ISRAEL JOURNAL OF MATHEMATICS **236** (2020), 711–726
DOI: 10.1007/s11856-020-1986-z

# ON THE NUMBER OF SETS WITH A GIVEN DOUBLING CONSTANT

BY

Marcelo Campos*

*Instituto de Matemática Pura e Aplicada*
*Estrada Dona Castorina 110, Jardim Botânico*
*CEP 22460-320 Rio de Janeiro, RJ, Brasil*
*e-mail: marcelo.campos@impa.br*

ABSTRACT

We study the number of $s$-element subsets $J$ of a given abelian group $G$, such that $|J+J| \leq K|J|$. Proving a conjecture of Alon, Balogh, Morris and Samotij, and improving a result of Green and Morris, who proved the conjecture for $K$ fixed, we provide an upper bound on the number of such sets which is tight up to a factor of $2^{o(s)}$, when $G = \mathbb{Z}$ and $K = o(s/(\log n)^3)$. We also provide a generalization of this result to arbitrary abelian groups which is tight up to a factor of $2^{o(s)}$ in many cases. The main tool used in the proof is the asymmetric container lemma, introduced recently by Morris, Samotij and Saxton.

## 1. Introduction

In additive combinatorics one of the main objectives of the field is, given an abelian group $G$ and a finite subset $A \subset G$, to understand the relation between the sumset $A+A$ and $A$. In this direction, a fundamental result of Freiman [6] says that for $G = \mathbb{Z}$, if $|A+A| \leq K|A|$ (we say that $A$ has **doubling constant** $K$), then there is a generalized arithmetic progression $P$ such that $A \subset P$, the dimension of $P$ is at most $f(K)$, and $|P| \leq f(K)|A|$ for some function $f$.

* Research partially supported by CNPq.
Received November 16, 2018 and in revised form April 16, 2019

This was later generalized to the setting of arbitrary abelian groups by Green and Ruzsa [9], but many fundamental questions remain open, for example, whether $f$ can be a polynomial.

Another famous problem in additive combinatorics is the Cameron–Erdős conjecture about the number of sum-free subsets of $[n]$, which was solved independently by Green [7] and Sapozhenko [14]. More recently Alon, Balogh, Morris and Samotij [1] obtained a refinement of the Cameron–Erdős conjecture using an early form of the method of hypergraph containers. In order to prove this refinement of the Cameron–Erdős conjecture, they needed a bound on the number of $s$-sets $A \subset [n]$ with doubling constant $K$. They moreover conjectured that the following stronger (and, if true, best possible) bound holds.

CONJECTURE 1.1 (Alon, Balogh, Morris and Samotij): *For every $\delta > 0$, there exists $C > 0$ such that the following holds. If $s \geq C \log n$ and if $K \leq s/C$, then there are at most*

$$2^{\delta s} \binom{\frac{1}{2}Ks}{s}$$

*sets $J \subset [n]$ with $|J| = s$ and $|J + J| \leq K|J|$.*

The conjecture was later confirmed for $K$ constant by Green and Morris [8]; in fact they proved a slightly more general result: for each fixed $K$ and as $s \to \infty$, the number of sets $J \subset [n]$ with $|J| = s$ and $|J + J| \leq K|J|$ is at most

$$2^{o(s)} \binom{\frac{1}{2}Ks}{s} n^{\lfloor K + o(1) \rfloor}.$$

The authors of [8] used this result to bound the size of the largest clique in a random Cayley graph and recently the result was also applied by Balogh, Liu, Sharifzadeh and Treglown [2] to determine the number of maximal sum-free sets in $[n]$.

Our main theorem confirms Conjecture 1.1 for all $K = o(s/(\log n)^3)$.

THEOREM 1.2: *Let $s, n$ be integers and $2 \leq K \leq o(\frac{s}{(\log n)^3})$. The number of sets $J \subset [n]$ with $|J| = s$ such that $|J + J| \leq K|J|$ is at most*

$$2^{o(s)} \binom{\frac{1}{2}Ks}{s}.$$

We will in fact prove stronger bounds on the error term than those stated above, see Theorem 4.1. Nevertheless, we are unable to prove the conjecture in the range $K = \Omega(s/(\log n)^3)$, and actually the conjecture is false for a certain

range of values of $s$ and $K \gg s/\log n$. More precisely, for any integers $n, s$, and any positive numbers $K, \epsilon$ with

$$\min\{s, n^{1/2-\epsilon}\} \geq K \geq \frac{4\log(24C)s}{\epsilon \log n},$$

there are at least

$$\binom{\frac{n}{2}}{\frac{K}{4}}\binom{\frac{Ks}{8}}{s-\frac{K}{4}} \geq \binom{CKs}{s}$$

sets $J \subset [n]$ with $|J| = s$ and $|J + J| \leq Ks$. The construction[1] is very simple: let $P$ be an arithmetic progression of size $Ks/8$ and set $J = J_0 \cup J_1$, where $J_0$ is any subset of $P$ of size $s - K/4$, and $J_1$ is any subset of $[n] \setminus P$ of size $K/4$. For convenience we provide the details in the appendix of [5].

Our methods also allow us to characterize the typical structure of an $s$-set with doubling constant $K$, and obtain the following result.

THEOREM 1.3: *Let $s, n$ be integers and $2 \leq K \leq o(\frac{s}{(\log n)^3})$. For almost all sets $J \subset [n]$ with $|J| = s$ such that $|J+J| \leq K|J|$, there is a set $T \subset J$ such that $J \setminus T$ is contained in an arithmetic progression of size $\frac{1+o(1)}{2}Ks$ and $|T| = o(s)$.*

In the case $s = \Omega(n)$ (and hence $K = O(1)$), this result was proved by Mazur [11]. We will provide better bounds for the error terms in Theorem 5.1, below.

