## Distribution agreement

In presenting this dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to copy from, or to publish, this dissertation may be granted by the professor under whose direction it was written, or, in his absence, by the Dean of the Graduate School when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from, or publication of, this dissertation which involves potential financial gain will not be allowed without written permission.

_____

Sangjune Lee

# Results on Sidon and $B_h$ Sequences

By

Sangjune Lee
Doctor of Philosophy

Department of Mathematics and Computer Science

---

Vojtěch Rödl, Ph.D.
Advisor

---

Dwight Duffus, Ph.D.
Committee Member

---

Ronald J. Gould, Ph.D.
Committee Member

Accepted:

---

Lisa A. Tedesco, Ph.D.
Dean of the Graduate School

---

Date

# Results on Sidon and $B_h$ Sequences

By

Sangjune Lee
B.S. Mathematics, Seoul National University, 2002

Advisor: Vojtěch Rödl, Ph.D.

An Abstract of
A dissertation submitted to the Faculty of the Graduate
School of Emory University in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Department of Mathematics and Computer Science

2012

**Abstract**

# Results on Sidon and $B_h$ Sequences

Sangjune Lee

A set $A$ of non-negative integers is a *Sidon set* if all the sums $a_1 + a_2$, with $a_1 \leq a_2$ and $a_1$, $a_2 \in A$, are distinct. In this dissertation, we deal with results on the number of Sidon sets in $[n] = \{0, 1, \cdots, n-1\}$ and the maximum size of Sidon sets in sparse random subsets of $[n]$ or $\mathbb{N}$ (the set of natural numbers). We also consider a natural generalization of Sidon sets called $B_h$-sets with $h \geq 2$. A set $A$ of non-negative integers is called a $B_h$-set if all the sums $a_1 + a_2 + \cdots + a_h$, with $a_1 \leq a_2 \leq \cdots \leq a_h$ and $a_i \in A$, are distinct.

The first question in this dissertation was suggested by Cameron–Erdős in 1990. They proposed the problem of estimating the number of Sidon sets contained in $[n]$. We obtain an upper bound $2^{c\sqrt{n}}$ on the number of Sidon sets which is sharp up to a constant factor in the exponent when compared to the previous lower bound $2^{(1+o(1))\sqrt{n}}$.

Next, we investigate the maximum size of Sidon sets contained in sparse random sets $R \subset [n]$. Let $R = [n]_m$ be a uniformly chosen, random $m$-element subset of $[n]$. Let $F([n]_m) = \max\{|S| \colon S \subset [n]_m \text{ is Sidon}\}$. Fix a constant $0 \leq a \leq 1$ and suppose $m = (1 + o(1))n^a$. We show that there is a constant $b = b(a)$ for which

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

almost surely and we obtain what $b = b(a)$ is. Surprisingly, between two points $a = 1/3$ and $a = 2/3$, the function $b = b(a)$ is constant.

Next, we deal with infinite Sidon sets in sparse random subsets of $\mathbb{N}$. Fix $0 < \delta \leq 1$, and let $R = R_\delta$ be the set obtained by choosing each element $i \subset \mathbb{N}$ independently with probability $i^{-1+\delta}$. We show that for every $0 < \delta \leq 2/3$ there exists a constant $c = c(\delta)$ such that a random set $R$ satisfies the following with probability 1:

- Every Sidon set $S \subset R$ satisfies that $|S \cap [n]| \leq n^{c+o(1)}$ for every sufficiently large $n$.
- There exists a large Sidon set $S \subset R$ such that $|S \cap [n]| \geq n^{c+o(1)}$ for every sufficiently large $n$.

# Results on Sidon and $B_h$ Sequences

By

Sangjune Lee
B.S. Mathematics, Seoul National University, 2002

Advisor: Vojtěch Rödl, Ph.D.

A dissertation submitted to the Faculty of the Graduate
School of Emory University in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy

Department of Mathematics and Computer Science

2012

**Acknowledgments**

I thank God and his only Son, Jesus.

I am heartily thankful to my advisor, Vojtěch Rödl, who has shared his knowledge with me and has taught me how to research. His supervision and support from the preliminary to the concluding level enabled me to complete this dissertation.

I owe my deepest gratitude to Yoshiharu Kohayakawa for his valuable discussions and collaborations. Without the help of Yoshi, this work would not have been possible. Next, I am grateful to committee members, Ronald Gould and Dwight Duffus for their careful reading and valuable comments. I appreciate the fruitful conversations and help of Daniel Martin and Domingos Dellamonica.

I am also grateful to many people for their support during my studies at Emory University. First, I would like to thank professors at Emory University. Special thanks to David Borthwick, Eric Brussel, Dwight Duffus, Ronald Gould, James Nagy, Raman Parimala, and Andrzej Ruciński for their teaching and guidance. Next, I thank to many friends for their help and care. Special thanks to Alvin Grissom, Eric Choi, Jongwhan Lee, Donghyun Koo, and Kwangjoo Choi.

I express my gratitude to all people I met in church. Thanks to Seung-woo Lee and Hee-seung Lee who introduced Jesus to me, and thanks to Pastor Se-kwan Jin and many friends at Norcross Korean Church.

Last, I would like to take this opportunity to thank my wife, Kei Hyun Shin, who trusted and encouraged me all the time and gave a birth of our first boy Noah Yoonwoo Lee. Also, I would like to show my gratitude to my parents and parents-in-law for their support and trust.

*To my wife Kei Hyun Shin and our son Noah Yoonwoo Lee*

# Contents

# List of Figures

# Chapter 1

# Introduction

A set $A$ of non-negative integers is called a *Sidon set* if all the sums $a_1 + a_2$, with $a_1 \leq a_2$ and $a_1$, $a_2 \in A$, are distinct. In 1930s the analyst S. Sidon asked Erdős the maximum size of Sidon sets in $[n] = \{0, 1, \cdots, n-1\}$. Erdős was captivated by the problem. Since then, the study of Sidon sets has been one of the main concerns in additive number theory. In this dissertation, we deal with results on the number of Sidon sets in $[n]$ and the maximum size of Sidon sets in sparse random subsets of $[n]$ or $\mathbb{N}$ (the set of natural numbers). We also consider a natural generalization of Sidon sets called $B_h$-sets with $h \geq 2$. A set $A$ of non-negative integers is called a $B_h$-set if all the sums $a_1 + a_2 + \cdots + a_h$, with $a_1 \leq a_2 \leq \cdots \leq a_h$ and $a_i \in A$, are distinct. Thus, Sidon sets are $B_2$-sets.

The first question in this dissertation was suggested by Cameron–Erdős in 1990. Let $\mathcal{Z}_n$ be the family of Sidon sets contained in $[n] = \{0, 1, \cdots, n-1\}$. They proposed the problem of estimating $|\mathcal{Z}_n|$. This problem is related to the problem of determining the maximum size $F(n)$ of Sidon sets in $[n]$. Results of Chowla, Erdős, Singer, and Turán from the 1940s imply that $F(n) = \sqrt{n}(1 + o(1))$. From this result, one can trivially obtain the following:

$$2^{F(n)} \leq |\mathcal{Z}_n| \leq \sum_{i=1}^{F(n)} \binom{n}{i}, \quad \text{that is,} \quad 2^{(1+o(1))\sqrt{n}} \leq |\mathcal{Z}_n| \leq 2^{c\sqrt{n}\log n},$$

where $c$ is an absolute constant. One of the main results in Chapter 2 improves the above upper bound on $|\mathcal{Z}_n|$ to $2^{c\sqrt{n}}$, which is sharp up to a constant factor in the exponent.

In Chapter 2, we also investigate the maximum size of Sidon subsets of sparse, *random* subsets of $[n]$. Recall that the maximum size of Sidon sets in $[n]$ is $\sqrt{n}(1+o(1))$. We are interested in replacing the 'environment' $[n]$ by a sparse, random subset $R$ of $[n]$.

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 [20], where Szemerédi's classical theorem on arithmetic progressions [43] is transferred to certain sparse, pseudorandom sets of integers and to the set of primes themselves (see [37, 38, 44] for more in this direction). Much closer examples to our setting are a version of Roth's theorem on 3-term arithmetic progressions [39] for random subsets of integers [29], and the far reaching generalizations due to Conlon and Gowers [12] and Schacht [40]. For the sake of brevity, we shall not discuss this further and refer the reader to [12], [40], [22, Chapter 8] and [27, Section 4].

Now we introduce several definitions. Let $F(R)$ be the maximum size of Sidon subsets contained in $R$. Let $[n]_m$ be a random subset of $[n]$ of cardinality $m$, with all the $\binom{n}{m}$ subsets of $[n]$ equiprobable. We study the behavior of the random variable $F([n]_m)$.

We first state the previous results about $F([n]_m)$. First, if $m = m(n) \ll n^{1/3}$, then almost surely $F([n]_m) = (1-o(1))m$. It can be shown by the usual deletion method. On the other hand, if $m = m(n) \gg n^{1/3}$, then almost surely $F([n]_m) = o(m)$. It is based on the recent results of Schacht [40] and Conlon and Gowers [12]. Hence, we know that $m = n^{1/3}$ is the threshold for the property that the maximum size of Sidon sets in a random set $[n]_m$ of $[n]$ is much smaller than the size of a random set $[n]_m$.

Now we state a weak version of our results which give more precise infor-

mation about the behavior of $F([n]_m)$. Fix a constant $0 < a \leq 1$ and suppose $m = (1 + o(1))n^a$. We show that there exists a constant $b = b(a)$ for which

$$F([n]_m) = n^{b+o(1)} \tag{1.1}$$

almost surely and we obtain the value of $b = b(a)$. Surprisingly, between two points $a = 1/3$ and $a = 2/3$, the function $b = b(a)$ is constant. Our results in fact determine $F([n]_m)$ up to a *constant* multiplicative factor except for $n^{2/3-\delta} < m < n^{2/3}(\log n)^{8/3}$ for any fixed $\delta > 0$. For the missing range of $m$, around $n^{2/3}$, our lower and upper bounds differ by a factor of $O((\log n)/\log\log n)$.

In Chapter 3, we consider the analogous results on $B_h$-sets with $h \geq 3$ which are natural generalizations of Sidon sets. Recall that a set $A$ of non-negative integers is called a $B_h$-set if all the sums $a_1+a_2+\cdots+a_h$, with $a_1 \leq a_2 \leq \cdots \leq a_h$ and $a_i \in A$, are distinct. Note that a $B_2$-set is a Sidon set. First, we consider a natural generalization of the problem of Cameron–Erdős. Let $\mathcal{Z}_n^h$ be the family of $B_h$-sets contained in $[n]$. We show that there exists a constant $c_h$, only depending on $h$, such that $|\mathcal{Z}_n^h| \leq 2^{c_h n^{1/h}}$, which is sharp up to a constant factor in the exponent when compared to the previous known lower bound $2^{(1+o(1))n^{1/h}}$. Second, we investigate the maximum size of $B_h$-sets in a sparse random set $[n]_m \subset [n]$. For a simpler explanation, we focus on $B_3$-sets. Let $a$ be a constant with $m = (1 + o(1))n^a$. We show that if $0 < a \leq 2/5$ or $3/4 \leq a \leq 1$, then there exists $b_3 = b_3(a)$ such that the maximum size of $B_3$-sets contained in $[n]_m$ is $n^{b_3+o(1)}$ almost surely and we obtain the value of $b_3 = b_3(a)$. The existence remains open if $2/5 < a < 3/4$.

The final investigation in this dissertation concerns infinite Sidon sets contained in sparse random subsets of $\mathbb{N}$. We begin with the previous results in the dense environment $\mathbb{N}$. While the maximum size of Sidon subsets of $[n]$ is known to be $\sqrt{n}(1 + o(1))$, the quantification of dense infinite Sidon subsets of $\mathbb{N}$ is not completely understood.

For $S \subset \mathbb{N}$, let $S[n] := S \cap [n]$. The above result on the finite case implies that any Sidon set $S \subset \mathbb{N}$ satisfies that for every sufficiently large $n$, depending

on $S$,

$$|S[n]| \leq \sqrt{n}(1 + o(1)).$$

On the other hand, a greedy algorithm gives that there exists an infinite Sidon subset $S \subset \mathbb{N}$ satisfying that for every sufficiently large $n$,

$$|S[n]| \geq n^{1/3}.$$

In 1998 Ruzsa showed that there exists an infinite Sidon subset $S \subset \mathbb{N}$ satisfying that for every sufficiently large $n$,

$$|S[n]| \geq n^{\gamma+o(1)},$$

where $\gamma := \sqrt{2} - 1 = 0.414\cdots$ . Indeed, the problem of estimating the largest possible density of an infinite Sidon subset of $\mathbb{N}$ is a well-known hard problem.

