

# ON THE NUMBER OF $B_h$ -SETS

DOMINGOS DELLAMONICA JR., YOSHIHARU KOHAYAKAWA, SANG JUNE LEE, VOJTĚCH RÖDL,  
AND WOJCIECH SAMOTIJ

ABSTRACT. A set  $A$  of positive integers is a  $B_h$ -set if all the sums of the form  $a_1 + \dots + a_h$ , with  $a_1, \dots, a_h \in A$  and  $a_1 \leq \dots \leq a_h$ , are distinct. We provide asymptotic bounds for the number of  $B_h$ -sets of a given cardinality contained in the interval  $[n] = \{1, \dots, n\}$ . As a consequence of our results, we address a problem of Cameron and Erdős (1990) in the context of  $B_h$ -sets. We also use these results to estimate the maximum size of a  $B_h$ -set contained in a typical (random) subset of  $[n]$  with a given cardinality.

## 1. INTRODUCTION

We deal with a natural extension of the concept of *Sidon sets*: For a positive integer  $h \geq 2$ , a set  $A$  of integers is called a  $B_h$ -set if all sums of the form  $a_1 + \dots + a_h$ , where  $a_i \in A$  and  $a_1 \leq \dots \leq a_h$ , are distinct. We obtain Sidon sets letting  $h = 2$ . A central classical problem on  $B_h$ -sets is the determination of the maximum size  $F_h(n)$  of a  $B_h$ -set contained in  $[n] := \{1, \dots, n\}$ . Results of Chowla, Erdős, Singer, and Turán [5, 9, 10, 29] from the 1940s yield that  $F_2(n) = (1 + o(1))\sqrt{n}$ , where  $o(1)$  is a function that tends to 0 as  $n \rightarrow \infty$ . In 1962, Bose and Chowla [2] showed that  $F_h(n) \geq (1 + o(1))n^{1/h}$  for  $h \geq 3$ . On the other hand, an easy argument gives that for every  $h \geq 3$ ,

$$F_h(n) \leq (h \cdot h! \cdot n)^{1/h} \leq h^2 n^{1/h}. \quad (1)$$

Successively better bounds of the form  $F_h(n) \leq c_h n^{1/h}$  were given in [4, 6, 8, 15, 21, 22, 23, 28]. Currently, the best known upper bound on the constant  $c_h$  is given by Green [11], who proved that

$$c_3 < 1.519, \quad c_4 < 1.627, \quad \text{and} \quad c_h \leq \frac{1}{2e} \left( h + \left( \frac{3}{2} + o(1) \right) \log h \right),$$

where  $o(1) \rightarrow 0$  as  $h \rightarrow \infty$ . The interested reader is referred to the classical monograph of Halberstam and Roth [12] and to a recent survey by O'Bryant [24] and the references therein.

We study two problems related to the classical question of estimating  $F_h(n)$ . The first one is a natural generalization, to  $B_h$ -sets, of the problem of estimating the *number* of Sidon sets contained in  $[n]$ , proposed by Cameron and Erdős [3]. Second, we investigate the *maximum size* of a  $B_h$ -set contained in a *random subset* of  $[n]$ , in the spirit of [18, 19]. This second problem belongs to the study of extremal properties of random subsets contained in  $[n]$  (see, e.g., [7, 20, 25, 27, 30]). We present and discuss our results in detail in Section 2.

---

*Date:* 2014/11/26, 11.03pm.

The second author was partially supported by FAPESP (2013/03447-6, 2013/07699-0), CNPq (308509/2007-2 and 477203/2012-4), NSF (DMS 1102086) and NUMEC/USP (Project MaCLinC/USP). The third author was supported by Basic Science Research Program through the National Research Foundation of Korea (NRF) funded by the Ministry of Science, ICT & Future Planning (NRF-2013R1A1A1059913), and also by the National Research Foundation of Korea (NRF) Grant funded by the Korea Government (MSIP)(No. 2013042157). The fourth author was supported by the NSF grants DMS 0800070, 1301698, and 1102086. The fifth author was partially supported by ERC Advanced Grant DMMCA and a Trinity College JRF.

Our notation is standard. We write  $a \ll b$  as shorthand for the statement  $a/b \rightarrow 0$  as  $n \rightarrow \infty$ . We omit floor  $\lfloor \cdot \rfloor$  and ceiling  $\lceil \cdot \rceil$  symbols when they are not essential. We are mostly interested in large  $n$ ; in our statements and inequalities we often tacitly assume that  $n$  is larger than a suitably large constant.

## 2. THE MAIN RESULTS

Our main results are presented in two separate sections. We first discuss enumeration results and then we move on to their probabilistic consequences.

**2.1. A generalization of a problem of Cameron and Erdős.** Let  $\mathcal{Z}_n^h$  be the family of  $B_h$ -sets contained in  $[n]$ . In 1990, Cameron and Erdős [3] proposed the problem of estimating  $|\mathcal{Z}_n^2|$ , that is, the number of Sidon sets contained in  $[n]$ . We investigate the problem of estimating  $|\mathcal{Z}_n^h|$  for arbitrary  $h \geq 2$ . Recalling that  $F_h(n)$  is the maximum size of a  $B_h$ -set contained in  $[n]$ , one trivially has

$$2^{F_h(n)} \leq |\mathcal{Z}_n^h| \leq \sum_{i=0}^{F_h(n)} \binom{n}{i} \leq (1 + F_h(n)) \binom{n}{F_h(n)}.$$

Since  $(1 + o(1))n^{1/h} \leq F_h(n) \leq c_h n^{1/h}$  for some constant  $c_h$ , we have

$$2^{(1+o(1))n^{1/h}} \leq |\mathcal{Z}_n^h| \leq n^{c'_h n^{1/h}}, \quad (2)$$

for some constant  $c'_h$ . We improve the upper bound on  $|\mathcal{Z}_n^h|$  in (2) as follows.

**Theorem 2.1.** *For every  $h \geq 2$ , we have  $|\mathcal{Z}_n^h| \leq 2^{C_h n^{1/h}}$ , where  $C_h$  is a constant that depends only on  $h$ .*

The case  $h = 2$  in Theorem 2.1 was established in [18] and later given another proof in [26]. Our proof of Theorem 2.1 is based on the solution of a refined version of the question. Let  $\mathcal{Z}_n^h(t)$  be the family of  $B_h$ -sets contained in  $[n]$  with  $t$  elements. Theorem 2.1 is obtained from the following result, which estimates  $|\mathcal{Z}_n^h(t)|$  for all  $t \geq n^{1/(h+1)}(\log n)^2$ .

**Theorem 2.2.** *For every  $h \geq 2$  and any  $t \geq n^{1/(h+1)}(\log n)^2$ ,*

$$|\mathcal{Z}_n^h(t)| \leq \left( \frac{c_h n}{t^h} \right)^t, \quad (3)$$

where  $c_h = e^6(2h)^{2h}$ .

The derivation of Theorem 2.1 from Theorem 2.2 is given in Section 3 and Theorem 2.2 is proved in Section 4.2.

We now turn to lower bounds for  $|\mathcal{Z}_n^h(t)|$ . The bound in (5) in Proposition 2.3(ii) below complements (3) in Theorem 2.2. On the other hand, Proposition 2.3(i) shows that for small  $t$ , say,  $t \ll n^{1/(2h-1)}$ , the  $B_h$ -sets in  $[n]$  form a much larger proportion of the total number  $\binom{n}{t}$  of  $t$ -element sets (see (4)). Note that for large  $t$ , namely,  $t \geq n^{1/(h+1)}(\log n)^2$ , Theorem 2.2 tells us that this proportion is, very roughly speaking, of the order of  $(n/t^h)^t \binom{n}{t}^{-1} \leq (n/t^h)^t / (n/t)^t = t^{-(h-1)t}$ .

**Proposition 2.3.** *The following bounds hold for every  $h \geq 2$ .*

(i) For any  $\delta > 0$ , there exists an  $\varepsilon > 0$  such that for any  $t \leq \varepsilon n^{1/(2h-1)}$ ,

$$|\mathcal{Z}_n^h(t)| \geq (1 - \delta)^t \binom{n}{t}. \quad (4)$$

(ii) There are constants  $c'_h, \varepsilon' = \varepsilon'(h) > 0$  such that for all  $t \leq \varepsilon' n^{1/h}$ ,

$$|\mathcal{Z}_n^h(t)| \geq \left( \frac{c'_h n}{t^h} \right)^t. \quad (5)$$

The combination of the lower bounds in Proposition 2.3 and the upper bound of Theorem 2.2 naturally partitions the range of  $t$  into three intervals.

- For  $t \ll n^{1/(2h-1)}$ , Proposition 2.3(i) tells us that  $|\mathcal{Z}_n^h(t)|$  is, up to a multiplicative factor of  $(1 - o(1))^t$ , equal to the number  $\binom{n}{t}$  of all  $t$ -element subsets of  $[n]$ . In this range, one might therefore say that  $B_h$ -sets are ‘relatively abundant’.
- For  $t$  between  $n^{1/(2h-1)}$  and  $n^{1/(h+1)}(\log n)^2$ , a trivial, although loose upper bound follows from the monotonicity of  $|\mathcal{Z}_n^h(t)|$ , that is,  $|\mathcal{Z}_n^h(t)| \leq |\mathcal{Z}_n^h(n^{1/(h+1)}(\log n)^2)|$ , which is then bounded by Theorem 2.2. We note that the lower bound, given by Proposition 2.3(ii) is quite far from the upper bound. In the final section of this paper we present Conjecture 7.1, which states that the upper bound should essentially match the lower bound of Proposition 2.3(ii).
- For  $t \geq n^{1/(h+1)}(\log n)^2$ , Theorem 2.2 and Proposition 2.3(ii) determine  $|\mathcal{Z}_n^h(t)|$  up to a multiplicative factor of the form  $c^t$ . In this range of  $t$ ,  $B_h$ -sets are therefore much scarcer than in the first range.

