{m}.$$

Now, using (4) to maximize over k', we obtain

$$\sum_{S \in \mathcal{S}^{(c)}} \binom{n/2 - |S'+S'|}{m-\ell} \leqslant 2^{O(\delta\ell)} \left(\frac{2}{3} \cdot \frac{m}{m-\ell}\right)^{\ell'} \left(\frac{em}{2\ell''}\right)^{\ell''} \binom{n/2}{m},$$

and hence, using (29), it follows that

$$\sum_{S \in \mathcal{S}^{(c)}} \max\left\{e^{-\delta|S''|m^2/n}, e^{-\delta m}\right\} \binom{n/2 - |S'+S'|}{m-\ell} \leqslant 2^{O(\delta\ell)} \left(\frac{2}{3}\right)^{\ell} \binom{n/2}{m}. \qquad (35)$$

We apply (35) with $c = \delta^3(1+\delta)^t$ for each $0 \leqslant t \leqslant n/\delta m$.[10] Summing over such integers t, using (32) and (33) to bound the remaining terms in the sum, and using our assumption that $\ell \geqslant (2C\delta)n/m$, which implies that $n/m \leqslant 2^{\delta\ell}$, it follows that

$$\sum_{S \in \mathcal{S}} \max\left\{e^{-\delta|S''|m^2/n}, e^{-\delta m}\right\} \binom{n/2 - |S'+S'|}{m-\ell} \leqslant 2^{O(\delta\ell)} \left(\frac{2}{3}\right)^{\ell} \binom{n/2}{m},$$

Hence, by (28), the number of sum-free m-sets $I \subset [n]$ with $S(I) \in \mathcal{S}$ is at most

$$2^{O(\delta\ell)} \left(\frac{2}{3}\right)^{\ell} \binom{n/2}{m} + e^{-\delta m} \binom{n/2}{m}.$$

Since $2/3 < 3/2\sqrt{e}$, this proves the claim in the case $C^{1/3}\sqrt{n} \leqslant m \leqslant \delta n$.

Now we turn to the case $m = \Theta(n)$. The calculation in this case is similar to that in Case 1, except we shall use Theorem 1.4 in place of Theorem 1.3. To simplify the calculations, we

[10] In fact we only need to go up to $t \approx \log_{1+\delta}(n/\delta^4 m) < n/\delta m$, since $m \leqslant \delta n$.

shall assume first that $m \leqslant n/4$, and then (in Case 3) show how the result for $m > n/4$ follows by the same argument.

Case 2: $\delta n \leqslant m \leqslant n/4$.

We shall again bound (28) by breaking up $\mathcal{S}$ according to the size of $|S' + S'|$. However, 'small' will now mean at most $3|S'| - 4$ and 'large' will mean at least $|S'|/\delta^3$; in between these bounds we shall again partition into small intervals. The main difference compared with Case 1 is that when $|S' + S'|$ is small, we shall use Freiman's $3k - 4$ Theorem.

For each $\lambda > 0$, let $\mathcal{S}_{(\lambda)}$ denote the collection of sets $S \in \mathcal{S}$ such that

$$\lambda|S'| \leqslant |S' + S'| \leqslant \big(1+\delta\big)\lambda|S'|. \tag{36}$$

(a) $2|S'| - 1 \leqslant |S' + S'| \leqslant 3|S'| - 4$.

Let $\lambda \in \mathbb{R}$ be such that $S \in \mathcal{S}_{(\lambda)}$, and note that $2 - \delta \leqslant \lambda \leqslant 3$, since $|S'| \geqslant \ell/2 \geqslant 1/\delta$. By Freiman's $3k - 4$ Theorem, and using our upper bound on $|S' + S'|$, it follows that S' is contained in an arithmetic progression of length at most

$$|S' + S'| - |S'| + 1 \leqslant \big(\lambda - 1 + 3\delta\big)|S'|.$$

Since there are at most k^2 arithmetic progressions in $\{n/2 - k + 1, \ldots, n/2\}$ and $k = O(\ell^3)$, it follows that

$$|\mathcal{S}_{(\lambda)}| \leqslant k^2 \binom{(\lambda - 1 + 3\delta)\ell'}{\ell'}\binom{n/4}{\ell''} \leqslant 2^{O(\delta\ell)}\binom{(\lambda-1)\ell'}{\ell'}\binom{n/4}{\ell''}.$$