In order to consider infinite Sidon sets in a sparse random subset of $\mathbb{N}$, we first define a sparse random set in $\mathbb{N}$. Fix $0 < \delta \leq 1$, and let $R = R_\delta$ be a random set obtained by choosing each element $i \in \mathbb{N}$ independently with probability $i^{-1+\delta}$.

Our result is as follows. Let $a = a(\delta)$ and $b = b(\delta)$ be such that with probability 1, a random set $R_\delta$ satisfies the following:

- Every Sidon set $S$ contained in $R_\delta$ satisfies that $|S \cap [n]| \leq n^{b+o(1)}$ for every sufficiently large $n$.
- There exists a Sidon set $S$ contained in $R_\delta$ such that $|S \cap [n]| \geq n^{a+o(1)}$ for every sufficiently large $n$.

We show that our constant $b$ is the same as constant $b$ in (1.1) for every $0 < \delta \leq 1$, and our constants $a$ and $b$ above are identical if $0 < \delta \leq 2/3$. It remains open whether constant $a$ can be the same as constant $b$ if $2/3 < \delta \leq 1$.

# Chapter 2

# Finite Sidon sets

## 2.1 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, for example, [45]). For a modern treatment of the subject, the reader is referred to [44].

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 $a_1 \leq a_2$ 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 $[n] = \{0, 1, \ldots, n-1\}$. In 1941, Erdős and Turán [17] showed that $F(n) \leq \sqrt{n} + O(n^{1/4})$. In 1944, Chowla [10] and Erdős [16], independently of each other, observed that a result of Singer [42] implies that $F(n) \geq \sqrt{n} - O(n^{5/16})$. 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 [21] and to a recent survey by O'Bryant [36] 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 [8].

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

### 2.1.1 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 [8] proposed the problem of estimating $|\mathcal{Z}_n|$. Observe that one trivially has

$$2^{F(n)} \leq |\mathcal{Z}_n| \leq \sum_{1 \leq i \leq F(n)} \binom{n}{i} = n^{(1/2+o(1))\sqrt{n}}. \tag{2.1}$$

Cameron and Erdős [8] improved the lower bound in (2.1) 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 2.1.1** (Kohayakawa, Lee, Rödl, and Samotij [31]) *There is a constant $c$ for which $|\mathcal{Z}_n| \leq 2^{cF(n)}$.*

Our proof method gives that the constant $c$ in Theorem 2.1.1 may be taken to be arbitrarily close to $\log_2(32e) = 6.442\cdots$ (for large enough $n$). We do not make any attempts to optimize this constant as it seems that our approach cannot yield a sharp estimate for $\log_2 |\mathcal{Z}_n|$. It remains an interesting open question whether $\log_2 |\mathcal{Z}_n| = (1 + o(1))F(n)$.

### 2.1.2 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.

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 $F([n]_m)$.

The usual deletion method gives that, almost surely, that is, with probability tending to 1 as $n \to \infty$, we have $F([n]_m) = (1-o(1))m$ if $m = m(n) \ll n^{1/3}$. On the other hand, the results of Schacht [40] and Conlon and Gowers [12] imply that, if $m = m(n) \gg n^{1/3}$, then, almost surely, we have

$$F([n]_m) = o(m). \tag{2.2}$$

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 2.1.2** (Kohayakawa, Lee, Rödl, and Samotij [31]) *Let $0 \le a \le 1$ be a fixed constant. Suppose $m = m(n) = (1 + o(1))n^a$. There exists a constant $b = b(a)$ such that almost surely*

$$F([n]_m) = n^{b+o(1)}. \tag{2.3}$$

*Furthermore,*

$$b(a) = \begin{cases} a & \text{if } 0 \le a \le 1/3, \\ 1/3 & \text{if } 1/3 \le a \le 2/3, \\ a/2 & \text{if } 2/3 \le a \le 1. \end{cases} \tag{2.4}$$

Thus, the function $b = b(a)$ is piecewise linear. The graph of $b = b(a)$ is

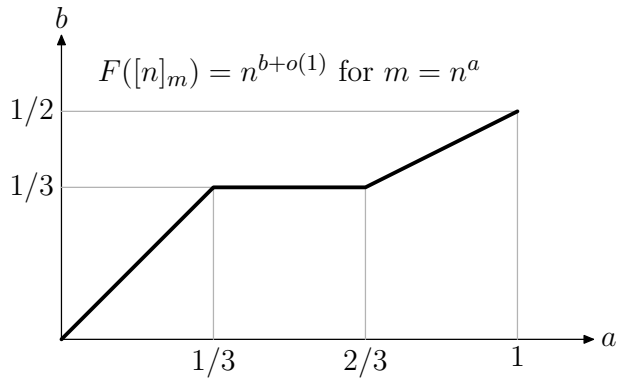


Figure 2.1: The graph of $b = b(a)$

given in Figure 2.1. The point $(a, b) = (1, 1/2)$ in the graph is clear from the Erdős–Turán and Chowla results [10, 16, 17] mentioned above. The behaviour of $b = b(a)$ in the interval $0 \leq a \leq 1/3$ 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 [40] and Conlon and Gowers [12]. 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 2.2.1. Our results in fact determine $F([n]_m)$ up to a *constant* multiplicative factor for $m \leq n^{2/3-\delta}$ for any fixed $\delta > 0$ and for $m \geq n^{2/3}(\log n)^{8/3}$. For the missing range of $m$, around $n^{2/3}$, our lower and upper bounds differ by a factor of $O((\log n)/\log\log n)$.

## 2.2 Main results

### 2.2.1 Statement of the main results

We prove a more detailed result than Theorem 2.1.1. Let $\mathcal{Z}_n(t)$ be the family of Sidon sets of cardinality $t$ contained in $[n]$.

**Theorem 2.2.1** *Let $0 < \sigma < 1$ be a real number. For any large enough $n$ and $t \geq 2s_0$, where $s_0 = (2(1-\sigma)^{-1} n \log n)^{1/3}$, we have*

$$|\mathcal{Z}_n(t)| \leq n^{3s_0} \left( \frac{32en}{\sigma t^2} \right)^t. \tag{2.5}$$

Theorem 2.1.1 follows from Theorem 2.2.1 by summing over all $t$ (see Section 2.3.2). Our next result covers values of $t$ smaller than the ones covered in Theorem 2.2.1.

**Theorem 2.2.2** *Let $n$ and $t$ be integers with*

$$3 \cdot 2^3 n^{1/3} \leq t \leq 4(n \log n)^{1/3}. \tag{2.6}$$

*Then*

$$|\mathcal{Z}_n(t)| \leq \left( \frac{6 \cdot 2^{3/2} n}{t} \exp\left( -\frac{t^3}{3 \cdot 2^7 n} \right) \right)^t. \tag{2.7}$$

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 2.2.2 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 $0 \leq a \leq 1/3$ in Theorem 2.1.2.

**Theorem 2.2.3** *For $n^{-1} \ll p = p(n) \ll n^{-2/3}$, we almost surely have*

$$F([n]_p) = (1 + o(1))np. \tag{2.8}$$

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

$$\left(\frac{1}{3} + o(1)\right) np \leq F([n]_p) \leq (1 + o(1))np, \tag{2.9}$$

**Remark 2.2.4.** *One may in fact prove the following result: if $p = \gamma n^{-2/3}$ for some constant $\gamma$, then*

$$\left(1 - \frac{1}{12}\gamma^3 + o(1)\right) np \leq F([n]_p) \leq \left(1 - \frac{1}{12}\gamma^3 + \frac{1}{12}\gamma^6 + o(1)\right) np. \tag{2.10}$$

Our next result covers the range $1/3 \leq a < 2/3$ in Theorem 2.1.2.

**Theorem 2.2.5** *For any $\delta > 0$, there is a positive constant $c_2 = c_2(\delta)$ such that if $2n^{-2/3} \leq p = p(n) \leq n^{-1/3-\delta}$, then we almost surely have*

$$c_1\big(n \log(n^2p^3)\big)^{1/3} \leq F([n]_p) \leq c_2\big(n \log(n^2p^3)\big)^{1/3}, \tag{2.11}$$

*where $c_1$ is a positive absolute constant.*

We now turn to the point $a = 2/3$ in Theorem 2.1.2.

**Theorem 2.2.6** *For any $0 \leq \delta < 1/3$, there is a positive constant $c_3 = c_3(\delta)$ such that if $1 \leq \alpha = \alpha(n) \leq n^{\delta}$ and $p = p(n) = \alpha^{-1}n^{-1/3}(\log n)^{2/3}$, then we almost surely have*

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

*where $c_4$ is an absolute constant.*

We remark that Theorems 2.2.5 and 2.2.6 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 $2/3 \leq a \leq 1$ in Theorem 2.1.2.

**Theorem 2.2.7** *There exist positive absolute constants $c_5$ and $c_6$ for which the following holds. If $1 \leq \alpha = \alpha(n) \leq (\log n)^2$ 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,*

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

### 2.2.2 The uniform model and the binomial model

We now discuss how to translate Theorems 2.2.3, 2.2.5–2.2.7 on $[n]_p$ in Section 2.2.1 to the corresponding results on $[n]_m$. Before we proceed, let us make the following definition.

**Definition 2.2.8.** *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, for example, [6, p. 35] and [23, p. 17]) will suffice for translating results on $[n]_p$ to results on $[n]_m$.

**Lemma 2.2.9** *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 $P$. We shall show that, for any constant $C > 0$, there exists a constant $C' > 0$, where $C' \to \infty$ as $C \to \infty$, such that the following holds. If $\mathbb{P}\Big[[n]_p \text{ is in } Q\Big] = O(n^{-C})$, then $\mathbb{P}\Big[[n]_m \text{ is in } Q \cap \binom{[n]}{m}\Big] = O(n^{-C'})$.

Pittel's inequality (see [6, p. 35] and [23, p. 17]) states that

$$\mathbb{P}\Big[[n]_m \text{ is in } Q \cap \binom{[n]}{m}\Big] = O(\sqrt{m}) \cdot \mathbb{P}\Big[[n]_p \text{ is in } Q\Big]. \tag{2.12}$$

Since, by hypothesis, $\mathbb{P}\Big[[n]_p \text{ is in } Q\Big] = O(n^{-C})$ holds for any constant $C > 0$, inequality (2.12) implies that

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

which completes the proof of Lemma 2.2.9. □

Each of Theorems 2.2.5–2.2.7 will be proved with 'w.o.p.' rather than with 'almost surely'. By Lemma 2.2.9, we can translate each such result on $[n]_p$ to the corresponding result on $[n]_m$, where $p = m/n$. For example, Theorem 2.2.5 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 2.2.3 holds almost surely.

### 2.2.3 Organization

Our results on the number of Sidon sets are proved in Section 2.3. In Section 2.4, we consider the upper bounds in Theorems 2.2.5–2.2.7. Section 2.5 contains some preparatory lemmas for the proof of Theorem 2.2.3 and for the proofs of the lower bounds in Theorems 2.2.5–2.2.7. The proof of Theorem 2.2.3 is given in Section 2.6. In Section 2.7, we give the proofs of the lower bounds in

Theorems 2.2.5–2.2.7.

For simplicity, we omit 'floor' and 'ceiling' symbols in our formulae, when they are not essential. For the sake of clarity of the presentation, we write $a/bc$ instead of the less ambiguous $a/(bc)$.

## 2.3 The number of Sidon sets

The proofs of Theorems 2.2.1 and 2.2.2 are based on a method introduced by Kleitman and Winston [26] (see [2, 4, 5, 18, 28] for other applications of this method).

### 2.3.1 Independent sets in dense graphs

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

**Lemma 2.3.1** (Kohayakawa, Lee, Rödl, and Samotij [31]) *Let $G$ be a graph on $N$ vertices, let $q$ be an integer and let $0 \leq \beta \leq 1$ and $R$ be real numbers with*

$$R \geq e^{-\beta q} N. \tag{2.13}$$

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

$$e(U) \geq \beta \binom{|U|}{2}. \tag{2.14}$$

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

$$\binom{N}{q}\binom{R}{r}. \tag{2.15}$$

**Proof** Fix an integer $r \geq 0$. 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 $|A| \leq R$. Furthermore, if, for some inputs $I$ and $I'$, the algorithm outputs $(S_0, A)$ and $(S_0', A')$ with $S_0 = S_0'$, then $A = A'$. Hence, the number of independent sets in $G$ with $q + r$ elements is at most as given in (2.15), 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 $A = \{v_1, \ldots, v_a\}$, with the vertex labelling convention described above. Let $i$ be the smallest index such that $v_i$ belongs to our independent set $I$, move $v_1, \ldots, v_{i-1}$ 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 $U = \{v_i, \ldots, v_a\}$. 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, that is, if $|U| < R$, consider the first $q - |S|$ members $v_{i_1}, \ldots, v_{i_{q-|S|}}$ of $I \cap A$ and move them from $A$ to $S$ (note that $i < i_1 < \cdots < i_{q-|S|} \leq a$ and we now have $|S| = q$).