**2.2. Almost  $B_h$ -sets.** We now consider a generalization of the notion of a  $B_h$ -set. For a set  $S$  of integers and an integer  $z$ , let

$$r_{S,h}(z) = \left| \left\{ (a_1, \dots, a_h) \in S^h : a_1 + \dots + a_h = z \text{ and } a_1 \leq \dots \leq a_h \right\} \right|. \quad (6)$$

**Definition 2.4.** A set  $S$  is called a  $B_h[g]$ -set if  $r_{S,h}(z) \leq g$  for all integers  $z$ .

Observe that a  $B_h[1]$ -set is simply a  $B_h$ -set and hence this definition extends the notion of  $B_h$ -sets. Let  $F_{h,g}(n)$  denote the maximum size of a  $B_h[g]$ -set contained in  $[n]$ . It is not hard to see that

$$(1 + o(1))n^{1/h} \leq F_h(n) \leq F_{h,g}(n) \leq (gh \cdot h!)^{1/h} n^{1/h}. \quad (7)$$

Our final result in this section gives a lower bound for the cardinality of  $\mathcal{Z}_n^{h,g}(t)$ , the family of  $t$ -element  $B_h[g]$ -sets contained in  $[n]$ .

**Theorem 2.5.** Fix an integer  $h \geq 2$  and a function  $g = g(n)$ . For every fixed  $\delta > 0$  and integer  $1 \leq t \ll (n^{1-h!/g})^{1/h}$ , we have

$$(1 - \delta)^t \binom{n}{t} \leq |\mathcal{Z}_n^{h,g}(t)| \leq \binom{n}{t}. \quad (8)$$

Notice that the bounds of Theorem 2.5 and Proposition 2.3(i) are the same, however, the ranges of  $t$  for which each applies differ drastically. Indeed, for  $g \gg h!$ , one can take  $t$  quite close to  $n^{1/h}$  in Theorem 2.5 and of course, this is essentially best possible as can be seen from (7). In effect, unlike in the case of  $B_h$ -sets, apart from a very narrow range of  $t$ ,  $B_h[g]$ -sets with  $t$ -elements are either ‘relatively abundant’ or simply do not exist.

The proof of Theorem 2.5 is given in Section 6.

**2.3. Probabilistic results.** Let  $[n]_m$  be an  $m$ -element subset of  $[n]$  chosen uniformly at random. We are interested in estimating the cardinality of the largest  $B_h$ -sets contained in  $[n]_m$ . Our bounds for the size of the families  $\mathcal{Z}_n^h(t)$  presented in Section 2.1 will be useful in investigating this problem. It will be convenient to have the following definition.

**Definition 2.6.** For an integer  $h \geq 2$  and a set  $R$ , let  $F_h(R)$  denote the maximum size of a  $B_h$ -set contained in  $R$ .

The asymptotic behavior of the random variable  $F_2([n]_m)$  was investigated in [18, 19]. Our goal here is to study  $F_h([n]_m)$  for arbitrary  $h \geq 3$ . A standard deletion argument implies that, with probability tending to 1 as  $n \rightarrow \infty$ , or *asymptotically almost surely* (**a.a.s.** for short), we have

$$F_h([n]_m) = (1 + o(1))m \quad \text{if } m = m(n) \ll n^{1/(2h-1)},$$

where  $o(1)$  denotes some function that tends to 0 as  $n \rightarrow \infty$ . On the other hand, if we apply the results of Schacht [27] and Conlon and Gowers [7] to  $B_h$ -sets, we have that **a.a.s.**

$$F_h([n]_m) = o(m) \quad \text{if } m = m(n) \gg n^{1/(2h-1)}.$$

Thus  $n^{1/(2h-1)}$  is the threshold for the property that  $F_h([n]_m) = o(m)$ .

The following abridged version of our results gives us quite precise information about  $F_h([n]_m)$  for a wide range of  $m$  and non-trivial but looser bounds for  $n^{1/(2h-1)} \leq m \leq n^{h/(h+1)}$ ; see also Figure 1.

**Theorem 2.7.** Fix  $h \geq 3$  and let  $0 \leq a \leq 1$  be a fixed constant. Suppose  $m = m(n) = (1 + o(1))n^a$ . Then **a.a.s.**

$$n^{b_1+a(1)} \leq F_h([n]_m) \leq n^{b_2+a(1)}, \quad (9)$$

where

$$b_1(a) = \begin{cases} a, & \text{for } 0 \leq a \leq 1/(2h-1); \\ 1/(2h-1), & \text{for } 1/(2h-1) < a \leq h/(2h-1); \\ a/h, & \text{for } h/(2h-1) < a \leq 1; \end{cases} \quad (10)$$

and

$$b_2(a) = \begin{cases} a, & \text{for } 0 \leq a \leq 1/(h+1); \\ 1/(h+1), & \text{for } 1/(h+1) < a \leq h/(h+1); \\ a/h, & \text{for } h/(h+1) < a \leq 1. \end{cases} \quad (11)$$

We prove the upper bounds in Theorem 2.7 (that is, (9) and (11)) in Section 3. The lower bounds (that is, (9) and (10)) are proved in Section 5. Theorem 2.7 determines  $b = b(a)$  for which  $F_h([n]_m) = n^{b+a(1)}$  when  $m = (1 + o(1))n^a$  whenever  $a \leq 1/(2h-1)$  or  $a \geq h/(h+1)$ . An interesting open question is the existence and determination of  $b = b(a)$  such that  $F_h([n]_m) = n^{b+a(1)}$  for  $1/(2h-1) \leq a \leq h/(h+1)$ , see Conjecture 7.2 in Section 7.

As in the previous section, we now move on to consider  $B_h[g]$ -sets.

**Definition 2.8.** For integers  $h \geq 2$  and  $g \geq 1$  and a set  $R$ , denote by  $F_{h,g}(R)$  the maximum size of a  $B_h[g]$ -set contained in  $R$ .

As a natural extension of Theorem 2.7, we investigate the random variable  $F_{h,g}([n]_m)$ . Trivially, one has

$$F_{h,g}([n]_m) \leq \min\{m, F_{h,g}(n)\}. \quad (12)$$

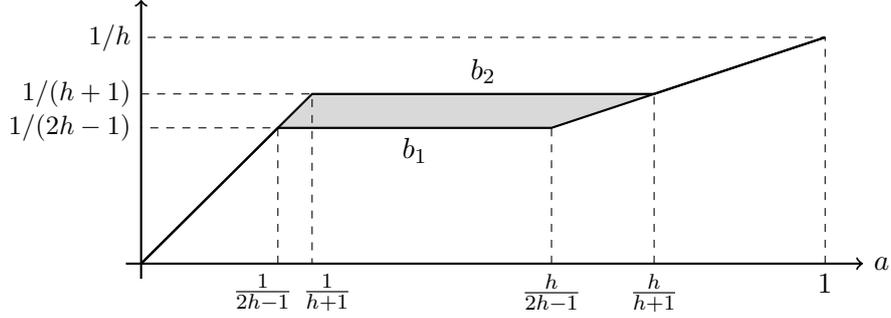


FIGURE 1. The graphs of  $b_1 = b_1(a)$  and  $b_2 = b_2(a)$  from the statement of Theorem 2.7

Surprisingly, as our next result shows, one can obtain a matching lower bound to this trivial upper bound, up to an  $n^{o(1)}$  factor, as long as one allows  $g$  to grow with  $n$ , however slowly.

**Theorem 2.9.** *Let  $h \geq 2$  be an integer and suppose  $g(n) \rightarrow \infty$  as  $n \rightarrow \infty$ . Let  $0 \leq a \leq 1$  be a fixed constant and suppose  $m = m(n) = (1 + o(1))n^a$ . Then **a.a.s.***

$$F_{h,g}([n]_m) = n^{b+o(1)}, \quad (13)$$

where

$$b(a) = \begin{cases} a, & \text{for } 0 \leq a \leq 1/h; \\ 1/h, & \text{for } 1/h \leq a \leq 1. \end{cases} \quad (14)$$

We remark that, by monotonicity, the assumption  $m = (1 + o(1))n^a$  can be replaced by  $m = n^{a+o(1)}$ . The upper bound on  $F_{h,g}([n]_m)$  contained in Theorem 2.9 follows from (12). The lower bound follows from the following more precise result, which is proved in Section 6.

**Theorem 2.10.** *Fix an integer  $h \geq 2$  and a function  $g = g(n)$ . For every fixed  $\varepsilon > 0$  and  $1 \leq m \leq (\varepsilon/3h) (n^{1-h!/g})^{1/h}$ , we **a.a.s.** have  $F_{h,g}([n]_m) \geq (1 - \varepsilon)m$ .*

We remark that Theorem 2.10 above is closely related to Theorem 2.5 in the previous section. Indeed, we shall derive the latter from the former at the end of Section 6.

### 3. PROOF OF THEOREM 2.1 AND PROOF OF THE UPPER BOUNDS IN THEOREM 2.7

We first derive Theorem 2.1 from Theorem 2.2.

*Proof of Theorem 2.1.* The total number of subsets of  $[n]$  having fewer than  $n^{1/(h+1)}(\log n)^2$  elements is  $2^{o(n^{1/h})}$ . Therefore, we may focus on  $B_h$ -sets of cardinality at least  $n^{1/(h+1)}(\log n)^2$ . In particular, by Theorem 2.2,

$$|\mathcal{Z}_n^h| \leq 2^{o(n^{1/h})} + \sum_{t \geq n^{1/(h+1)}(\log n)^2} \left( \frac{c_h n}{t^h} \right)^t. \quad (15)$$

Since the function  $t \mapsto (c_h n/t^h)^t$  is maximized when  $t = (c_h n)^{1/h}/e$ , it follows from (15) that, for an appropriate choice of the constant  $C_h$ ,

$$|\mathcal{Z}_n^h| \leq 2^{o(n^{1/h})} + n \cdot \left( \frac{c_h n}{c_h n/e^h} \right)^{(c_h n)^{1/h}/e} \leq 2^{o(n^{1/h})} + n \cdot \exp\left( \frac{h(c_h n)^{1/h}}{e} \right) \leq 2^{C_h n^{1/h}}. \quad \square$$

We now turn to the proof of the upper bound on  $F_h([n]_m)$  contained in Theorem 2.7. We start with the following easy remark.

