The Number of Sidon Sets and the Maximum Size of Sidon Sets Contained in a Sparse Random Set of Integers*

Yoshiharu Kohayakawa,[1,2,†] Sang June Lee,[3,‡] Vojtěch Rödl,[2,§] Wojciech Samotij[4,5,¶]

[1] *Instituto de Matemática e Estatística, Universidade de São Paulo, São Paulo 05508–090, Brazil; e-mail: yoshi@ime.usp.br*

[2] *Department of Mathematics and Computer Science, Emory University, Atlanta, Georgia 30322; e-mail: rodl@mathcs.emory.edu*

[3] *Department of Mathematical Sciences, Korea Advanced Institute of Science and Technology (KAIST), Daejeon 305-701, South Korea; e-mail: sjlee242@gmail.com*

[4] *School of Mathematical Sciences, Tel Aviv University, Tel Aviv 69978, Israel*

[5] *Trinity College, Cambridge CB2 1TQ, UK; e-mail: ws299@cam.ac.uk*

Received 1 June 2012; accepted 19 October 2012
Published online in Wiley Online Library (wileyonlinelibrary.com).
DOI 10.1002/rsa.20496

ABSTRACT: A set A of non-negative integers is called a *Sidon set* if all the sums a_1+a_2, with $a_1 \le a_2$ and $a_1, a_2 \in A$, are distinct. A well-known problem on Sidon sets is the determination of the maximum possible size $F(n)$ of a Sidon subset of $[n] = \{0, 1, \dots, n-1\}$. Results of Chowla, Erdős, Singer and Turán from the 1940s give that $F(n) = (1 + o(1))\sqrt{n}$. We study Sidon subsets of sparse random sets of integers, replacing the 'dense environment' $[n]$ by a sparse, random subset R of $[n]$, and ask how large a subset $S \subset R$ can be, if we require that S should be a Sidon set.

Let $R = [n]_m$ be a random subset of $[n]$ of cardinality $m = m(n)$, with all the $\binom{n}{m}$ subsets of $[n]$ equiprobable. We investigate the random variable $F([n]_m) = \max |S|$, where the maximum is taken over all Sidon subsets $S \subset [n]_m$, and obtain quite precise information on $F([n]_m)$ for the whole range of m, as illustrated by the following abridged version of our results. Let $0 \le a \le 1$ be a fixed constant

Correspondence to: S. J. Lee

*Parts of this work appeared in preliminary form in SODA 2011.

†Partially supported by CNPq (Proc. 308509/2007-2 and Proc. 484154/2010-9); NUMEC/USP (Project MaCLinC/USP); NSF (DMS 1102086).

‡Supported by the Korea Institute for Advanced Study (KIAS) grant and by the National Research Foundation of Korea (NRF) grant (No. 20120000798) funded by the Korea government (MEST).

§Supported by NSF (DMS 0800070 and DMS 1102086).

¶Partially supported by ERC Advanced Grant DMMCA and a Trinity College JRF.

and suppose $m = m(n) = (1 + o(1))n^a$. We show that there is a constant $b = b(a)$ such that, almost surely, we have $F([n]_m) = n^{b+o(1)}$. As it turns out, the function $b = b(a)$ is a continuous, piecewise linear function of a that is non-differentiable at two 'critical' points: $a = 1/3$ and $a = 2/3$. Somewhat surprisingly, between those two points, the function $b = b(a)$ is constant.

Our approach is based on estimating the number of Sidon sets of a given cardinality contained in $[n]$. Our estimates also directly address a problem raised by Cameron and Erdős (On the number of sets of integers with various properties, Number theory (Banff, AB, 1988), de Gruyter, Berlin, 1990, pp. 61–79). Random Struct. Alg., 00, 000–000, 2013

Keywords: Sidon sets, random sets of integers, probabilistic extremal problems, additive combinatorics

1. INTRODUCTION

Recent years have witnessed vigorous research in the classical area of additive combinatorics. An attractive feature of these developments is that applications in theoretical computer science have motivated some of the striking research in the area (see, e.g., [35]). For a modern treatment of the subject, the reader is referred to [34].

Among the best known concepts in additive number theory is the notion of a *Sidon set*. A set A of non-negative integers is called a *Sidon set* if all the sums $a_1 + a_2$, with $a_1 \le a_2$ and $a_1, a_2 \in A$, are distinct. A well-known problem on Sidon sets is the determination of the maximum possible size $F(n)$ of a Sidon subset of $[n] = \{0, 1, \dots, n-1\}$. In 1941, Erdős and Turán [14] showed that $F(n) \le \sqrt{n} + O(n^{1/4})$. In 1944, Chowla [8] and Erdős [11], independently of each other, observed that a result of Singer [32] implies that $F(n) \ge \sqrt{n} - O(n^{5/16})$. Consequently, it is known that $F(n) = (1+o(1))\sqrt{n}$. For a wealth of related material, the reader is referred to the classical monograph of Halberstam and Roth [17] and to a recent survey by O'Bryant [26] and the references therein.

We investigate Sidon sets contained in *random sets of integers*, and obtain essentially tight bounds on their relative density. Our approach is based on finding upper bounds for the number of Sidon sets of a given cardinality contained in $[n]$. Besides being the key to our probabilistic results, our upper bounds also address a problem of Cameron and Erdős [7].

We discuss our bounds on the number of Sidon sets and our probabilistic results in the next two subsections.

1.1. A Problem of Cameron and Erdős

Let $\mathcal{Z}_n$ be the family of Sidon sets contained in $[n]$. Over two decades ago, Cameron and Erdős [7] proposed the problem of estimating $|\mathcal{Z}_n|$. Observe that one trivially has

$$2^{F(n)} \le |\mathcal{Z}_n| \le \sum_{0 \le i \le F(n)} \binom{n}{i} = n^{(1/2+o(1))\sqrt{n}}. \tag{1}$$

Cameron and Erdős [7] improved the lower bound in (1) by showing that $\limsup_n |\mathcal{Z}_n| 2^{-F(n)} = \infty$ and asked whether the upper bound could also be strengthened. Our result is as follows.

Theorem 1.1. *There is a constant c for which $|\mathcal{Z}_n| \le 2^{cF(n)}$ for all large enough n.*

Our proof method gives that the constant c in Theorem 1.1 may be taken to be arbitrarily close to $\log_2(32e) = 6.442\cdots$. We do not make any attempts to optimize this constant as

it seems that our approach cannot yield a sharp estimate for $\log_2 |\mathcal{Z}_n|$ (in particular, we give the proof for constants arbitrarily close to $\log_2(33e) = 6.487\cdots$). Very recently, Saxton and Thomason [30] derived Theorem 1.1 (for c arbitrarily close to 55) from a more general theorem bounding the number of independent sets in certain hypergraphs. They also proved that $\log_2 |\mathcal{Z}_n| \geq (1.16 + o(1))F(n)$.

1.2. Probabilistic Results

We investigate Sidon subsets of sparse, *random* sets of integers, that is, we replace the 'environment' $[n]$ by a sparse, random subset R of $[n]$, and ask how large a subset $S \subset R$ can be, if we require that S should be a Sidon set.

Investigating how classical extremal results in 'dense' environments transfer to 'sparse' settings has proved to be a deep line of research. A fascinating example along these lines occurs in the celebrated work of Tao and Green [16], where Szemerédi's classical theorem on arithmetic progressions [33] is transferred to certain sparse, pseudorandom sets of integers and to the set of primes themselves (see [27, 28, 34] for more in this direction). Much closer examples to our setting are a version of Roth's theorem on 3-term arithmetic progressions [29] for random subsets of integers [24], and the far reaching generalizations due to Conlon and Gowers [9] and Schacht [31] (for details, the interested reader is referred to [9], [31], [18, Chapter 8] and [22, Section 4]). Before we proceed, we mention that additive properties of random sets of integers were already exploited by Erdős in the 50s to address a problem due to Sidon [12, 13] (see also [17, Chapter III] and [34, Chapter 1]).

Let us now state a weak, but less technical version of our main probabilistic results. Let $F(R) = \max |S|$, where the maximum is taken over all Sidon subsets $S \subset R$. Let $[n]_m$ be a random subset of $[n]$ of cardinality $m = m(n)$, with all the $\binom{n}{m}$ subsets of $[n]$ equiprobable. We are interested in the random variable $F([n]_m)$.

Standard methods give that, almost surely, that is, with probability tending to 1 as $n \to \infty$, we have $F([n]_m) = (1 - o(1))m$ if $m = m(n) \ll n^{1/3}$ (here and throughout we write $f \ll g$ to mean $f = o(g)$). On the other hand, the results of Schacht [31] and Conlon and Gowers [9] imply that, if $m = m(n) \gg n^{1/3}$, then, almost surely, we have

$$F([n]_m) = o(m). \tag{2}$$

Thus $F([n]_m)$ undergoes a sudden change of behaviour at $m = n^{1/3+o(1)}$. The following abridged version of our results already gives us quite precise information on $F([n]_m)$ for the whole range of m.

Theorem 1.2. *Let $0 \leq a \leq 1$ be a fixed constant. Suppose $m = m(n) = (1 + o(1))n^a$. There exists a constant $b = b(a)$ such that almost surely*

$$F([n]_m) = n^{b+o(1)}. \tag{3}$$

Furthermore,

$$b(a) = \begin{cases} a & \text{if } 0 \leq a \leq 1/3, \\ 1/3 & \text{if } 1/3 \leq a \leq 2/3, \\ a/2 & \text{if } 2/3 \leq a \leq 1. \end{cases} \tag{4}$$

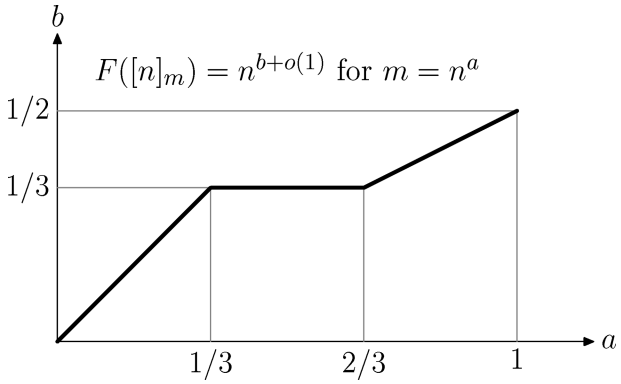

Fig. 1. The graph of $b = b(a)$.

Thus, the function $b = b(a)$ is piecewise linear. The graph of $b = b(a)$ is given in Fig. 1. The point $(a, b) = (1, 1/2)$ in the graph is clear from the Erdős–Turán and Chowla results [8, 11, 14] mentioned above. The behaviour of $b = b(a)$ in the interval $0 \le a \le 1/3$ is not hard to establish. The fact that the point $(1/3, 1/3)$ could be an interesting point in the graph is suggested by the results of Schacht [31] and Conlon and Gowers [9]. It is somewhat surprising that, besides the point $a = 1/3$, there is a second value at which $b = b(a)$ is 'critical', namely, $a = 2/3$. Finally, we find it rather interesting that $b = b(a)$ should be *constant* between those two critical points. We state our results in full in Section 2.1. Our results in fact determine $F([n]_m)$ up to a *constant* multiplicative factor for $m \le n^{2/3-\delta}$ for any fixed $\delta > 0$ and for $m \ge n^{2/3}(\log n)^{8/3}$. For the missing range of m, around $n^{2/3}$, our lower and upper bounds differ by a factor of $O((\log n)/\log\log n)$.