1.1. ABELIAN GROUPS. Notice that the doubling constant is defined for finite subsets of any abelian group. So, given a finite subset $Y$ of an abelian group, one might ask: how many subsets of $Y$ of size $s$ with doubling constant $K$ are there? We are also able to provide an answer to this more general question. From now on, fix an arbitrary abelian group $G$ throughout the paper. To state our main result formally in the context of general abelian groups we define, for each positive real number $t$, the quantity $\beta(t)$ to be the size of the biggest subgroup of $G$ of size at most $t$, that is,

$$\beta(t) = \max\{|H| : H \leqslant G,\ |H| \leq t\}. \tag{1}$$

THEOREM 1.4: *Let $s, n$ be integers, $2 \leq K \leq o(\frac{s}{(\log n)^3})$, and $Y \subset G$ with $|Y| = n$. The number of sets $J \subset Y$ with $|J| = s$ such that $|J + J| \leq K|J|$ is at most*

$$2^{o(s)}\binom{\frac{1}{2}(Ks+\beta)}{s},$$

*where $\beta := \beta((1 + o(1))Ks)$.*

[1] We would like to thank Rob Morris for pointing out this construction.

Again we will actually prove somewhat stronger (although slightly more convoluted) bounds for Theorem 1.4; see Theorem 4.1. We remark that Theorem 1.4 implies Theorem 1.2, since the only finite subgroup of $\mathbb{Z}$ is the trivial one, so in this case $\beta(t) = 1$ for all $t$. Finally, let us remark that Theorem 1.4 is best possible in many cases. Indeed suppose for some integers $l, m$ that the largest subgroup $H \leqslant G$ with $|H| \leq m \leq |G|$ is of size

$$\beta = \frac{m}{2l-1};$$

then there are at least

$$\binom{\frac{m+\beta}{2}}{s}$$

sets $J \subset G$ of size $s$ such that $|J + J| \leq m$. To see this, take an arithmetic progression $P \subset G/H$ of size $l$ (there exists one because of the choice of $H$) and consider $B = P + H$. Since $|B + B| \leq |P + P||H| = m$, for every set $J \subset B$ of size $s$ we have

$$|J + J| \leq |B + B| \leq m.$$

Therefore, there are at least

$$\binom{\frac{lm}{2l-1}}{s} = \binom{\frac{m+\beta}{2}}{s}$$

sets $J \subset B$ of size $s$ with $|J + J| \leq m$.

1.2. The method of hypergraph containers. Before diving into the proof of the main results, let us briefly mention the main tool used in the proof of Theorem 1.2. The method of hypergraph containers, introduced by Balogh, Morris and Samotij [3] and independently by Saxton and Thomason [15], has proven to be a very useful tool in counting problems that involve forbidden structures; for a general overview of the method and its applications see [4]. More recently, Morris, Samotij and Saxton [12] introduced asymmetric containers, a generalization of hypergraph containers for forbidden structures with some sort of asymmetry, and applied the method to give a structural characterization of almost all graphs with a given number of edges free of an induced $C_4$. A variant of the asymmetric container lemma, which follows essentially from a minor modification of the proof in [12], will be our main tool in this article; we give more details in the next section.

## 2. The asymmetric container lemma

In this section we will state our main tool and give a brief explanation of how we will apply it to our problem. Let $Y \subset G$, with $|Y| = n$, and observe that when trying to count sets $J \subset Y$ with $|J| = s$ and $|J + J| \leq Ks$, one may instead count sets $J \subset Y$ such that there is a set $I \subset Y$ with $J + J \subset I$ and $|I| \leq Ks$. Keeping this in mind, the following definition will be useful.

*Definition 2.1:* Given disjoint copies of $Y+Y$ and $Y$, namely $Y_0,\ Y_1$ respectively, and $A \subset Y_0$ and $B \subset Y_1$, we define $\mathcal{H}(A, B)$ to be the hypergraph with vertex set

$$V(\mathcal{H}(A, B)) := (Y_0 \setminus A) \cup B$$

and edge set

$$E(\mathcal{H}(A, B)) := \{(\{c\}, \{a, b\}) : c \in Y_0 \setminus A,\ a, b \in B,\ a + b = c\}.$$

Sometimes when $A$ and $B$ are clear from the context we will denote $\mathcal{H}(A, B)$ simply by $\mathcal{H}$. Notice that $\mathcal{H}(A, B)$ is not uniform since there are edges $(\{c\}, \{a\})$ corresponding to $a + a = c$, but these will not be a problem. The usefulness of Definition 2.1 is that now for every pair of sets $(I, J)$ with $J + J \subset I$ we know that $(Y_0 \setminus I) \cup J$ doesn't contain any edges of $\mathcal{H}(A, B)$, so $(Y_0 \setminus I) \cup J$ would usually be called an independent set, but instead we will call the pair $(I, J)$ independent for convenience. Since we have a method for counting what are usually called independent sets in hypergraphs, and each of those is in correspondence to what we call an independent pair, we can obtain a theorem for counting independent pairs.