Therefore, by (3) and since $|S' + S'| \geqslant \lambda|S'|$,

$$\sum_{S \in \mathcal{S}_{(\lambda)}} \binom{n/2 - |S' + S'|}{m - \ell} \leqslant 2^{O(\delta\ell)}\binom{(\lambda-1)\ell'}{\ell'}\left(\frac{2m}{n-2m}\right)^{\ell}\left(\frac{n - 2\lambda\ell'}{n}\right)^{m}\binom{n/4}{\ell''}\binom{n/2}{m}.$$

Now, observe that, via simple calculus,[11]

$$\binom{(\lambda-1)\ell'}{\ell'}\left(\frac{2m}{n-2m}\right)^{\ell'}\left(\frac{n-2\lambda\ell'}{n}\right)^{m} \leqslant e^{-3\ell'/2+\delta\ell}\binom{2\ell'}{\ell'} \leqslant 2^{O(\delta\ell)}\left(\frac{4}{e^{3/2}}\right)^{\ell}.$$

Moreover, recall from (29) that

$$\max\left\{e^{-\delta\ell'' m^2/n}, e^{-\delta m}\right\}\left(\frac{2m}{n-2m}\right)^{\ell''}\binom{n/4}{\ell''} \leqslant 2^{\delta\ell},$$

since $m \leqslant n/4$. It follows that

$$\sum_{S \in \mathcal{S}_{(\lambda)}} \max\left\{e^{-\delta|S''|m^2/n}, e^{-\delta m}\right\}\binom{n/2 - |S' + S'|}{m - \ell} \leqslant 2^{O(\delta\ell)}\left(\frac{4}{e^{3/2}}\right)^{\ell}\binom{n/2}{m}, \tag{37}$$

and thus, summing over $O(1)$ values[12] of λ, we obtain a bound for the entire range.

[11]The left-hand side is increasing in m for $m \leqslant \big(1/4 - \delta^2\big)n$ and in λ on $(2-\delta, 3)$ since $\frac{\lambda-1}{\lambda-2} > 2 > \sqrt{e}$.

[12]Consider $\lambda = (2-\delta)(1+\delta)^j$ for $0 \leqslant j \leqslant 1/\delta$, say.

(b) $|S' + S'| \geqslant |S'|/\delta^3$.

The proof is almost identical to Case 1(b). Indeed, let

$$\mathcal{S}^d = \big\{ S \in \mathcal{S} : |S' + S'| \geqslant |S'|/\delta^3 \big\},$$

and observe that

$$|\mathcal{S}^d| \,\leqslant\, \left(\frac{e^2 k'}{\ell'^2}\right)^{\ell'} \binom{n/4}{\ell''} \,\leqslant\, \left(\frac{e^3 k'}{\ell^2}\right)^{\ell'} \left(\frac{n}{\ell''}\right)^{\ell''},$$

by Lemma 5.1 and (27). Thus, by (3) and since $|S' + S'| \geqslant |S'|/\delta^3$ and $\delta n \leqslant m \leqslant n/4$,

$$\sum_{S \in \mathcal{S}^d} \binom{n/2 - |S' + S'|}{m - \ell} \,\leqslant\, \left(\frac{e^3 k'}{\ell^2}\right)^{\ell'} \left(\frac{n}{\ell''}\right)^{\ell''} \left(1 - \frac{2|S'|}{\delta^3 n}\right)^{m-\ell} \left(\frac{2m}{n-2m}\right)^{\ell} \binom{n/2}{m}$$
$$\leqslant\, \left(\frac{e^3 k'}{\ell^2} \cdot \frac{4m}{n}\right)^{\ell'} \left(\frac{4m}{\ell''}\right)^{\ell''} \exp\left(-\frac{\ell'}{\delta^2}\right) \binom{n/2}{m}.$$

But since $k' \leqslant \ell^2 n/\delta m$, and using (27) and (29), we have

$$\left(\frac{e^3 k'}{\ell^2} \cdot \frac{4m}{n}\right)^{\ell'} \left(\frac{4m}{\ell''}\right)^{\ell''} \max\left\{ e^{-\delta \ell'' m^2/n}, e^{-\delta m} \right\} \,\leqslant\, 2^{\delta \ell} \left(\frac{4e^3}{\delta}\right)^{\ell'} \,\leqslant\, e^{\ell'/\delta},$$

and hence

$$\sum_{S \in \mathcal{S}^d} \max\left\{ e^{-\delta |S''| m^2/n}, e^{-\delta m} \right\} \binom{n/2 - |S' + S'|}{m - \ell} \,\leqslant\, \delta^{\ell} \binom{n/2}{m}. \tag{38}$$

(c) $3|S'| - 4 < |S' + S'| < |S'|/\delta^3$.