2. MAIN RESULTS

2.1. Statement of the Main Results

We prove a more detailed result than Theorem 1.1. Let $\mathcal{Z}_n(t)$ be the family of Sidon sets of cardinality t contained in $[n]$.

Theorem 2.1. *Let $0 < \sigma < 1$ be a real number. For any large enough n and $t \ge 2s_0$, where $s_0 = \big(2(1-\sigma)^{-1} n \log n\big)^{1/3}$, we have*

$$|\mathcal{Z}_n(t)| \le n^{3s_0} \left(\frac{32en}{\sigma t^2}\right)^t. \tag{5}$$

Theorem 1.1 follows from Theorem 2.1 by summing over all t (see Section 3.2). Our next result covers values of t smaller than the ones covered in Theorem 2.1.

Theorem 2.2. *Let n and t be integers with*

$$30n^{1/3} \le t \le 5(n \log n)^{1/3}. \tag{6}$$

Then

$$|\mathcal{Z}_n(t)| \le \left(\frac{22n}{t}\exp\left(-\frac{t^3}{6\cdot 5^3 n}\right)\right)^t. \tag{7}$$

Let us now turn to our probabilistic results. Instead of working with the *uniform model* $[n]_m$ of random subsets of $[n]$, it will be more convenient to work with the so called *binomial model* $[n]_p$, in which each element of $[n]$ is put in $[n]_p$ with probability p, independently of all other elements. Routine methods allow us to translate our results on $[n]_p$ below to the corresponding results on $[n]_m$, where $p = m/n$ (see Section 2.2 for details).

We state our results on $F([n]_p)$ split into theorems covering different ranges of $p = p(n)$. Our first result corresponds to the range $0 \le a \le 1/3$ in Theorem 1.2.

Theorem 2.3. *For $n^{-1} \ll p = p(n) \ll n^{-2/3}$, we almost surely have*

$$F([n]_p) = (1 + o(1))np. \tag{8}$$

For $n^{-1} \ll p \le 2n^{-2/3}$, we almost surely have

$$\left(\frac{1}{3} + o(1)\right)np \le F([n]_p) \le (1 + o(1))np, \tag{9}$$

Remark 2.4. One may in fact prove the following result: if $p = \gamma n^{-2/3}$ for some constant γ, then almost surely

$$\left(1 - \frac{1}{12}\gamma^3 + o(1)\right)np \le F([n]_p) \le \left(1 - \frac{1}{12}\gamma^3 + \frac{1}{12}\gamma^6 + o(1)\right)np. \tag{10}$$

Our next result covers the range $1/3 \le a < 2/3$ in Theorem 1.2.

Theorem 2.5. *For any $\delta > 0$, there is a positive constant $c_2 = c_2(\delta)$ such that if $2n^{-2/3} \le p = p(n) \le n^{-1/3-\delta}$, then we almost surely have*

$$c_1(n\log(n^2p^3))^{1/3} \le F([n]_p) \le c_2(n\log(n^2p^3))^{1/3}, \tag{11}$$

where c_1 is a positive absolute constant.

We now turn to the point $a = 2/3$ in Theorem 1.2.

Theorem 2.6. *For any $0 \le \delta < 1/3$, there is a positive constant $c_3 = c_3(\delta)$ such that if $1 \le \alpha = \alpha(n) \le n^\delta$ and $p = p(n) = \alpha^{-1}n^{-1/3}(\log n)^{2/3}$, then we almost surely have*

$$c_3(n\log n)^{1/3} \le F([n]_p) \le c_4(n\log n)^{1/3}\frac{\log n}{\log(\alpha + \log n)},$$

where c_4 is an absolute constant.

We remark that Theorems 2.5 and 2.6 consider ranges that overlap (functions $p = p(n)$ of the form $n^{-1/3-\delta'}$ for some $0 < \delta' < 1/3$ are covered by both theorems). Finally, we consider the range $2/3 \le a \le 1$ in Theorem 1.2.

Theorem 2.7. *There exist positive absolute constants c_5 and c_6 for which the following holds. If $1 \le \alpha = \alpha(n) \le (\log n)^2$ and $p = p(n) = \alpha^{-1} n^{-1/3} (\log n)^{8/3}$, then we almost surely have*

$$c_5 \sqrt{np} \le F([n]_p) \le c_6 \sqrt{np} \cdot \frac{\sqrt{\alpha}}{1 + \log \alpha}.$$

Furthermore, if $n^{-1/3} (\log n)^{8/3} \le p = p(n) \le 1$, then, almost surely,

$$c_5 \sqrt{np} \le F([n]_p) \le c_6 \sqrt{np}.$$

2.2. The Uniform Model and the Binomial Model

We now discuss how to translate Theorems 2.3, 2.5–2.7 on $[n]_p$ in Section 2.1 to the corresponding results on $[n]_m$. Before we proceed, let us make the following definition.

Definition 2.8. We shall say that an event in the probability space of the random sets $[n]_p$ or in the probability space of the random sets $[n]_m$ holds *with overwhelming probability*, abbreviated as *w.o.p.*, if the probability of failure of that event is $O(n^{-C})$ for any constant C, that is, if the probability of failure is 'superpolynomially' small.

For us, the following consequence of Pittel's inequality (see, e.g., [6, p. 35] and [19, p. 17]) will suffice for translating results on $[n]_p$ to results on $[n]_m$.

Lemma 2.9. *Let $1 \le m = m(n) < n$ and $p = p(n)$ be such that $p = m/n$. Let P be an event in the probability space of the random sets $[n]_p$. If $[n]_p$ is in P w.o.p., then $[n]_m$ is in $P \cap \binom{[n]}{m}$ w.o.p.*

Proof. Let Q be the complement of P. Pittel's inequality (see [6, p. 35] and [19, p. 17]) states that

$$\mathbb{P}\left[[n]_m \text{ is in } Q \cap \binom{[n]}{m} \right] = O(\sqrt{m}) \cdot \mathbb{P}\left[[n]_p \text{ is in } Q \right]. \tag{12}$$

Since, by hypothesis, $\mathbb{P}[[n]_p \text{ is in } Q] = O(n^{-C})$ holds for any constant $C > 0$, inequality (12) implies that

$$\mathbb{P}\left[[n]_m \text{ is in } Q \cap \binom{[n]}{m} \right] = O(\sqrt{m} \cdot n^{-C}) = O(\sqrt{n} \cdot n^{-C}) = O(n^{-C+1/2}),$$

which completes the proof of Lemma 2.9. ∎

Every result in Theorems 2.5–2.7 will be proved with 'w.o.p.' rather than with 'almost surely'. By Lemma 2.9, we can translate each such result on $[n]_p$ to the corresponding result on $[n]_m$, where $p = m/n$. For example, Theorem 2.5 implies the following uniform version: *For any $\delta > 0$, there is a positive constant $c_2 = c_2(\delta)$ such that if $2n^{1/3} \le m = m(n) \le n^{2/3-\delta}$, then, with overwhelming probability, we have*

$$c_1 \left(n \log \frac{m^3}{n} \right)^{1/3} \le F([n]_m) \le c_2 \left(n \log \frac{m^3}{n} \right)^{1/3},$$

where c_1 is a positive absolute constant.

Finally, we remark that one may use the usual deletion method to prove that the result on $[n]_m$ corresponding to Theorem 2.3 holds almost surely.

2.3. Organization and Notation

Our results on the number of Sidon sets are proved in Section 3. In Section 4, we consider the upper bounds in Theorems 2.5–2.7. Section 5 contains some preparatory lemmas for the proof of Theorem 2.3 and for the proofs of the lower bounds in Theorems 2.5–2.7. The proof of Theorem 2.3 is given in Section 6. In Section 7, we give the proofs of the lower bounds in Theorems 2.5–2.7.

More in line with the combinatorics literature and deviating from the number-theoretic usage, we write $f \ll g$ and $g \gg f$ to mean $f = o(g)$. For simplicity, we omit 'floor' and 'ceiling' symbols in our formulae, when they are not essential. For simplicity, we often write a/bc instead of the less ambiguous $a/(bc)$.

3. THE NUMBER OF SIDON SETS

The proofs of Theorems 2.1 and 2.2 are based on a method introduced by Kleitman and Winston [21] (see [2, 4, 5, 15, 23] for other applications of this method).

3.1. Independent Sets in Locally Dense Graphs

We start with the following lemma, which gives an upper bound for the number of independent sets in graphs that are 'locally dense'.

Lemma 3.1. *Let G be a graph on N vertices, let q be an integer and let $0 \le \beta \le 1$ and R be real numbers with*

$$R \ge e^{-\beta q} N. \tag{13}$$

Suppose the number of edges $e(U)$ induced in G by any set $U \subset V(G)$ with $|U| \ge R$ satisfies

$$e(U) \ge \beta \binom{|U|}{2}. \tag{14}$$

Then, for all integers $r \ge 0$, the number of independent sets in G of cardinality $q + r$ is at most

$$\binom{N}{q}\binom{R}{r}. \tag{15}$$

Proof. Fix an integer $r \ge 0$. We describe a deterministic algorithm that associates to every independent set I of size $q+r$ in G a pair (S_0, A) of disjoint sets with $S_0 \subset I \subset S_0 \cup A \subset V(G)$ and with $|S_0| = q$ and $|A| \le R$. Furthermore, if, for some inputs I and I', the algorithm outputs (S_0, A) and (S_0', A') with $S_0 = S_0'$, then $A = A'$. A moment's thought now reveals that the number of independent sets in G with $q + r$ elements is at most as given in (15), as claimed. We now proceed to describe the algorithm.

At all times, our algorithm maintains a partition of $V(G)$ into sets S, X, and A (short for *selected*, *excluded*, and *available*). As the algorithm evolves, S increases, X increases and A decreases. The vertices in A will be labelled $v_1, \ldots, v_{|A|}$, where, for every i, the vertex v_i has maximum degree in $G[\{v_i, \ldots, v_{|A|}\}]$ (the graph induced by $\{v_i, \ldots, v_{|A|}\}$ in G); we break ties

arbitrarily by giving preference to vertices that come earlier in some arbitrary predefined ordering of $V(G)$.

We start the algorithm with $A = V(G)$ and $S = X = \emptyset$. Crucially, at all times we maintain $S \subset I \subset S \cup A$. The algorithm works as follows. While $|S| < q$, we repeat the following. Let $a = |A|$ and suppose $A = \{v_1, \dots, v_a\}$, with the vertex labelling convention described above. Let i be the smallest index such that v_i belongs to our independent set I, move $v_1, \dots, v_{i-1}$ from A to X (they are not in I by the choice of i), and move v_i from A to S (v_i is in I). Observe that A has already lost i members in this iteration and S has gained one. Let $U = \{v_i, \dots, v_a\}$. If $|U| \geq R$, we further move all neighbours of v_i in A to X (since I is an independent set and $v_i \in I$). Otherwise, i.e., if $|U| < R$, consider the first $q - |S|$ members $v_{i_1}, \dots, v_{i_{q-|S|}} \in I \cap A$ ($i < i_1 < \cdots < i_{q-|S|} \leq a$) and move them from A to S. Note that this is possible because $|I \cap A| = q + r - |S| \geq q - |S|$, and note that we now have $|S| = q$ (we do this because it is convenient to have S of cardinality q).