To state the main tool in this article we will need to go into some more slightly technical definitions. We first define a useful generalization of uniform hypergraphs, that includes the hypergraph presented in Definition 2.1. Given disjoint finite sets $V_0,\ V_1$ we define an $(r_0, r_1)$-bounded hypergraph $\mathcal{H}$ on the vertex set $V = V_0 \cup V_1$ to be a set of edges $E(\mathcal{H}) \subset \binom{V_0}{\leq r_0} \times \binom{V_1}{\leq r_1}$. Note that the hypergraph in Definition 2.1 is $(1, 2)$-bounded. Given a pair $(W_0, W_1) \in 2^{V_0} \times 2^{V_1}$, we say $(W_0, W_1)$ **violates** $(e_0, e_1) \in E(\mathcal{H})$ if $e_0 \subset V_0 \setminus W_0$ and $e_1 \subset W_1$. If a set $(W_0, W_1)$ doesn't violate any $(e_0, e_1) \in E(\mathcal{H})$ then we call $(W_0, W_1)$ **independent** with respect to $\mathcal{H}$. Let $\mathcal{F}_{\leq m}(\mathcal{H}) \subset 2^{V(\mathcal{H})}$ be the family of independent pairs $(W_0, W_1)$ such that $|W_0| \leq m$, and observe that for any pair of sets $(I, J)$, with $|I| \leq m$ and $J + J \subset I$, we have $(I, J) \in \mathcal{F}_{\leq m}(\mathcal{H}(\emptyset, Y))$. We define

the codegree $d_{(L_0,L_1)}(\mathcal{H})$ of $L_0 \subset V_0,\ L_1 \subset V_1$ to be the size of the set

$$\{(e_0,e_1) \in E(\mathcal{H}) :\ L_0 \subset e_0,\ L_1 \subset e_1\}$$

and we define the maximum $(\ell_0,\ell_1)$-codegree of $\mathcal{H}$ to be

$$\Delta_{(\ell_0,\ell_1)} := \max\{d_{(L_0,L_1)}(\mathcal{H}) :\ L_0 \subset V_0,\ L_1 \subset V_1,\ |L_0| = \ell_0,\ |L_1| = \ell_1\}.$$

With all of this in mind we introduce a variant of the asymmetric container lemma of Morris, Samotij and Saxton [12] that we can, once we have a suitable supersaturation theorem to check the codegree condition, apply iteratively and prove Theorem 1.2.

THEOREM 2.2: *For all non-negative integers $r_0, r_1$, not both zero, and each $R > 0$, the following holds. Suppose that $\mathcal{H}$ is a non-empty $(r_0,r_1)$-bounded hypergraph with $V(\mathcal{H}) = V_0 \cup V_1$, and $b$, $m$, and $q$ are integers with $b \le \min\{m, |V_1|\}$, satisfying*

$$\Delta_{(\ell_0,\ell_1)}(\mathcal{H}) \le R\frac{b^{\ell_0+\ell_1-1}}{m^{\ell_0}|V_1|^{\ell_1}}e(\mathcal{H})\Big(\frac{m}{q}\Big)^{1[\ell_0>0]} \tag{2}$$

*for every pair $(\ell_0,\ell_1) \in \{0,1,\dots,r_0\} \times \{0,1,\dots,r_1\} \setminus \{(0,0)\}$. Then there exists a family $\mathcal{S} \subset \binom{V_0}{\le r_0 b} \times \binom{V_1}{\le r_1 b}$ and functions $f\colon \mathcal{S} \to 2^{V_0} \times 2^{V_1}$ and $g\colon \mathcal{F}_{\le m}(\mathcal{H}) \to \mathcal{S}$, such that, letting $\delta = 2^{-(r_0+r_1+1)(r_0+r_1)}R^{-1}$:*

(i) *If $f(g(I,J)) = (A,B)$ with $A \subset V_0$ and $B \subset V_1$, then $A \subset I$ and $J \subset B$.*
(ii) *For every $(A,B) \in f(\mathcal{S})$ either $|A| \ge \delta q$ or $|B| \le (1-\delta)|V_1|$.*
(iii) *If $g(I,J) = (S_0,S_1)$ and $f(g(I,J)) = (A,B)$, then $S_0 \subset V_0 \setminus I$ and $S_1 \subset J$, and $|S_0| > 0$ only if $|A| \ge \delta q$.*

The proof of this variant of the asymmetric container lemma is virtually identical to that in [12], but, for the sake of completeness, it is provided in the appendix of [5]. Let us remark that the main difference between this statement of the asymmetric container lemma and the one in [12] is that we partition the vertex set in two parts and treat them differently, which is essential in our application. More specifically, we will apply the container lemma iteratively in such a way that $V_1$ will shrink much more than $V_0$, and to account for this imbalance we must differentiate between the two sets of the partition. Another small difference is that the hypergraph $\mathcal{H}$ doesn't need to be uniform. Finally we observe that if $S_0$ is non-empty, where $g(I,J) = (S_0,S_1)$, then we must have $|A| \ge \delta q$, where $f(g(I,J)) = (A,B)$.

## 3. The Supersaturation results

We would like to remind the reader that $G$ will always be a fixed abelian group throughout the paper. To apply Theorem 2.2 to our setting we will need, for sets $A, B \subset G$, bounds on the number of pairs $(b_1, b_2) \in B \times B$ such that $b_1 + b_2 \notin A$. In the case $G = \mathbb{Z}$, one such result is Pollard's theorem [13], which tells us that if $|B| \geq (1/2 + \epsilon)|A|$ and $\epsilon < 1/2$ then at least an $\epsilon^2$ proportion of all pairs $(b_1, b_2) \in B \times B$ are such that $b_1 + b_2 \notin A$. To prove similar results for arbitrary abelian groups one has to have some control on the structure of the group. With this in mind, we define the following quantity.

*Definition 3.1:* Given finite sets $U, V \subset G$, we define

$$\alpha(U, V) = \max\{|V'| : V' \subset G,\ |V'| \leq |V|,\ |\langle V' \rangle| \leq |U| + |V| - |V'|\}.$$

Given $U, V \subset G$ and $x \in G$ we will use the notation $1_U * 1_V(x)$ to denote the number of pairs $(u, v) \in U \times V$ such that $u + v = x$. The following theorem is the generalization we want of Pollard's theorem for arbitrary abelian groups. It is a simple variant of a result of Hamidoune and Serra [10], but for completeness we provide a proof in the appendix of [5].