**Remark 3.1.** At times, it will be convenient to work with the binomial random set  $[n]_p$ , which is a random subset of  $[n]$ , with each element of  $[n]$  included independently with probability  $p$ . The models  $[n]_m$  and  $[n]_p$ , with  $p = m/n$ , are fairly similar: If some property holds for  $[n]_p$  with probability  $1 - o(1/\sqrt{pn})$  then the same property holds **a.a.s.** for  $[n]_m$  (this follows from Pittel's inequality; see [14, p. 17]).

The following theorem is a direct corollary of Theorem 2.2.

**Theorem 3.2.** *There is a constant  $C = C(h)$  such that for every  $p \geq n^{-1/(h+1)}(\log n)^{2h}$ , **a.a.s.**,*

$$F_h([n]_p) \leq C(pn)^{1/h}.$$

Moreover, the probability that the inequality above fails is at most  $\exp(-c(pn)^{1/h})$  for some constant  $c = c(h) > 0$ .

*Proof.* The probability that  $F_h([n]_p) \geq t$  is the same as the probability that there exists a  $B_h$ -set of cardinality  $t$  in  $[n]_p$ . Hence, the union bound yields

$$\mathbf{P}[F_h([n]_p) \geq t] \leq p^t |\mathcal{Z}_n^h(t)|.$$

Let  $t = (2c_h np)^{1/h}$  where  $c_h = e^6(2h)^{2h}$  is the constant given in Theorem 2.2. By the assumption  $p \geq n^{-1/(h+1)}(\log n)^{2h}$ , we have that

$$\begin{aligned} t &\geq \left(2c_h n^{h/(h+1)}(\log n)^{2h}\right)^{1/h} = (2c_h)^{1/h} n^{1/(h+1)}(\log n)^2 \\ &\geq n^{1/(h+1)}(\log n)^2, \end{aligned}$$

which satisfies the assumption of Theorem 2.2. Theorem 2.2 gives that

$$\mathbf{P}[F_h([n]_p) \geq t] \leq p^t |\mathcal{Z}_n^h(t)| \leq \left(\frac{pc_h n}{t^h}\right)^t = \left(\frac{1}{2}\right)^t = \exp\left(-(\log 2)(2c_h)^{1/h}(np)^{1/h}\right),$$

which completes the proof of Theorem 3.2.  $\square$

We now prove the upper bound on  $F_h([n]_m)$  given in Theorem 2.7 (see (9) and (11)). Let us first recall that Remark 3.1 links the binomial random set  $[n]_p$ , appearing in Theorem 3.2, to the random set  $[n]_m$  that appears in Theorem 2.7. In what follows, we establish (9) and (11) in Theorem 2.7 using Theorem 3.2. We analyse three ranges of  $a$  separately.

- (i)  $0 \leq a \leq 1/(h+1)$ : From the trivial bound  $F_h([n]_m) \leq m$ , we see that we may take  $b_2(a) = a$ .
- (ii)  $1/(h+1) < a \leq h/(h+1)$ : It is clear that, in probability,  $F_h([n]_m)$  is non-decreasing in  $m$ . Hence,  $b_2(a)$  may be taken to be non-decreasing in  $a$  as well. Since, as we show next, we may take  $b_2(h/(h+1)) = 1/(h+1)$ , this monotonicity lets us take  $b_2(a) = 1/(h+1)$  in this range of  $a$ .
- (iii)  $h/(h+1) < a \leq 1$ : In this range,  $b_2(a) = a/h$  follows from Theorem 3.2. Indeed, if  $p \geq n^{-1/(h+1)}(\log n)^{2h}$ , then with probability at least  $1 - \exp(-c(pn)^{1/h}) \geq 1 - o(1/\sqrt{pn})$  we have  $F_h([n]_p) \leq C(pn)^{1/h}$  for some constant  $C > 0$ . Remark 3.1 implies that, **a.a.s.**,  $F_h([n]_m) \leq Cm^{1/h}$  for all  $m \geq n^{h/(h+1)}(\log n)^{2h}$ , giving that we may take  $b_2(a) = a/h$  for  $a > h/(h+1)$ , as claimed.

#### 4. UPPER BOUNDS FOR THE NUMBER OF $B_h$ -SETS OF A GIVEN CARDINALITY

We prove Theorem 2.2 in this section. For the case where  $h = 2$ , Theorem 2.2 was shown in [18] (see Theorem 2.1 of [18]). Hence, we assume that  $h \geq 3$  in this section. We follow a strategy that may be described very roughly as follows. Suppose a  $B_h$ -set  $S \subset [n]$  of cardinality  $s$  is given and one would like to extend it to a larger  $B_h$ -set of cardinality  $s'$ . We shall show that if  $s$  is not too small, then the number of such extensions is very small. To prove Theorem 2.2, we shall apply this fact iteratively, considering a sequence of cardinalities  $s < s' < s'' < \dots$ .

**4.1. Bounding the number of extensions of  $B_h$ -sets.** We use a graph-based approach to bounding the number of extensions of a large  $B_h$ -set to a larger  $B_h$ -set. This approach is inspired by the work of Kleitman and Winston [17] and Kleitman and Wilson [16]. We start with the following simple observation. If two distinct elements  $x, y \in [n] \setminus S$  satisfy

$$\begin{aligned} x + a_1 + \dots + a_{h-1} &= y + b_1 + \dots + b_{h-1} \\ \text{for some } \{a_1, \dots, a_{h-1}\}, \{b_1, \dots, b_{h-1}\} &\in \binom{S}{h-1}, \end{aligned} \tag{16}$$

then  $S \cup \{x, y\}$  is clearly not a  $B_h$ -set. This motivates our next definition.

**Definition 4.1.** The *collision graph*  $\text{CG}_S$  is a graph on the vertex set  $[n] \setminus S$  whose edges are all pairs of distinct elements  $x, y \in [n] \setminus S$  that satisfy (16).

Clearly, by the construction of  $\text{CG}_S$ , we have that if  $I \subseteq [n] \setminus S$  is such that  $I \cup S$  is a  $B_h$ -set, then  $I$  is an independent set in  $\text{CG}_S$ .

One of our main tools is the following lemma, implicit in the work of Kleitman and Winston [17], which provides an upper bound on the number of independent sets in graphs that have many edges in each sufficiently large vertex subset (see (18)). Lemma 4.2 in the version presented below is stated and proved in [18, 19], where it is used to bound the number of Sidon subsets of  $[n]$ . For other applications of this lemma to problems in additive combinatorics, we refer the reader to [1].

**Lemma 4.2.** *Let  $\delta$  and  $\beta > 0$  and  $q \in \mathbb{N}$  be numbers satisfying*

$$e^{\beta q} \delta > 1. \tag{17}$$

*Suppose that  $G = (V, E)$  is a graph satisfying*

$$e_G(A) \geq \beta |A|^2 \text{ for all } A \subset V \text{ with } |A| \geq \delta |V|. \tag{18}$$

*Then, for every  $m \geq 1$ , there are at most*

$$\binom{|V|}{q} \binom{\delta |V|}{m} \tag{19}$$

*independent sets in  $G$  of size  $q + m$ .*

**Remark 4.3.** When we apply Lemma 4.2 to  $\text{CG}_S$ , we shall take  $m \gg q$  to take advantage of the upper bound (19). In condition (18), there is a trade-off between  $\beta$  (larger is better) and  $\delta$  (smaller is better) which needs to be optimized.

We wish to show that  $\text{CG}_S$  satisfies (18) with good parameters  $\beta$  and  $\delta$ . To that end, we shall make use of another auxiliary graph, which we now define.

**Definition 4.4.** Let  $\widetilde{\text{CG}}_S$  be a multigraph version of  $\text{CG}_S$ , where the multiplicity of a pair  $\{x, y\}$  of distinct  $x, y \in [n] \setminus S$  is given by the number of pairs  $(\{a_1, \dots, a_{h-1}\}, \{b_1, \dots, b_{h-1}\}) \in \binom{S}{h-1}^2$  that satisfy (16).

**Lemma 4.5.** For every  $B_h$ -set  $S$  with  $s \geq h$  elements and every  $A \subset [n] \setminus S$  with  $|A| \geq h^{2h}n/s^{h-1}$ , we have

$$e_{\widetilde{\text{CG}}_S}(A) \geq \frac{s^{2h-2}}{h^{2h}n} |A|^2, \quad (20)$$

where the edges in  $\widetilde{\text{CG}}_S$  are counted with multiplicity.

The proof of Lemma 4.5 will be given in Section 4.3. In view of Lemma 4.5, if the maximal multiplicity of an edge in  $\widetilde{\text{CG}}_S$  is at most  $r$ , then the graph  $\text{CG}_S$  satisfies the conditions of Lemma 4.2 with  $V = [n]$ ,  $\beta = s^{2h-2}/(h^{2h}rn)$  and  $\delta = h^{2h}/s^{h-1}$ . Consequently, we are interested in bounding the multiplicity of the edges of  $\widetilde{\text{CG}}_S$ .

**Proposition 4.6.** For every  $B_h$ -set  $S$  of cardinality  $s$ , the maximal multiplicity of an edge in  $\widetilde{\text{CG}}_S$  does not exceed  $s^{h-2}$ .

We postpone the proof of Proposition 4.6 to Section 4.4. The following is an immediate corollary of Lemma 4.5 and Proposition 4.6.