The procedure above defines an increasing sequence of sets S. Once we obtain a set S with $|S| = q$, we let $S_0 = S$, output (S_0, A) and stop the algorithm. Inspection shows that A depends only on S_0 and not on I, i.e., if (S_0, A) and (S_0, A') are both outputs of the algorithm (for some inputs I and I'), then $A = A'$. We now use our assumption on G to show that $|A| \leq R$.

We consider two cases: The first case is the case in which the body of the while loop of the algorithm is executed with $|U| < R$ at an iteration. The second case is the case in which we have $|U| \geq R$ during the q iterations of the while loop. Observe that one of two cases must occur.

First, we consider the first case. At the iteration with $|U| < R$, the set A lost the first i vertices (and possibly others) and hence at the end of this iteration we have $|A| \leq a - i = |U| - 1 < R$. Moreover, $|S|$ becomes of cardinality q and the algorithm stops.

Next, we consider the second case in which we have $|U| \geq R$ during the q iterations of the while loop. In each iteration, consider an execution of the body of the while loop of the algorithm when $|U| \geq R$ and (only) the vertex v_i is moved to S. In this execution, A loses, in total, $i + d(v_i, U)$ vertices, where $d(v_i, U)$ is the degree of v_i in the graph $G[U]$. Recall that we are considering the case $|U| \geq R$ and that v_i has maximum degree in the graph $G[U]$. Applying (14), we see that $d(v_i, U) \geq \beta(|U| - 1)$. Therefore, at the end of this iteration, A has cardinality

$$a - (i + d(v_i, U)) \leq a - (a - |U| + 1 + \beta(|U| - 1)) \leq |U| - \beta|U| \leq (1 - \beta)a.$$

In the second case, the cardinality of A decreases by a factor of $1 - \beta$ in the q iterations of the while loop and, at the end, A has at most $N(1 - \beta)^q \leq Ne^{-\beta q} \leq R$ elements. ∎

3.2. Proof of Theorem 2.1

We derive Theorem 2.1 from the following lemma.

Lemma 3.2. *Let n, s and q be integers and let $0 < \sigma < 1$ be a real number such that*

$$\frac{s^2 q}{n} \geq \frac{2}{1 - \sigma} \log \frac{\sigma s}{2}. \tag{16}$$

Then, for any integer $r \geq 0$, we have

$$|\mathcal{Z}_n(s+q+r)| \leq |\mathcal{Z}_n(s)| \binom{n}{q} \binom{2n/\sigma s}{r}. \tag{17}$$

To obtain the bound for $|\mathcal{Z}_n(t)|$ in Theorem 2.1, we apply Lemma 3.2 iteratively (in the calculations below, we omit 'floor' and 'ceiling' symbols when they are not essential).

Proof of Theorem 2.1. Fix integers n and t, with $t \geq 2s_0$, where s_0 is as given in the statement of our theorem, that is, $s_0 = \left(2(1-\sigma)^{-1} n \log n\right)^{1/3}$. We may clearly suppose that $t \leq F(n) = (1+o(1))\sqrt{n}$, as otherwise $\mathcal{Z}_n(t) = \emptyset$. Let K be the largest integer satisfying $t2^{-K} \geq 2s_0$. We define three sequences $(s_k)_{0\leq k\leq K}$, $(q_k)_{0\leq k\leq K}$ and $(r_k)_{0\leq k\leq K}$ as follows. We let $q_0 = s_0$ and $r_0 = t2^{-K} - s_0 - q_0$. Moreover, we let $s_1 = t2^{-K} \geq 2s_0$, $q_1 = q_0/4$ and $r_1 = t2^{-K+1} - s_1 - q_1$. For $k = 2, \ldots, K$, we let $s_k = 2s_{k-1} = t2^{-K+k-1}$, $q_k = q_{k-1}/4 = q_0 4^{-k}$ and $r_k = t2^{-K+k} - s_k - q_k$. Note that $q_k + r_k = s_k$ for $k \geq 1$ since $s_{k+1} = 2s_k$. We apply Lemma 3.2 with parameters s_k, q_k and r_k for $k = 0, \ldots, K$, to obtain from (17) that

$$|\mathcal{Z}_n(t2^{-K+k})| = |\mathcal{Z}_n(s_k + q_k + r_k)| \leq |\mathcal{Z}_n(s_k)| \binom{n}{q_k} \binom{2n/\sigma s_k}{r_k} \tag{18}$$

for all k. It suffices to check (16) to justify these applications of Lemma 3.2. Since $s_k^2 q_k \geq s_0^2 q_0 = 2(1-\sigma)^{-1} n \log n > 2(1-\sigma)^{-1} n \log(\sigma s_k/2)$ for all $0 \leq k \leq K$, inequality (16) holds for n, s_k and q_k. Using that $s_{k-1} + q_{k-1} + r_{k-1} = t2^{-K+k-1} = s_k$ for $k \geq 1$ and that $|\mathcal{Z}_n(s_0)| \leq \binom{n}{s_0}$, we obtain from (18) that

$$|\mathcal{Z}_n(t)| \leq \binom{n}{s_0} \prod_{0\leq k\leq K} \binom{n}{q_k} \prod_{0\leq k\leq K} \binom{2n/\sigma s_k}{r_k}. \tag{19}$$

Note that

$$\binom{n}{s_0} \leq \left(\frac{en}{s_0}\right)^{s_0} \leq n^{2s_0/3} \tag{20}$$

and that

$$\prod_{0\leq k\leq K} \binom{n}{q_k} \leq n^{\sum_{0\leq k\leq K} q_k} \leq n^{q_0 \sum_{0\leq k\leq K} 1/4^k} \leq n^{4q_0/3} = n^{4s_0/3}. \tag{21}$$

We now proceed to estimate the last factor of the right-hand side of (19). First note that, by the choice of K, we have $(r_0 + s_0 + q_0)/2 = t2^{-K-1} < 2s_0$, and hence $r_0 < 2s_0$. Therefore, we have that

$$\binom{2n/\sigma s_0}{r_0} \leq \begin{cases} 1 \leq n^{s_0} & \text{if } r_0 = 0 \\ \left(\frac{2en}{\sigma s_0 r_0}\right)^{r_0} \leq n^{s_0} & \text{if } 0 < r_0 \leq s_0 \\ \left(\frac{2en}{\sigma s_0 r_0}\right)^{r_0} \leq n^{r_0/3} \leq n^{2s_0/3} \leq n^{s_0} & \text{if } s_0 < r_0 < 2s_0 \end{cases} \tag{22}$$

for all large n. We now note that

$$\prod_{1\leq k\leq K} \binom{2n/\sigma s_k}{r_k} = \prod_{1\leq k\leq K} \binom{2n/\sigma s_{K-k+1}}{r_{K-k+1}} \leq \prod_{1\leq k\leq K} \binom{2n/\sigma s_{K-k+1}}{r_{K-k+1} + q_{K-k+1}}. \tag{23}$$

To justify the inequality in (23) above, we check that

$$r_{K-k+1} + q_{K-k+1} \le \frac{2n}{3\sigma s_{K-k+1}}. \tag{24}$$

Recalling that $r_{K-k+1} + q_{K-k+1} = s_{K-k+1} = t2^{-k}$, we see that (24) is equivalent to $t2^{-k} \le \sqrt{2n/3\sigma}$. However,

$$\frac{t}{2^k} \le \frac{t}{2} \le \frac{1}{2}F(n) = \left(\frac{1}{2} + o(1)\right)\sqrt{n} \le \sqrt{\frac{2n}{3}} \le \sqrt{\frac{2n}{3\sigma}} \tag{25}$$

for all large enough n. We continue (23) by noticing that

$$\begin{aligned}\prod_{1\le k\le K} \binom{2n/\sigma s_{K-k+1}}{r_{K-k+1}+q_{K-k+1}} &= \prod_{1\le k\le K} \binom{2n/\sigma t2^{-k}}{t2^{-k}} \le \prod_{1\le k\le K} \left(\frac{2^{2k+1}en}{\sigma t^2}\right)^{t2^{-k}} \\ &\le \left(\frac{2en}{\sigma t^2}\right)^{t\sum_{k\ge 1}2^{-k}} 2^{2t\sum_{k\ge 1}k2^{-k}} = \left(\frac{2en}{\sigma t^2}\right)^{t} 2^{4t} = \left(\frac{32en}{\sigma t^2}\right)^{t}.\end{aligned} \tag{26}$$

Inequality (5) now follows from (19), (20), (21), (22) and (26). ∎

It now remains to prove Lemma 3.2.

Proof of Lemma 3.2. Let $S_0 \subset [n]$ be an arbitrary Sidon set with $|S_0| = s$. We show that the number of Sidon sets $S \subset [n]$ with $S_0 \subset S$ and $|S| = s + q + r$ is at most $\binom{n}{q}\binom{2n/\sigma s}{r}$, whence our lemma will follow.

Let G be the graph on $V = [n] \setminus S_0$ satisfying that $\{a_1, a_2\}$ $(a_1 \ne a_2)$ is an edge in G if and only if there are b_1 and $b_2 \in S_0$ such that $a_1 + b_1 = a_2 + b_2$. Observe that if $S \subset [n]$ is a Sidon set containing S_0, then $S \setminus S_0$ is an independent set in G. Let $N = |V| = n - s$, $\beta = (1-\sigma)s^2/2n$ and $R = 2n/\sigma s$. We wish to apply Lemma 3.1 to G with β and R as just defined, to obtain an upper bound for the number of independent sets of cardinality $q + r$. Note that (13) follows from (16). Now let $U \subset V$ with $|U| \ge R$ be given. We check (14) as follows.

Let J be the bipartite graph with (disjoint) vertex classes $[2n]$ and U, with $w \in [2n]$ adjacent to $a \in U$ in J if and only if $w = a + b$ for some $b \in S_0$. Note that a_1 and $a_2 \in U$ have a common neighbour $w \in [2n]$ if and only if there are b_1 and $b_2 \in S_0$ with $a_1 + b_1 = w = a_2 + b_2$, that is, if and only if $\{a_1, a_2\}$ is an edge of G.

Now note that J contains no 4-cycle: if $a_1, a_2 \in U$ with $a_1 \ne a_2$ are both adjacent to both w and $w' \in [2n]$ with $w \ne w'$, then $a_1 + b_1 = w = a_2 + b_2$ for some b_1 and $b_2 \in S_0$ and $a_1 + b_1' = w' = a_2 + b_2'$ for some b_1' and $b_2' \in S_0$. But then $b_1 - b_1' = b_2 - b_2'$, and hence $b_1 + b_2' = b_1' + b_2$. As b_1, b_1', b_2 and $b_2' \in S_0$ and S_0 is a Sidon set, we have $\{b_1, b_2'\} = \{b_1', b_2\}$. Since $a_1 \ne a_2$, we have $b_1 \ne b_2$, whence $b_1 = b_1'$, implying that $w = a_1 + b_1 = a_1 + b_1' = w'$.