THEOREM 3.2: *Let $t$ be a positive integer and $U, V \subset G$ with $t \leq |V| \leq |U| < \infty$. Then*

$$\sum_{x \in G} \min(1_U * 1_V(x), t) \geq t(|U| + |V| - t - \alpha), \tag{3}$$

*where $\alpha := \alpha(U, V)$.*

This implies the following corollary.

COROLLARY 3.3: *Let $A, B \subset G$ be finite and non-empty sets, let $0 < \epsilon < \frac{1}{2}$ and set $\beta := \beta((1 + 4\epsilon)|A|)$. If $|B| \geq (\frac{1}{2} + \epsilon)(|A| + \beta)$ then there are at least $\epsilon^2|B|^2$ pairs $(b_1, b_2) \in B^2$ such that $b_1 + b_2 \notin A$.*

*Proof.* Note first that if $|B| \geq (1 + \epsilon)|A|$ then the result is trivial, since for each element $a \in A$ there are at most $|B|$ pairs $(b_1, b_2) \in B^2$ with $b_1 + b_2 = a$, and therefore there are at least $|B|^2 - |A||B| \geq \epsilon^2|B|^2$ pairs in $B$ whose sum is not in $A$. When $|B| \leq (1 + \epsilon)|A|$ we will apply Theorem 3.2 with $U = V = B$ and $t = \epsilon|B|$. We first observe that

$$\alpha(B, B) \leq \max(\beta, 2|B| - (1 + 4\epsilon)|A|).$$

Indeed, suppose that $B' \subset G$ satisfies $|\langle B'\rangle| \leq 2|B|-|B'|$. If $|\langle B'\rangle| > (1+4\epsilon)|A|$ then $|B'| \leq 2|B|-|\langle B'\rangle| \leq 2|B|-(1+4\epsilon)|A|$. Otherwise, if $|\langle B'\rangle| \leq (1+4\epsilon)|A|$, then by the definition (1) of $\beta$, we have $|B'| \leq |\langle B'\rangle| \leq \beta$.

Now by Theorem 3.2, we have

$$\sum_{x\in G} \min(1_B * 1_B(x), \epsilon|B|) \geq \epsilon|B|((2-\epsilon)|B| - \max(\beta, 2|B| - (1+4\epsilon)|A|)).$$

By subtracting from both sides the sum over $x \in A$, we obtain

$$\sum_{x\in G\setminus A} \min(1_B * 1_B(x), \epsilon|B|) \geq \epsilon|B|((2-\epsilon)|B| - \max(\beta, 2|B| - (1+4\epsilon)|A|) - |A|).$$

Now, if $2|B| - (1+4\epsilon)|A| \geq \beta$, then, using that $|B| \leq 2|A|$,

$$\sum_{x\in G\setminus A} 1_B * 1_B(x) \geq \epsilon|B|(4\epsilon|A| - \epsilon|B|) \geq \epsilon^2|B|^2$$

as required. Otherwise, if $\beta \geq 2|B| - (1+4\epsilon)|A|$, then

$$\sum_{x\in G\setminus A} 1_B * 1_B(x) \geq \epsilon|B|((2-\epsilon)|B| - \beta - |A|) \geq \epsilon^2|B|^2,$$

since $|B| \geq (\frac{1}{2}+\epsilon)(|A|+\beta)$ and $0 < \epsilon < \frac{1}{2}$, so $(2-\epsilon) - \frac{2}{1+2\epsilon} \geq \epsilon$. ∎

To prove a stability theorem for almost all sets with a given size and doubling constant we will also need the following result of Mazur [11].

THEOREM 3.4: *Let $l$ and $t$ be positive integers, with $t \leq l/40$, and let $B \subset \mathbb{Z}$ be a set of size $l$. Suppose that*

$$\sum_{x\in\mathbb{Z}} \min(1_B * 1_B(x), t) \leq (2+\delta)lt,$$

*for some $0 < \delta \leq 1/8$. Then there is an arithmetic progression $P$ of length at most $(1+2\delta)l + 6t$ containing all but at most $3t$ points of $B$.*

From Theorem 3.4 we can easily deduce the following corollary:

COROLLARY 3.5: *Let $s$ be an integer, $K > 0$, and $0 < \epsilon < 2^{-10}$. If $A, B \subset \mathbb{Z}$, with $(1-\epsilon)\frac{Ks}{2} \leq |B| \leq (1+2\epsilon)\frac{Ks}{2}$ and $|A| \leq Ks$, then one of the following holds:*

(a) *There are at least $4\epsilon^2K^2s^2$ pairs $(b_1, b_2) \in B^2$ such that $b_1 + b_2 \notin A$.*
(b) *There is an arithmetic progression $P$ of size at most $\frac{Ks}{2} + 32\epsilon Ks$ containing all but at most $8\epsilon Ks$ points of $B$.*

*Proof.* Suppose first that

$$\sum_{x\in\mathbb{Z}} \min(1_B * 1_B(x), t) \leq (2+8\epsilon)2\epsilon|B|Ks. \tag{4}$$

In this case we apply Theorem 3.4 with $l := |B|$, $\delta := 8\epsilon$, and $t = 2\epsilon Ks \leq l/40$, and deduce that (b) holds. Therefore suppose (4) doesn't hold; in this case

$$\sum_{x\in\mathbb{Z}\setminus A} \min(1_B * 1_B(x), t) \geq (2+8\epsilon)(1-\epsilon)\epsilon K^2 s^2 - t|A|,$$

since $|B| \geq (1-\epsilon)\frac{1}{2}Ks$. Noting that $t|A| \leq 2\epsilon K^2 s^2$ it follows that

$$\sum_{x\in\mathbb{Z}\setminus A} 1_B * 1_B(x) \geq ((2+8\epsilon)(1-\epsilon)-2)\epsilon K^2 s^2 \geq 4\epsilon^2 K^2 s^2,$$

since $\epsilon < 2^{-10}$, so (a) holds as required. ∎

## 4. The number of sets with a given doubling

In this section we prove the following statement which implies Theorems 1.2 and 1.4.