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 $I$, that is, if $(S_0, A)$ and $(S_0, A')$ are both outputs by 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.

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 (2.14), 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. □

### 2.3.2 Proof of Theorem 2.2.1

We derive Theorem 2.2.1 from the following lemma.

**Lemma 2.3.2** *Let $n$, $s$ and $q$ be integers and let $0 < \sigma < 1$ be a real number such that*

$$\frac{s^2 q}{n} \geq \frac{2}{1 - \sigma} \log \frac{\sigma s}{2}. \tag{2.16}$$

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

$$|\mathcal{Z}_n(s + q + r)| \leq |\mathcal{Z}_n(s)| \binom{n}{q} \binom{2n/\sigma s}{r}. \tag{2.17}$$

To obtain the bound for $|\mathcal{Z}_n(t)|$ in Theorem 2.2.1, we apply Lemma 2.3.2 iteratively.

**Proof of Theorem 2.2.1** Fix integers $n$ and $t$, with $t \geq 2s_0$, where $s_0$ is as given in the statement of our theorem, that is, $s_0 = (2(1-\sigma)^{-1} n \log n)^{1/3}$. 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 $r_k = t2^{-K+k} - s_k - q_k$. We apply Lemma 2.3.2 with parameters $s_k$, $q_k$ and $r_k$ for $k = 0, \ldots, K$, to obtain from (2.17) that

$$|\mathcal{Z}_n(t2^{-K+k})| = |\mathcal{Z}_n(s_k + q_k + r_k)| \leq |\mathcal{Z}_n(s_k)| \binom{n}{q_k} \binom{2n/\sigma s_k}{r_k} \tag{2.18}$$

for all $k$. It suffices to check (2.16) to justify these applications of Lemma 2.3.2. 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 (2.16) holds for $n$, $s_k$ and $q_k$. Using that $s_k = s_{k-1} + q_{k-1} + r_{k-1} = t2^{-K+k-1}$ for $k \geq 1$ and that $|\mathcal{Z}_n(s_0)| \leq \binom{n}{s_0}$, we obtain from (2.18) that

$$|\mathcal{Z}_n(t)| \leq \binom{n}{s_0} \prod_{0\leq k\leq K} \binom{n}{q_k} \prod_{0\leq k\leq K} \binom{2n/\sigma s_k}{r_k}. \tag{2.19}$$

Note that

$$\binom{n}{s_0} \leq \left(\frac{en}{s_0}\right)^{s_0} \leq n^{2s_0/3} \tag{2.20}$$

and that

$$\prod_{0\leq k\leq K} \binom{n}{q_k} \leq n^{\sum_{0\leq k\leq K} q_k} \leq n^{q_0 \sum_{0\leq k\leq K} 1/4^k} \leq n^{4q_0/3} = n^{4s_0/3}. \tag{2.21}$$

We now proceed to estimate the last factor of the right-hand side of (2.19). 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

$$\binom{2n/\sigma s_0}{r_0} \leq \left(\frac{2en}{\sigma s_0 r_0}\right)^{r_0} \leq n^{r_0/3} \leq n^{2s_0/3} \leq n^{s_0} \tag{2.22}$$

for all large $n$. We now note that

$$\prod_{1\leq k\leq K} \binom{2n/\sigma s_k}{r_k} = \prod_{1\leq k\leq K} \binom{2n/\sigma s_{K-k+1}}{r_{K-k+1}} \leq \prod_{1\leq k\leq K} \binom{2n/\sigma s_{K-k+1}}{r_{K-k+1}+q_{K-k+1}}. \tag{2.23}$$

To justify the inequality in (2.23) above, we check that

$$r_{K-k+1} + q_{K-k+1} \leq \frac{2n}{3\sigma s_{K-k+1}}. \tag{2.24}$$

Recalling that $r_{K-k+1}+q_{K-k+1} = s_{K-k+1} = t2^{-k}$, we see that (2.24) is equivalent to $t2^{-k} \leq \sqrt{2n/3\sigma}$. However,

$$\frac{t}{2^k} \leq \frac{t}{2} \leq \frac{1}{2}F(n) = \left(\frac{1}{2}+o(1)\right)\sqrt{n} \leq \sqrt{\frac{2n}{3}} \leq \sqrt{\frac{2n}{3\sigma}} \tag{2.25}$$

for all large enough $n$. We continue (2.23) by noticing that

$$\prod_{1\leq k\leq K} \binom{2n/\sigma s_{K-k+1}}{r_{K-k+1}+q_{K-k+1}} = \prod_{1\leq k\leq K} \binom{2n/\sigma t2^{-k}}{t2^{-k}} \leq \prod_{1\leq k\leq K} \left(\frac{2^{2k+1}en}{\sigma t^2}\right)^{t2^{-k}} \leq \left(\frac{2en}{\sigma t^2}\right)^{t\sum_{k\geq 1} 2^{-k}} 2^{2t\sum_{k\geq 1} k2^{-k}} = \left(\frac{2en}{\sigma t^2}\right)^{t} 2^{4t} = \left(\frac{32en}{\sigma t^2}\right)^{t}. \tag{2.26}$$

Inequality (2.5) now follows from (2.19), (2.20), (2.21), (2.22) and (2.26). □

It now remains to prove Lemma 2.3.2.

**Proof of Lemma 2.3.2** 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 $\{a_1, a_2\}$ $(a_1 \neq a_2)$ 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 2.3.1 to $G$ with $\beta$ and $R$ as just defined, to obtain an upper bound for the number of independent sets of cardinality $q+r$. Note that (2.13) follows from (2.16). Now let $U \subset V$ with $|U| \geq R$ be given. We check (2.14) 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 $a_1 \neq a_2$ are both adjacent to both $w$ and $w' \in [2n]$ with $w \neq w'$, 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 $a_1 \neq a_2$, we have $b_1 \neq b_2$, 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 $|U| \geq R = 2n/\sigma s$, that is, $1 \leq \sigma \frac{|U|s}{2n}$, 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 (2.14). 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 (2.15) for the number of independent sets of cardinality $q + r$ in $G$. □

We conclude this section by deriving Theorem 2.1.1 from Theorem 2.2.1.

**Proof of Theorem 2.1.1** Let $\sigma = 32/33$ in Theorem 2.2.1. Then $s_0 = (2(1-\sigma)^{-1} n \log n)^{1/3} = (66n \log n)^{1/3}$. For large enough $n$, we have

$$|\mathcal{Z}_n| = \sum_{0 \le t \le F(n)} |\mathcal{Z}_n(t)| \le \sum_{0 \le t < 2s_0} \binom{n}{t} + \sum_{2s_0 \le t \le F(n)} n^{3s_0} \left(\frac{33en}{t^2}\right)^t. \tag{2.27}$$

Note that

$$\sum_{0 \le t < 2s_0} \binom{n}{t} \le 2s_0 \binom{n}{2s_0} \le n^{2s_0}, \tag{2.28}$$

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

$$\begin{aligned} \sum_{2s_0 \le t \le F(n)} n^{3s_0} \left(\frac{33en}{t^2}\right)^t &\le \sqrt{n} \cdot n^{3s_0} (33e)^{\sqrt{n}(1+o(1))} \\ &\le (33e)^{\sqrt{n}(1+o(1))} \le (33e)^{F(n)(1+o(1))}. \end{aligned} \tag{2.29}$$

Combining (2.27) together with (2.28) and (2.29) implies that $|\mathcal{Z}_n| \le 2^{cF(n)}$ for a suitable constant $c$. □

### 2.3.3 Proof of Theorem 2.2.2

We derive Theorem 2.2.2 from the following more general but technical estimate.

**Lemma 2.3.3** *Let $n$ and $t$ be integers. Suppose $s$ is an integer and $\sigma$ is a real number such that, letting $\omega = t/s$, we have*

$$\omega \ge 4, \quad 0 < \sigma < 1 \quad \textit{and} \quad \frac{s^3}{n} \ge \frac{2}{1-\sigma} \log \frac{\sigma s}{2}. \tag{2.30}$$

*Then*

$$|\mathcal{Z}_n(t)| \le \left(\frac{12\omega n}{(t\sigma)^{1-2/\omega}t}\right)^t. \tag{2.31}$$

**Proof** We invoke Lemma 2.3.2 with $q = s$. Note that, then, (2.30) implies (2.16). We now let $r$ in Lemma 2.3.2 be $t - 2s$ and obtain that

$$|\mathcal{Z}_n(t)| \le \binom{n}{s}\binom{n}{s}\binom{2n/\sigma s}{t-2s}. \tag{2.32}$$

The right-hand side of (2.32) 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 2.3.3. □

**Proof of Theorem 2.2.2** We shall apply Lemma 2.3.3. Let $\omega = 4$ and $s = t/\omega = t/4$. Let $\lambda = \exp\left(t^3/(3\cdot 2^6 n)\right)$ and set $\sigma = 2\lambda/s = 8\lambda/t \le 1/3$, where the last inequality follows from (2.6). It follows that $2/(1-\sigma) \le 3$, and hence

$$\frac{s^3}{n} = \frac{t^3}{4^3 n} = 3\log\lambda \ge \frac{2}{1-\sigma}\log\lambda,$$

hence the third condition in (2.30) holds. We thus conclude that (2.31) holds. Let us now estimate the right-hand side of (2.31).

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

$$\frac{12\omega n}{(t\sigma)^{1-2/\omega}t} = \frac{12\cdot 4n}{(8\lambda)^{1/2}t} = \frac{6\cdot 8n}{8^{1/2}\lambda^{1/2}t} = \frac{6\cdot 2^{3/2}n}{\lambda^{1/2}t}$$
$$= \frac{6\cdot 2^{3/2}n}{t}\exp\left(-\frac{t^3}{3\cdot 2^7 n}\right). \quad (2.33)$$

Inequality (2.7) follows from (2.31) and (2.33), and Theorem 2.2.2 is proved. □

## 2.4 The upper bounds in Theorems 2.2.5–2.2.7

We shall apply Lemma 2.3.3 and Theorem 2.2.1 in order to prove the upper bounds in Theorem 2.2.5 and Theorems 2.2.6–2.2.7, respectively.

### 2.4.1 Proof of the upper bound in Theorem 2.2.5

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

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

To this end, we apply Lemma 2.3.3. We first define several auxiliary constants used to set $t$, $\omega$ and $\sigma$ in Lemma 2.3.3. Choose $\eta > 0$ small enough so that

$$(1-3\delta)\left(\frac{1}{3}+\eta\right) < \frac{1}{3}. \quad (2.34)$$

Choose $\omega \ge 4$ so that

$$\left(\frac{1}{3}+\eta\right)\left(1-\frac{2}{\omega}\right) > \frac{1}{3}. \quad (2.35)$$

Finally, choose $c = c_2$ so that

$$\left(\frac{c}{\omega}\right)^3 > 3\left(\frac{1}{3}+\eta\right) \quad \text{and} \quad c > \frac{24\omega}{2^{(1+3\eta)(1-2/\omega)}}. \quad (2.36)$$

Now set

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

Note that

$$t \geq c(n \log 8)^{1/3} \geq cn^{1/3} \qquad \text{and} \qquad \xi < 1. \tag{2.37}$$

We first check that condition (2.30) holds for large enough $n$. We have $\omega \geq 4$ by the choice of $\omega$. Moreover, we have $\sigma \to 0$ as $n \to \infty$ because of (2.34). Finally, from (2.36) 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 (2.30). Hence Lemma 2.3.3 implies that

$$\mathbb{P}\left([n]_p \text{ contains a Sidon set of size } t\right) \leq |\mathcal{Z}_n(t)|p^t \leq \left(\frac{12\omega np}{t(t\sigma)^{1-2/\omega}}\right)^t. \tag{2.38}$$

Making use of the first equation of (2.37) and the fact that $t\sigma = \omega s\sigma = 2\omega(n^2p^3)^{1/3+\eta}$, we see that the upper bound in (2.38) is at most

$$\begin{aligned}\left(\frac{12\omega np}{cn^{1/3}(2\omega)^{1-2/\omega}(n^2p^3)^{(1/3+\eta)(1-2/\omega)}}\right)^t &\leq \left(\frac{12\omega}{c(2\omega)^{1-2/\omega}}\cdot\frac{n^{2/3}p}{(n^2p^3)^{(1/3+\eta)(1-2/\omega)}}\right)^t \\ &= \left(\frac{12\omega^{2/\omega}}{2^{1-2/\omega}c(n^2p^3)^{(1/3+\eta)(1-2/\omega)-1/3}}\right)^t, \end{aligned} \tag{2.39}$$

which, by (2.35) and the assumption $p \geq 2n^{-2/3}$, is at most

$$\left(\frac{12\omega}{2^{1/2}c(2^3)^{(1/3+\eta)(1-2/\omega)-1/3}}\right)^t \leq \left(\frac{24\omega}{c2^{(1+3\eta)(1-2/\omega)}}\right)^t = \xi^t. \tag{2.40}$$

To complete the proof, it suffices to recall (2.37).