The remarks above give that $e(U) = \sum_{w\in[2n]} \binom{d_J(w)}{2}$, where $d_J(w)$ denotes the degree of w in J. Note that $\sum_{w\in[2n]} d_J(w) = \sum_{a\in U} d_J(a) = |U||S_0| = |U|s$. Using the convexity of

the function $f(x) = \binom{x}{2}$ and Jensen's inequality and recalling that $|U| \geq R = 2n/\sigma s$, i.e., $1 \leq \sigma \frac{|U|s}{2n}$, we obtain

$$e(U) = \sum_{w\in[2n]} \binom{d_J(w)}{2} \geq 2n\binom{|U|s/2n}{2} = \frac{|U|s}{2}\left(\frac{|U|s}{2n} - 1\right)$$
$$\geq \frac{1}{4}(1-\sigma)\frac{s^2}{n}|U|^2 \geq \beta\binom{|U|}{2},$$

as required in (14). Recall that a Sidon set $S \subset [n]$ containing S_0 is such that $S \setminus S_0$ is an independent set in G. Therefore, our required bound for the number of such S with $|S| = s+q+r$ follows from the upper bound (15) for the number of independent sets of cardinality $q+r$ in G. ∎

We conclude this section by deriving Theorem 1.1 from Theorem 2.1.

Proof of Theorem 1.1. Let $\sigma = 32/33$ in Theorem 2.1. Then $s_0 = (2(1-\sigma)^{-1}n\log n)^{1/3} = (66n\log n)^{1/3}$. For large enough n, we have

$$|\mathcal{Z}_n| = \sum_{0\leq t\leq F(n)} |\mathcal{Z}_n(t)| \leq \sum_{0\leq t<2s_0}\binom{n}{t} + \sum_{2s_0\leq t\leq F(n)} n^{3s_0}\left(\frac{33en}{t^2}\right)^t. \tag{27}$$

Note that

$$\sum_{0\leq t<2s_0}\binom{n}{t} \leq 2s_0\binom{n}{2s_0} \leq n^{2s_0}, \tag{28}$$

and that since $f(t) = (33en/t^2)^t$ is increasing on the interval $\left(0, \sqrt{33n/e}\right)$,

$$\sum_{2s_0\leq t\leq F(n)} n^{3s_0}\left(\frac{33en}{t^2}\right)^t \leq \sqrt{n}\cdot n^{3s_0}(33e)^{\sqrt{n}(1+o(1))} \leq (33e)^{\sqrt{n}(1+o(1))} \leq (33e)^{F(n)(1+o(1))}. \tag{29}$$

Combining (27) together with (28) and (29) implies that $|\mathcal{Z}_n| \leq 2^{cF(n)}$ for a suitable constant c. ∎

3.3. Proof of Theorem 2.2

We derive Theorem 2.2 from the following more general but technical estimate.

Lemma 3.3. *Let n and t be integers. Suppose s is an integer and σ is a real number such that, letting $\omega = t/s$, we have*

$$\omega \geq 4, \quad 0 < \sigma < 1 \quad \text{and} \quad \frac{s^3}{n} \geq \frac{2}{1-\sigma}\log\frac{\sigma s}{2}. \tag{30}$$

Then

$$|\mathcal{Z}_n(t)| \leq \left(\frac{12\omega n}{(t\sigma)^{1-2/\omega}t}\right)^t. \tag{31}$$

Proof. We invoke Lemma 3.2 with $q = s$. Note that, then, (30) implies (16). We now let r in Lemma 3.2 be $t - 2s$ and obtain that

$$|\mathcal{Z}_n(t)| \le \binom{n}{s}\binom{n}{s}\binom{2n/\sigma s}{t-2s}. \tag{32}$$

The right-hand side of (32) is

$$\begin{aligned}\binom{n}{s}^2\binom{2n/\sigma s}{t-2s} &\le \left(\frac{en}{s}\right)^{2s}\left(\frac{2en}{\sigma s(t-2s)}\right)^{t-2s} = \left(\frac{en}{s}\right)^{2s}\left(\frac{en}{s}\right)^{t-2s}\left(\frac{2}{\sigma(t-2s)}\right)^{t-2s}\\ &= \left(\frac{e\omega n}{t}\right)^{t}\left(\frac{2}{\sigma t(1-2/\omega)}\right)^{t(1-2/\omega)} = \left(C\frac{n}{t^{2-2/\omega}\sigma^{1-2/\omega}}\right)^{t},\end{aligned}$$

where $C = 2^{1-2/\omega}e\omega/(1-2/\omega)^{1-2/\omega} = 2^{1-2/\omega}e\omega^{2-2/\omega}/(\omega-2)^{1-2/\omega}$. As $\omega \ge 4$, we have $\omega - 2 \ge \omega/2$, and hence $C \le e\omega 4^{1-2/\omega} < 12\omega$, completing the proof of Lemma 3.3. ∎

Proof of Theorem 2.2. We shall apply Lemma 3.3. Let $s = \lfloor t/4 \rfloor$ and let $\omega = t/s \ge 4$. Note that $\omega \le 5$ by our assumption on t. Let $\lambda = \exp\left(t^3/(3\cdot 5^3 n)\right)$ and set $\sigma = 2\lambda/s$. It follows from (6) that $\lambda \le n^{1/3}$ and $\sigma \le \frac{10\lambda}{t} \le 1/3$. Therefore, $2/(1-\sigma) \le 3$, and hence

$$\frac{s^3}{n} \ge \frac{t^3}{5^3 n} = 3\log\lambda \ge \frac{2}{1-\sigma}\log\lambda,$$

whence the third condition in (30) holds. We thus conclude that (31) holds. Let us now estimate the right-hand side of (31).

Note that $t\sigma \ge 4s\sigma = 8\lambda$, and therefore $(t\sigma)^{1-2/\omega} \le (8\lambda)^{1-2/\omega} = (8\lambda)^{1/2}$ and

$$\frac{12\omega n}{(t\sigma)^{1-2/\omega}t} = \frac{60n}{(8\lambda)^{1/2}t} = \frac{15\cdot 2^{1/2}n}{\lambda^{1/2}t} \le \frac{22n}{t}\exp\left(-\frac{t^3}{6\cdot 5^3 n}\right). \tag{33}$$

Inequality (7) follows from (31) and (33), and Theorem 2.2 is proved. ∎

4. THE UPPER BOUNDS IN THEOREMS 2.5–2.7

We shall apply Lemma 3.3 and Theorem 2.1 in order to prove the upper bounds in Theorem 2.5 and Theorems 2.6–2.7, respectively.

4.1. Proof of the Upper Bound in Theorem 2.5

Let $\delta > 0$ be given. We show that there is a constant $c_2 = c_2(\delta)$ such that if $2n^{-2/3} \le p = p(n) \le n^{-1/3-\delta}$, then w.o.p. we have

$$F([n]_p) \le c_2(n\log n^2p^3)^{1/3}.$$

To this end, we apply Lemma 3.3. We first define several auxiliary constants used to set t, ω and σ in Lemma 3.3. Choose $\eta > 0$ small enough so that

$$(1-3\delta)\left(\frac{1}{3}+\eta\right) < \frac{1}{3}. \tag{34}$$

Choose $\omega \geq 4$ so that

$$\left(\frac{1}{3}+\eta\right)\left(1-\frac{2}{\omega}\right) > \frac{1}{3}. \tag{35}$$

Finally, choose $c = c_2$ so that

$$\left(\frac{c}{\omega}\right)^3 > 3\left(\frac{1}{3}+\eta\right) \quad \text{and} \quad c > \frac{24\omega}{2^{(1+3\eta)(1-2/\omega)}}. \tag{36}$$

Now set $t = c(n\log n^2p^3)^{1/3}$, $s = t/\omega$, $\sigma = 2(n^2p^3)^{1/3+\eta}/s$ and $\xi = 24\omega/c2^{(1+3\eta)(1-2/\omega)}$. Note that

$$t \geq c(n\log 8)^{1/3} \geq cn^{1/3} \qquad \text{and} \qquad \xi < 1. \tag{37}$$

We first check that condition (30) holds for large enough n. We have $\omega \geq 4$ by the choice of ω. Moreover, we have $\sigma \to 0$ as $n \to \infty$ because of (34). Finally, from (36) and the fact that $\sigma \to 0$, we have

$$\frac{s^3}{n} = \left(\frac{c}{\omega}\right)^3 \log n^2p^3 \geq 3\left(\frac{1}{3}+\eta\right)\log n^2p^3 \geq \frac{2(1/3+\eta)}{1-\sigma}\log n^2p^3 = \frac{2}{1-\sigma}\log\frac{\sigma s}{2},$$

which completes the verification of (30). Hence Lemma 3.3 implies that

$$\mathbb{P}\left([n]_p \text{ contains a Sidon set of size } t\right) \leq |\mathcal{Z}_n(t)|p^t \leq \left(\frac{12\omega np}{t(t\sigma)^{1-2/\omega}}\right)^t. \tag{38}$$

Making use of the first equation of (37) and the fact that $t\sigma = \omega s\sigma = 2\omega(n^2p^3)^{1/3+\eta}$, we see that the upper bound in (38) is at most

$$\begin{aligned}\left(\frac{12\omega np}{cn^{1/3}(2\omega)^{1-2/\omega}(n^2p^3)^{(1/3+\eta)(1-2/\omega)}}\right)^t &\leq \left(\frac{12\omega}{c(2\omega)^{1-2/\omega}}\cdot\frac{n^{2/3}p}{(n^2p^3)^{(1/3+\eta)(1-2/\omega)}}\right)^t \\ &= \left(\frac{12\omega^{2/\omega}}{2^{1-2/\omega}c(n^2p^3)^{(1/3+\eta)(1-2/\omega)-1/3}}\right)^t,\end{aligned} \tag{39}$$

which, by (35) and the assumption $p \geq 2n^{-2/3}$, is at most

$$\left(\frac{12\omega}{2^{1/2}c(2^3)^{(1/3+\eta)(1-2/\omega)-1/3}}\right)^t \leq \left(\frac{24\omega}{c2^{(1+3\eta)(1-2/\omega)}}\right)^t = \xi^t. \tag{40}$$

To complete the proof, it suffices to recall (37).

4.2. Proof of the Upper Bound in Theorem 2.6

Suppose $1 \leq \alpha = \alpha(n) \leq n^{1/3}$, and let $p = p(n) = \alpha^{-1}n^{-1/3}(\log n)^{2/3}$. We show that w.o.p.

$$F([n]_p) \leq c_4(n\log n)^{1/3}\frac{\log n}{\log(\alpha+\log n)} \tag{41}$$

for some absolute constant c_4. To this end, we use Theorem 2.1. Let $\sigma = 3/4$, $s_0 = 2(n \log n)^{1/3}$ and $t = \omega s_0$, where

$$\omega = 11e \frac{\log n}{\log(\alpha + \log n)}, \tag{42}$$

and note that $\omega \geq 2$ for sufficiently large n. Hence, by Theorem 2.1 and the union bound, the probability that $[n]_p$ contains a Sidon set with at least t elements can be bounded as follows:

$$\mathbb{P}\left(F([n]_p) \geq t\right) \leq |\mathcal{Z}_n(t)|p^t \leq n^{3s_0}\left(\frac{44enp}{t^2}\right)^t = n^{3s_0}\left(\frac{44enp}{\omega^2 s_0^2}\right)^{\omega s_0} \leq \left[\left(\frac{11e}{\alpha\omega^2}\right)^{\omega} n^3\right]^{s_0}, \tag{43}$$

where the last inequality follows from $p = \alpha^{-1}n^{-1/3}(\log n)^{2/3}$ and $s_0 = 2(n \log n)^{1/3}$.