**Corollary 4.7.** If  $S$  is a  $B_h$ -set with  $s$  elements, then for every  $A \subset [n] \setminus S$  with  $|A| \geq h^{2h}n/s^{h-1}$ ,

$$e_{\text{CG}_S}(A) \geq \frac{s^h}{h^{2h}n} |A|^2.$$

**4.2. Proof of Theorem 2.2.** The case  $h = 2$  of Theorem 2.2 is proved in [18] and we therefore restrict ourselves to  $h \geq 3$  here. We shall in fact prove the following: for every  $h \geq 3$  and

$$t \geq h^2 n^{1/(h+1)} (\log n)^{1+1/(h+1)}, \quad (21)$$

we have that

$$|\mathcal{Z}_n^h(t)| \leq \left( \frac{2^{2h} e^6 h^{2h} n}{t^h} \right)^t.$$

In view of (1), we have  $\mathcal{Z}_n^h(t) = 0$  for  $t > h^2 n^{1/h}$ . Hence we assume

$$t \leq h^2 n^{1/h}, \quad (22)$$

that is,  $h^2 n^{1/(h+1)} (\log n)^{1+1/(h+1)} \leq t \leq h^2 n^{1/h}$ . Let

$$s_0 = h^2 (n \log n)^{1/(h+1)} \quad (23)$$

and let  $K$  be the largest integer satisfying  $t 2^{-K} \geq 2s_0$ . We define three sequences  $(s_k)_{0 \leq k \leq K}$ ,  $(q_k)_{0 \leq k \leq K}$  and  $(m_k)_{0 \leq k \leq K}$  as follows. We let

$$q_0 = s_0/2 \quad \text{and} \quad m_0 = t 2^{-K} - s_0 - q_0. \quad (24)$$

For  $k = 1, \dots, K$ , we let

$$s_k = t 2^{-K+k-1}, \quad (25)$$

$$q_k = q_0 2^{-hk}, \quad (26)$$

$$m_k = t 2^{-K+k} - s_k - q_k. \quad (27)$$

We will bound the number of sequences  $S_0 \subset \dots \subset S_K \subset S_{K+1}$  of  $B_h$ -sets with  $|S_{K+1}| = t$  and  $|S_k| = s_k$  for all  $k = 0, \dots, K$ , from which a bound on  $|\mathcal{Z}_n^h(t)|$  will easily follow. Although we will only use the trivial bound  $\binom{n}{s_0}$  for the number of choices for  $S_0$ , we will then employ Lemma 4.2 to obtain a non-trivial bound on the number of extensions of  $S_k$  to  $S_{k+1}$  for all  $k$ .

Let us now estimate the number of extensions of a  $B_h$ -set  $S_k$  to a larger  $B_h$ -set  $S_{k+1}$  for some  $k = 0, \dots, K$ . By Corollary 4.7, the graph  $\text{CG}_{S_k}$  is such that for all  $A \subset [n] \setminus S_k$  with  $|A| \geq h^{2h}n/s_k^{h-1}$ ,

$$e_{\text{CG}_{S_k}}(A) \geq \beta_k |A|^2, \quad \text{where} \quad \beta_k = \frac{s_k^h}{h^{2h}n}.$$

Let

$$\delta_k = h^{2h}/s_k^{h-1} \geq 1/n \tag{28}$$

and observe that

$$e^{\beta_k q_k} = \exp\left(\frac{s_k^h}{h^{2h}n} \cdot \frac{q_0}{2^{hk}}\right) \stackrel{(25)}{\geq} \exp\left(\frac{(2^k s_0)^h \cdot s_0}{h^{2h}n \cdot 2^{hk+1}}\right) \geq \exp\left(\frac{s_0^{h+1}}{2h^{2h}n}\right) \stackrel{(23)}{\geq} n \stackrel{(28)}{\geq} \delta_k^{-1}.$$

Consequently,  $\text{CG}_{S_k}$ ,  $\delta_k$ ,  $\beta_k$  and  $q_k$  satisfy the conditions of Lemma 4.2. Note that  $S_{k+1} \setminus S_k$  must be an independent set in  $\text{CG}_{S_k}$  with cardinality  $s_{k+1} - s_k = q_k + m_k$ . Therefore, by Lemma 4.2, the number of extensions of  $S_k$  into a  $B_h$ -set  $S_{k+1}$  is at most

$$\binom{n}{q_k} \binom{\delta_k n}{m_k}. \tag{29}$$

In order to obtain an upper bound of (29), we first claim that

$$\binom{\delta_0 n}{m_0} \leq \binom{\delta_0 n}{3s_0} \tag{30}$$

and

$$\binom{\delta_k n}{m_k} \leq \binom{\delta_k n}{s_k} \tag{31}$$

for all  $1 \leq k \leq K$ . Indeed, inequality (30) follows from the fact that  $m_0 = s_1 - s_0 - q_0 \leq 4s_0 - s_0 \leq 3s_0$  and also  $3s_0 \leq \frac{\delta_0 n}{2}$ . Inequality (31) follows from the fact that for all  $1 \leq k \leq K$ ,  $m_k \leq s_k \leq \frac{\delta_k n}{2}$  as

$$\frac{s_k}{\delta_k} \stackrel{(28)}{=} \frac{s_k^h}{h^{2h}} \leq \frac{s_K^h}{h^{2h}} \stackrel{(25)}{=} \frac{(t/2)^h}{h^{2h}} \stackrel{(22)}{\leq} \frac{n}{2^h}.$$

Hence,

$$\binom{n}{q_0} \binom{\delta_0 n}{m_0} \leq \binom{n}{q_0} \binom{\delta_0 n}{3s_0} \leq \binom{n}{q_0} \binom{n}{3s_0} \leq n^{q_0} n^{3s_0},$$

and for all  $1 \leq k \leq K$

$$\binom{n}{q_k} \binom{\delta_k n}{m_k} \leq \binom{n}{q_k} \binom{\delta_k n}{s_k} \leq n^{q_k} \left(\frac{e\delta_k n}{s_k}\right)^{s_k} \leq n^{q_k} \left(\frac{eh^{2h}n}{s_k^h}\right)^{s_k}.$$

Applying (29) iteratively implies that

$$|\mathcal{Z}_n^h(t)| \leq \binom{n}{s_0} \prod_{k=0}^K \binom{n}{q_k} \binom{\delta_k n}{m_k} \leq n^{4s_0 + \sum_{k=0}^K q_k} \prod_{k=1}^K \left(\frac{eh^{2h}n}{s_k^h}\right)^{s_k}. \tag{32}$$

Finally, since

$$\sum_{k=0}^K q_k \stackrel{(26)}{=} q_0 \sum_{k=0}^K 2^{-hk} \leq 2q_0 \stackrel{(24)}{=} s_0 \stackrel{(21),(23)}{\leq} \frac{t}{\log n}$$

and

$$\prod_{k=1}^K \left( \frac{eh^{2h}n}{s_k^h} \right)^{s_k} \stackrel{(25)}{\leq} \prod_{k=1}^K \left( \frac{eh^{2h}n}{(t2^{-k})^h} \right)^{t2^{-k}} \leq \left[ 2^{h \sum_{k=1}^{\infty} k2^{-k}} \cdot \left( \frac{eh^{2h}n}{t^h} \right)^{\sum_{k=1}^{\infty} 2^{-k}} \right]^t = \left( \frac{2^{2h}eh^{2h}n}{t^h} \right)^t,$$

Theorem 2.2 follows from (32).  $\square$

**4.3. Proof of Lemma 4.5.** Let  $S$  be a  $B_h$ -set with  $s$  elements. Let  $A \subset [n] \setminus S$  be an arbitrary subset with  $|A| \geq h^{2h}n/s^{h-1}$ . Consider the auxiliary bipartite graph  $\Gamma$  defined as follows. The vertex classes of  $\Gamma$  are  $A$  and a disjoint copy of  $[hn]$ . The edge set of  $\Gamma$  is defined as

$$E(\Gamma) = \left\{ (x, u) \in A \times [hn] : u = x + a_1 + \cdots + a_{h-1} \text{ for some } \{a_1, \dots, a_{h-1}\} \in \binom{S}{h-1} \right\}.$$

Note that, because  $S$  is a  $B_h$ -set, for fixed  $x$  and  $u$ , there is at most one solution to  $u = x + a_1 + \cdots + a_{h-1}$  with  $\{a_1, \dots, a_{h-1}\} \in \binom{S}{h-1}$ . We will now argue that the multiplicity of a pair  $\{x, y\} \in \binom{A}{2}$  in the multigraph  $\widetilde{\text{CG}}_S$  is the number of paths of length two connecting  $x$  to  $y$  in  $\Gamma$ . Indeed, there is a bijection between pairs  $(\{a_1, \dots, a_{h-1}\}, \{b_1, \dots, b_{h-1}\}) \in \binom{S}{h-1}^2$  that satisfy (16) and paths  $xuy$  in  $\Gamma$ , where

$$u = x + a_1 + \cdots + a_{h-1} = y + b_1 + \cdots + b_{h-1}.$$

Consequently,  $e_{\widetilde{\text{CG}}_S}(A)$  is the number of paths of length two in  $\Gamma$  containing two vertices in the class  $A$ . By Jensen's inequality applied to the convex function  $f(\alpha) = \binom{\alpha}{2} = \alpha(\alpha-1)/2$ ,

$$e_{\widetilde{\text{CG}}_S}(A) \geq \sum_{u \in [hn]} \binom{\deg_{\Gamma}(u)}{2} \geq hn \binom{e(\Gamma)/hn}{2}.$$

On the other hand,

$$e(\Gamma) = \sum_{x \in A} \deg_{\Gamma}(x) = |A| \binom{s}{h-1} \geq \left( \frac{s}{h} \right)^{h-1} |A|.$$

It follows that  $e(\Gamma) \geq h^h n$  and thus,

$$e_{\widetilde{\text{CG}}_S}(A) \geq hn \binom{e(\Gamma)/hn}{2} \geq e(\Gamma) \left( \frac{e(\Gamma) - hn}{2hn} \right) \geq \frac{e(\Gamma)^2}{hn} \left( \frac{h^h - h}{2h^h} \right) \geq \frac{e(\Gamma)^2}{3hn} \geq \frac{s^{2h-2}}{h^{2h}n} |A|^2.$$