### 2.4.2 Proof of the upper bound in Theorem 2.2.6

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.

$$F([n]_p) \leq c_4 (n \log n)^{1/3} \frac{\log n}{\log(\alpha + \log n)} \tag{2.41}$$

for some absolute constant $c_4$. To this end, we use Theorem 2.2.1. Let $\sigma = 3/4$, $s_0 = 2(n \log n)^{1/3}$ and $t = \omega s_0$, where

$$\omega = 11e \, \frac{\log n}{\log(\alpha + \log n)}, \tag{2.42}$$

and note that $\omega \geq 2$ for sufficiently large $n$. Hence, by Theorem 2.2.1 and the union bound, the probability that $[n]_p$ contains a Sidon set with at least $t$ elements can be bounded as follows:

$$\begin{aligned} \mathbb{P}\left(F([n]_p) \geq t\right) \leq |\mathcal{Z}_n(t)| p^t &\leq n^{3s_0} \left(\frac{44enp}{t^2}\right)^t \\ &= n^{3s_0} \left(\frac{44enp}{\omega^2 s_0^2}\right)^{\omega s_0} \leq \left[\left(\frac{11e}{\alpha\omega^2}\right)^{\omega} n^3\right]^{s_0}, \end{aligned} \tag{2.43}$$

where the last inequality follows from

$$p = \alpha^{-1} n^{-1/3} (\log n)^{2/3} \text{ and } s_0 = 2(n \log n)^{1/3}.$$

For the proof of (2.41), it suffices to show that the base of the exponential in the right-hand side of (2.43) is bounded away from 1, that is, whether

$$\left(\frac{11e}{\alpha\omega^2}\right)^{\omega} n^3 < 1 - \varepsilon \tag{2.44}$$

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

$$\left(\frac{\alpha\omega^2}{11e}\right)^{\omega} \geq (\alpha\omega)^{\omega} = \exp\left(\omega \log(\alpha\omega)\right). \tag{2.45}$$

We claim that

$$2\log(\alpha\omega) \geq \log(\alpha + \log n). \tag{2.46}$$

Observe that since $\omega \geq 2$, then (2.46) 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 (2.42), (2.45) and (2.46) 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 (2.44) holds, completing the proof of (2.41).

### 2.4.3 Proof of the upper bounds in Theorem 2.2.7

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 2.2.6, see (2.43), using Theorem 2.2.1, we estimate

$$\mathbb{P}\left(F([n]_p) \geq t\right) \leq |\mathcal{Z}_n(t)|p^t \leq \left[\left(\frac{11e\beta}{\omega^2}\right)^{\omega} n^3\right]^{s_0}. \tag{2.47}$$

We split into two cases, depending on the order of magnitude of $\beta$.

(*Case I*) If $\beta(n) \leq (\log n)^2$, then we let

$$\alpha = \beta^{-1}(\log n)^2 \text{ and } \omega = (11e\log n)/\log(e\alpha)$$

so that $t = \omega s_0 = 22e\sqrt{np}\cdot\sqrt{\alpha}/\log(e\alpha)$. Note that

$$\left(\frac{11e\beta}{\omega^2}\right)^{\omega} = \left(\frac{11e(\log n)^2}{\alpha\omega^2}\right)^{\omega} = \left(\frac{(\log(e\alpha))^2}{11e\alpha}\right)^{11e(\log(e\alpha))^{-1}\log n}. \tag{2.48}$$

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 (2.48) 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 (2.47), 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 (2.47), 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.

## 2.5 Nontrivial solutions in random sets

### 2.5.1 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 2.5.1.** *Let*

$$\mathcal{S} = \big\{\{a_1, a_2, a_3, a_4\} \colon (a_1, a_2, a_3, a_4) \text{ is a nontrivial solution of } x_1 + x_2 = y_1 + y_2\big\}. \quad (2.49)$$

*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 $\big|\mathcal{S}\big[[n]_p\big]\big|$, that is, the number of hyperedges of $\mathcal{S}$ induced by $[n]_p$.*

In Lemma 2.5.4 below, we give an estimate for $X$ that will be used in the proof of Theorem 2.2.3 and in the proofs of the lower bounds in Theorems 2.2.5–2.2.7.

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

We clearly have

$$\mathcal{S} = \mathcal{S}_4 \cup \mathcal{S}_3 \quad \text{and} \quad \mathcal{S}_4 \cap \mathcal{S}_3 = \emptyset \quad (2.50)$$

and hence

$$X = X_4 + X_3. \tag{2.51}$$

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

**Lemma 2.5.3** *Fix $\delta > 0$. The following assertions hold w.o.p.*

*(i) If $p \geq n^{-3/4+\delta}$, 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 2.5.3 is based on a concentration result due to Kim and Vu [25]. We shall introduce the Kim–Vu polynomial concentration result in Section 2.5.2 and prove Lemma 2.5.3 in Section 2.5.3. Assuming Lemma 2.5.3, we are ready to estimate $X$.

**Lemma 2.5.4** *Fix $\delta > 0$ and suppose $p \geq n^{-3/4+\delta}$. Then, w.o.p., $X = n^3p^4(1/12 + o(1))$.*

**Proof** Let $X = X([n]_p)$ be as defined in Definition 2.5.1 and recall (2.51). From the assumption that $p \geq n^{-3/4+\delta}$, we see that the estimates for $X_4$ and $X_3$ given in Lemma 2.5.3(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 $n^{3\delta} \ll n^{4\delta} \leq n^3p^4$, because $p \geq n^{-3/4+\delta}$, 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 (2.51) and the estimate in Lemma 2.5.3(i) that $X = n^3p^4(1/12+o(1))$ holds w.o.p. □

It now remains to prove Lemma 2.5.3. We first introduce the main tool we shall use in the proof of that lemma, due to Kim and Vu [25].

## 2.5.2 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 $\mathcal{H}\big[[n]_p\big]$ be the subhypergraph of $\mathcal{H}$ induced on $[n]_p$. Let $Y$ be the sum of the weights of all the hyperedges in $\mathcal{H}\big[[n]_p\big]$, that is, $Y = \sum_{e\in\mathcal{H}[[n]_p]} w(e)$. Kim and Vu obtained a concentration result for the random variable $Y$. We now proceed to present their result [25] (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 $\mathcal{H}\big[[n]_p\big]$ containing $A$, that is, $Y_A = \sum_{A\subset e\in\mathcal{H}[[n]_p]} w(e)$. 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

$$E' = \max\{E_i \colon 1 \le i \le k\} \qquad \text{and} \qquad E = \max\{E', \mu\}. \tag{2.52}$$

**Theorem 2.5.5** (Kim–Vu polynomial concentration inequality) *With the above notation, we have, for every $\lambda > 1$,*

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

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

## 2.5.3 Proof of Lemma 2.5.3

We prove (i) and (ii) of Lemma 2.5.3 separately.

**Proof of Lemma 2.5.3**(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,

$$|\mathcal{S}_4| = \sum_{i=0}^{n-1} \sum_{j=i}^{(n+i)/2} \sum_{k=j}^{n+i-j-1} 1 = \sum_{i=0}^{n-1} \sum_{j=i}^{(n+i)/2} (n+i-2j)$$
$$\sim n^3 \int_0^1 \int_x^{(1+x)/2} (1+x-2y)dydx \sim \frac{1}{12}n^3,$$

where for the first identity we ignore the floors or ceilings. Hence

$$\mu(X_4) = |\mathcal{S}_4|p^4 = \left(\frac{1}{12} + o(1)\right) n^3 p^4. \tag{2.53}$$

Next we apply Theorem 2.5.5 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 (2.52). 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 $E_{\{a\}} = O(n^2p^3)$. Consequently, $E_1 = \max\{E_A \colon |A| = 1\} = O(n^2p^3)$. A similar argument gives that $E_i = \max\{E_A \colon |A| = i\} = O(n^{3-i}p^{4-i})$ for all $1 \le i < 4$. Therefore, since $np \gg 1$, we have $E_i = O(n^2p^3)$ for all $1 \le i < 4$. Also, clearly, $E_4 = \max\{E_A \colon |A| = 4\} = 1$. Thus

$$E' = \max\{E_i \colon 1 \le i \le 4\} = O(\max\{n^2p^3, 1\}), \tag{2.54}$$

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

$$E = O(n^3p^4). \tag{2.55}$$

In view of (2.54) and (2.55), a simple computation implies the following:

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

$$E' = O(1) \quad \text{and} \quad E = O(n^3 p^4). \tag{2.56}$$

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

$$E' = O(n^2 p^3) \quad \text{and} \quad E = O(n^3 p^4). \tag{2.57}$$

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, (2.56) implies that

$$(EE')^{1/2} = O(n^3 p^4 \cdot 1)^{1/2} = O(n^3 p^4)^{1/2}. \tag{2.58}$$

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

$$\lambda = (n^3 p^4)^{1/12} \geq n^{\delta/3}. \tag{2.59}$$

Also $n^3 p^4 \geq n^{4\delta} \gg 1$, and hence combining (2.58) and $\lambda = (n^3 p^4)^{1/12}$ implies that

$$(EE')^{1/2} \lambda^4 = O(n^3 p^4)^{1/2} (n^3 p^4)^{1/3} = O(n^3 p^4)^{5/6} = o(n^3 p^4). \tag{2.60}$$

Theorem 2.5.5 together with (2.59) then yields that

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

where $a_4 = 8^4 (4!)^{1/2}$. Given (2.60), we have that w.o.p.

$$X_4 = \mu(X_4) + o(n^3 p^4). \tag{2.61}$$

(*Case II*) *Suppose* $p \geq n^{-2/3}$. In this case, (2.57) yields that

$$(EE')^{1/2} = O(n^3p^4n^2p^3)^{1/2} = O\Big(\frac{n^3p^4}{(np)^{1/2}}\Big). \tag{2.62}$$

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

$$\lambda \geq \big(n^{1/3}\big)^{1/12} = n^{1/36}. \tag{2.63}$$

Since $np \gg 1$, combining (2.62) and $\lambda = (np)^{1/12}$ implies that

$$(EE')^{1/2}\lambda^4 = O\Big(\frac{n^3p^4}{(np)^{1/2}}\Big)(np)^{1/3} = O\Big(\frac{n^3p^4}{(np)^{1/6}}\Big) = o(n^3p^4). \tag{2.64}$$

Theorem 2.5.5 together with (2.63) then yields that

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

where $a_4 = 8^4(4!)^{1/2}$. Given (2.64), we have that w.o.p.

$$X_4 = \mu(X_4) + o(n^3p^4). \tag{2.65}$$

In view of (2.53), it follows from (2.61) and (2.65) 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 2.5.3. □

**Proof of Lemma 2.5.3**(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

$$\mu(X_3) = O(n^2p^3). \tag{2.66}$$

Next, we prove a concentration result for $X_3$ applying Theorem 2.5.5. To

this end, we estimate the quantities $E_i$ $(1 \le i \le 3)$. As in the proof of Lemma 2.5.3(i), one may check that $E' = \max_{1\le i\le 3} E_i = O(\max\{np^2, p, 1\})$ 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

$$E' = O(\max\{np^2, 1\}) \quad \text{and} \quad E = O(\max\{n^2p^3, 1\}). \tag{2.67}$$

Based on (2.67), we consider the cases $p \ge n^{-2/3+\delta}$ and $n^{-1} \ll p \le n^{-2/3+\delta}$ separately.