For the proof of (41), it suffices to show that the base of the exponential in the right-hand side of (43) is bounded away from 1, that is, whether

$$\left(\frac{11e}{\alpha\omega^2}\right)^{\omega} n^3 < 1 - \varepsilon \tag{44}$$

for some absolute constant $\varepsilon > 0$. Since $\omega \geq 11e$ for sufficiently large n, then we have

$$\left(\frac{\alpha\omega^2}{11e}\right)^{\omega} \geq (\alpha\omega)^{\omega} = \exp\left(\omega \log(\alpha\omega)\right). \tag{45}$$

We claim that

$$2\log(\alpha\omega) \geq \log(\alpha + \log n). \tag{46}$$

Observe that since $\omega \geq 2$, then (46) is trivially satisfied if $\alpha \geq \log n$. On the other hand, if $\alpha \leq \log n$, then $\omega \geq (\log n)/\log\log n$ and hence

$$2\log(\alpha\omega) \geq 2\log\omega \geq 2\log\log n - 2\log\log\log n \geq \log(2\log n) \geq \log(\alpha + \log n).$$

It follows from (42), (45) and (46) that

$$\left(\frac{\alpha\omega^2}{11e}\right)^{\omega} \geq \exp\left(\omega \log(\alpha\omega)\right) \geq \exp\left(5e \log n\right) \geq 2n^3$$

and hence (44) holds, completing the proof of (41).

4.3. Proof of the Upper Bounds in Theorem 2.7

Suppose that $\beta = \beta(n) \geq 1$ and let $p = p(n) = \beta n^{-1/3}(\log n)^{2/3}$. Let $\sigma = 3/4$, $s_0 = 2(n \log n)^{1/3}$ and $t = \omega s_0$ for some $\omega \geq 2$. Similarly as in the proof of the upper bound in Theorem 2.6, see (43), using Theorem 2.1, we estimate

$$\mathbb{P}\left(F([n]_p) \geq t\right) \leq |\mathcal{Z}_n(t)|p^t \leq \left[\left(\frac{11e\beta}{\omega^2}\right)^{\omega} n^3\right]^{s_0}. \tag{47}$$

We split into two cases, depending on the order of magnitude of β.

(*Case I*) If $\beta(n) \le (\log n)^2$, then we let $\alpha = \beta^{-1}(\log n)^2$ and $\omega = (11e\log n)/\log(e\alpha)$ so that $t = \omega s_0 = 22e\sqrt{np}\cdot\sqrt{\alpha}/\log(e\alpha)$. Note that

$$\left(\frac{11e\beta}{\omega^2}\right)^{\omega} = \left(\frac{11e(\log n)^2}{\alpha\omega^2}\right)^{\omega} = \left(\frac{(\log(e\alpha))^2}{11e\alpha}\right)^{11e(\log(e\alpha))^{-1}\log n}. \tag{48}$$

Since the function $f(x) = \left(\frac{x^2}{11e^x}\right)^{1/x} = \frac{1}{e}\left(\frac{x^2}{11}\right)^{1/x}$ is bounded by $e^{\sqrt{4/11}/e-1} = 0.459\cdots$ on the interval $[1,\infty)$, it follows from (48) that (we let $x = \log(e\alpha)$)

$$\left(\frac{11e\beta}{\omega^2}\right)^{\omega} \le \left(\frac{1}{2}\right)^{11e\log n} \le n^{-4},$$

which, together with (47), proves that w.o.p. we have

$$F([n]_p) \le t = c_6\sqrt{np}\cdot\frac{\sqrt{\alpha}}{1+\log\alpha},$$

where c_6 is an absolute constant.

(*Case II*) If $\beta(n) \ge (\log n)^2$, then we let $\omega = 11e\sqrt{\beta}$ so that $t = \omega s_0 = 22e\sqrt{np}$. By (47), we have

$$\mathbb{P}\left(F([n]_p) \ge t\right) \le \left[(11e)^{-11e\sqrt{\beta}}n^3\right]^{s_0} \le \left[(11e)^{-\log n}n^3\right]^{s_0} \le e^{-s_0},$$

which proves that w.o.p. we have

$$F([n]_p) \le t = c_6\sqrt{np},$$

where c_6 is an absolute constant.

5. NONTRIVIAL SOLUTIONS IN RANDOM SETS

5.1. Estimating the Number of Nontrivial Solutions

A *solution* of the equation $x_1 + x_2 = y_1 + y_2$ is a quadruplet $(a_1,a_2,b_1,b_2) \in [n]^4 = [n]\times[n]\times[n]\times[n]$ with $a_1 + a_2 = b_1 + b_2$. A solution (a_1,a_2,b_1,b_2) of $x_1 + x_2 = y_1 + y_2$ is called *trivial* if it is of the form (a_1,a_2,a_1,a_2) or (a_1,a_2,a_2,a_1). Otherwise, it is called a *nontrivial* solution. Let us define a hypergraph and a random variable that will be important for us.

Definition 5.1. Let

$$\mathcal{S} = \{\{a_1,a_2,a_3,a_4\} : (a_1,a_2,a_3,a_4) \text{ is a nontrivial solution of } x_1 + x_2 = y_1 + y_2\}. \tag{49}$$

We think of $\mathcal{S}$ as a hypergraph on the vertex set $[n]$. As usual, for $R \subset [n]$, we let $\mathcal{S}[R]$ denote the subhypergraph of $\mathcal{S}$ induced on R. Let X be the random variable $|\mathcal{S}[[n]_p]|$, that is, the number of hyperedges of $\mathcal{S}$ induced by $[n]_p$.

In Lemma 5.4 below, we give an estimate for X that will be used in the proof of Theorem 2.3 and in the proofs of the lower bounds in Theorems 2.5–2.7.

To estimate X, we have to deal with the issue of 'repeated entries' in a hyperedge $\{a_1, a_2, b_1, b_2\} \in \mathcal{S}$. Indeed, if $\{a_1, a_2, a_3, a_4\} \in \mathcal{S}$, with $a_1 \le a_2 \le a_3 \le a_4$, we may have $a_2 = a_3$, but no other equality can occur. Hence the hypergraph $\mathcal{S}$ has hyperedges of size 4 and 3. Based on this, we make the following definition.

Definition 5.2. For $i = 3$ and 4, let $\mathcal{S}_i$ be the subhypergraph of $\mathcal{S}$ with all the hyperedges of size i. Furthermore, let $X_i := |\mathcal{S}_i[[n]_p]|$.

We clearly have

$$\mathcal{S} = \mathcal{S}_4 \cup \mathcal{S}_3 \quad \text{and} \quad \mathcal{S}_4 \cap \mathcal{S}_3 = \emptyset \tag{50}$$

and hence

$$X = X_4 + X_3. \tag{51}$$

In order to estimate X, we estimate X_4 and X_3 separately.

Lemma 5.3. *Fix $\delta > 0$. The following assertions hold w.o.p.*

(i) If $p \ge n^{-3/4+\delta}$, then $X_4 = n^3p^4(1/12 + o(1))$.
(ii) If $p \gg n^{-1}$, then $X_3 = O(\max\{n^2p^3, n^{3\delta}\})$.

We remark that the constant implicit in the big-O notation in (ii) above is an absolute constant. The proof of Lemma 5.3 is based on a concentration result due to Kim and Vu [20]. We shall introduce the Kim–Vu polynomial concentration result in Section 5.2 and prove Lemma 5.3 in Section 5.3. Assuming Lemma 5.3, we are ready to estimate X.

Lemma 5.4. *Fix $\delta > 0$ and suppose $p \ge n^{-3/4+\delta}$. Then, w.o.p., $X = n^3p^4(1/12 + o(1))$.*

Proof. Let $X = X([n]_p)$ be as defined in Definition 5.1 and recall (51). From the assumption that $p \ge n^{-3/4+\delta}$, we see that the estimates for X_4 and X_3 given in Lemma 5.3(i) and (ii) do hold w.o.p. Since the inequality $np \gg 1$ yields $n^2p^3 \ll n^3p^4$ and we also have $n^{3\delta} \ll n^{4\delta} \le n^3p^4$, because $p \ge n^{-3/4+\delta}$, we infer $\max\{n^2p^3, n^{3\delta}\} \ll n^3p^4$, and hence, w.o.p., $X_3 \ll X_4$. It follows from (51) and the estimate in Lemma 5.3(i) that $X = n^3p^4(1/12 + o(1))$ holds w.o.p. ∎

It now remains to prove Lemma 5.3. We first introduce the main tool we shall use in the proof of that lemma, due to Kim and Vu [20].

5.2. The Kim–Vu Polynomial Concentration Result

Let $\mathcal{H} = (V, E)$ be a hypergraph on the vertex set $V = [n]$. We assume each hyperedge $e \in E(\mathcal{H})$ has a real weight $w(e)$. Let $[n]_p$ be a random subset of $[n]$ obtained by choosing each element $i \in [n]$ independently with probability p and let $\mathcal{H}[[n]_p]$ be the subhypergraph of $\mathcal{H}$ induced on $[n]_p$. Let Y be the sum of the weights of all the hyperedges in $\mathcal{H}[[n]_p]$, i.e., $Y = \sum_{e \in \mathcal{H}[[n]_p]} w(e)$. Kim and Vu obtained a concentration result for the random variable Y. We now proceed to present their result [20] (see also Theorem 7.8.1 in Alon and Spencer [3]).

We start by introducing basic definitions and notation (we follow [3]). Let k be the maximum cardinality of the hyperedges in $\mathcal{H}$. For a set $A \subset [n]$ ($|A| \le k$), let Y_A be the sum of the weights of all the hyperedges in $\mathcal{H}[[n]_p]$ containing A, i.e., $Y_A = \sum_{A \subset e \in \mathcal{H}[[n]_p]} w(e)$. Let $E_A = \mathbb{E}(Y_A \mid A \subset [n]_p)$ be the expectation of Y_A conditioned on the event that A should be contained in $[n]_p$. Let E_i be the maximum value of E_A over all $A \subset [n]$ with $|A| = i$. Note that $E_0 = \mathbb{E}(Y)$. Let $\mu = \mathbb{E}(Y)$ and set

$$E' = \max\{E_i : 1 \le i \le k\} \quad \text{and} \quad E = \max\{E', \mu\}. \tag{52}$$

Theorem 5.5 (Kim–Vu polynomial concentration inequality)**.** *With the above notation, we have, for every* $\lambda > 1$,

$$\mathbb{P}\left[|Y - \mu| > a_k (EE')^{1/2} \lambda^k\right] < 2e^2 e^{-\lambda} n^{k-1},$$

where $a_k = 8^k (k!)^{1/2}$.

5.3. Proof of Lemma 5.3

We prove (i) and (ii) of Lemma 5.3 separately.

Proof of Lemma 5.3(i). We need to show that, for $p \ge n^{-3/4+\delta}$, where $\delta > 0$ is fixed, we have $X_4 = n^3 p^4 (1/12 + o(1))$ w.o.p. We first estimate the expectation $\mu(X_4)$ of X_4.