THEOREM 4.1: *Let $s, n$ be integers, let $2 \leq K < 2^{-36}\frac{s}{(\log n)^3}$, and let $Y \subset G$ with $|Y| = n$. The number of sets $J \subset Y$ with $|J| = s$ such that $|J+J| \leq K|J|$ is at most*

$$\exp(2^9\lambda K^{1/6}s^{5/6}\sqrt{\log n})\binom{\frac{1}{2}(Ks+\beta)}{s},$$

*where*

$$\beta := \beta(Ks + 2^6K^{7/6}s^{5/6}\sqrt{\log n}) \quad \textit{and} \quad \lambda := \min\left\{\frac{K}{K-2}, \log s\right\}.$$

Theorem 4.1 will follow easily from the following container theorem combined with Corollary 3.3. We will also use it together with Corollary 3.5 to prove Theorem 5.1.

THEOREM 4.2: *Let $m, n$ be integers with $m \geq (\log n)^2$, let $Y \subset G$ with $|Y| = n$, and let $0 < \epsilon < \frac{1}{4}$. There is a family $\mathcal{A} \subset 2^{Y+Y} \times 2^Y$ of pairs of sets $(A, B)$, of size*

$$|\mathcal{A}| \leq \exp\left(2^{16}\frac{1}{\epsilon^2}\sqrt{m}(\log n)^{3/2}\right), \tag{5}$$

*such that:*

(i) *For every pair of sets $J \subset Y$, $I \subset Y+Y$, with $J+J \subset I$ and $|I| \le m$ there is $(A,B) \in \mathcal{A}$ such that $A \subset I$ and $J \subset B$.*

(ii) *For every $(A,B) \in \mathcal{A}$, $|A| \le m$ and either $|B| \le \frac{m}{\log n}$ or there are at most $\epsilon^2|B|^2$ pairs $(b_1,b_2) \in B \times B$ such that $b_1+b_2 \notin A$.*

*Proof that Theorem 4.2 implies Theorem 4.1.* Let $\mathcal{A}$ be a family given by Theorem 4.2 applied with $m := Ks$ and $\epsilon > 0$ to be chosen later. Then by condition (i), for every $s$-set $J$ with doubling constant $K$ there is a pair $(A,B) \in \mathcal{A}$ such that $J \subset B$ and $A \subset J+J$. Define $\mathcal{B}$ to be the family of all sets $B$ that are in some container pair, that is

$$\mathcal{B} = \{B \subset Y : \exists A \text{ such that } (A,B) \in \mathcal{A}\}.$$

Observe that, by Corollary 3.3 and condition (ii) on $\mathcal{A}$, for every $B \in \mathcal{B}$ we have $|B| \le (\frac{1}{2}+\epsilon)(m+\beta)$, where $\beta := \beta((1+4\epsilon)m)$, since the number of pairs $(b_1,b_2) \in B^2$ such that $b_1+b_2 \notin A$ is at most $\epsilon^2|B|^2$ and $\frac{m}{\log n} \le (\frac{1}{2}+\epsilon)(m+\beta)$. Therefore the number of sets of size $s$ with doubling constant $K$ is at most

$$|\mathcal{B}| \max_{B \in \mathcal{B}} \binom{|B|}{s} \le \exp\left(2^{16}\frac{1}{\epsilon^2}\sqrt{Ks}(\log n)^{3/2}\right)\binom{(\frac{1+2\epsilon}{2})(Ks+\beta)}{s}. \tag{6}$$

Let $\lambda := \min\{\frac{K}{K-2}, \log s\}$; suppose first that $\frac{K}{K-2} \le \log s$. By applying the inequality $\binom{cn}{k} \le (\frac{cn-k}{n-k})^k \binom{n}{k}$ with $k=s$, $c=1+2\epsilon$ and $n = \frac{Ks+\beta}{2}$, it follows that in this case (6) is at most

$$\exp\left(2^{16}\frac{1}{\epsilon^2}\sqrt{Ks}(\log n)^{3/2} + 2\epsilon\lambda s\right)\binom{\frac{Ks+\beta}{2}}{s}.$$

Now choosing $\epsilon := 2^4(\frac{K}{s})^{1/6}\sqrt{\log n}$, by our restrictions on $K$ we see that

$$\epsilon < 2^4\Big(\frac{1}{2^{36}(\log n)^3}\Big)^{1/6}\sqrt{\log n} = \frac{1}{4}.$$

It follows that there are at most $\exp\left(2^9\lambda K^{1/6}s^{5/6}\sqrt{\log n}\right)\binom{\frac{1}{2}(Ks+\beta)}{s}$ sets of size $s$ with doubling constant $K$, when $\frac{K}{K-2} \le \log s$. If $\log s \le \frac{K}{K-2}$ we use the binomial estimate

$$\binom{(\frac{1+2\epsilon}{2})(Ks+\beta)}{s} \le \exp\left(4\epsilon s \log\frac{1}{\epsilon}\right)\binom{\frac{Ks+\beta}{2}}{s}$$

and the result follows by a similar calculation. Since

$$\beta(m+4\epsilon m) = \beta(Ks + 2^6K^{7/6}s^{5/6}\sqrt{\log n}),$$

this proves the theorem. ∎

Before we proceed with the proof of Theorem 4.2, let us give a brief overview of how we will deduce it from Theorem 2.2. We fix from now on a finite subset $Y \subset G$ with $|Y| = n$, and recall that the $(1,2)$-bounded hypergraph $\mathcal{H}(A,B)$ in Definition 2.1 was defined to have as edges pairs $(\{c\},\{a,b\})$ where $a+b=c$, with $a,b \in B$ and $c \notin A$. Note that condition (ii) in Theorem 4.2 implies that $\mathcal{H}(A,B)$ has at most $\frac{\epsilon^2}{2}|B|^2$ edges, as long as $|B| > \frac{m}{\log n}$. We remind the reader that a pair of sets $I \subset Y+Y$ and $J \subset Y$ with $J+J \subset I$ correspond to an independent set in $\mathcal{H}(A,B)$ for any $A \subset Y+Y$ and $B \subset Y$, since there are no $c \notin I$ and $a,b \in J$ such that $a+b=c$. If we additionally assume that $(I,J) \in \mathcal{F}_{\leq m}(\mathcal{H})$, then we know that every $J$ that is in such an independent pair satisfies $|J+J| \leq m$.