We first suppose $p \ge n^{-2/3+\delta}$. From (2.67), we have $E' = O(\max\{np^2, 1\})$ and $E = O(n^2p^3)$. A proof similar to the proofs of (2.61) and (2.65) shows that, for $p \ge n^{-2/3+\delta}$, w.o.p., $X_3 = \mu(X_3) + o(n^2p^3)$. This together with (2.66) implies that for $p \ge n^{-2/3+\delta}$, w.o.p.,

$$X_3 = O(n^2p^3). \tag{2.68}$$

We now suppose $n^{-1} \ll p \le n^{-2/3+\delta}$. In this case, (2.67) yields that $E' = O(1)$ and $E = O(n^{3\delta})$ and hence, setting $\lambda = n^{\delta/2}$, we have

$$(EE')^{1/2}\lambda^3 = O(n^{(3/2)\delta})n^{(3/2)\delta} = O(n^{3\delta}). \tag{2.69}$$

Theorem 2.5.5 with $\lambda = n^{\delta/2}$ yields

$$\mathbb{P}\Big[|X_3 - \mu(X_3)| > a_3(EE')^{1/2}\lambda^3\Big] < 2e^2e^{-\lambda}n^2 \le 2e^2e^{-n^{\delta/2}}n^2, \tag{2.70}$$

where $a_3 = 8^3(3!)^{1/2}$. Inequality (2.70) together with (2.69) implies that, for $n^{-1} \ll p \le n^{-2/3+\delta}$, w.o.p., $X_3 = \mu(X_3) + O(n^{3\delta})$. Since, under the assumption $p \le n^{-2/3+\delta}$, we have $\mu(X_3) = O(n^2p^3) = O(n^{3\delta})$, we infer that, for $n^{-1} \ll p \le n^{-2/3+\delta}$, w.o.p.,

$$X_3 = O(n^{3\delta}). \tag{2.71}$$

Combining (2.68) and (2.71) completes the proof of (ii) of Lemma 2.5.3. □

## 2.6 Proof of Theorem 2.2.3

### 2.6.1 Theorem 2.2.3 for smaller $p = p(n)$

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

**Proof of** (2.8) **in Theorem 2.2.3** Suppose $n^{-1} \ll p \ll n^{-2/3}$. We show that (2.8) holds almost surely, using the usual deletion method. Let $\mathcal{S}$, $\mathcal{S}\big[[n]_p\big]$ and $X$ be as in Definition 2.5.1. If we delete one vertex from each hyperedge in $\mathcal{S}\big[[n]_p\big]$, the remaining vertex set is an independent set of $\mathcal{S}\big[[n]_p\big]$, and hence it is a Sidon set contained in $[n]_p$. Consequently, $F([n]_p) \geq \big|[n]_p\big| - \big|\mathcal{S}\big[[n]_p\big]\big| = \big|[n]_p\big| - X$. Since trivially $F([n]_p) \leq \big|[n]_p\big|$, we have $\big|[n]_p\big| - X \leq F([n]_p) \leq \big|[n]_p\big|$. Note that the Chernoff bound gives that, for $p \gg n^{-1}$, we almost surely have $\big|[n]_p\big| = np + o(np)$. Therefore, in order to show (2.8), it only remains to show that $X = o(np)$ almost surely. Recall that $X_i$ is the number of edges of cardinality $i$ in $\mathcal{S}\big[[n]_p\big]$ ($i \in \{3, 4\}$), and that $X = X_3 + X_4$ (see Definition 2.5.2 and (2.51)). Equations (2.53) and (2.66), 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. □

### 2.6.2 Theorem 2.2.3 for larger $p = p(n)$

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

**Proof of** (2.9) **in Theorem 2.2.3** 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 (2.9) holds if, for example, $n^{-2/3}/\log n \leq p \leq 2n^{-2/3}$. We proceed as in the proof of (2.8), given in Section 2.6.1 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 2.5.4 tells us that, w.o.p.,

$$X = \frac{1}{12}n^3p^4 + o(n^3p^4) = \frac{1}{12}n^3p^4 + o(np) \leq \left(\frac{2}{3} + o(1)\right)np, \tag{2.72}$$

as required. □

## 2.7 The lower bounds in Theorems 2.2.5–2.2.7

Let us first state a simple monotonicity result (see, for example, [23, Lemma 1.10]) that will be used a few times in this section.

**Fact 2.7.1** *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 2.7.1 are, in fact, equivalent. We state them both explicitly just for convenience.

### 2.7.1 Proofs of the lower bounds in Theorems 2.2.5 and 2.2.6

The lower bounds in Theorems 2.2.5 and 2.2.6 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 Rödl [14].

**Lemma 2.7.1** *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*

$$c\,\frac{(\log t)^{1/(r-1)}}{t}\,N, \tag{2.73}$$

*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 2.7.1. Let $\mathcal{S}\big[[n]_p\big]$ be the hypergraph in Definition 2.5.1. Since an independent set of $\mathcal{S}\big[[n]_p\big]$ is a Sidon set contained in $[n]_p$, independent sets in $\mathcal{S}\big[[n]_p\big]$ give lower bounds for $F([n]_p)$. To apply Lemma 2.7.1, we shall obtain a simple 4-uniform subhypergraph $\mathcal{S}^*$ of $\mathcal{S}\big[[n]_p\big]$ by deleting suitable vertices from $\mathcal{S}\big[[n]_p\big]$. Lemma 2.7.1 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 2.7.2** *There is an absolute constant $d > 0$ such that, for $p \geq 2n^{-2/3}$, w.o.p. $F([n]_p) \geq d\,(n\log(n^2p^3))^{1/3}$ holds.*

Lemma 2.7.2 easily implies the lower bounds in Theorems 2.2.5 and 2.2.6. The proof of Lemma 2.7.2 will be given in Section 2.7.3.

## 2.7.2 Proof of the lower bound in Theorem 2.2.7

For larger $p = p(n)$, it turns out that, instead of using Lemma 2.7.1, it is better to make use of the fact that $[n]$ contains a Sidon set of cardinality $(1+o(1))\sqrt{n}$ (see Section 2.1). 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 [33].

**Lemma 2.7.3** *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 2.7.3 together with $F([m]) \geq (1 + o(1))\sqrt{m}$ gives

the lower bound in Theorem 2.2.7. 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 following fact:

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

$$F([n]_p) \geq \left(\frac{1}{3\sqrt{2}} + o(1)\right)\sqrt{np}. \tag{2.74}$$

Fact 2.7.1 then 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** (*) Let $(\log n)^2/n \ll p \leq 1/3$. We shall show that (2.74) holds w.o.p. We define a partition of $[n] = \{0, \dots, n-1\}$ 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 (2.74) holds w.o.p. The details are as follows.

Let $\mathcal{I} = \{I_i \colon 0 \leq i < \lceil n/x \rceil\}$ 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 $\mathcal{I}_{\text{even}} = \{I_0, I_2, I_4, \dots\} \subset \mathcal{I}$ 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 [10, 16], there exists a Sidon subset $S$ of $[y]$ with

$$|S| = (1 + o(1))\sqrt{y} = \left(\frac{1}{\sqrt{2}} + o(1)\right)\sqrt{np}. \tag{2.75}$$

We "identify" $[y]$ and $\mathcal{I}_{\text{even}}$ by the bijection $i \mapsto I_{2i}$. Let $\{a_i \colon i \in S\}$ be a set of integers with $a_i \in I_{2i}$ for all $i \in S$. We claim that $\{a_i \colon i \in S\}$ 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

$$a_{i_1} + a_{i_2} \in I_{2i_1+2i_2} \cup I_{2i_1+2i_2+1} \quad \text{and} \quad a_{j_1} + a_{j_2} \in I_{2j_1+2j_2} \cup I_{2j_1+2j_2+1}, \tag{2.76}$$

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 $\{a_i \colon i \in S\}$ 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

$$\tilde{p} = 1-(1-p)^x = 1-(1-p)^{\lfloor 1/p \rfloor} \geq 1-e^{-p(1/p-1)} \geq 1-e^{-1+p} \geq 1-e^{-2/3} > 1/3, \tag{2.77}$$

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, (2.75) and (2.77) give that, w.o.p.,

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

Recalling that $F([n]_p) \geq |\mathcal{I}_{\mathrm{occ}}|$, statement (*) follows. □

### 2.7.3 Proof of Lemma 2.7.2

In Lemma 2.7.4 below, we prove Lemma 2.7.2 for a narrower range of $p$. We shall then invoke monotonicity (Fact 2.7.1) to obtain Lemma 2.7.2 in full.

**Lemma 2.7.4** *There is an absolute constant $d > 0$ such that, for $2n^{-2/3} \leq p \ll n^{-2/3+1/15}$, we have $F([n]_p) \geq d(n \log n^2 p^3)^{1/3}$ w.o.p.*

**Proof** Let $\mathcal{S}\big[[n]_p\big]$, $\mathcal{S}_i\big[[n]_p\big]$, $X$ and $X_i$ be as in Definitions 2.5.1 and 2.5.2. Recall that the size of an independent set of $\mathcal{S}\big[[n]_p\big]$ gives a lower bound on $F([n]_p)$.

We wish to apply Lemma 2.7.1. However, since $\mathcal{S}\big[[n]_p\big]$ may be neither simple nor uniform, we consider a suitable *induced* subhypergraph $\mathcal{S}^* \subset \mathcal{S}\big[[n]_p\big]$, as discussed just after the statement of Lemma 2.7.1. We have $\mathcal{S}\big[[n]_p\big] = \mathcal{S}_3\big[[n]_p\big] \cup \mathcal{S}_4\big[[n]_p\big]$. Let $\widetilde{\mathcal{S}}_4$ be the set of all hyperedges in $\mathcal{S}_4\big[[n]_p\big]$ that share at least two vertices with some other hyperedge in $\mathcal{S}_4\big[[n]_p\big]$. If we delete one vertex from each hyperedge of $\mathcal{S}_3\big[[n]_p\big] \cup \widetilde{\mathcal{S}}_4$, the remaining induced subhypergraph $\mathcal{S}^*$ of $\mathcal{S}\big[[n]_p\big]$ is both simple and 4-uniform. To apply Lemma 2.7.1 to $\mathcal{S}^*$, we now estimate $|V(\mathcal{S}^*)|$ and the average degree of $\mathcal{S}^*$.

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

**Fact 2.7.2** *Fix $\delta > 0$ and suppose $n^{-1+\delta} \ll p \ll n^{-1/2}$. We have, w.o.p., $X_3 = o(np)$.*

**Fact 2.7.3** *Fix $\delta > 0$ and suppose $n^{-1+\delta} \ll p \ll n^{-2/3+1/15}$. We have, w.o.p., $\big|\widetilde{\mathcal{S}}_4\big| = o(np)$.*

Since the Chernoff bound gives that $\big|[n]_p\big| = np + o(np)$ w.o.p. for $p \gg (\log n)/n$, Facts 2.7.2 and 2.7.3 imply that, w.o.p., we have

$$|V(\mathcal{S}^*)| = np(1 + o(1)). \tag{2.78}$$

Next we consider the average degree of $\mathcal{S}^*$. Owing to $\mathcal{S}^* \subset \mathcal{S}\big[[n]_p\big]$, (2.78) and Lemma 2.5.4, 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 2.7.1. In view of our average degree estimate above, we set $t = (n^2p^3)^{1/3}$. Given (2.78), Lemma 2.7.1 implies that,

w.o.p., the hypergraph $\mathcal{S}^*$, and thus $\mathcal{S}\big[[n]_p\big]$, has an independent set of size

$$c\,\frac{(\log t)^{1/3}}{t}|V(\mathcal{S}^*)| \geq c\,\frac{\big[(1/3)\log(n^2p^3)\big]^{1/3}}{(n^2p^3)^{1/3}}np(1+o(1)) \geq d\big(n\log(n^2p^3)\big)^{1/3}, \tag{2.79}$$

for, say, $d = c/2$. This completes the proof of Lemma 2.7.4. □

In order to finish the proof of Lemma 2.7.4, it remains to prove Facts 2.7.2 and 2.7.3.

**Proof of Fact 2.7.2** Lemma 2.5.3(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 2.7.3** 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 $\mathcal{S}_4\big[[n]_p\big]$ with $|E_1 \cap E_2| \geq 2$. Observe that

$$\big|\widetilde{\mathcal{S}}_4\big| \leq 2|\mathcal{P}|. \tag{2.80}$$

An argument similar to one in the proof of Lemma 2.5.3(ii), based on the Kim–Vu polynomial concentration result, tells us that $|\mathcal{P}| = O(\max\{\mathbb{E}\big[|\mathcal{P}|\big], n^\delta\}) = O\big(\max\{n^4p^6, n^\delta\}\big)$ 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 (2.80), we have, w.o.p., $\big|\widetilde{\mathcal{S}}_4\big| = o(np)$. □

In order to establish Lemma 2.7.2, we need to expand the range of $p$ in Lemma 2.7.4 from $2n^{-2/3} \leq p \ll n^{-2/3+1/15} = n^{-3/5}$ to $p \geq 2n^{-2/3}$.