This concludes the proof of Lemma 4.5.  $\square$

**4.4. Proof of Proposition 4.6.** Let  $S$  be a  $B_h$ -set of cardinality  $s$  and let  $x \neq y \in [n]$  be arbitrary. By definition, the multiplicity of  $\{x, y\}$  in  $\widetilde{\text{CG}}_S$  is the number of pairs of sets  $\{a_1, \dots, a_{h-1}\}, \{b_1, \dots, b_{h-1}\} \in \binom{S}{h-1}$  such that

$$y - x = a_1 + \cdots + a_{h-1} - (b_1 + \cdots + b_{h-1}).$$

Since  $x \neq y$ , we clearly have  $\{a_1, a_2, \dots, a_{h-1}\} \neq \{b_1, b_2, \dots, b_{h-1}\}$ . Hence, we may assume, without loss of generality, that  $a_{h-1} \notin \{b_1, \dots, b_{h-1}\}$ . Let us now bound the number of possible sets in the following way: first, pick arbitrary values of  $a_1, \dots, a_{h-2} \in S$ , then find values (if any exist)  $a_{h-1}, b_1, \dots, b_{h-1}$ , with  $a_{h-1} \notin \{b_1, \dots, b_{h-1}\}$ , that satisfy

$$y - x - a_1 - a_2 - \cdots - a_{h-2} = a_{h-1} - (b_1 + \cdots + b_{h-1}).$$

We claim that for each fixed sequence  $a_1, \dots, a_{h-2} \in S$  there is at most one such completion (up to the order of elements  $b_i$ ) that satisfies the above equality. Indeed, suppose that we also

have  $a'_{h-1}, b'_1, \dots, b'_{h-1} \in S$  such that

$$a'_{h-1} - (b'_1 + \dots + b'_{h-1}) = a_{h-1} - (b_1 + \dots + b_{h-1}).$$

Then, since  $S$  is a  $B_h$ -set, the following is a (multi)set equality:

$$\{a_{h-1}, b'_1, \dots, b'_{h-1}\} = \{a'_{h-1}, b_1, \dots, b_{h-1}\}.$$

Moreover, since  $a_{h-1} \notin \{b_1, \dots, b_{h-1}\}$ , we have  $a'_{h-1} = a_{h-1}$ , which implies that we also have  $\{b'_1, \dots, b'_{h-1}\} = \{b_1, \dots, b_{h-1}\}$ .

In conclusion, fixing any of the  $s^{h-2}$  choices for  $a_1, \dots, a_{h-2} \in S$  completely determines both  $a_{h-1}$  and  $\{b_1, \dots, b_{h-1}\}$ . The proposition follows.

## 5. LOWER BOUNDS

In this section, we establish the lower bounds in Theorem 2.7 and prove Proposition 2.3. For conciseness, we shall be somewhat sketchy when dealing with routine arguments.

First, we show that a simple deletion argument (given in Lemma 5.1 below) yields that if  $m \ll n^{1/(2h-1)}$ , then  $F_h([n]_m) = (1 - o(1))m$ . This immediately implies that in Theorem 2.7, for  $0 \leq a < 1/(2h-1)$ , one may take  $b_1(a) = a$  (see (9) and (10)). Since  $F_3([n]_m)$  is non-decreasing in probability with respect to  $m$ , for  $a \geq 1/(2h-1)$ , we may take  $b_1(a) = 1/(2h-1)$ . Moreover, as an easy corollary of Lemma 5.1, we will also derive Proposition 2.3(i).

In the second part of this section, following the strategy of [18, 19], for every  $t = o(n^{1/h})$ , we will describe a deterministic construction of a large subfamily of  $\mathcal{Z}_n^h(t)$ . The existence of such a subfamily will immediately imply Proposition 2.3(ii). Moreover, we shall show that if  $1 \ll m \leq n$ , then **a.a.s.** the set  $[n]_m$  contains a  $B_h$ -set with  $\Omega(m^{1/h})$  elements from the constructed subfamily. This yields that in Theorem 2.7, we may take  $b_1(a) = a/h$  for all  $0 \leq a \leq 1$ . Note that, in the range  $1/(2h-1) \leq a \leq h/(2h-1)$ , this is superseded by the bound obtained in the first part, that is,  $b_1(a) = 1/(2h-1)$ .

**Lemma 5.1.** *If  $1 \leq m = o(n^{1/(2h-1)})$ , then we **a.a.s.** have  $m \geq F_h([n]_m) \geq (1 - o(1))m$ .*

*Proof.* Let  $1 \leq m \ll n^{1/(2h-1)}$  and let  $X$  be the random variable that counts the number of solutions to

$$a_1 + \dots + a_h = b_1 + \dots + b_h \quad \text{with} \quad \{a_1, \dots, a_h\} \neq \{b_1, \dots, b_h\} \quad (33)$$

and  $a_i, b_i \in [n]_m$  for all  $i \in [h]$ . Let  $p = m/n$ . It follows from the linearity of expectation that

$$\mathbf{E}[X] = O\left(\sum_{k=2}^{2h-1} p^{k+1} n^k\right) = O(p^{2h} n^{2h-1}) = o(m).$$

Hence, by Markov's inequality, we **a.a.s.** have  $X = o(m)$ . Since deleting from  $[n]_m$  one element from the set  $\{a_1, b_1, \dots, a_h, b_h\}$  for each of the  $X$  solutions to (33) yields a  $B_h$ -set, the lemma follows.  $\square$

*Proof of Proposition 2.3(i).* Fix a constant  $\delta > 0$ . Choose  $\beta > 0$  small enough so that  $(1 - \beta)(1 - \delta/2) \geq 1 - \delta$  and  $\binom{1+\beta}{\beta t} \leq (1 + \delta/2)^t$  for all  $t$ . Let  $\varepsilon > 0$  be a small constant. Assume that  $t \leq \varepsilon n^{1/(2h-1)}$ . Lemma 5.1 with  $m = (1 + \beta)t$  implies that if  $\varepsilon$  is sufficiently small, then

$F_h([n]_m) \geq t$  with probability at least  $1 - \beta$ . It follows that, for large enough  $n$ , we have

$$\begin{aligned} |\mathcal{Z}_n^h(t)| &\geq (1 - \beta) \binom{n}{(1 + \beta)t} \binom{n - t}{\beta t}^{-1} = (1 - \beta) \binom{n}{t} \binom{(1 + \beta)t}{\beta t}^{-1} \\ &\geq (1 - \beta)(1 - \delta/2)^t \binom{n}{t} \geq (1 - \delta)^t \binom{n}{t}, \end{aligned} \quad (34)$$

as required.  $\square$

In order to construct a large family of  $B_h$ -sets for larger  $t$ , we will use the following theorem of Bose and Chowla [5] (with the statement adapted for our purposes).

**Theorem 5.2.** *Fix an integer  $h \geq 2$ . For every  $m$ , there exists a  $B_h$ -set  $Y \subset \mathbb{Z}_m$  with  $|Y| = \Omega(m^{1/h})$ .*

From Theorem 5.2 we obtain the following corollary.

**Corollary 5.3.** *Let  $n$  and  $m$  satisfy  $n \geq 3hm$  and suppose that  $Y \subset \mathbb{Z}_m$  is a  $B_h$ -set. Then, there exists at least*

$$\left( \frac{n}{3hm} \right)^{|Y|}$$

$B_h$ -sets of cardinality  $|Y|$  in  $[n]$ .

More precisely, there are pairwise disjoint sets  $I_0, \dots, I_{m-1} \subset [n]$ , each of cardinality  $\ell \geq n/(3hm)$ , with the following property: For any  $B_h$ -set  $Y \subset \mathbb{Z}_m$ , all sets  $S \subset [n]$  with  $|S| = Y$  such that  $|S \cap I_j| = 1$  for all  $j \in Y$  are  $B_h$ -sets.

*Proof.* Let  $k = \lfloor n/m \rfloor$  and  $\ell = \lfloor n/(2hm) \rfloor \geq n/(3hm) \geq 1$ . Define the integer intervals

$$I_j = [jk + 1, jk + \ell], \quad j = 0, 1, \dots, m - 1$$

and note that by construction they are all pairwise disjoint subsets of  $[n]$ . Let  $S$  be an arbitrary set which contains a single element of each  $I_j$  with  $j \in Y$  and no additional elements.

We claim that  $S$  is a  $B_h$ -set. Indeed, suppose that

$$a_1 + \dots + a_h = b_1 + \dots + b_h$$

for some  $a_1, b_1, \dots, a_h, b_h \in S$  with  $a_1 \leq a_2 \leq \dots \leq a_h$ ,  $b_1 \leq b_2 \leq \dots \leq b_h$ . For  $i = 1, \dots, h$ , let  $u_i \in Y$  be the unique index such that  $a_i \in I_{u_i}$ . Set  $u = \sum_{i=1}^h u_i$  and notice that

$$a_1 + \dots + a_h \in [ku + h, ku + \ell h] \subset [ku + 1, k(u + 1)].$$

Now for  $i = 1, \dots, h$ , let  $v_i \in Y$  be the unique index such that  $b_i \in I_{v_i}$ , and set  $v = \sum_{i=1}^h v_i$ . Notice that the same as argument as above yields

$$b_1 + \dots + b_h \in [kv + 1, k(v + 1)],$$

which thus means that  $u_1 + \dots + u_h = u = v = v_1 + \dots + v_h$ . Since  $Y$  is a  $B_h$ -set, we have  $\{u_1, \dots, u_h\} = \{v_1, \dots, v_h\}$ . Since the elements  $a_i$  and  $b_i$  are in increasing order, the same holds for the  $u_i$  and  $v_i$  and thus  $u_i = v_i$  for all  $i$ . Moreover, by construction,  $|S \cap I_j| = 1$  for all  $j \in Y$ , which means that  $\{a_i\} = S \cap I_{u_i} = S \cap I_{v_i} = \{b_i\}$  for all  $i = 1, \dots, h$ .

The above argument shows that  $S$  is a  $B_h$ -set. Since there are  $\ell^{|Y|}$  choices for the construction of  $S$ , the corollary follows.  $\square$

The proof of Proposition 2.3(ii) easily follows from Corollary 5.3.