The proof is similar to Case 1(c), except we shall use Theorem 1.4 in place of Theorem 1.3 in order to bound the size of $\mathcal{S}_{(\lambda)}$ for each $3(1-\delta) \leqslant \lambda \leqslant 1/\delta^3$. First, let us check that the conditions of Theorem 1.4 hold, i.e., that ℓ' is sufficiently large (as a function of δ), and that $k' \leqslant \ell'^2/\delta$. Recalling that $k \leqslant \ell^2 n/\delta m$, $\ell \geqslant (2C\delta)n/m$ and $\ell'' \leqslant \delta \ell$, we obtain

$$k' \,\leqslant\, k \,\leqslant\, \frac{\ell^2 n}{\delta m} \,\leqslant\, \frac{2}{\delta^2} \cdot \ell'^2 \qquad \text{and} \qquad \ell' \,\geqslant\, \frac{(C\delta)n}{m} \,\geqslant\, 4C\delta,$$

since $\delta n \leqslant m \leqslant n/4$. Thus we may apply Theorem 1.4, and deduce that

$$|\mathcal{S}_{(\lambda)}| \,\leqslant\, 2^{3\delta\ell'} \left(\frac{(4\lambda - 3)e}{6}\right)^{\ell'} \binom{n/4}{\ell''}$$

for every $2 \leqslant \lambda \leqslant 2/\delta^3$, since $|S' + S'| \leqslant (1+\delta)\lambda|S'|$ for every $S \in \mathcal{S}_{(\lambda)}$. Moreover,

$$\binom{n/2 - |S' + S'|}{m - \ell} \,\leqslant\, \left(1 - \frac{2\lambda|S'|}{n}\right)^{m-\ell} \left(\frac{2m}{n-2m}\right)^{\ell} \binom{n/2}{m}$$

by (3), and because $|S' + S'| \geqslant \lambda|S'|$ for every $S \in \mathcal{S}_{(\lambda)}$.

Now, by simple calculus[13], it is straightforward to show that if $\lambda \geqslant 3(1-\delta)$ and $m \leqslant n/4$ then

$$\left(\frac{(4\lambda-3)e}{6}\right)^{\ell'} \exp\left(-\frac{2\lambda\ell' m}{n}\right)\left(\frac{2m}{n-2m}\right)^{\ell'} \leqslant 2^{O(\delta\ell)}\left(\frac{3}{2\sqrt{e}}\right)^{\ell},$$

It follows that

$$\sum_{S\in\mathcal{S}_{(\lambda)}} \binom{n/2-|S'+S'|}{m-\ell} \leqslant 2^{O(\delta\ell)}\binom{n/4}{\ell''}\left(\frac{2m}{n-2m}\right)^{\ell''}\left(\frac{3}{2\sqrt{e}}\right)^{\ell}\binom{n/2}{m},$$

and hence, using (29),

$$\sum_{S\in\mathcal{S}_{(\lambda)}} \max\left\{e^{-\delta|S''|m^2/n}, e^{-\delta m}\right\}\binom{n/2-|S'+S'|}{m-\ell} \leqslant 2^{O(\delta\ell)}\left(\frac{3}{2\sqrt{e}}\right)^{\ell}\binom{n/2}{m}. \tag{39}$$