**Proof of Lemma 2.7.2** To complement the range of $p$ covered by Lemma 2.7.4, it is enough to show that, say, for $p \geq n^{-2/3+1/16}$, we have, w.o.p., $F([n]_p) \geq$

$d'\left(n\log(n^2p^3)\right)^{1/3}$ for some absolute constant $d'>0$. Lemma 2.7.4 implies that, for $p=n^{-2/3+1/16}$, we have, w.o.p.,

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

where $d'=d(1/16)^{1/3}$. By Fact 2.7.1, we infer that, for $p\geq n^{-2/3+1/16}$, we have, w.o.p., $F([n]_p)\geq d'\big[n\log n^2\big]^{1/3}\geq d'\big[n\log(n^2p^3)\big]^{1/3}$, completing the proof of Lemma 2.7.2. □

# Chapter 3

# Finite $B_h$-sets

## 3.1 Introduction

Let $h \geq 2$ be a fixed integer. A set $S$ of positive integers is called a $B_h$-set if all the sums $a_1 + a_2 + \cdots + a_h$, with $a_1 \leq a_2 \leq \cdots \leq a_h$ and $a_i \in S$, are distinct. Note that a $B_2$-set is also called a Sidon set. A well-known problem on $B_h$-sets is the determination of $F_h(n)$, where $F_h(n)$ denotes the maximum possible size of $B_h$-sets of $[n] = \{1, \cdots, n\}$. The results of Chowla, Erdős, Singer, and Turán [10, 16, 17, 42] imply that $F_2(n) = (1 + o(1))\sqrt{n}$. In 1962 Bose and Chowla [7] showed that $F_h(n) \geq (1 + o(1))n^{1/h}$ for $h \geq 3$. On the other hand, various improvements on constants $c_h = c_h(h)$ for which $F_h(n) \leq c_h n^{1/h}$ were given in [9, 11, 15, 24, 32, 34, 35, 41]. Currently, the smallest known upper bound of $c_h$ is given by Green [19] as

$$c_3 < 1.519, \;\; c_4 < 1.627, \text{ and } \; c_h \leq (1/2e)\big(h + (3/2)\log h + o_h(\log h)\big).$$

For a wealth of related material, the reader is referred to the classical monograph of Halberstam and Roth [21] and to a recent survey by O'Bryant [36] and the references therein.

We investigate $B_h$-sets contained in *random sets of integers*. Though our

results can be obtained for $B_h$-sets with $h \geq 3$, we focus on $B_3$-sets for a simpler explanation. We obtain upper and lower bounds on their relative density, and gain essentially tight bounds for some range of the size of a random set. Our approach is based on finding upper bounds for the number of $B_3$-sets of a given size contained in $[n]$. Besides being the key lemma to our probabilistic results, our upper bounds also address a natural generalization of a problem of Cameron and Erdős [8].

We discuss our bounds on the number of $B_3$-sets and our probabilistic results in the next two subsections.

### 3.1.1 A generalization of a problem of Cameron and Erdős

Let $\mathcal{Z}_n^h$ be the family of $B_h$-sets contained in $[n]$. An interesting question is to estimate $|\mathcal{Z}_n^h|$. Observe that one trivially has

$$2^{F_h(n)} \leq |\mathcal{Z}_n^h| \leq \sum_{i=1}^{F_h(n)} \binom{n}{i} \leq F_h(n) \binom{n}{F_h(n)}. \tag{3.1}$$

From the result that $(1 + o(1))n^{1/h} \leq F_h(n) \leq cn^{1/h}$ for a constant $c$ only depending on $h$, we have

$$2^{(1+o(1))n^{1/h}} \leq |\mathcal{Z}_n^h| \leq n^{c'n^{1/h}}, \tag{3.2}$$

where $c'$ is a constant only depending on $h$. In this paper we improve the upper bound on $|\mathcal{Z}_n^3|$ in (3.2) as follows:

**Theorem 3.1.1** (Dellamonica, Kohayakawa, Lee, Rödl, and Samotij [13]) *There exists an absolute constant $C$ such that $|\mathcal{Z}_n^3| \leq 2^{Cn^{1/3}}$.*

The proof of Theorem 3.1.1 is given in Section 3.2.1.

**Remark 3.1.2.** *One may in fact prove the following about $|\mathcal{Z}_n^h|$ with a similar argument:*
*There exists a constant $c_h$, only depending on $h$, such that $|\mathcal{Z}_n^h| \leq 2^{c_h n^{1/h}}$.*

### 3.1.2 Probabilistic results

We investigate $B_3$-sets contained in a sparse, *random* set of $[n]$, that is, we replace the 'environment' $[n]$ by a sparse, random subset $[n]_m$ of $[n]$, where $[n]_m$ denotes a random subset of $[n]$ of cardinality $m = m(n)$, with all the subsets of $[n]$ with size $m$ having the same probability. Then we ask how large a subset $S \subset [n]_m$ can be, if we require that $S$ should be a $B_3$-set. The following definition will provide a notation suitable for this problem.

**Definition 3.1.3.** *For a set $R \subset [n]$, denote by $F_3(R)$ the maximum size of a $B_3$-set contained in $R$.*

We are therefore interested in the random variable $F_3([n]_m)$. The deletion methods yield that almost surely $F_3([n]_m) = (1 - o(1))m$ if $m = m(n) \ll n^{1/5}$. On the other hand, the result of Schacht [40] and Conlon and Gowers [12] yield that almost surely $F_3([n]_m) = o(m)$ if $m = m(n) \gg n^{1/5}$. Thus $F_3([n]_m)$ undergoes a sudden change of its behavior at $m = n^{1/5+o(1)}$. The following abridged version of our results gives us quite precise information on $F_3([n]_m)$ for a large range of $m$, and non-trivial, but loose bounds for $n^{2/5} \le m \le n^{3/4}$— see Figure 3.1.

**Theorem 3.1.4** (Dellamonica, Kohayakawa, Lee, Rödl, and Samotij [13]) *Let $0 \le a \le 1$ be a fixed constant. Suppose $m = m(n) = (1 + o(1))n^a$. Then almost surely*

$$n^{b_1+o(1)} \le F_3([n]_m) \le n^{b_2+o(1)}, \tag{3.3}$$

*where*

$$b_1(a) = \begin{cases} a, & \text{for } 0 \le a < 1/5 \\ 1/5, & \text{for } 1/5 \le a < 3/5 \\ a/3, & \text{for } 3/5 \le a \le 1 \end{cases} \tag{3.4}$$

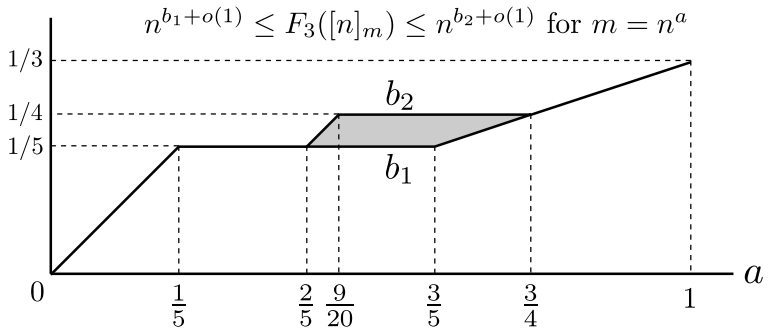


Figure 3.1: The graphs of $b_1 = b_1(a)$ and $b_2 = b_2(a)$.

*and*

$$b_2(a) = \begin{cases} a & \textit{for } 0 \le a \le 1/5 \\ 1/5 & \textit{for } 1/5 \le a \le 2/5 \\ a - 1/5 & \textit{for } 2/5 \le a \le 9/20 \\ 1/4 & \textit{for } 9/20 \le a \le 3/4 \\ a/3 & \textit{for } 3/4 \le a \le 1. \end{cases} \tag{3.5}$$

**Remark 3.1.5.** *One may in fact show a similar result on $F_h([n]_m)$ for any $h \ge 3$.*

The graphs of $b_1 = b_1(a)$ and $b_2 = b_2(a)$ are given in Figure 3.1. The point $(1, 1/3)$ in the graph is clear from the above result that $(1+o(1))n^{1/3} \le F_3(n) \le cn^{1/3}$, for some absolute constant $c$. The behavior of $b_1 = b_2$ in the interval $0 \le a \le 1/5$ is not hard to establish. The fact that the point $(1/5, 1/5)$ could be an interesting point in the graph is suggested by the results of Schacht [40] and Conlon and Gowers [12]. We determined $b = b(a)$ for which $F_3([n]_m) = n^{b+o(1)}$, where $m = (1 + o(1))n^a$, in the intervals $1/5 \le a \le 2/5$ and $3/4 \le a \le 1$. It is somewhat surprising that $b = b_1 = b_2$ should be constant for $1/5 \le a \le 2/5$. An interesting open question is the existence and determination of $b = b(a)$ such that $F_3([n]_m) = n^{b+o(1)}$ for $2/5 \le a \le 3/4$. We state our results in full in Section 3.2. The upper bounds of Theorem 3.1.4 are proved in Section 3.2.2

and the lower bounds are proved in Section 3.4.

**Remark 3.1.6.** *In some of the proofs below it will be convenient to use the Binomial random model $[n]_p$, which is a random subset of $[n]$ where each element is selected to be in the set independently with probability $p$. The models $[n]_m$ and $[n]_p$, with $p = m/n$, are fairly similar: if a property holds for $[n]_p$ with probability $1 - o(1/\sqrt{m})$ then the same property holds almost surely for $[n]_m$ (this is known as Pittel's inequality, see [23, p. 17]).*

## 3.2 Refined results

### 3.2.1 A refinement of Theorem 3.1.1

The results in this paper are obtained by estimating the number of $B_3$-sets of given cardinality, in other words, by bounding $|\mathcal{Z}_n(t)|$. Since a threshold phenomenon occurs at $t \sim n^{1/5}$, we are particularly interested in bounding $|\mathcal{Z}_n(t)|$ for $t \sim n^{1/5}$. The following theorem compiles the upper bounds we obtain here. These bounds can be used to establish Theorem 3.1.1 and prove the upper bounds on Theorem 3.1.4.

**Theorem 3.2.1** *The following bounds apply to $|\mathcal{Z}_n(\cdot)|$:*

(*i*) *There is an absolute constant $c > 0$ such that for any $t \geq n^{1/4}(\log n)^2$, $|\mathcal{Z}_n(t)| \leq \left(\frac{cn}{t^3}\right)^t$.*

(*ii*) *For any $0 < \varepsilon < 1/5$ there exists $C_\varepsilon > 0$ such that for $t = C_\varepsilon(n \log n)^{1/5}$, we have $|\mathcal{Z}_n(t)| \leq n^{t(3/5+\varepsilon)}$.*

Part (*i*) is stated as Lemma 3.3.7 below and Part (*ii*) is Lemma 3.3.13. Let us show how Theorem 3.2.1(*i*) implies Theorem 3.1.1.

**Proof of Theorem 3.1.1** The total number of subsets of $[n]$ having fewer than $n^{1/4}(\log n)^2$ elements is $2^{o(n^{1/3})}$. Therefore we can focus on $B_3$-sets of

size $n^{1/4}(\log n)^2 \le t < n^{1/3}$. In particular, by Theorem 3.2.1$(i)$,

$$|\mathcal{Z}_n^3| \le 2^{o(n^{1/3})} + \sum_{t=n^{1/4}(\log n)^2}^{n^{1/3}} \left(\frac{cn}{t^3}\right)^t. \tag{3.6}$$

Now note that we have

$$\left(\frac{cn}{t^3}\right)^t \Big/ \left(\frac{cn}{(t-1)^3}\right)^{t-1} = \frac{cn}{t^3}\left(1-\frac{1}{t}\right)^{3(t-1)} \ge \frac{cn}{e^3 t^3}.$$

The ratio above is at least 2 provided $t^3 < cn/(2e^3)$, thus, for an appropriate choice of the absolute constant $C > 0$,

$$\sum_{t=1}^{(cn/2e^3)^{1/3}} \left(\frac{cn}{t^3}\right)^t \le 2\left(\frac{cn}{cn/2e^3}\right)^{(cn/2e^3)^{1/3}} = 2\cdot\left(2e^3\right)^{(cn/2e^3)^{1/3}} \ll 2^{Cn^{1/3}}.$$

On the other hand,

$$\sum_{t=(cn/2e^3)^{1/3}}^{n^{1/3}} \left(\frac{cn}{t^3}\right)^t \le \sum_{t=(cn/2e^3)^{1/3}}^{n^{1/3}} (2e^3)^t \le n^{1/3}(2e^3)^{n^{1/3}} \ll 2^{Cn^{1/3}}.$$

Theorem 3.1.1 follows from (3.6) as we bounded the sum on the R-H-S of (3.6). □

### 3.2.2 A refinement of Theorem 3.1.4

The following Theorem is a direct corollary of Theorem 3.2.1:

**Theorem 3.2.2** *There is an absolute constant $C > 1$ such that for any value $n^{-1/4}(\log n)^6 < p \le 1$, almost surely,*

$$F_3([n]_p) \le C(pn)^{1/3}. \tag{3.7}$$

*For any $\varepsilon > 0$ there exists $C_\varepsilon > 0$ such that for any $3n^{-4/5} \le p \le \frac{1}{2}n^{-3/5-\varepsilon}$,*

*almost surely,*

$$F_3([n]_p) \leq C_\varepsilon (n \log n)^{1/5}. \tag{3.8}$$

*Moreover, the probability that the inequality above is not satisfied is* $o\big(2^{-n^{1/5}}\big)$.