Suppose $\{i, j, k, l\} \in \mathcal{S}_4$ with $0 \le i < j < k < l \le n-1$. Note that $i + l = j + k$. Let us fix $0 \le i \le n-1$. If $j \ge (n+i)/2$, then $l = j + k - i > 2j - i \ge n + i - i = n$, which contradicts $l \le n-1$. Hence we have $i < j < (n+i)/2$. For fixed i and j, if $k > n+i-j-1$, then $l = j+k-i > n-1$, which contradicts $l \le n-1$. Therefore we have $j < k \le n+i-j-1$. Once i, j and k are chosen, the value of l is determined by the condition $i + l = j + k$. Consequently,

$$\begin{aligned} |\mathcal{S}_4| &\sim \sum_{i=0}^{n-1} \sum_{j=i}^{(n+i)/2} \sum_{k=j}^{n+i-j-1} 1 = \sum_{i=0}^{n-1} \sum_{j=i}^{(n+i)/2} (n + i - 2j) \\ &\sim n^3 \int_0^1 \int_x^{(1+x)/2} (1 + x - 2y) \mathrm{d}y \mathrm{d}x \sim \frac{1}{12} n^3. \end{aligned}$$

Hence

$$\mu(X_4) = |\mathcal{S}_4| p^4 = \left(\frac{1}{12} + o(1)\right) n^3 p^4. \tag{53}$$

Next we apply Theorem 5.5 to prove that X_4 is concentrated around its expectation $\mu(X_4)$. To this end, we compute the quantities E_i ($1 \le i \le 4$) and E' and E defined in (52). We first estimate E_1. For $a \in [n]$, consider the quantity $E_{\{a\}}$. The number of hyperedges in $\mathcal{S}_4$ containing a is $O(n^2)$ and the probability that one such hyperedge is in $[n]_p$, conditioned on $a \in [n]_p$, is p^3. We conclude that, for any $a \in [n]$, we have $E_{\{a\}} = O(n^2 p^3)$. Consequently, $E_1 = \max\{E_A : |A| = 1\} = O(n^2 p^3)$. A similar argument gives that $E_i = \max\{E_A : |A| = i\} = O(n^{3-i} p^{4-i})$ for all $1 \le i < 4$. Therefore, since $np \gg 1$, we have $E_i = O(n^2 p^3)$ for all $1 \le i < 4$. Also, clearly, $E_4 = \max\{E_A : |A| = 4\} = 1$. Thus

$$E' = \max\{E_i : 1 \le i \le 4\} = O(\max\{n^2 p^3, 1\}), \tag{54}$$

and $E = \max\{E', \mu(X_4)\} = O(\max\{n^2p^3, 1, n^3p^4\})$. Since $p \geq n^{-3/4+\delta} > n^{-3/4}$, we have

$$E = O(n^3p^4). \tag{55}$$

In view of (54) and (55), a simple computation implies the following:

(*Case I*) If $n^{-3/4+\delta} \leq p \leq n^{-2/3}$, then

$$E' = O(1) \quad \text{and} \quad E = O(n^3p^4). \tag{56}$$

(*Case II*) If $p \geq n^{-2/3}$, then

$$E' = O(n^2p^3) \quad \text{and} \quad E = O(n^3p^4). \tag{57}$$

We now estimate X_4 for each case separately.
(*Case I*) *Suppose* $n^{-3/4+\delta} \leq p \leq n^{-2/3}$. In this case, (56) implies that

$$(EE')^{1/2} = O(n^3p^4 \cdot 1)^{1/2} = O(n^3p^4)^{1/2}. \tag{58}$$

Set $\lambda = (n^3p^4)^{1/12}$. By the assumption $p \geq n^{-3/4+\delta}$, we have

$$\lambda = (n^3p^4)^{1/12} \geq n^{\delta/3}. \tag{59}$$

Also $n^3p^4 \geq n^{4\delta} \gg 1$, and hence combining (58) and $\lambda = (n^3p^4)^{1/12}$ implies that

$$(EE')^{1/2}\lambda^4 = O(n^3p^4)^{1/2}(n^3p^4)^{1/3} = O(n^3p^4)^{5/6} = o(n^3p^4). \tag{60}$$

Theorem 5.5 together with (59) then yields that

$$\mathbb{P}\left[|X_4 - \mu(X_4)| > a_4(EE')^{1/2}\lambda^4\right] < 2e^2e^{-\lambda}n^3 \leq 2e^2e^{-n^{\delta/3}}n^3,$$

where $a_4 = 8^4(4!)^{1/2}$. Given (60), we have that w.o.p.

$$X_4 = \mu(X_4) + o(n^3p^4). \tag{61}$$

(*Case II*) Suppose $p \geq n^{-2/3}$. In this case, (57) yields that

$$(EE')^{1/2} = O(n^3p^4n^2p^3)^{1/2} = O\left(\frac{n^3p^4}{(np)^{1/2}}\right). \tag{62}$$

Set $\lambda = (np)^{1/12}$. By the assumption $p \geq n^{-2/3}$,

$$\lambda \geq (n^{1/3})^{1/12} = n^{1/36}. \tag{63}$$

Since $np \gg 1$, combining (62) and $\lambda = (np)^{1/12}$ implies that

$$(EE')^{1/2}\lambda^4 = O\left(\frac{n^3p^4}{(np)^{1/2}}\right)(np)^{1/3} = O\left(\frac{n^3p^4}{(np)^{1/6}}\right) = o(n^3p^4). \tag{64}$$

Theorem 5.5 together with (63) then yields that

$$\mathbb{P}\left[|X_4 - \mu(X_4)| > a_4(EE')^{1/2}\lambda^4\right] < 2e^2e^{-\lambda}n^3 \leq 2e^2e^{-n^{1/36}}n^3,$$

where $a_4 = 8^4(4!)^{1/2}$. Given (64), we have that w.o.p.

$$X_4 = \mu(X_4) + o(n^3p^4). \tag{65}$$

In view of (53), it follows from (61) and (65) that, for $p \geq n^{-3/4+\delta}$, we have $X_4 = n^3p^4(1/12 + o(1))$ w.o.p. This completes the proof of (i) of Lemma 5.3. ∎

Proof of Lemma 5.3(ii). Fix $\delta > 0$. We show that, w.o.p., $X_3 = O(\max\{n^2p^3, n^{3\delta}\})$ for $p \gg n^{-1}$. First we estimate the expectation $\mu(X_3)$ of X_3. Since $|\mathcal{S}_3| = O(n^2)$, we have

$$\mu(X_3) = O(n^2p^3). \tag{66}$$

Next, we prove a concentration result for X_3 applying Theorem 5.5. To this end, we estimate the quantities E_i $(1 \leq i \leq 3)$. As in the proof of Lemma 5.3(i), one may check that $E' = \max_{1\leq i\leq 3} E_i = O(\max\{np^2, p, 1\})$ and hence $E = \max\{E', \mu(X_3)\} = O(\max\{np^2, p, 1, n^2p^3\})$. By the assumption $np \gg 1$, we infer

$$E' = O(\max\{np^2, 1\}) \quad \text{and} \quad E = O(\max\{n^2p^3, 1\}). \tag{67}$$

Based on (67), we consider the cases $p \geq n^{-2/3+\delta}$ and $n^{-1} \ll p \leq n^{-2/3+\delta}$ separately.

We first suppose $p \geq n^{-2/3+\delta}$. From (67), we have $E' = O(\max\{np^2, 1\})$ and $E = O(n^2p^3)$. A proof similar to the proofs of (61) and (65) shows that, for $p \geq n^{-2/3+\delta}$, w.o.p., $X_3 = \mu(X_3) + o(n^2p^3)$. This together with (66) implies that for $p \geq n^{-2/3+\delta}$, w.o.p.,

$$X_3 = O(n^2p^3). \tag{68}$$

We now suppose $n^{-1} \ll p \leq n^{-2/3+\delta}$. In this case, (67) yields that $E' = O(1)$ and $E = O(n^{3\delta})$ and hence, setting $\lambda = n^{\delta/2}$, we have

$$(EE')^{1/2}\lambda^3 = O(n^{(3/2)\delta})n^{(3/2)\delta} = O(n^{3\delta}). \tag{69}$$

Theorem 5.5 with $\lambda = n^{\delta/2}$ yields

$$\mathbb{P}\left[|X_3 - \mu(X_3)| > a_3(EE')^{1/2}\lambda^3\right] < 2e^2e^{-\lambda}n^2 \leq 2e^2e^{-n^{\delta/2}}n^2, \tag{70}$$

where $a_3 = 8^3(3!)^{1/2}$. Inequality (70) together with (69) implies that, for $n^{-1} \ll p \leq n^{-2/3+\delta}$, w.o.p., $X_3 = \mu(X_3) + O(n^{3\delta})$. Since, under the assumption $p \leq n^{-2/3+\delta}$, we have $\mu(X_3) = O(n^2p^3) = O(n^{3\delta})$, we infer that, for $n^{-1} \ll p \leq n^{-2/3+\delta}$, w.o.p.,

$$X_3 = O(n^{3\delta}). \tag{71}$$

Combining (68) and (71) completes the proof of (ii) of Lemma 5.3. ∎

6. PROOF OF THEOREM 2.3

6.1. Theorem 2.3 for Smaller $p = p(n)$

We first consider the case in which $n^{-1} \ll p \ll n^{-2/3}$.

Proof of (8) *in Theorem 2.3.* Suppose $n^{-1} \ll p \ll n^{-2/3}$. We show that (8) holds almost surely, using the usual deletion method. Let $\mathcal{S}$, $\mathcal{S}[[n]_p]$ and X be as in Definition 5.1.

If we delete one vertex from each hyperedge in $\mathcal{S}[[n]_p]$, the remaining vertex set is an independent set of $\mathcal{S}[[n]_p]$, and hence it is a Sidon set contained in $[n]_p$. Consequently, $F([n]_p) \geq |[n]_p| - |\mathcal{S}[[n]_p]| = |[n]_p| - X$. Since trivially $F([n]_p) \leq |[n]_p|$, we have $|[n]_p| - X \leq F([n]_p) \leq |[n]_p|$. Note that the Chernoff bound gives that, for $p \gg n^{-1}$, we almost surely have $|[n]_p| = np + o(np)$. Therefore, in order to show (8), it only remains to show that $X = o(np)$ almost surely. Recall that X_i is the number of edges of cardinality i in $\mathcal{S}[[n]_p]$ ($i \in \{3, 4\}$), and that $X = X_3 + X_4$ (see Definition 5.2 and (51)). Equations (53) and (66), together with $n^{-1} \ll p \ll n^{-2/3}$, imply that $\mathbb{E}(X) = \Theta(n^3p^4) + O(n^2p^3) = \Theta(n^3p^4) = o(np)$. Hence Markov's inequality gives that we almost surely have $X = o(np)$, and our result follows. ∎

6.2. Theorem 2.3 for Larger $p = p(n)$

We now consider the wider range $n^{-1} \ll p \leq 2n^{-2/3}$.