Finally, we apply the bound (37) with $\lambda = (2-\delta)(1+\delta)^t$ for each $0 \leqslant t \leqslant 1/\delta$, and the bound (39) with $\lambda = 3(1-\delta)(1+\delta)^t$ for each $0 \leqslant t \leqslant 1/\delta^2$. Summing over such integers t, using (38) to bound the remaining terms in the sum, and noting that $4/e^{3/2} < 3/2\sqrt{e}$, we obtain

$$\sum_{S\in\mathcal{S}} \max\left\{e^{-\delta|S''|m^2/n}, e^{-\delta m}\right\}\binom{n/2-|S'+S'|}{m-\ell} \leqslant 2^{O(\delta\ell)}\left(\frac{3}{2\sqrt{e}}\right)^{\ell}\binom{n/2}{m},$$

Hence, by (28), the number of sum-free m-sets $I \subset [n]$ with $S(I) \in \mathcal{S}$ is at most

$$2^{O(\delta\ell)}\left(\frac{3}{2\sqrt{e}}\right)^{\ell}\binom{n/2}{m} + e^{-\delta m}\binom{n/2}{m}.$$

This proves the claim in the case $\delta n \leqslant m \leqslant n/4$.

The proof is essentially complete; all that remains is to show that case $m \geqslant n/4$ can be deduced easily from the case above.

Case 3: $n/4 \leqslant m \leqslant (1/2-\delta)n$.

We shall reduce this case to the previous one. Indeed, set $t = n/2 - m$ and note that $\delta n \leqslant t \leqslant n/4$ and that

$$\binom{n/2-|S'+S'|}{m-\ell} = \binom{n/2-|S'+S'|}{t-|S'+S'|+\ell} \leqslant 2^{O(\delta\ell)}\binom{n/2-|S'+S'|}{t-\ell+3\delta\ell}, \tag{40}$$

since $|S'+S'| \geqslant 2|S'|-1 \geqslant (2-3\delta)\ell$, by (27). It follows that (28) is at most

$$2^{O(\delta\ell)}\sum_{S\in\mathcal{S}} \max\left\{e^{-\delta|S''|t^2/n}, e^{-\delta t}\right\}\binom{n/2-|S'+S'|}{t-\ell} + e^{-\delta t}\binom{n/2}{t}$$

[13]The function $\frac{(4x-3)y}{3(1-2y)}e^{1-2xy}$ on $[3(1-\delta),\infty)\times(0,1/4]$ is maximized at $(3(1-\delta),1/4)$.

if $|S' + S'| \geqslant (2 + 2\delta)\ell' \geqslant 2\ell$, in which case the argument of Case 2 applies (with m replaced by t), and the upper bound proved there still holds. On the other hand, if $|S'+S'| \leqslant (2+2\delta)\ell'$ then the argument of Case 2(a), combined with (40), gives an upper bound of

$$\left(\frac{n}{t-\ell}\right)^{3\delta\ell} 2^{O(\delta\ell)} \left(\frac{4}{e^{3/2}}\right)^{\ell} \binom{n/2}{m} < \left(\frac{3}{2\sqrt{e}}\right)^{\ell} \binom{n/2}{m}$$

since $t \geqslant \delta n$, $\ell \leqslant \delta^3 n$ and $4/e^{3/2} < 3/2\sqrt{e}$. This completes the proof of Claim 6. $\square$

Finally, let us put together the pieces and show that Claims 1–6 prove Theorem 1.1. By Claim 2, there are $O\big(\binom{n/2}{m}\big)$ sum-free m-subsets of $[n]$ for every $m \geqslant \big(\frac{1}{2} - \delta\big)n$. By Claim 3, if $m \leqslant \big(\frac{1}{2} - \delta\big)n$ then there are at most $\delta \cdot 2^{Cn/m}\binom{n/2}{m}$ such sets with at least $\delta^3 m$ of its elements less than $n/2$. By Claim 4, there are at most $2^{Cn/2m}\binom{n/2}{m}$ such sets I with $|S(I)| = \ell(I) \leqslant (2C\delta)n/m$, and by Claim 5 there are at most $O\big(\binom{n/2}{m}\big)$ such sets such that $k(I) \geqslant \ell^2 n/\delta m$ and $\ell(I) \geqslant n/4m$. Finally, by Claim 6, there are at most $O\big(\binom{n/2}{m}\big)$ such sets which were not contained in any of the previous cases, i.e., such that $\ell(I) \geqslant (2C\delta)n/m$ and $k(I) \leqslant \ell^2 n/\delta m$. The induction step, and hence Theorem 1.1, now follows. $\square$

7. The proof of Theorem 1.2

In this section we shall sketch how the proof in the previous section may be adapted in order to prove Theorem 1.2.