The following proposition allows us to use estimates on $|\mathcal{Z}_n(t)|$ in Theorem 3.2.1 to immediately establish Theorem 3.2.2.

**Proposition 3.2.3** *The expected number of* $B_3$*-sets of cardinality* $t$ *in* $[n]_p$ *is*

$$p^t \, |\mathcal{Z}_n(t)|.$$

*In particular,* $\mathbf{P}\big[F_3([n]_p) \geq t\big] \leq p^t \, |\mathcal{Z}_n(t)|$. □

We are now able to give a proof of the upper bound on $F_3([n]_m)$ given by Theorem 3.1.4 (see eqs. (3.3), (3.5)).

**Proof of the upper bound in Theorem 3.1.4** In this proof, we will use Theorem 3.2.2 to establish the non-trivial upper bounds of Theorem 3.1.4, namely, the upper bound function $b_2(a)$. We recall that Remark 3.1.6 links the Binomial random model $[n]_p$, appearing in Theorem 3.2.2, with the random model $[n]_m$ which appears in Theorem 3.1.4.

- Suppose $0 \leq a \leq 1/5$: $b_2(a) = a$ is a trivial upper bound as $F_3([n]_m) \leq m$ with probability 1.

- Suppose $1/5 \leq a \leq 2/5$: the upper bound $b_2(a) = 1/5$ follows from the second part of Theorem 3.2.2 (see eq. (3.8)). Indeed, we may fix $\varepsilon > 0$ arbitrarily small, and by Theorem 3.2.2, we have almost surely $F_3([n]_p) < n^{1/5+o(1)}$ for $3n^{-4/5} < p < \frac{1}{2}n^{-3/5-\varepsilon}$. Consequently, almost surely $F_3([n]_m) < n^{1/5+o(1)}$ for all $m \leq n^{2/5}$.

- Suppose $2/5 < a \leq 9/20$: for $p > \frac{1}{2}n^{-3/5-\varepsilon}$, we cannot directly apply Theorem 3.2.2 and for this reason we will employ the following strategy. In addition to sampling $[n]_p$, also pick randomly a map $\phi\colon [n] \to [k]$,

where $k = 2pn^{3/5+\varepsilon}$. For each $j = 1, \ldots, k$, let $R_j = \phi^{-1}(j) \cap [n]_p$. Note that each set $R_j$ follows the distribution $[n]_{p/k}$ and that any $B_3$-set $S \subset [n]_p$ induces $B_3$-sets $S_j = S \cap R_j$, $j = 1, \ldots, k$. Since $p/k = \frac{1}{2}n^{-3/5-\varepsilon}$, by Theorem 3.2.2, the probability that a given $R_j$ contains a $B_3$-set $S_j$ with more than $C_\varepsilon(n \log n)^{1/5}$ elements is $o\big(2^{-n^{1/5}}\big)$. Consequently, by the union bound over $j = 1, \ldots, k$, almost surely,

$$\begin{aligned} F_3([n]_p) &\leq \sum_{j=1}^{k} F_3(R_j) \leq kC_\varepsilon(n \log n)^{1/5} \\ &= pn \cdot 2n^{-2/5+\varepsilon} C_\varepsilon(n \log n)^{1/5} \\ &= pn \cdot n^{-1/5+o(1)}. \end{aligned} \tag{3.9}$$

In other words, almost surely, $F_3([n]_m) \leq mn^{-1/5+o(1)}$ and therefore $b_2(a) = a - 1/5$ is an upper bound for this range.

- Suppose $9/20 \leq a \leq 3/4$: since $F_3([n]_m)$ is monotone with respect to $m$, our upper bound $b_2(\cdot)$ should be monotone as well. Since $b_2(\cdot)$ satisfies $b_2(9/20) = 9/20 - 1/5 = 1/4$ by the previous statement and, as we show next, $b_2(3/4) = 1/4$, this monotonicity implies that $b_2(a) = 1/4$ in this range.

- Suppose $3/4 \leq a \leq 1$: the upper bound $b_2(a)$ follows immediately[1] the first part of Theorem 3.2.2 (see eq. (3.7)).

□

The lower bounds on Theorem 3.1.4 will be proved in Section 3.4.

[1]The probability that (3.7) holds is at least $1 - \exp\{-\Omega\big((pn)^{1/3}\big)\}$ and therefore Remark 3.1.6 applies here.

## 3.3 Proof of Theorem 3.2.1

The proof of Theorem 3.2.1 uses the following strategy. Suppose that a $B_3$-set $S \subset [n]$ of cardinality $s$ is given and one would like to extend it to a larger $B_3$-set. We will show that if $S$ satisfies a boundedness condition (see Definition 3.3.9 below), then the number of such extensions is fairly small. Moreover, we also prove that almost all $B_3$-sets of cardinality $s$ are sufficiently bounded in the sense of Definition 3.3.9. Consequently, in order to provide an upper bound for the number $\mathcal{Z}_n(t)$ of $B_3$-sets of size $t$ in $[n]$, for some $t > s$, we

($i$) estimate the number of $B_3$-sets which are not bounded in the sense of Definition 3.3.9 and account for at most $\binom{n}{t-s}$ possible extensions of each of those sets;

($ii$) for each bounded $B_3$-set $S$, we estimate the number of extensions of $S$ to a $B_3$-set of cardinality $t$.

To establish ($ii$) we will describe a graph-based approach for bounding the number of extensions of an arbitrary $B_3$-set $S$. This approach will shape Definition 3.3.9 below.

If two distinct elements $x, y \in [n] \setminus S$ satisfy

$$x + a_1 + a_2 = y + b_1 + b_2, \text{ for some } \{a_1, a_2\}, \{b_1, b_2\} \in \binom{S}{2}, \tag{3.10}$$

then $S \cup \{x, y\}$ is clearly not a $B_3$-set. This simple observation motivates our next definition.

**Definition 3.3.1.** *The* collision graph $\mathrm{CG}_S$ *has vertex set* $[n]\setminus S$ *and edges given by all pairs of distinct elements* $x, y \in [n] \setminus S$ *satisfying* (3.10). *By construction, any set of elements of* $[n] \setminus S$ *which extends* $S$ *to a larger* $B_3$*-set must induce an independent set in* $\mathrm{CG}_S$.

The following lemma provides an upper bound on the number of independent sets of graphs that have many edges in each sufficiently large vertex subset

(see (3.12)). The proof will be given in Section 3.3.1.

**Lemma 3.3.2** *Let $\delta, \beta > 0$, and $q \in \mathbb{N}$ be numbers satisfying*

$$(1+2\beta)^q \delta > 1. \tag{3.11}$$

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

$$e_G(A) \geq \beta |A|^2 \text{ for all } A \subset V, |A| \geq \delta |V|. \tag{3.12}$$

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

$$\binom{|V|}{q}\binom{\delta |V|}{m}. \tag{3.13}$$

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

When we apply Lemma 3.3.2 to $\mathrm{CG}_S$ we will take $m = Cq$, for a large constant $C$, to take advantage of the upper bound (3.13). In condition (3.12), there is a trade-off between $\beta$ (larger is better) and $\delta$ (smaller is better) which needs to be optimized.

In order to apply Lemma 3.3.2 to $\mathrm{CG}_S$ we first need to characterize the $B_3$-sets $S$ which ensure that condition (3.12) holds for $\mathrm{CG}_S$. The next lemma will be a step in this direction. First, we need a definition.

**Definition 3.3.3.** *Let $\widetilde{\mathrm{CG}}_S$ denote a multigraph version of $\mathrm{CG}_S$ where the multiplicity of a pair of distinct $x, y \in [n] \setminus S$ is given by the number of pairs $\big(\{a_1, a_2\}, \{b_1, b_2\}\big) \in \binom{S}{2}^2$ that satisfy* (3.10).

**Lemma 3.3.4** *For every $A \subset [n] \setminus S$ with $|A| \geq 16n/s^2$, we have*

$$e_{\widetilde{\mathrm{CG}}_S}(A) \geq \frac{s^4}{100n}|A|^2, \tag{3.14}$$

*where the edges are counted with multiplicity.*

The proof of Lemma 3.3.4 will be given in Section 3.3.2.

In view of Lemma 3.3.4, if the maximum multiplicity of an edge in $\widetilde{\mathrm{CG}_S}$ is at most $r$ then the graph $\mathrm{CG}_S$ satisfies the conditions of Lemma 3.3.2 with $\beta = \frac{s^4}{100rn}$ and $\delta = 16/s^2$. Consequently, we are interested in bounding the multiplicity of the edges of $\widetilde{\mathrm{CG}_S}$.

For any $z \in \mathbb{Z}$, let

$$R_S(z) = \left| \left\{ \big(\{a_1, a_2\}, \{b_1, b_2\}\big) \in \binom{S}{2}^2 \ : \ z = a_1 + a_2 - b_1 - b_2 \right\} \right|. \tag{3.15}$$

By construction, the multiplicity of a pair $\{x, y\}$ in the graph $\widetilde{\mathrm{CG}_S}$ is given by $R_S(x - y) = R_S(y - x)$.

We define

$$\mathbf{R}_S^* = \max_{0 \neq z \in \mathbb{Z}} \big\{ R_S(z) \big\} \tag{3.16}$$

and show that $\mathbf{R}_S^* = \Theta(s)$ in the following proposition.

**Proposition 3.3.5** *A $B_3$-set $S$ of cardinality $s$ satisfies $s - 2 \leq \mathbf{R}_S^* \leq 2s - 2$.*

**Proof** Let $S^-$ be the set of all differences of two elements of $S$. Note that for $z \in S^-$ (say, $z = a_1 - b_1$), we have $R_S(z) \geq s - 2$ since $z = a_1 + c - b_1 - c$ for all $c \in S \setminus \{a_1, b_1\}$. This shows the lower bound for $\mathbf{R}_S^*$.

For an arbitrary $z \neq 0$, let us estimate $R_S(z)$ by bounding the number of solutions to

$$z = a_1 + a_2 - b_1 - b_2, \quad \big(\{a_1, a_2\}, \{b_1, b_2\}\big) \in \binom{S}{2}^2.$$

We start by considering solutions of *first type*, namely, solutions with $\{a_1, a_2\} \cap \{b_1, b_2\} \neq \emptyset$. Without loss of generality, assume that $a_2 = b_2 = c$, $c \neq a_1, b_1$. In this case, $z = a_1 - b_1$. Given that $S$ is a $B_3$-set, there is at most one pair $(a_1, b_1) \in S^2$ which satisfies $z = a_1 - b_1$. It follows that the number of solutions of first type is at most $s - 2$ (the number of choices for $c$).

Now consider solutions of *second type*, namely, solutions with $\{a_1, a_2\} \cap \{b_1, b_2\} = \emptyset$. For each fixed value of $b_2 \in S$, the fact that $S$ is a $B_3$-set implies

that there is at most one solution $(\{a_1, a_2\}, b_1) \in \binom{S}{2} \times S$, $b_1 \neq a_1, a_2$, to the equation $a_1 + a_2 - b_1 = z - b_2$. Hence the number of solutions of second type is at most $s$ (the number of choices for $b_2$). The proposition follows from the bounds on the number of solutions of the first and second types. □

The following is an immediate corollary of Lemma 3.3.4 and Proposition 3.3.5. Indeed, Proposition 3.3.5 implies that the multiplicity of any edge of $\widetilde{\mathrm{CG}_S}$ is at most $2s - 2 < 2s$.

**Corollary 3.3.6** *If $S$ is a $B_3$-set then for every $A \subset [n] \setminus S$ with $|A| \geq 16n/s^2$, we have*

$$e_{\mathrm{CG}_S}(A) \geq \frac{s^3}{200n}|A|^2.$$

□

We are now ready to obtain the following result for $B_3$-sets of size $t \geq n^{1/4}$.