Proof of (9) *in Theorem 2.3.* We have already shown that, if $n^{-1} \ll p \ll n^{-2/3}$, then $F([n]_p) = (1 + o(1))np$ holds almost surely. Therefore, it suffices to show that (9) holds if, e.g., $n^{-2/3}/\log n \leq p \leq 2n^{-2/3}$. We proceed as in the proof of (8), given in Section 6.1 above. We have already observed that $|[n]_p| = np(1 + o(1))$ almost surely as long as $p \gg n^{-1}$, and therefore $F([n]_p) \leq np(1 + o(1))$ almost surely in this range of p. It now suffices to recall that $F([n]_p) \geq |[n]_p| - X$ and to prove that, almost surely, we have $X \leq (2/3 + o(1))np$ if $n^{-2/3}/\log n \leq p \leq 2n^{-2/3}$. But with this assumption on p, Lemma 5.4 tells us that, w.o.p.,

$$X = \frac{1}{12}n^3p^4 + o(n^3p^4) = \frac{1}{12}n^3p^4 + o(np) \leq \left(\frac{2}{3} + o(1)\right)np, \tag{72}$$

as required. ∎

7. THE LOWER BOUNDS IN THEOREMS 2.5–2.7

Let us first state a simple monotonicity result (see, e.g., [19, Lemma 1.10]) that will be used a few times in this section.

Fact 7.1. *Let $p = p(n)$ and $q = q(n)$ be such that $0 \leq p < q \leq 1$, and let $a = a(n) > 0$ and $b = b(n) > 0$ be functions of n.*

(i) If $F([n]_p) \geq a$ holds w.o.p., then $F([n]_q) \geq a$ holds w.o.p.
(ii) If $F([n]_q) \leq b$ holds w.o.p., then $F([n]_p) \leq b$ holds w.o.p.

Statements (i) and (ii) in Fact 7.1 are, in fact, equivalent. We state them both explicitly just for convenience.

7.1. Proofs of the Lower Bounds in Theorems 2.5 and 2.6

The lower bounds in Theorems 2.5 and 2.6 rely on a result on independent sets in hypergraphs. Before stating the relevant result, we introduce some definitions. A hypergraph is called *simple* if any two of its hyperedges share at most one vertex. A hypergraph is

r-uniform if all its hyperedges have cardinality r. We shall use the following extension of a celebrated result due to Ajtai, Komlós, Pintz, Spencer and Szemerédi [1], obtained by Duke, Lefmann and the third author [10].

Lemma 7.2. *Let $\mathcal{H}$ be a simple r-uniform hypergraph, $r \geq 3$, with N vertices and average degree at most t^{r-1} for some t. Then $\mathcal{H}$ has an independent set of size at least*

$$c\,\frac{(\log t)^{1/(r-1)}}{t}\,N, \tag{73}$$

where $c = c(r)$ is a positive constant that depends only on r.

We now briefly discuss how to obtain a lower bound on $F([n]_p)$ using Lemma 7.2. Let $\mathcal{S}[[n]_p]$ be the hypergraph in Definition 5.1. Since an independent set of $\mathcal{S}[[n]_p]$ is a Sidon set contained in $[n]_p$, independent sets in $\mathcal{S}[[n]_p]$ give lower bounds for $F([n]_p)$. To apply Lemma 7.2, we shall obtain a simple 4-uniform subhypergraph $\mathcal{S}^*$ of $\mathcal{S}[[n]_p]$ by deleting suitable vertices from $\mathcal{S}[[n]_p]$. Lemma 7.2 will then tell us that $\mathcal{S}^*$ has a suitably large independent set, and this will yield our lower bound on $F([n]_p)$. In fact, we obtain the following result.

Lemma 7.3. *There is an absolute constant $d > 0$ such that, for $p \geq 2n^{-2/3}$, w.o.p. $F([n]_p) \geq d\left(n\log(n^2p^3)\right)^{1/3}$ holds.*

Lemma 7.3 easily implies the lower bounds in Theorems 2.5 and 2.6. The proof of Lemma 7.3 will be given in Section 7.2.

7.2. Proof of Lemma 7.3

In Lemma 7.4 below, we prove Lemma 7.3 for a narrower range of p. We shall then invoke monotonicity (Fact 7.1) to obtain Lemma 7.3 in full.

Lemma 7.4. *There is an absolute constant $d > 0$ such that, for $2n^{-2/3} \leq p \ll n^{-2/3+1/15}$, we have $F([n]_p) \geq d(n\log n^2p^3)^{1/3}$ w.o.p.*

Proof. Let $\mathcal{S}[[n]_p]$, $\mathcal{S}_i[[n]_p]$, X and X_i be as in Definitions 5.1 and 5.2. Recall that the size of an independent set of $\mathcal{S}[[n]_p]$ gives a lower bound on $F([n]_p)$.

We wish to apply Lemma 7.2. However, since $\mathcal{S}[[n]_p]$ may be neither simple nor uniform, we consider a suitable *induced* subhypergraph $\mathcal{S}^* \subset \mathcal{S}[[n]_p]$, as discussed just after the statement of Lemma 7.2. We have $\mathcal{S}[[n]_p] = \mathcal{S}_3[[n]_p] \cup \mathcal{S}_4[[n]_p]$. Let $\widetilde{\mathcal{S}}_4$ be the set of all hyperedges in $\mathcal{S}_4[[n]_p]$ that share at least two vertices with some other hyperedge in $\mathcal{S}_4[[n]_p]$. If we delete one vertex from each hyperedge of $\mathcal{S}_3[[n]_p] \cup \widetilde{\mathcal{S}}_4$, the remaining induced subhypergraph $\mathcal{S}^*$ of $\mathcal{S}[[n]_p]$ is both simple and 4-uniform. To apply Lemma 7.2 to $\mathcal{S}^*$, we now estimate $|V(\mathcal{S}^*)|$ and the average degree of $\mathcal{S}^*$.

First we consider $|V(\mathcal{S}^*)|$. Note that $|[n]_p| - X_3 - |\widetilde{\mathcal{S}}_4| = |[n]_p| - |\mathcal{S}_3[[n]_p]| - |\widetilde{\mathcal{S}}_4| \leq |V(\mathcal{S}^*)| \leq |[n]_p|$. We shall show the following two facts.

Fact 7.5. *Fix $\delta > 0$ and suppose $n^{-1+\delta} \ll p \ll n^{-1/2}$. We have, w.o.p., $X_3 = o(np)$.*

Fact 7.6. *Fix $\delta > 0$ and suppose $n^{-1+\delta} \ll p \ll n^{-2/3+1/15}$. We have, w.o.p., $|\widetilde{\mathcal{S}}_4| = o(np)$.*

Since the Chernoff bound gives that $|[n]_p| = np + o(np)$ w.o.p. for $p \gg (\log n)/n$, Facts 7.5 and 7.6 imply that, w.o.p., we have

$$|V(\mathcal{S}^*)| = np(1 + o(1)). \tag{74}$$

Next we consider the average degree of $\mathcal{S}^*$. Owing to $\mathcal{S}^* \subset \mathcal{S}[[n]_p]$, (74) and Lemma 5.4, the average degree $4|\mathcal{S}^*|/|V(\mathcal{S}^*)|$ of $\mathcal{S}^*$ is such that, w.o.p., $4|\mathcal{S}^*|/|V(\mathcal{S}^*)| \le 4X/|V(\mathcal{S}^*)| \le n^2p^3$.

We now are ready to apply Lemma 7.2. In view of our average degree estimate above, we set $t = (n^2p^3)^{1/3}$. Given (74), Lemma 7.2 implies that, w.o.p., the hypergraph $\mathcal{S}^*$, and thus $\mathcal{S}[[n]_p]$, has an independent set of size

$$c\,\frac{(\log t)^{1/3}}{t}|V(\mathcal{S}^*)| \ge c\,\frac{[(1/3)\log(n^2p^3)]^{1/3}}{(n^2p^3)^{1/3}}np(1 + o(1)) \ge d(n\log(n^2p^3))^{1/3}, \tag{75}$$

for, say, $d = c/2$. This completes the proof of Lemma 7.4. ∎

In order to finish the proof of Lemma 7.4, it remains to prove Facts 7.5 and 7.6.

Proof of Fact 7.5. Lemma 5.3(ii) tells us that, w.o.p., $X_3 = O(\max\{n^2p^3, n^\delta\})$. From the assumption $n^{-1+\delta} \ll p \ll n^{-1/2}$, we have both $n^2p^3 \ll np$ and $n^\delta \ll np$, whence, w.o.p., $X_3 = o(np)$. ∎

Proof of Fact 7.6. We give a sketch of the proof. Let $\mathcal{P}$ be the family of the pairs $\{E_1, E_2\}$ of distinct members E_1 and E_2 of $\mathcal{S}_4[[n]_p]$ with $|E_1 \cap E_2| \ge 2$. Observe that

$$|\widetilde{\mathcal{S}}_4| \le 2|\mathcal{P}|. \tag{76}$$

An argument similar to one in the proof of Lemma 5.3(ii), based on the Kim–Vu polynomial concentration result, tells us that $|\mathcal{P}| = O(\max\{\mathbb{E}[|\mathcal{P}|], n^\delta\}) = O(\max\{n^4p^6, n^\delta\})$ holds w.o.p. From the assumption $n^{-1+\delta} \ll p \ll n^{-2/3+1/15} = n^{-3/5}$, we have both $n^4p^6 \ll np$ and $n^\delta \ll np$, and hence $|\mathcal{P}| = o(np)$ holds w.o.p. Given (76), we have, w.o.p., $|\widetilde{\mathcal{S}}_4| = o(np)$. ∎

In order to establish Lemma 7.3, we need to expand the range of p in Lemma 7.4 from $2n^{-2/3} \le p \ll n^{-2/3+1/15} = n^{-3/5}$ to $p \ge 2n^{-2/3}$.

Proof of Lemma 7.3. To complement the range of p covered by Lemma 7.4, it is enough to show that, say, for $p \ge n^{-2/3+1/16}$, we have, w.o.p., $F([n]_p) \ge d'\left(n\log(n^2p^3)\right)^{1/3}$ for some absolute constant $d' > 0$. Lemma 7.4 implies that, for $p = n^{-2/3+1/16}$, we have, w.o.p.,

$$\begin{aligned} F([n]_p) &\ge d[n\log(n^2n^{-2+3/16})]^{1/3} = d[n\log(n^{3/16})]^{1/3} = d[n(3/16)\log n]^{1/3} \\ &> d(1/16)^{1/3}[n(2\log n)]^{1/3} = d'[n\log n^2]^{1/3}, \end{aligned}$$

where $d' = d(1/16)^{1/3}$. By Fact 7.1, we infer that, for $p \ge n^{-2/3+1/16}$, we have, w.o.p., $F([n]_p) \ge d'[n\log n^2]^{1/3} \ge d'[n\log(n^2p^3)]^{1/3}$, completing the proof of Lemma 7.3. ∎

7.3. Proof of the Lower Bound in Theorem 2.7

For larger $p = p(n)$, it turns out that, instead of using Lemma 7.2, it is better to make use of the fact that $[n]$ contains a Sidon set of cardinality $(1 + o(1))\sqrt{n}$ (see Section 1). An immediate use of this fact gives the lower bound $(1 + o(1))p\sqrt{n}$, but one can, in fact, do better. The following is a particular case of a very general theorem of Komlós, Sulyok and Szemerédi [25].