Our strategy will be to iteratively apply the container lemma until either there are few edges in the hypergraph $\mathcal{H}(A,B)$, or $|A| > m$, in which case the container doesn't contain any elements of $\mathcal{F}_{\leq m}(\mathcal{H})$. More precisely we will build a rooted tree $\mathcal{T}$ with root $\mathcal{H}(\emptyset, Y)$ whose vertices correspond to hypergraphs $\mathcal{H}(A,B)$ and whose leaves correspond to a family $\mathcal{A}$ satisfying the conclusion of Theorem 4.2. Given a vertex $\mathcal{H}(A,B)$ of the tree, such that

$$|A| \leq m, \quad |B| > \frac{m}{\log n}$$

and

$$e(\mathcal{H}(A,B)) > \frac{\epsilon^2}{2}|B|^2, \tag{7}$$

we will generate its children by applying the following procedure:

(a) Apply the asymmetric container lemma (Theorem 2.2) to $\mathcal{H} := \mathcal{H}(A,B)$ setting

$$R := \frac{2}{\epsilon^2}, \quad q := \frac{m}{\log n}, \quad b := \sqrt{\frac{m}{\log n}}.$$

Notice that the co-degrees of $\mathcal{H}$ satisfy

$$\max\{\Delta_{(1,0)}(\mathcal{H}), \Delta_{(0,1)}(\mathcal{H})\} \leq |B| = \frac{2}{\epsilon^2}\frac{\epsilon^2|B|^2}{2|B|} \leq R\frac{e(\mathcal{H})}{|B|}$$

and

$$\begin{aligned}\Delta_{(0,2)}(\mathcal{H}) =& \Delta_{(1,1)}(\mathcal{H}) = \Delta_{(1,2)}(\mathcal{H}) = 1\\ =& \frac{2}{\epsilon^2}\frac{b^2}{q|B|^2}\frac{\epsilon^2}{2}|B|^2\\ \leq& R\frac{b^2}{q|B|^2}e(\mathcal{H}),\end{aligned}$$

since (7) holds. Since $b < q < |B|$, it follows that

$$\Delta_{(0,2)}(\mathcal{H}) \leq R\frac{b^2}{q|B|^2}e(\mathcal{H}) \leq R\frac{b}{|B|^2}e(\mathcal{H}),$$

$$\Delta_{(1,1)}(\mathcal{H}) \leq R\frac{b^2}{q|B|^2}e(\mathcal{H}) \leq R\frac{b}{q|B|}e(\mathcal{H})$$

and

$$\Delta_{(1,0)}(\mathcal{H}) \leq R\frac{e(\mathcal{H})}{|B|} \leq R\frac{e(\mathcal{H})}{q},$$

as required.

(b) By Theorem 2.2, there exists a family $\mathcal{C} \subset 2^{(Y+Y)\setminus A} \times 2^B$ of at most

$$\binom{n^2}{b}\binom{|B|}{2b} \leq n^{4b} \leq e^{4\sqrt{m\log n}} \tag{8}$$

pairs of sets $(C, D)$ that satisfies the conditions of the container lemma. That is for each independent pair $(I, J) \in \mathcal{F}_{\leq m}(\mathcal{H})$, with $I \subset Y + Y$ and $J \subset Y$, there is $(C, D) \in \mathcal{C}$ such that $C \subset I$ and $J \subset D$, and either $|C| \geq \delta\frac{m}{\log n}$ or $D \leq (1-\delta)|B|$.

(c) For each $(C, D) \in \mathcal{C}$, let $\mathcal{H}(A\cup C, D)$ be a child of $\mathcal{H}(A, B)$ in the tree $\mathcal{T}$.

Now to count the number of leaves of $\mathcal{T}$ we will first bound its depth.

LEMMA 4.3: *The tree $\mathcal{T}$ has depth at most $d = 2^{14}\epsilon^{-2}\log n$.*

*Proof.* We will prove that after $d$ iterations either $|A| > m$ or

$$|B| \leq \frac{m}{\log n}e(\mathcal{H}(A, B)) \leq \frac{\epsilon^2}{2}|B|^2.$$

Notice that the $\delta$ provided by Theorem 2.2 in this application is $2^{-13}\epsilon^2$ and in each iteration either we increase the size of $A$ by $\delta q$ or we decrease the size of $B$ by $\delta|B|$. After $d$ iterations, either we would have increased the size of $A$ more than $\frac{d}{2}$ times, in which case

$$|A| > \frac{d}{2}\delta q = \frac{2^{13}\log n}{\epsilon^2}2^{-13}\epsilon^2\frac{m}{\log n} = m,$$

or we would have reduced the size of $B$ at least $\frac{d}{2}$ times, in which case

$$|B| \leq (1-\delta)^{\frac{d}{2}}n < e^{-\frac{\delta d}{2}}n \leq e^{-\log n}n = 1.$$

In either case, we would have stopped already by this point because we only generate children of $\mathcal{H}(A, B)$ if $|A| \leq m$, $|B| > \frac{m}{\log n}$ and (7) holds. ∎

*Proof of Theorem 4.2.* Let $\mathcal{L}$ be the set of leaves of the tree $\mathcal{T}$ constructed above, and define

$$\mathcal{A} := \{(A,B) : A \subset Y + Y,\ B \subset Y,\ \mathcal{H}(A,B) \in \mathcal{L},\ |A| \leq m\}.$$

Notice that for every $(A,B) \in \mathcal{A}$, we have either the bound $e(\mathcal{H}(A,B)) \leq \frac{\epsilon^2}{2}|B|^2$ or $|B| \leq \frac{m}{\log n}$, since they come from the leaves of $\mathcal{T}$ and $|A| \leq m$. Since the edges of $\mathcal{H}(A,B)$ correspond exactly to pairs $a, b \in B$ such that $a + b \notin A$, it follows that $\mathcal{A}$ has property (ii).