Proof of Theorem 1.2. Let $\delta > 0$ be a sufficiently small constant, and let $\omega = \omega(n)$ be an arbitrary function such that $\omega \to \infty$ as $n \to \infty$. Let $C > 0$ be sufficiently large, let $m, n \in \mathbb{N}$ satisfy $m \geqslant C\sqrt{n \log n}$, and consider the sum-free m-sets $I \subset [n]$ such that $I \not\subset O_n$. For simplicity, given such a set I let us write $\ell(I) = |S(I)|$, $k(I) = \sum_{a \in S(I)}(n/2 - a)$ and $a(I) = n/2 - \min(S(I))$.

Suppose first that $m \geqslant \big(\frac{1}{2} - \delta\big)n$. Then, by the proof of Claim 2, there are $o\big(\binom{n/2}{m}\big)$ such sets with $a(I) \geqslant \sqrt{\omega} = \sqrt{\omega(n)}$, which implies that $|S(I)| \leqslant \sqrt{\omega}$ and $k(I) \leqslant \omega$ for almost every sum-free m-set in $[n]$, as required. Hence we may assume that $m \leqslant \big(\frac{1}{2} - \delta\big)n$.

Next, we observe the following strengthening of Claim 3 when $m \geqslant C\sqrt{n \log n}$.

Claim 3′. *If $m \geqslant C\sqrt{n \log n}$, then there are $o\big(\binom{n/2}{m}\big)$ sum-free subsets $I \subset [n]$ of size m with $I \not\subset O_n$ and at least δm elements less than $n/2$.*

Proof of Claim 3′. The proof is almost identical to that of Claim 3. The only difference is that when we bound the number of sum-free m-sets I such that $1 \leqslant |I \setminus O_n| \leqslant \delta m$, we replace Proposition 4.1 by Proposition 4.2, which holds for $m \geqslant C\sqrt{n \log n}$, and implies that there are at most $o\big(\binom{n/2}{m}\big)$ such sets. When bounding the size of the collection $\mathcal{X}$ of sum-free m-sets $I \subset [n]$ that satisfy $|I \setminus B| \leqslant \delta^3 m$ for some interval B of length $n/2$, we use (20) and note that

$$m \cdot e^{-\delta^2 m} \left(\frac{4e}{\delta^2}\right)^{\delta^3 m} 2^{Cn/m} \binom{n/2}{m} \leqslant 2^{-\delta^3 m} \binom{n/2}{m},$$

since $\delta^3 m > Cn/m$ for $m \geqslant C\sqrt{n \log n}$. The rest of the proof is exactly the same. $\square$

By Claim 3′, we may restrict our attention to sum-free m-sets $I \subset [n]$ such that $I \not\subset O_n$ and $|S(I)| \leqslant \delta m$. By Claims 5 and 6, there are $o\big(\binom{n/2}{m}\big)$ such sets with $\ell(I) \geqslant \frac{Cn}{m} + \omega(n)$. However, if $\ell(I) \leqslant \frac{Cn}{m} + \omega(n)$ and

$$k(I) \,\geqslant\, \frac{C^4 n^3}{m^3} + \omega(n)^4 \,\geqslant\, \max\left\{\frac{\ell^2 n}{\delta m}, \frac{\ell n^2}{\delta m^2}\right\}, \tag{41}$$

then combining the proofs of Claims 4 and 5 proves the theorem. Indeed, first note that by Claim 5 there are $o\big(\binom{n/2}{m}\big)$ such sum-free m-sets I with $\ell(I) \geqslant \max\{n/4m, \omega\}$. Next, if $\ell(I) \leqslant n/4m$, then by (23) in the proof of Claim 4, there are at most

$$\sum_{k \geqslant C^4n^3/m^3 + \omega^4} C \left(\frac{e^2 k}{\ell^2}\right)^{\ell} e^{-km^2/2n^2} \left(\frac{2m}{n-2m}\right)^{\ell} \binom{n/2}{m} \tag{42}$$

such sets. Note that (42) is decreasing exponentially[14] in k. Thus if $m = o(n)$, then (42) is at most[15]

$$\frac{C^2n^2}{m^2}\left(\frac{e^2}{\ell^2}\cdot\frac{C^4n^3}{m^3}\cdot\frac{4m}{n}\right)^{\ell} e^{-C^4 n/2m}\binom{n/2}{m} \,\leqslant\, \frac{n^2}{m^2}\cdot e^{-C^3 n/m}\binom{n/2}{m} \,=\, o(1)\cdot\binom{n/2}{m}.$$