Lemma 7.7. *There is an absolute constant $d > 0$ such that, for every sufficiently large m and every set of integers A with $|A| = m$, we have*

$$F(A) \geq d \cdot F([m]).$$

Since the Chernoff bound gives that, for $p \gg 1/n$, we almost surely have $|[n]_p| = (1 + o(1))np$, Lemma 7.7 together with $F([m]) \geq (1 + o(1))\sqrt{m}$ gives the lower bound in Theorem 2.7. Clearly, to have this result with 'w.o.p.', it suffices to assume $p \gg (\log n)/n$.

There is an alternative, simple proof of the lower bound in Theorem 2.7, based on the following lemma.

Lemma 7.8. *If $(\log n)^2/n \ll p \leq 1/3$, then, w.o.p.,*

$$F([n]_p) \geq \left(\frac{1}{3\sqrt{2}} + o(1)\right)\sqrt{np}. \tag{77}$$

Combining Lemma 7.8 and Fact 7.1 implies that, for $p \gg (\log n)^2/n$, we have, w.o.p., $F([n]_p) \geq (1/3\sqrt{6} + o(1))\sqrt{np}$.

Proof of Lemma 7.8. Let $(\log n)^2/n \ll p \leq 1/3$. We shall show that (77) holds w.o.p. We define a partition of $[n] = \{0, \ldots, n-1\}$ into equal length intervals, and consider a family of intervals in the partition satisfying the property that, if we choose an arbitrary element from each interval, the set of chosen elements forms a Sidon set. We shall choose the length of the intervals so that $[n]_p$ will intersect each interval in a constant number of elements on average. A simple analysis of this construction yields that (77) holds w.o.p. The details are as follows.

Let $\mathcal{I} = \{I_i : 0 \leq i < \lceil n/x \rceil\}$ be the partition of $[n]$ into consecutive intervals with $x = \lfloor 1/p \rfloor$ elements each. More precisely, let $I_i = [xi, x(i+1) - 1] \cap [n]$ for all $0 \leq i < \lceil n/x \rceil$. In what follows, we ignore $I_{\lceil n/x \rceil - 1}$ if this interval has fewer than x elements. Let $\mathcal{I}_{\text{even}} = \{I_0, I_2, I_4, \ldots\} \subset \mathcal{I}$ be the set of all intervals with even indices and let $y = |\mathcal{I}_{\text{even}}|$. Note that $y \geq (1/2)\lfloor n/x \rfloor - 1 \geq (1/2)\lfloor np \rfloor - 1 = (1/2 + o(1))np$. By the Chowla–Erdős result [8, 11], there exists a Sidon subset S of $[y]$ with

$$|S| = (1 + o(1))\sqrt{y} = \left(\frac{1}{\sqrt{2}} + o(1)\right)\sqrt{np}. \tag{78}$$

We "identify" $[y]$ and $\mathcal{I}_{\text{even}}$ by the bijection $i \mapsto I_{2i}$. Let $\{a_i : i \in S\}$ be a set of integers with $a_i \in I_{2i}$ for all $i \in S$. We claim that $\{a_i : i \in S\}$ is a Sidon set. Suppose $a_{i_1} + a_{i_2} = a_{j_1} + a_{j_2}$, where i_1, i_2, j_1 and $j_2 \in S$. Observe that

$$a_{i_1} + a_{i_2} \in I_{2i_1+2i_2} \cup I_{2i_1+2i_2+1} \quad \text{and} \quad a_{j_1} + a_{j_2} \in I_{2j_1+2j_2} \cup I_{2j_1+2j_2+1}, \tag{79}$$

which, together with the assumption that $a_{i_1} + a_{i_2} = a_{j_1} + a_{j_2}$, implies that $i_1 + i_2 = j_1 + j_2$. Since S is a Sidon set, we have $\{i_1, i_2\} = \{j_1, j_2\}$, whence $\{a_{i_1}, a_{i_2}\} = \{a_{j_1}, a_{j_2}\}$. This shows that $\{a_i : i \in S\}$ is indeed a Sidon set.

We now consider a random set $[n]_p$. An interval I_{2i} ($i \in S$) is said to be *occupied* if I_{2i} contains at least one element of $[n]_p$. Let $\mathcal{I}_{\mathrm{occ}}$ be the family of occupied intervals. By the above claim, we have $F([n]_p) \geq |\mathcal{I}_{\mathrm{occ}}|$. Let us estimate $|\mathcal{I}_{\mathrm{occ}}|$. Note that each interval I_{2i} ($i \in S$) is independently occupied with probability

$$\tilde{p} = 1 - (1-p)^x = 1 - (1-p)^{\lfloor 1/p \rfloor} \geq 1 - e^{-p(1/p-1)} \geq 1 - e^{-1+p} \geq 1 - e^{-2/3} > 1/3, \tag{80}$$

where the third inequality follows from the assumption $p \leq 1/3$. Thus, under the assumption $(\log n)^2/n \ll p \leq 1/3$, the Chernoff bound, (78) and (80) give that, w.o.p.,

$$\begin{aligned} |\mathcal{I}_{\mathrm{occ}}| &= (1 + o(1))\mathbb{E}(|\mathcal{I}_{\mathrm{occ}}|) = (1 + o(1))|S|\tilde{p} \geq \left(\frac{1}{\sqrt{2}} + o(1)\right)\sqrt{np} \cdot \frac{1}{3} \\ &= \left(\frac{1}{3\sqrt{2}} + o(1)\right)\sqrt{np}. \end{aligned}$$

To complete the proof of Lemma 7.8, it now suffices to recall that $F([n]_p) \geq |\mathcal{I}_{\mathrm{occ}}|$. ∎

ACKNOWLEDGMENTS

The fourth author is indebted to Tomasz Schoen for drawing his attention to the problem of counting Sidon sets. The second author thanks Jaigyoung Choe for his support throughout a visiting program at Korea Institute for Advanced Study (KIAS). The authors thank the referees for their comments and suggestions.

REFERENCES

[1] M. Ajtai, J. Komlós, J. Pintz, J. Spencer, and E. Szemerédi, Extremal uncrowded hypergraphs, J Comb Theory Ser A 32 (1982), 321–335.

[2] N. Alon, J. Balogh, R. Morris, and W. Samotij, Counting sum-free sets in Abelian groups, Israel J Math. (in press).

[3] N. Alon and J. H. Spencer, The probabilistic method, 2nd edition, Wiley-Interscience series in discrete mathematics and optimization, Wiley-Interscience [Wiley], New York, 2000. With an appendix on the life and work of Paul Erdős.

[4] J. Balogh and W. Samotij, The number of $K_{m,m}$-free graphs, Combinatorica 31 (2011), 131–150.

[5] J. Balogh and W. Samotij, The number of $K_{s,t}$-free graphs, J Lond Math Soc 83 (2011), 368–388.

[6] B. Bollobás, Random graphs, Academic Press [Harcourt Brace Jovanovich Publishers], London, 1985.

[7] P. J. Cameron and P. Erdős, On the number of sets of integers with various properties, Proceedings of the First Conference of the Canadian Number Theory Association held in Banff, Alberta, April 17–27, 1988, R. A. Mollin (Editor), Walter de Gruyter & Co., Berlin, 1990, pp. 61–79.

[8] 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.

[9] D. Conlon and W. T. Gowers, Combinatorial theorems in sparse random sets. (Submitted for publication), 2010.

[10] R. A. Duke, H. Lefmann, and V. Rödl, On uncrowded hypergraphs, Random Struct Algorithms 6 (1995), 209–212.

[11] P. Erdős, On a problem of Sidon in additive number theory and on some related problems. Addendum, J Lond Math Soc 19 (1944), 208.

[12] P. Erdős, On a problem of Sidon in additive number theory, Acta Sci Math Szeged 15 (1954), 255–259.

[13] P. Erdős, Problems and results in additive number theory, In Colloque sur la Théorie des Nombres, Bruxelles, 1955, George Thone, Liège, 1956, pp. 127–137.

[14] P. Erdős and P. Turán, On a problem of Sidon in additive number theory, and on some related problems, J Lond Math Soc 16 (1941), 212–215.

[15] Z. Füredi, Random Ramsey graphs for the four-cycle, Discrete Math 126 (1994), 407–410.

[16] B. Green and T. Tao, The primes contain arbitrarily long arithmetic progressions, Ann Math 167 (2008), 481–547.

[17] H. Halberstam and K. F. Roth, Sequences, 2nd edition, Springer-Verlag, New York, 1983.

[18] S. Janson, T. Łuczak, and A. Ruciński, An exponential bound for the probability of nonexistence of a specified subgraph in a random graph, In Random graphs '87 (Poznań, 1987), Wiley, Chichester, 1990, pp. 73–87.

[19] S. Janson, T. Łuczak, and A. Ruciński, Random graphs, Wiley-Interscience, New York, 2000.

[20] J. H. Kim and V. H. Vu, Concentration of multivariate polynomials and its applications, Combinatorica 20 (2000), 417–434.

[21] D. J. Kleitman and K. J. Winston, On the number of graphs without 4-cycles, Discrete Math 41 (1982), 167–172.

[22] Y. Kohayakawa, Szemerédi's regularity lemma for sparse graphs, In Foundations of computational mathematics (Rio de Janeiro, 1997), F. Cucker and M. Shub (Editors), Springer, Berlin, 1997, pp. 216–230.

[23] Y. Kohayakawa, B. Kreuter, and A. Steger, An extremal problem for random graphs and the number of graphs with large even-girth, Combinatorica 18 (1998), 101–120.

[24] Y. Kohayakawa, T. Łuczak, and V. Rödl, Arithmetic progressions of length three in subsets of a random set, Acta Arith 75 (1996), 133–163.

[25] J. Komlós, M. Sulyok, and E. Szemeredi, Linear problems in combinatorial number theory, Acta Math Acad Sci Hungar 26 (1975), 113–121.

[26] K. O'Bryant, A complete annotated bibliography of work related to Sidon sequences, Electron J Comb (2004). Dynamic surveys 11, 39 pp. (electronic).

[27] O. Reingold, L. Trevisan, M. Tulsiani, and S. P. Vadhan, Dense subsets of pseudorandom sets, In FOCS (2008), 76–85.

[28] O. Reingold, L. Trevisan, M. Tulsiani, and S. P. Vadhan, Dense subsets of pseudorandom sets, Electronic Colloquium on Computational Complexity (ECCC) TR08-045 (2008), 33pp (electronic).

[29] K. F. Roth, On certain sets of integers, J Lond Math Soc 28 (1953), 104–109.

[30] D. Saxton and A. Thomason, Hypergraph containers. arXiv:1204.6595, 2012.

[31] M. Schacht, Extremal results for random discrete structures. (Submitted for publication), 2009.

[32] J. Singer, A theorem in finite projective geometry and some applications to number theory, Trans Am Math Soc 43 (1938), 377–385.

[33] E. Szemerédi, On sets of integers containing no k elements in arithmetic progression, Acta Arith 27 (1975), 199–245. Collection of articles in memory of Juriĭ Vladimirovič Linnik.

[34] T. Tao and V. Vu, Additive combinatorics, In Cambridge studies in advanced mathematics, Vol. 105, Cambridge University Press, Cambridge, 2006.

[35] L. Trevisan, Guest column: Additive combinatorics and theoretical computer science, SIGACT News 40 (2009), 50–66.