To bound the size of $\mathcal{A}$, notice that the number of leaves of the tree $\mathcal{T}$ is at most $Z^d$ where $Z$ denotes the maximum number of children of a vertex of the tree and $d$ denotes its depth. Thus, by (8) and Lemma 4.3,

$$|\mathcal{A}| \leq |\mathcal{L}| \leq Z^d \leq \exp\left(2^{16}\frac{1}{\epsilon^2}\sqrt{m}(\log n)^{3/2}\right),$$

so $\mathcal{A}$ satisfies (5), as required.

Finally, observe that for every pair of sets $J \subset Y,\ I \subset Y + Y$ with $J + J \subset I$ and $|I| \leq m$, there is $(A,B) \in \mathcal{A}$ such that $A \subset I$ and $J \subset B$. Indeed $(I,J) \in \mathcal{F}_{\leq m}(\mathcal{H}(\emptyset, Y))$ and therefore, by property (b) of our containers, there exists a path from the root to a leaf of $\mathcal{T}$ such that $A \subset I$ and $J \subset B$ for every vertex $\mathcal{H}(A,B)$ of the path, so (i) holds. ∎

## 5. A typical structure result

In this section we use Theorem 4.2 to determine the typical structure of a set $J \subset [n]$ of a given size with doubling constant $K$.

THEOREM 5.1: *Let $s, n$ be integers, let $2 \leq K \leq \frac{s}{2^{120}(\log n)^3}$, and let $J \subset [n]$ be a uniformly chosen random set with $|J| = s$ and $|J + J| \leq K|J|$. With probability at least $1 - \exp(-K^{1/6}s^{5/6}\sqrt{\log n})$ the following holds: there is a set $T \subset J$, of size*

$$|T| \leq 2^{15}K^{1/6}s^{5/6}\sqrt{\log n},$$

*such that $J \setminus T$ is contained in an arithmetic progression of size*

$$\frac{Ks}{2} + 2^{17}K^{7/6}s^{5/6}\sqrt{\log n}.$$

The proof of Theorem 5.1 is similar to that of Theorem 4.1, but we use Corollary 3.5 as well as Corollary 3.3.

*Proof of Theorem 5.1.* Let $G := \mathbb{Z}$ and apply Theorem 4.2 to the set $Y := [n]$ with $m := Ks$ and $\epsilon > 0$ to be chosen later. We say $B \subset [n]$ is $(\epsilon, Ks)$**-close to an arithmetic progression** if there is an arithmetic progression $P$ with $|P| \le \frac{Ks}{2} + 2^5\epsilon Ks$, and a set $T \subset B$ with $|T| \le 2^5\epsilon|B|$ such that $B \setminus T \subset P$. We claim that if $\mathcal{A}$ is the family provided by Theorem 4.2, then for every pair $(A, B) \in \mathcal{A}$ either

(I) $|B| \le (1-\epsilon)\frac{Ks}{2}$ or

(II) $B$ is $(\epsilon, Ks)$-close to an arithmetic progression.

To see this, note first that, by condition (ii) in Theorem 4.2, for every pair $(A, B) \in \mathcal{A}$ either there are at most $\epsilon^2|B|^2$ pairs $b_1, b_2 \in B$ with $b_1 + b_2 \notin A$ or $|B| \le \frac{m}{\log n}$, and so, by Corollary 3.3, $|B| \le (1+2\epsilon)\frac{Ks}{2}$. Now, if (I) doesn't hold, that is $|B| \ge (1-\epsilon)\frac{Ks}{2}$, then, by Corollary 3.5, (II) holds, since there are at most $\epsilon^2|B|^2 < 4\epsilon^2K^2s^2$ pairs $b_1, b_2 \in B$ such that $b_1 + b_2 \notin A$.

Now we will count the number of sets $J$ of size $s$ and doubling constant $K$ such that $J$ is not $(2^4\epsilon, Ks)$-close to an arithmetic progression. Recall from Theorem 4.2 (i) that, for any such set, there exists $(A, B) \in \mathcal{A}$ such that $J \subset B$. Now, observe that there are at most $|\mathcal{A}|\binom{(1-\epsilon)\frac{Ks}{2}}{s}$ sets $J$ of size $s$ that are contained in a set $B$ such that $(A, B) \in \mathcal{A}$ and $|B| \le (1-\epsilon)\frac{Ks}{2}$. Choosing

$$\epsilon := 2^6\Big(\frac{K}{s}\Big)^{1/6}\sqrt{\log n} < 2^{-10}$$

and using the bound (5) on the size of $\mathcal{A}$, we obtain

$$\begin{aligned}|\mathcal{A}|\binom{(1-\epsilon)\frac{Ks}{2}}{s} &\le \exp(2^{16}\epsilon^{-2}\sqrt{Ks}(\log n)^{3/2} - \epsilon s)\binom{\frac{Ks}{2}}{s}\\ &\le \exp(-2^5K^{1/6}s^{5/6}(\log n)^{1/2})\binom{\frac{Ks}{2}}{s}.\end{aligned} \tag{9}$$