*Proof of Proposition 2.3(ii).* From Theorem 5.2 we obtain a constant  $C$  satisfying that for all values of  $t$ , there exists  $m = m(t) \leq Ct^h$  such that there exists a  $B_h$ -set of cardinality  $t$  in  $\mathbb{Z}_m$ . We may also assume that  $m(\cdot)$  is monotone.

Let  $\varepsilon' > 0$  be such  $m(\varepsilon'n^{1/h}) \leq n/(3h)$ . It follows that for any  $t \leq \varepsilon'n^{1/h}$ , there is a  $B_h$ -set  $Y \subset \mathbb{Z}_{m(t)}$  with  $|Y| = t$ . Since  $m(t) \leq n/(3h)$ , applying Corollary 5.3 to  $Y$  shows that there are at least

$$\left(\frac{n}{3hm(t)}\right)^t \geq \left(\frac{n}{3hCt^h}\right)^t$$

$B_h$ -sets of cardinality  $t$  in  $[n]$ , which establishes the proposition with  $c'_h = 1/(3hC)$ .  $\square$

Next, we show that Corollary 5.3 also yields the lower bound in Theorem 2.7.

**Lemma 5.4.** *For any  $1 \ll m \leq n$ , we **a.a.s.** have  $F_h([n]_m) = \Omega(m^{1/h})$ .*

*Proof.* Lemma 5.1 implies that  $F_h([n]_m) = \Omega(m^{1/h})$  for  $m \ll n^{1/(2h-1)}$ . We now assume that  $m \gg n^{1/(2h)}$ , which covers the remaining range of  $m$  (with plenty to spare). It will be convenient for us to use the model  $[n]_p$  with  $p = m/n$  rather than  $[n]_m$  (recall Remark 3.1).

Without loss of generality we may assume that  $n \geq 3hm$ , since we just need to adjust the constant hidden in the  $\Omega$  in the bound from the statement of the lemma. Let  $I_j$ ,  $j = 0, \dots, m-1$ , be sets obtained from Corollary 5.3. From Theorem 5.2, we may obtain a  $B_h$ -set  $Y \subset \mathbb{Z}_m$  with  $|Y| = \Omega(m^{1/h})$ .

Consider, for each  $j \in Y$ , the intersection of the random set  $[n]_p$  and  $I_j$ . The probability  $q$  that this intersection is empty satisfies

$$q = (1-p)^{|I_j|} \leq \exp(-p|I_j|) = \exp\left(-p\frac{n}{3hm}\right) = \exp\left(-\frac{1}{3h}\right).$$

Notice that  $q$  is bounded away from 1 by a constant depending only on  $h$ . Let  $r$  be the random variable denoting the number of sets  $I_j$ ,  $j \in Y$ , that intersect  $[n]_p$ , namely

$$r = \left| \left\{ j \in Y : I_j \cap [n]_p \neq \emptyset \right\} \right|.$$

Since the sets  $I_j$  are disjoint,  $r$  is a binomial random variable with parameters  $|Y|$  and  $1-q$ . Also note that by collecting an element from each set  $I_j$ ,  $j \in Y$ , that intersects  $[n]_p$ , we have a  $B_h$ -set which is a subset of the random set, thus

$$F_h([n]_p) \geq r.$$

By Chernoff's bound, we have that  $r \geq (1-q)|Y|/2$  with probability

$$1 - \exp(-c|Y|) \geq 1 - \exp(-c'm^{1/h}) = 1 - o(\sqrt{pn}),$$

for some constants  $c, c' > 0$ . In particular, with probability  $1 - o(\sqrt{pn})$  there is a  $B_h$ -set  $S \subset [n]_p$  with cardinality  $r = \Omega(|Y|) = \Omega(m^{1/h})$ . The lemma follows from Remark 3.1.  $\square$

## 6. PROOFS OF THEOREMS 2.5 AND 2.10

We need some preparations for the proofs of Theorems 2.5 and 2.10. For the remainder of this section, we fix an integer  $h \geq 2$  and a function  $g = g(n)$ . Since we are only proving asymptotic results, we shall make the technical assumption that  $n$  is relatively prime to  $h!$ . Furthermore, it will be more convenient for us to work with modular arithmetic, that is, we consider addition modulo  $n$ . Clearly, any modular  $B_h[g]$ -subset of  $\mathbb{Z}_n$  naturally corresponds to a  $B_h[g]$ -subset

of  $[n]$  and hence the claimed lower bound results for  $[n]$  follows from the corresponding results for  $\mathbb{Z}_n$ .

Recall the definition of  $r_{S,h}$  (see (6) in Section 2.3). Observe that  $S$  is a  $B_h[g]$ -set if and only if  $r_{S,h}(z) \leq g$  for every  $z \in \mathbb{Z}_n$ . In order to show that  $r_{S,h}(z) \leq g$  for every  $z \in \mathbb{Z}_n$ , we define the following:

For every  $1 \leq \ell \leq h$  and  $\lambda > 0$  and  $S \subset \mathbb{Z}_n$ , let

$$E_{S,\ell}(\lambda) = \sum_{z \in \mathbb{Z}_n} \exp(\lambda r_{S,\ell}(z)). \quad (35)$$

Clearly, for every  $z \in \mathbb{Z}_n$ , we have that  $r_{S,h}(z) \leq \lambda^{-1} \log(E_{S,h}(\lambda))$ . Hence, in order to bound  $r_{S,h}(z)$  for every  $z \in \mathbb{Z}_n$ , it suffices to bound  $E_{S,h}(\lambda)$  for some appropriate choice of  $\lambda$ . We remark that the definition (35) is heavily inspired by *moment generating functions* studied in Probability Theory. Indeed, if  $z$  is sampled uniformly over  $\mathbb{Z}_n$ , then  $r_{S,\ell}$  becomes a random variable whose moment generating function is  $\mathbf{E}[e^{\lambda r_{S,\ell}}] = \frac{1}{n} E_{S,\ell}(\lambda)$ .

The following claim bounds the average increase of  $E_{S,\ell}(\lambda)$  as we add some  $y \in \mathbb{Z}_n$  to  $S$ .

**Claim 6.1.** *Let  $h \geq 2$  be fixed,  $n$  be a number relatively prime to  $h!$  and  $\ell \in [h]$ . Then, for any  $\emptyset \neq S \subset \mathbb{Z}_n$  and  $\lambda > 0$ , we have*

$$\mathbf{E}_{y \in \mathbb{Z}_n} [E_{S \cup \{y\}, \ell}(\lambda) - E_{S, \ell}(\lambda)] \leq \frac{1}{n} E_{S, \ell}(\lambda) (E_{S, \ell-1}(\ell\lambda) - n). \quad (36)$$

*Proof.* Note first that

$$r_{S \cup \{y\}, \ell}(z) \leq r_{S, \ell}(z) + \mathbf{1}[z = \ell y] + \sum_{i=1}^{\ell-1} r_{S, \ell-i}(z - iy). \quad (37)$$

Hence,

$$\begin{aligned} \sum_{y \in \mathbb{Z}_n} E_{S \cup \{y\}, \ell}(\lambda) &= \sum_{y \in \mathbb{Z}_n} \sum_{z \in \mathbb{Z}_n} \exp(\lambda r_{S \cup \{y\}, \ell}(z)) \\ &\leq \sum_{y \in \mathbb{Z}_n} \sum_{z \in \mathbb{Z}_n} \exp \left\{ \lambda \left( r_{S, \ell}(z) + \mathbf{1}[z = \ell y] + \sum_{i=1}^{\ell-1} r_{S, \ell-i}(z - iy) \right) \right\} \\ &= \sum_{z \in \mathbb{Z}_n} \underbrace{\left\{ \exp(\lambda r_{S, \ell}(z)) \sum_{y \in \mathbb{Z}_n} \exp(\lambda \mathbf{1}[z = \ell y]) \prod_{i=1}^{\ell-1} \exp(\lambda r_{S, \ell-i}(z - iy)) \right\}}_{Q(z, \lambda)}. \end{aligned} \quad (38)$$

We will now use a variation<sup>1</sup> of Hölder's inequality (see, e.g. [13, p. 22]):

$$\sum_{i=1}^m \left| \prod_{j=1}^{\ell} a_{ij} \right| \leq \prod_{j=1}^{\ell} \left( \sum_{i=1}^m |a_{ij}|^{\ell} \right)^{1/\ell}. \quad (39)$$

<sup>1</sup>This form can be obtained from the original as follows: for  $\ell = 2$ , it is a special case of Hölder's inequality with reciprocals  $1/2 + 1/2 = 1$ . For  $\ell \geq 3$  it follows by induction using Hölder's inequality with reciprocals  $\ell^{-1} + (\ell/(\ell-1))^{-1} = 1$  applied to  $\sum_{i=1}^m |x_i y_i|$ , where  $x_i = a_{i,1}$  and  $y_i = \prod_{j=2}^{\ell} a_{ij}$ .