On the other hand, if $m = \Theta(n)$ then, since $n - 2m \geqslant \delta m$, it follows that (42) is at most

$$\omega \cdot \left(\frac{\omega^5}{\ell^2}\right)^{\ell} e^{-\omega^3}\binom{n/2}{m} \,\leqslant\, e^{-\omega^3/2}\binom{n/2}{m} \,=\, o(1)\cdot\binom{n/2}{m}.$$

Finally, if $n/4m \leqslant \ell(I) \leqslant \omega$ then, by (24), and again using the exponential decay in k, it follows that there are at most

$$\sum_{k \geqslant \omega^4} C\left(\frac{e^2k}{\ell^2}\right)^{\ell} e^{-km/8n\ell}\left(\frac{2m}{n-2m}\right)^{\ell}\binom{n/2}{m} \,\leqslant\, \omega\cdot\left(\frac{\omega^5}{\ell^2}\right)^{\ell} e^{-\omega^2/32}\binom{n/2}{m} \,=\, o(1)\cdot\binom{n/2}{m}$$

such sum-free m-sets. This completes the proof of Theorem 1.2. □

Finally, we prove that the bounds on $\ell(I) = |S(I)|$ and $k(I) = \sum_{a \in S(I)}(n/2 - a)$ given by Theorem 1.2 are best possible up to a constant factor. More precisely, we will show that if $\varepsilon = \varepsilon(C) > 0$ is sufficiently small and $C\sqrt{n \log n} \leqslant m = o(n)$, then almost every sum-free m-subset $I \subset [n]$ satisfies $\ell(I) \geqslant \varepsilon n/m$ and $k(I) \geqslant \varepsilon n^3/m^3$.

First, let us count the m-sets $I \subset [n]$ satisfying $\ell(I) \leqslant \varepsilon n/m$ and $k(I) \leqslant 2Cn^3/m^3$. By Lemma 5.1 the number of such sets is at most

$$\begin{aligned}\frac{2Cn^3}{m^3}\cdot\sum_{\ell=0}^{\varepsilon n/m}\left(\frac{e^2}{\ell^2}\cdot\frac{2Cn^3}{m^3}\right)^{\ell}\binom{n/2}{m-\ell} &\leqslant \frac{2Cn^3}{m^3}\cdot\sum_{\ell=0}^{\varepsilon n/m}\left(\frac{e^2}{\ell^2}\cdot\frac{2Cn^3}{m^3}\cdot\frac{3m}{n}\right)^{\ell}\binom{n/2}{m} \\ &\leqslant \frac{Cn^4}{m^4}\left(\frac{6e^2C}{\varepsilon^2}\right)^{\varepsilon n/m}\binom{n/2}{m} \,\leqslant\, 2^{O(\sqrt{\varepsilon} n/m)}\binom{n/2}{m}.\end{aligned}$$

[14] More precisely, if k increases to $k + 2n^2/m^2$, then (42) decreases by a factor of about e, by (41).

[15] Note, by differentiating $y = (c/x^2)^x$, that the left-hand side is maximized by setting $\ell^2 = 4C^4n^2/m^2$.

Indeed, the first inequality follows from (2) and the fact that $m = o(n)$, the second inequality follows from the fact that the summand is increasing for $\ell \in (0, \varepsilon n/m]$, and the third follows since $\varepsilon = \varepsilon(C)$ was chosen sufficiently small.

On the other hand, it follows from Proposition 3.1 that there are at least $2^{cn/m}\binom{n/2}{m}$ sum-free m-sets in $[n]$. Moreover, since $C\sqrt{n \log n} \leqslant m = o(n)$, by Theorem 1.2 almost all of them satisfy $k(I) \leqslant 2Cn^3/m^3$. Thus almost all sum-free m-sets $I \subset [n]$ satisfy $\ell(I) \geqslant \varepsilon n/m$, as claimed.