Finally, we count the number of sets $J$ of size $s$ that are not $(2^4\epsilon, Ks)$-close to an arithmetic progression and are contained in a set $B$ such that $(A, B) \in \mathcal{A}$ and $B$ is $(\epsilon, Ks)$-close to an arithmetic progression. For each such $B$, let $P$ be an arithmetic progression with $|P| \le \frac{Ks}{2} + 2^5\epsilon Ks$, and $T \subset B$ be a set with $|T| \le 2^5\epsilon|B| \le 2^5\epsilon Ks$, such that $B \setminus T \subset P$. Observe that there at most

$$\sum_{s' \ge 2^9\epsilon s}\binom{(1+2\epsilon)\frac{Ks}{2}}{s-s'}\binom{2^5\epsilon Ks}{s'} \tag{10}$$

$s$-sets $J \subset B$ that are not $(2^4\epsilon, Ks)$-close to an arithmetic progression, since they must have $s - s'$ elements in $B \setminus T$ and $s'$ elements in $T$ for some $s' \ge 2^9\epsilon s$.

Indeed, otherwise $J \setminus T \subset P$, with $|P| \le Ks + 2^9\epsilon Ks$ and $|J \cap T| < 2^9\epsilon|J|$. To bound this we will use

$$\binom{a}{c-d}\binom{b}{d} \le \binom{a}{c}\left(\frac{4bc}{ad}\right)^d,$$

valid for $d \le c \le a/4$. Note that, by our choice of $\epsilon$, we have $|\mathcal{A}| \le e^{\epsilon s}$ (cf. (9)). Hence summing (10) over $(A, B) \in \mathcal{A}$ we obtain[2]

$$\begin{aligned}|\mathcal{A}| \cdot s \max_{s' \ge 2^9\epsilon s} (1+4\epsilon)^s \binom{\frac{Ks}{2}}{s-s'}\binom{2^5\epsilon Ks}{s'} &\le |\mathcal{A}| \cdot s \max_{s' \ge 2^9\epsilon s} (1+4\epsilon)^s \binom{\frac{Ks}{2}}{s}\left(\frac{2^8\epsilon s}{s'}\right)^{s'} \\ &\le \left(\frac{2^8\epsilon s}{2^9\epsilon s}\right)^{2^9\epsilon s} 2^{6\epsilon s}\binom{\frac{Ks}{2}}{s} \\ &\le \exp(-2^{11}K^{1/6}s^{5/6}\sqrt{\log n})\binom{\frac{Ks}{2}}{s}.\end{aligned} \tag{11}$$

Finally observe that the bound (9) and (11) imply the probability we claimed in the statement since, by taking all subsets of size $s$ of an arithmetic progression of length $\frac{Ks}{2}$, there are at least $\binom{\frac{Ks}{2}}{s}$ sets of size $s$ and doubling constant $K$. ∎

Acknowledgements. The author would like to thank Rob Morris for his thorough comments on the manuscript and many helpful discussions. We would also like to thank Mauricio Collares, Victor Souza and Natasha Morrison for interesting discussions and comments on early versions of this manuscript. The author is also very grateful to Hoi Nguyen for spotting a mistake in and suggesting various improvements to the first version of this manuscript.

## References

[1] N. Alon, J. Balogh, R. Morris and W. Samotij, *A refinement of the Cameron–Erdős conjecture*, Proceedings of the London Mathematical Society **108** (2013), 44–72.

[2] J. Balogh, H. Liu, M. Sharifzadeh and A. Treglown, *Sharp bound on the number of maximal sum-free subsets of integers*, Journal of the European Mathematical Society **20** (2018), 1885–1911.

[3] J. Balogh, R. Morris and W. Samotij, *Independent sets in hypergraphs*, Journal of the American Mathematical Society **28** (2015), 669–709.

[2] We remark that if $K < 16$ then $\binom{2^5\epsilon Ks}{s'} = 0$ for all $s' \ge 2^9\epsilon s$, so we may suppose that $K \ge 16$.

[4] J. Balogh, R. Morris and W. Samotij, *The method of hypergraph containers*, in *Proceedings of the International Congress of Mathematicians–Rio de Janeiro 2018. Vol. IV*, World Scientific, Hackensack, NJ, 2018, pp. 3059–3092.

[5] M. Campos, *On the number of sets with a given constant*, https://arxiv.org/abs/1811.05793.

[6] G. A. Freiman, *The addition of finite sets I*, Izvestiya Vysshikh Uchebnykh Zavedenii. Matematika (1959), no. 6, 202–213.

[7] B. Green, *The Cameron–Erdős conjecture*, Bulletin of the London Mathematical Society **36** (2004), 769–778.

[8] B. Green and R. Morris, *Counting sets with small sumset and applications*, Combinatorica **36** (2016), 129–159.

[9] B. Green and I. Z. Ruzsa, *Freiman's theorem in an arbitrary abelian group*, Journal of the London Mathematical Society **75** (2007), 163–175.

[10] Y. O. Hamidoune and O. Serra, *A note on Pollard's theorem*, https://arxiv.org/abs/0804.2593,

[11] P. Mazur, *A structure theorem for sets of small popular doubling*, Acta Arithmetica **171** (2015), 221–239.

[12] R. Morris, W. Samotij and D. Saxton, *An asymmetric container lemma and the structure of graphs with no induced* 4*-cycle*, https://arxiv.org/abs/1806.03706.

[13] J. M. Pollard, *A generalisation of the theorem of Cauchy and Davenport*, Journal of the London Mathematical Society **2** (1974), 460–462.

[14] A. A. Sapozhenko, *The Cameron–Erdős conjecture*, Rossiĭskaya Akademiya Nauk. Doklady Akademii Nauk **393** (2003), 749–752.

[15] D. Saxton and A. Thomason, *Hypergraph containers*, Inventiones Mathematicae **201** (2015), 925–992.