It follows from the above Hölder's inequality that for every  $z \in \mathbb{Z}_n$ , the inner sum  $Q(z, \lambda)$  on the right-hand side of (38) is bounded from above by

$$Q(z, \lambda) \leq \left( \sum_{y \in \mathbb{Z}_n} \exp(\lambda \mathbf{1}[z = \ell y])^\ell \right)^{1/\ell} \prod_{i=1}^{\ell-1} \left( \sum_{y \in \mathbb{Z}_n} \exp(\lambda r_{S, \ell-i}(z - iy))^\ell \right)^{1/\ell}.$$

Recalling that we require that  $h!$  and  $n$  are co-prime and thus that each  $i \in [\ell]$  is co-prime with  $n$ , it follows that for fixed  $z$  and  $i$ , the map  $y \mapsto z - iy$  is a permutation of  $\mathbb{Z}_n$ . In particular, in the rightmost sum above, we can substitute  $w$  for  $z - iy$ , simplifying the expression to  $\sum_{w \in \mathbb{Z}_n} \exp(\lambda r_{S, \ell-i}(w))^\ell = E_{S, \ell-i}(\ell\lambda)$ . For the same reason, there exists only a single value  $y \in \mathbb{Z}_m$  such that  $z = \ell y$  and thus all but one term of the sum  $\sum_{y \in \mathbb{Z}_n} \exp(\lambda \mathbf{1}[z = \ell y])^\ell$  are equal to 1, the only other term being equal to  $e^{\ell\lambda}$ . Consequently, we have

$$Q(z, \lambda) \leq \left( (n + e^{\ell\lambda} - 1) \prod_{i=1}^{\ell-1} E_{S, \ell-i}(\ell\lambda) \right)^{1/\ell}.$$

Note that the bound we have obtained on  $Q(z, \lambda)$  is independent of  $z$ , hence (38) implies that

$$\sum_{y \in \mathbb{Z}_n} E_{S \cup \{y\}, \ell}(\lambda) \leq E_{S, \ell}(\lambda) \left( (n + e^{\ell\lambda} - 1) \prod_{i=1}^{\ell-1} E_{S, \ell-i}(\ell\lambda) \right)^{1/\ell}. \quad (40)$$

Observe that since  $S \neq \emptyset$ , then

$$E_{S, \ell-1}(\ell\lambda) = \max \left\{ n + e^{\ell\lambda} - 1, E_{S, 1}(\ell\lambda), \dots, E_{S, \ell-1}(\ell\lambda) \right\}. \quad (41)$$

Indeed, fix some arbitrary  $x \in S$  and notice that  $r_{S, \ell-1}(z) \geq r_{S, \ell-1-i}(z - ix)$  for all  $1 \leq i \leq \ell - 2$ , and since  $z \mapsto z - ix$  is a permutation of  $\mathbb{Z}_n$ , we have  $E_{S, \ell-1}(\ell\lambda) \geq E_{S, \ell-1-i}(\ell\lambda)$ . It is also clear that  $E_{S, \ell-1}(\ell\lambda) \geq n + e^{\ell\lambda} - 1$  since, for all  $x \in S$ , we have  $r_{S, \ell-1}((\ell - 1) \cdot x) \geq 1$ .

From (40) and (41) we conclude that for every non-empty  $S$  and all  $\lambda > 0$ ,

$$\sum_{y \in \mathbb{Z}_n} E_{S \cup \{y\}, \ell}(\lambda) \leq E_{S, \ell}(\lambda) E_{S, \ell-1}(\ell\lambda). \quad (42)$$

This gives inequality (36), and hence, the claim is proved.  $\square$

We now set

$$\lambda_\ell = \frac{h! \log(2n)}{\ell! g} \quad (43)$$

for each  $\ell \in [h]$ .

**Definition 6.2.** We shall call  $y \in \mathbb{Z}_n \setminus S$  an  $\varepsilon$ -good extension of a set  $S$  if, for all  $2 \leq \ell \leq h$ ,

$$E_{S \cup \{y\}, \ell}(\lambda_\ell) \leq E_{S, \ell}(\lambda_\ell) \left( 1 + \frac{2h}{\varepsilon} \frac{E_{S, \ell-1}(\lambda_{\ell-1}) - n}{n} \right). \quad (44)$$

**Claim 6.3.** Let  $h \geq 2$ ,  $\varepsilon > 0$ , and  $n$  be a number relatively prime to  $h!$ . Moreover, suppose that  $S \subset \mathbb{Z}_n$  satisfies  $1 \leq |S| \leq \varepsilon n/6$ . Then there are at least  $(1 - 2\varepsilon/3)n$  elements  $y \in \mathbb{Z}_n$  that are  $\varepsilon$ -good extensions of  $S$ .

*Proof.* Inequality (36) in Claim 6.1 implies that the number of elements  $y \in \mathbb{Z}_n$  which satisfy

$$E_{S \cup \{y\}, \ell}(\lambda_\ell) - E_{S, \ell}(\lambda_\ell) \geq \frac{2h}{\varepsilon} \cdot \frac{1}{n} E_{S, \ell}(\lambda_\ell) (E_{S, \ell-1}(\lambda_{\ell-1}) - n)$$

is at most  $\varepsilon n/(2h)$ . Together with the fact that  $\ell\lambda_\ell = \lambda_{\ell-1}$ , we have that the number of  $y \in \mathbb{Z}_n$  that violate (44) for a fixed  $\ell$  is at most  $\varepsilon n/(2h)$ . Since  $\ell$  can be any integer between 2 and  $h$ , there are at most  $\varepsilon n/2$  violators. Recalling that  $|S| \leq \varepsilon n/6$ , we obtain that the number of  $y \in \mathbb{Z}_n \setminus S$  that fail to be  $\varepsilon$ -good is at most  $\varepsilon n/2 + \varepsilon n/6 = (2\varepsilon/3)n$ .  $\square$

We are now in position to prove Theorem 2.10.

*Proof of Theorem 2.10.* Fix  $\varepsilon > 0$  and assume that  $1 \leq m \leq (\varepsilon/3h)(n^{1-h!/g})^{1/h}$ . We may and shall assume that  $m \geq \log n$ , since otherwise the random set  $[n]_m$  is **a.a.s.** a  $B_h$ -set and we are done.

Let  $R$  be a random  $m$ -element subset of  $\mathbb{Z}_n$ . We construct a subset  $S \subset R$  as follows. Let  $(x_1, \dots, x_m)$  be a random ordering of the elements of  $R$ . Let  $S_1 = \{x_1\}$  and for  $1 < j \leq m$ , let

$$S_j = \begin{cases} S_{j-1} \cup \{x_j\}, & \text{if } x_j \text{ is an } \varepsilon\text{-good extension of } S_{j-1}; \\ S_{j-1}, & \text{otherwise.} \end{cases}$$

We shall show that  $S = S_m$  is a  $B_h[g]$ -set and that **a.a.s.** it has at least  $(1 - \varepsilon)m$  elements. This will clearly suffice as  $S_m \leq F_{h,g}(R)$  for all  $m!$  orderings of  $R$ .

**Claim 6.4.** *The set  $S = S_m$  is a  $B_h[g]$ -set.*

*Proof.* We shall first prove by induction that for every  $1 \leq \ell \leq h$  and every  $1 \leq j \leq m$ , the following inequality holds:

$$\varphi(\ell, j) : \quad E_{S_j, \ell}(\lambda_\ell) \leq n + (2h/\varepsilon)^{\ell-1} e^{\lambda_1} |S_j|^\ell.$$

Before proving this, let us show how  $\varphi(h, m)$  implies this claim. By (6), for every  $z \in \mathbb{Z}_n$ , we have that

$$\exp(\lambda_h r_{S,h}(z)) \leq E_{S,h}(\lambda_h) \leq n + (2h/\varepsilon)^{h-1} e^{\lambda_1} m^h < 2n$$

since, recalling the definition of  $\lambda_1$  in (43) and the assumption on  $m$ , we have

$$(2h/\varepsilon)^{h-1} e^{\lambda_1} m^h \leq (2h/\varepsilon)^{h-1} (2n)^{h!/g} (\varepsilon/3h)^h n^{1-h!/g} < n. \quad (45)$$

Consequently, from the definition of  $\lambda_h$  in (43), we have  $r_{S,h}(z) \leq \lambda_h^{-1} \log(2n) = g$ . In other words,  $S$  is a  $B_h[g]$ -set.

We now resume the proof of the statements  $\varphi(\ell, j)$  by induction. Observe that regardless of  $x_1$ , for every  $\ell \in [h]$ ,

$$E_{S_1, \ell}(\lambda_\ell) = E_{\{x_1\}, \ell}(\lambda_\ell) = (n-1) + e^{\lambda_\ell} \leq n + e^{\lambda_1}$$

and hence  $\varphi(\ell, 1)$  holds for all  $\ell$ .

Next, we consider  $\varphi(1, j)$  for all  $j$ . Note that  $r_{S,1}(z) = \mathbf{1}[z \in S]$  and therefore, from the definition of  $E_{S,1}(\lambda)$ , we have that

$$E_{S,1}(\lambda) = n - |S| + |S|e^\lambda = n + (e^\lambda - 1)|S|. \quad (46)$$

Hence,  $\varphi(1, j)$  holds for all  $j$ .

Thus, it is enough to prove that if  $\ell \geq 2$ , then, assuming that  $\varphi(\ell', j')$  holds for all pairs  $(\ell', j')$  such that  $\ell' < \ell$  or  $j' < j$ , the inequality  $\varphi(\ell, j)$  is satisfied as well. If  $S_j = S_{j-1}$ , then there is nothing to show, and so we may assume that  $S_j = S_{j-1} \cup \{x_j\}$ , where  $x_j$  is an  $\varepsilon$ -good extension of  $S_{j-1}$ . In this case, letting  $s = |S_{j-1}|$ , recalling Definition 6.2, and invoking  $\varphi(\ell, j-1)$  and

$\varphi(\ell - 1, j - 1)$ , we have

$$\begin{aligned}
E_{S_j, \ell}(\lambda_\ell) &\leq E_{S_{j-1}, \ell}(\lambda_\ell) \left( 1 + \frac{2h}{\varepsilon} \frac{E_{S_{j-1}, \ell-1}(\lambda_{\ell-1}) - n}{n} \right) \\
&\leq \left( n + (2h/\varepsilon)^{\ell-1} e^{\lambda_1} s^\ell \right) \left( 1 + \frac{2h}{\varepsilon} \frac{(2h/\varepsilon)^{\ell-2} e^{\lambda_1} s^{\ell-1}}{n} \right) \\
&= n + (2h/\varepsilon)^{\ell-1} e^{\lambda_1} s^\ell + (2h/\varepsilon)^{\ell-1} e^{\lambda_1} s^{\ell-1} + \frac{(2h/\varepsilon)^{2\ell-2} e^{2\lambda_1} s^{2\ell-1}}{n} \\
&= n + (2h/\varepsilon)^{\ell-1} e^{\lambda_1} \left\{ s^\ell + s^{\ell-1} \left( 1 + \frac{(2h/\varepsilon)^{\ell-1} e^{\lambda_1} s^\ell}{n} \right) \right\}.
\end{aligned}$$