To show that almost every sum-free m-subset $I \subset [n]$ satisfies $k(I) \geqslant \varepsilon n^3/m^3$ we again apply Lemma 5.1. It follows that the number of m-sets $I \subset [n]$ satisfying $\ell(I) \leqslant 2Cn/m$ and $k(I) \leqslant \varepsilon n^3/m^3$ is at most

$$\frac{\varepsilon n^3}{m^3} \cdot \sum_{\ell=0}^{2Cn/m} \left(\frac{e^2}{\ell^2} \cdot \frac{\varepsilon n^3}{m^3} \right)^{\ell} \binom{n/2}{m-\ell} \leqslant \frac{\varepsilon n^3}{m^3} \cdot \sum_{\ell=0}^{2Cn/m} \left(\frac{e^2}{\ell^2} \cdot \frac{\varepsilon n^3}{m^3} \cdot \frac{4m}{n} \right)^{\ell} \binom{n/2}{m}$$
$$\leqslant \frac{Cn^4}{m^4} \cdot \exp\left(4\sqrt{\varepsilon} \cdot \frac{n}{m} \right) \binom{n/2}{m} \leqslant 2^{O(\sqrt{\varepsilon} n/m)} \binom{n/2}{m}.$$

Indeed, the first inequality follows from (2) and the fact that $m = o(n)$, and the second inequality follows since the summand is maximized by taking $\ell^2 = 4\varepsilon n^2/m^2$. By Proposition 3.1 and Theorem 1.2, there are at least $2^{cn/m}$ sum-free m-sets in $[n]$ and almost all of them satisfy $\ell(I) \leqslant 2Cn/m$. Therefore, if $\varepsilon > 0$ is sufficiently small, almost all sum-free m-sets $I \subset [n]$ satisfy $k(I) \geqslant \varepsilon n^3/m^3$, as claimed.

Acknowledgements

The third and fourth authors would like to thank Simon Griffiths and Gonzalo Fiz Pontiveros for several useful 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, J. Balogh, R. Morris and W. Samotij, Counting sum-free subsets of Abelian groups, to appear in *Israel J. Math.*
[4] 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.
[5] N. Alon and J. Spencer, The Probabilistic Method (3rd edition), Wiley Interscience, 2008.
[6] L. Babai, M. Simonovits and J. Spencer, Extremal subgraphs of random graphs, *J. Graph Theory*, **14** (1990), 599–622.
[7] J. Balogh, R. Morris and W. Samotij, Random sum-free subsets of Abelian groups, to appear in *Israel J. Math.*
[8] J. Balogh, R. Morris and W. Samotij, Independent sets in hypergraphs, submitted, arXiv:1204.6530v1.
[9] E. Breuillard, B. Green and T. Tao, The structure of approximate groups, to appear in *Pub. IHES.*
[10] N.J. Calkin, On the number of sum-free sets, *Bull. London Math. Soc.*, **22** (1990), 141–144.
[11] 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.