Since

$$(2h/\varepsilon)^{\ell-1} e^{\lambda_1} s^\ell \leq (2h/\varepsilon)^{h-1} e^{\lambda_1} m^h \stackrel{(45)}{\leq} n,$$

and  $(s+1)^\ell \geq s^\ell + 2s^{\ell-1}$ , it follows that

$$E_{S_j, \ell}(\lambda_\ell) \leq n + (2h/\varepsilon)^{\ell-1} e^{\lambda_1} (s+1)^\ell,$$

thus proving the  $\varphi(\ell, j)$  and concluding the induction step.  $\square$

Finally, we estimate the probability that  $|S| < (1 - \varepsilon)m$ . If this is the case, then there are more than  $\varepsilon m$  indices  $j$  for which  $x_j$  is not an  $\varepsilon$ -good extension of  $S_{j-1}$ . For each  $j$ , at least  $(1 - 2\varepsilon/3)n$  elements of  $\mathbb{Z}_n \setminus \{x_1, \dots, x_{j-1}\}$  are  $\varepsilon$ -good extensions of  $S_{j-1}$ . Since  $x_j$  is a uniformly chosen random element of  $\mathbb{Z}_n \setminus \{x_1, \dots, x_{j-1}\}$ , letting  $\text{Bin}(N, p)$  be a binomial random variable with parameters  $N$  and  $p$ , we have by Chernoff's bound,

$$\mathbf{P}\left(|S| < (1 - \varepsilon)m\right) \leq \mathbf{P}\left(\text{Bin}(m, 1 - 2\varepsilon/3) < (1 - \varepsilon)m\right) \leq \exp(-c_\varepsilon m)$$

for some constant  $c_\varepsilon > 0$ , and hence  $|S| \geq (1 - \varepsilon)m$  with probability  $1 - o(1)$ . This completes the proof of Theorem 2.10.  $\square$

We now derive Theorem 2.5 from Theorem 2.10 in the same way that we deduced Proposition 2.3(i) from Lemma 5.1.

*Proof of Theorem 2.5.* Fix  $\delta > 0$ . Let  $0 < \beta \leq 1/6$  be such that  $(1 - \beta)(1 - \delta/2) \geq 1 - \delta$  and  $\binom{(1+\beta)t}{\beta t} \leq (1 + \delta/2)^t$ . Now let  $m = (1 + \beta)t$ , and note that we may suppose that  $m \leq (\beta/6h) (n^{1-h/g})^{1/h}$ . It follows from Theorem 2.10 that  $F_{h,g}([n]_m) \geq (1 - \beta/2)m \geq t$  with probability at least  $1 - \beta$ . We conclude that

$$\begin{aligned}
|\mathcal{Z}_n^{h,g}(t)| &\geq (1 - \beta) \binom{n}{(1 + \beta)t} \binom{n-t}{\beta t}^{-1} = (1 - \beta) \binom{n}{t} \binom{(1 + \beta)t}{\beta t}^{-1} \\
&\geq (1 - \beta)(1 - \delta/2)^t \binom{n}{t} \geq (1 - \delta)^t \binom{n}{t},
\end{aligned} \tag{47}$$

The lower bound in (8) follows.  $\square$

## 7. CONCLUDING REMARKS

We close with two conjectures.

**Conjecture 7.1.** Fix an integer  $h \geq 3$  and  $\varepsilon > 0$ . For every  $t \geq n^{1/(2h-1)+\varepsilon}$  and every large enough  $n$ , we have

$$|\mathcal{Z}_n^h(t)| \leq \left( \frac{n}{t^{h-\varepsilon}} \right)^t. \quad (48)$$

Note that Proposition 2.3 implies that, if true, Conjecture 7.1 is basically optimal.

**Conjecture 7.2.** Let  $h \geq 3$  be an integer. Suppose  $0 \leq a \leq 1$  is a fixed constant and  $m = m(n) = (1 + o(1))n^a$ . Then **a.a.s.**  $F_h([n]_m) = n^{b+o(1)}$ , where  $b = b_1(a)$  and  $b_1(a)$  is as given in (10).

It is worth mentioning that an argument following the lines of the proof of the upper bound in Theorem 2.7 shows that Conjecture 7.1 implies Conjecture 7.2. At the time of writing, we strongly believe that we are able to prove Conjecture 7.1 for  $h = 3$ .

**Acknowledgement.** The authors thank the referee for valuable comments and suggestions.

#### REFERENCES

1. N. Alon, J. Balogh, R. Morris, and W. Samotij, *Counting sum-free sets in Abelian groups*, Israel J. Math. **199** (2014), 309–344.
2. R. C. Bose and S. Chowla, *Theorems in the additive theory of numbers*, Comment. Math. Helv. **37** (1962/1963), 141–147.
3. P. J. Cameron and P. Erdős, *On the number of sets of integers with various properties*, Number theory (Banff, AB, 1988), de Gruyter, Berlin, 1990, pp. 61–79.
4. S. Chen, *On the size of finite Sidon sequences*, Proc. Amer. Math. Soc. **121** (1994), no. 2, 353–356.
5. S. Chowla, *Solution of a problem of Erdős and Turán in additive-number theory*, Proc. Nat. Acad. Sci. India. Sect. A. **14** (1944), 1–2.
6. J. Cilleruelo, *New upper bounds for finite  $B_h$  sequences*, Adv. Math. **159** (2001), no. 1, 1–17.
7. D. Conlon and W. T. Gowers, *Combinatorial theorems in sparse random sets*, Submitted, 70pp, 2010.
8. A. G. D’yachkov and V. V. Rykov,  *$B_s$ -sequences*, Mat. Zametki **36** (1984), no. 4, 593–601.
9. P. Erdős, *On a problem of Sidon in additive number theory and on some related problems. Addendum*, J. London Math. Soc. **19** (1944), 208.
10. P. Erdős and P. Turán, *On a problem of Sidon in additive number theory, and on some related problems*, J. London Math. Soc. **16** (1941), 212–215.
11. B. Green, *The number of squares and  $B_h[g]$  sets*, Acta Arith. **100** (2001), no. 4, 365–390.
12. H. Halberstam and K. F. Roth, *Sequences*, second ed., Springer-Verlag, New York, 1983.
13. G. H. Hardy, J. E. Littlewood, and George Polya, *Inequalities*, by G. H. Hardy, J. E. Littlewood and G. Polya, The University press Cambridge, 1934 (English).
14. S. Janson, T. Łuczak, and A. Ruciński, *Random graphs*, Wiley-Interscience, New York, 2000.
15. X. D. Jia, *On finite Sidon sequences*, J. Number Theory **44** (1993), no. 1, 84–92.
16. D. J. Kleitman and D. B. Wilson, *On the number of graphs which lack small cycles*, manuscript, 15 pp, 1996.
17. D. J. Kleitman and K. J. Winston, *On the number of graphs without 4-cycles*, Discrete Math. **41** (1982), no. 2, 167–172.
18. Y. Kohayakawa, S. J. Lee, V. Rödl, and W. Samotij, *The number of Sidon sets and the maximum size of Sidon sets contained in a sparse random set of integers*, Random Structures Algorithms, to appear.
19. Y. Kohayakawa, S. Lee, and V. Rödl, *The maximum size of a Sidon set contained in a sparse random set of integers*, Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms (Philadelphia, PA), SIAM, 2011, pp. 159–171.
20. Y. Kohayakawa, T. Łuczak, and V. Rödl, *Arithmetic progressions of length three in subsets of a random set*, Acta Arith. **75** (1996), no. 2, 133–163.
21. M. N. Kolountzakis, *The density of  $B_h[g]$  sequences and the minimum of dense cosine sums*, J. Number Theory **56** (1996), no. 1, 4–11.

22. F. Krückeberg, *B<sub>2</sub>-Folgen und verwandte Zahlenfolgen*, J. Reine Angew. Math. **206** (1961), 53–60.
23. B. Lindström, *A remark on B<sub>4</sub>-sequences*, J. Combinatorial Theory **7** (1969), 276–277.
24. K. O’Bryant, *A complete annotated bibliography of work related to Sidon sequences*, Electron. J. Combin. (2004), Dynamic surveys 11, 39 pp. (electronic).
25. K. F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 104–109.
26. D. Saxton and A. Thomason, *Hypergraph containers*, arXiv:1204.6595, April 2012.
27. M. Schacht, *Extremal results for random discrete structures*, Submitted, 27pp, 2009.
28. I. E. Shparlinskiĭ, *On B<sub>s</sub>-sequences*, Combinatorial analysis, No. 7 (Russian), Moskov. Gos. Univ., Moscow, 1986, pp. 42–45, 163.
29. J. Singer, *A theorem in finite projective geometry and some applications to number theory*, Transactions of the American Mathematical Society **43** (1938), 377–385.
30. E. Szemerédi, *On sets of integers containing no k elements in arithmetic progression*, Acta Arith. **27** (1975), 199–245.

DEPARTMENT OF MATHEMATICS AND COMPUTER SCIENCE, EMORY UNIVERSITY, ATLANTA, GA 30322, USA  
(D. Dellamonica Jr., Y. Kohayakawa and V. Rödl)

*E-mail address:* `domingos.junior@gmail.com`, `rodl@mathcs.emory.edu`

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA, UNIVERSIDADE DE SÃO PAULO, RUA DO MATÃO 1010, 05508–090 SÃO PAULO, BRAZIL (Y. Kohayakawa)

*E-mail address:* `yoshi@ime.usp.br`

DEPARTMENT OF MATHEMATICS, DUKSUNG WOMEN’S UNIVERSITY, SEOUL, SOUTH KOREA (S. J. Lee)

*E-mail address:* `sanglee242@duksung.ac.kr`, `sjlee242@gmail.com`

SCHOOL OF MATHEMATICAL SCIENCES, TEL AVIV UNIVERSITY, TEL AVIV 69978, ISRAEL, AND TRINITY COLLEGE, CAMBRIDGE CB2 1TQ, UK (W. Samotij)

*E-mail address:* `samotij@post.tau.ac.il`