[12] 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.
[13] A. Cauchy, Recherches sur les nombres, *J. École Polytech*, **9** (1813), 99–116.
[14] D. Conlon and W.T. Gowers, Combinatorial Theorems in sparse random sets, submitted, arXiv:1011.4310.
[15] H. Davenport, On the addition of residue classes, *J. London Math. Soc.*, **10** (1935), 30–32.
[16] P.H. Diananda and H.P. Yap, Maximal sum-free sets of elements of finite groups, *Proc. Japan Academy*, **45** (1969), 1–5.
[17] 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.
[18] G.A. Freiman, The addition of finite sets. I. (Russian), *Izv. Vysš. Učebn. Zaved. Matematika*, **13** (1959), 202–213.
[19] G.A. Freiman, On the structure and the number of sum-free sets, *Journées Arithmétiques*, 1991 (Geneva), Astérique, **209** (1992) 195–201.
[20] 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.
[21] E. Friedgut, V. Rödl and M. Schacht, Ramsey properties of random discrete structures, *Random Structures Algorithms*, **37** (2010), 407–436.
[22] W.T. Gowers, A new proof of Szemerédi's theorem, *Geom. Funct. Anal.*, **11** (2001), 465–588.
[23] W.T. Gowers, Hypergraph regularity and the multidimensional Szemerédi theorem, *Annals of Math.*, **166** (2007), 897–946.
[24] R. Graham, V. Rödl and A. Ruciński, On Schur properties of random subsets of integers, *J. Number Theory*, **61** (1996), 388–408.
[25] B.J. Green, The Cameron-Erdős conjecture, *Bull. London Math. Soc.*, **36** (2004), 769–778.
[26] B.J. Green, Counting sets with small sumset, and the clique number of random Cayley graphs, *Combinatorica*, **25** (2005), 307–326.
[27] B.J. Green and I.Z. Ruzsa, Counting sumsets and sum-free sets modulo a prime, *Studia Sci. Math. Hungarica*, **41** (2004), 285–293.
[28] B.J. Green and I.Z. Ruzsa, Sum-free sets in abelian groups, *Israel J. Math.*, **147** (2005), 157–188.
[29] B.J. Green and T. Tao, The primes contain arbitrarily long arithmetic progressions, *Annals of Math.*, **167** (2008), 481–547.
[30] B.J. Green and T. Tao, Linear equations in primes, *Annals of Math.*, **171** (2010), 1753–1850.
[31] B.J. Green, T. Tao and T. Ziegler, An inverse theorem for the Gowers $U^{s+1}[N]$-norm, *Annals of Math.*, **176** (2012), 1231–1372.
[32] G.H. Hardy and S. Ramanujan, Asymptotic Formulae in Combinatory Analysis, *Proc. London Math. Soc.*, **17** (1918), 75–115.
[33] S. Janson, T. Łuczak and A. Ruciński, Random Graphs, Wiley, 2000.
[34] Y. Kohayakawa, T. Łuczak and V. Rödl, Arithmetic progressions of length three in subsets of a random set, *Acta Arith.*, **75** (1996), 133–163.
[35] V.F. Lev, T. Łuczak and T. Schoen, Sum-free sets in abelian groups, *Israel J. Math.*, **125** (2001), 347–367.
[36] D. Osthus, H.J. Prömel and A. Taraz, For which densities are random triangle-free graphs almost surely bipartite?, *Combinatorica*, **23** (2003), 201–250.
[37] R. Rado, Studien zur Kombinatorik, *Math. Zeit.*, **36** (1933), 424–480.
[38] V. Rödl and A. Ruciński, Threshold functions for Ramsey properties, *J. Amer. Math. Soc.*, **8** (1995), 917–942.
[39] V. Rödl and A. Ruciński, Rado partition theorem for random subsets of integers, *Proc. London Math. Soc.*, **74** (1997), 481–502.

[40] K.F. Roth, (1953), On certain sets of integers, *J. London Math. Soc.*, **28** (1952), 104–109.
[41] I.Z. Ruzsa, Generalized arithmetical progressions and sumsets, *Acta Math. Hungar.*, **65** (1994), 379–388.
[42] W. Samotij, Stability results for random discrete structures, to appear in *Random Structures Algorithms*.
[43] A.A. Sapozhenko, The Cameron-Erdős Conjecture, (Russian) *Dokl. Akad. Nauk.*, **393** (2003), no. 6, 749–752.
[44] A.A. Sapozhenko, Solution of the Cameron-Erdős problem for groups of prime order, *Zh. Vychisl. Mat. Mat. Fiz.*, **49** (2009), 1435–1441.
[45] D. Saxton and A. Thomason, Hypergraph containers, submitted, arXiv:1204.6595v2.
[46] M. Schacht, Extremal results for random discrete structures, submitted.
[47] I. Schur, Uber die Kongruenz $x^m + y^m \equiv z^m \pmod{p}$, *ber. Deutsch. Mat. Verein.*, **25** (1916), 114–117.
[48] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, *Acta Arith.*, **27** (1975), 199–245.
[49] T. Tao and V. Vu, Additive combinatorics, Cambridge University Press, 2006.
[50] H.P. Yap, Maximal sum-free sets of group elements, *J. London Math. Soc.*, **44** (1969), 131–136.

School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel
E-mail address: nogaa@post.tau.ac.il

Department of Mathematics, University of Illinois, 1409 W. Green Street, Urbana, IL 61801
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: ws299@cam.ac.uk