**Lemma 3.3.7** *There is an absolute constant $c > 1$ such that for any $t \geq n^{1/4}(\log n)^2$,*

$$|\mathcal{Z}_n(t)| \leq \left(\frac{cn}{t^3}\right)^t.$$

**Proof** Since for any $t$ satisfying $\binom{t}{3} > n$ there is no $B_3$-set of size $t$ in $[n]$, we can assume that $n^{1/4}(\log n)^2 \leq t < n^{1/3}$.

It will be convenient to first deal with the case where $t$ can be written as $t = 2^\ell s_0$, where $\ell \in \mathbb{N}$ and

$$s_0 \in \left[10(n \log n)^{1/4}, 20(n \log n)^{1/4}\right). \tag{3.17}$$

Notice that $\ell = \log_2(t/s_0)$ satisfies $\Omega(\log \log n) \leq \ell \leq O(\log n)$. We will deal with the general case at the end of this proof.

Set $s_i = 2^i s_0$ for $i = 1, \ldots, \ell$ and note that by definition, $s_\ell = t$. We will find an upper bound for $|\mathcal{Z}_n(t)|$ by constructing sequences of $B_3$-sets $S_0 \subset S_1 \subset \cdots \subset S_\ell$ with $|S_i| = s_i$ for all $i = 0, \ldots, \ell$. More precisely, we can bound the

number of choices of $S_0$ by $\binom{n}{s_0}$ and then use Lemma 3.3.2 to bound the number of extensions of $S_i$ into $S_{i+1}$ for all $i$.

Let us now estimate the number of extensions of a $B_3$-set $S_i$ into the larger $B_3$-set $S_{i+1}$, for $i = 0, \dots, \ell - 1$. By Corollary 3.3.6, the graph $\mathrm{CG}_{S_i}$ is such that for all $A \subset [n]$, $|A| \geq \delta_i n$, $\delta_i = 16/s_i^2$,

$$e_{\mathrm{CG}_{S_i}} \geq \beta_i \, |A|^2, \quad \text{with } \beta_i = \frac{s_i^3}{200n}.$$

Let

$$q_i = \frac{\log n}{\beta_i} = \frac{200}{s_i^3} n \log n = \frac{200}{8^i s_0^3} n \log n \overset{(3.17)}{\leq} \frac{s_0}{8^i \cdot 50}. \tag{3.18}$$

and observe that

$$(1 + 2\beta_i)^{q_i} \delta_i > \left( e^{\frac{2\beta_i}{1+2\beta_i}} \right)^{q_i} \delta_i > e^{\beta_i q_i} \delta_i = \delta_i n > 16n/t^2 > 16n^{1/3} > 1.$$

Consequently, $\mathrm{CG}_{S_i}$, $\delta_i$, $\beta_i$, and $q_i$ satisfy the conditions of Lemma 3.3.2. Note that $S_{i+1} \setminus S_i$ must be an independent set in $\mathrm{CG}_{S_i}$ with cardinality $s_{i+1} - s_i = s_i$. Therefore, applying Lemma 3.3.2 with $m = s_i - q_i$ shows that the number of extensions of $S_i$ into a $B_3$-set $S_{i+1}$ is at most

$$\begin{aligned} \binom{n}{q_i} \binom{\delta_i n}{s_i - q_i} &\leq n^{q_i} \left( \frac{e\delta_i n}{s_i - q_i} \right)^{s_i - q_i} = n^{q_i} \left( \frac{16en}{s_i^2 (s_i - q_i)} \right)^{s_i - q_i} \leq \\ &\overset{(3.18)}{\leq} n^{q_i} \left( \frac{32en}{s_i^3} \right)^{s_i - q_i} = n^{s_i} \left( \frac{32e}{s_i^3} \right)^{s_i - q_i}. \end{aligned} \tag{3.19}$$

Recalling that $t = s_0 + \sum_{i=0}^{\ell-1} s_i$, it follows that

$$\begin{aligned} |\mathcal{Z}_n(t)| &\leq n^{s_0 + \sum_{i=0}^{\ell-1} s_i} \prod_{i=0}^{\ell-1} \left( \frac{32e}{s_i^3} \right)^{s_i - q_i} \\ &= n^t \prod_{i=0}^{\ell-1} \left( \frac{32e \cdot 8^{\ell - i}}{t^3} \right)^{s_i - q_i} \\ &\leq n^t \cdot \left( t^{-3} \right)^{\sum_{i=0}^{\ell-1} (s_i - q_i)} \cdot \left( 2^{10} \right)^{\sum_{i=0}^{\ell-1} (\ell - i) s_i}. \end{aligned} \tag{3.20}$$

From (3.18), it follows that $\sum_{i=0}^{\ell-1} q_i < s_0/25$ and thus $\sum_{i=0}^{\ell-1}(s_i - q_i) \geq t - 1.04s_0$. On the other hand, $\sum_{i=0}^{\ell-1}(\ell - i)s_i = t\sum_{i=0}^{\ell-1}(\ell - i)2^{-(\ell-i)} < 3t$.

Notice that since $t \geq 2s_0 \log n$, we have

$$\left(t^{-3}\right)^{\sum_{i=0}^{\ell-1}(s_i-q_i)} = t^{3(1.04s_0-t)} \leq t^{3(t/\log n - t)} = \left(t^{3/\log n - 3}\right)^t < \left(e/t^3\right)^t.$$

Therefore,

$$|\mathcal{Z}_n(t)| \leq \left(\frac{2^{30}en}{t^3}\right)^t. \tag{3.21}$$

To conclude the lemma, we just need to deal with the case when $t$ cannot be expressed as $2^\ell s_0$ with $s_0$ as in (3.17). For this, let $t' \in (t/4, t/2]$ be such that $t' = 2^\ell s_0$ with $s_0$ as in (3.17). Then by using again Corollary 3.3.6 and Lemma 3.3.2 we have

$$|\mathcal{Z}_n(t)| \leq |\mathcal{Z}_n(t')| \binom{n}{q} \binom{16n/(t')^2}{t - t' - q},$$

where $q = t/\log n$. Applying (3.21) to $|\mathcal{Z}_n(t')|$ in the bound above yields the lemma. □

While the bound given by Proposition 3.3.5 is best possible up to a constant multiple, one can expect $R_S(z)$ to be significantly smaller than $s$ for $z \notin S - S$.

**Remark 3.3.8.** *With regards to $R_S(\cdot)$, elements $z \in S - S$ are considered* degenerate. *This is because for such elements, $R_S(z)$ is always large due to many trivial solutions of the form $z = a + c - b - c$ with $a, b, c \in S$.*

In view of Remark 3.3.8 we define

$$\mathbf{R}_S = \max\left\{R_S(z) \colon :: z \in \mathbb{Z} \setminus S^-\right\}, \tag{3.22}$$

recalling that $S^-$ denotes the set of all differences of two elements of $S$. We will use better bounds on $\mathbf{R}_S$ to obtain a stronger version of Corollary 3.3.6.

**Definition 3.3.9.** *A $B_3$-set $S \subset [n]$ is* $r$-bounded *if* $\mathbf{R}_S < r$.

The following corollary of Lemma 3.3.4 provides a conclusion that is stronger than Corollary 3.3.6 at the cost of requiring a better bound on $\mathbf{R}_S$.

**Corollary 3.3.10** *If $S$ is an $r$-bounded $B_3$-set then for every $A \subset [n] \setminus S$, $|A| \geq 10000rn/s^2$, we have*

$$e_{\mathrm{CG}_S}(A) \geq \frac{s^4}{100r^2n}|A|^2.$$

The proof of Corollary 3.3.10 will be given in Section 3.3.3. In view of Corollary 3.3.10, we are now primarily interested in $r$-bounded $B_3$-sets. This motivates our next definition.

**Definition 3.3.11.** *A $B_3$-set $S^\sharp$ is called* $(s^*, r)$-bad *if there is no $r$-bounded $B_3$-set $S^* \subset S^\sharp$ with $|S^*| = s^*$.*

The following lemma provides an upper bound on the number of $(s^*, r)$-bad sets of given size. The proof is given in Section 3.3.4.

**Lemma 3.3.12** *For every $\varepsilon > 0$ there exists $r = r(\varepsilon) \in \mathbb{N}$, $c_\varepsilon \geq \max\{100, 1/\varepsilon\}$, and $n_0 = n_0(\varepsilon)$ such that for all $n \geq n_0$ and $s^* = (n \log n)^{1/5}$, the number of $(s^*, r)$-bad sets $S^\sharp \subset [n]$, with $|S^\sharp| = m \geq c_\varepsilon s^*$, is at most*

$$\left(\frac{Cn^{4/5+\varepsilon}}{m}\right)^m,$$

*where $C > 0$ is an absolute constant.*

Using Lemma 3.3.12 we can obtain the following upper bound on $|\mathcal{Z}_n(t)|$ for $t = C(n \log n)^{1/5}$.

**Lemma 3.3.13** *For any $0 < \varepsilon < 1/5$, there exists $n_0 = n_0(\varepsilon)$, and $C_\varepsilon > 0$ such that for all $n \geq n_0$, and $t = C_\varepsilon \cdot (n \log n)^{1/5}$, we have*

$$|\mathcal{Z}_n(t)| \leq \frac{1}{2}n^{t(3/5+\varepsilon)} + n^{t(2/5+\varepsilon)}. \tag{3.23}$$

**Proof** Let $c_\varepsilon \geq \max\{100, 1/\varepsilon\}$, $C > 0$, $r = r(\varepsilon) \in \mathbb{N}$, and $s^* = (n \log n)^{1/5}$ as

in Lemma 3.3.12. We take $C_\varepsilon$ satisfying

$$C_\varepsilon \gg \max\{c_\varepsilon, r^4, C\}. \tag{3.24}$$

By Lemma 3.3.12, the number of $(s^*, r)$-bad sets $S^\sharp$ of cardinality $t$ is at most

$$\left(\frac{Cn^{4/5+\varepsilon}}{t}\right)^t \leq \left(\frac{Cn^{3/5+\varepsilon}}{C_\varepsilon}\right)^t \overset{(3.24)}{\leq} \frac{1}{2}n^{t(3/5+\varepsilon)},$$

which accounts for the first term in the bound (3.23). All other $B_3$-sets of $\mathcal{Z}_n(t)$ must contain an $r$-bounded set $S$ of size $s^*$. Hence, we may count those sets by starting with an $r$-bounded set $S$ of size $s^*$ and counting the number of extensions of $S$ to a $B_3$-set of size $t = C_\varepsilon s^*$.

To count the number of extensions, we first invoke Corollary 3.3.10, which guarantees that $\mathrm{CG}_S$ satisfies

$$e_{\mathrm{CG}_S}(A) \geq \frac{(s^*)^4}{100r^2n}|A|^2 \quad \text{for all } A \subset [n], |A| \geq 10000rn/(s^*)^2. \tag{3.25}$$

We now set

$$\begin{array}{rclcrcl} q & = & C_\varepsilon^{1/2}s^* & & m & = & t-q-s^* \\ \beta & = & \dfrac{(s^*)^4}{100r^2n} & & \delta & = & \dfrac{10000r}{(s^*)^2}. \end{array} \tag{3.26}$$

Note that for our choice of $C_\varepsilon$ in (3.24),

$$(1+2\beta)^q\delta \geq e^{\beta q}\delta = \delta \cdot \exp\left\{\frac{C_\varepsilon^{1/2}(s^*)^5}{100r^2n}\right\} = \delta \cdot \exp\left\{\frac{C_\varepsilon^{1/2}\log n}{100r^2}\right\} \gg 1. \tag{3.27}$$

Hence, $\delta$, $\beta$, and $q$ satisfy condition (3.11) of Lemma 3.3.2.

It follows by Lemma 3.3.2 that the number of independent sets in $\mathrm{CG}_S$ of cardinality $q+m = t-s^*$ is at most

$$\binom{n}{q}\binom{\delta n}{m} \leq \left(\frac{en}{q}\right)^q \Bigg(\underbrace{\frac{e\delta n}{t-q-s^*}}_{>s^*}\Bigg)^m \overset{(3.26)}{\leq} \left(\frac{en}{s^*}\right)^q \left(\frac{10000ern}{(s^*)^3}\right)^{t-q-s^*}.$$

Consequently, as there are at most $\binom{n}{s^*} \leq \left(\frac{en}{s^*}\right)^{s^*}$ choices for $S$, the number $N$ of $B_3$-sets of cardinality $t$ which are not $(s^*, r)$-bad satisfies

$$N \leq (10000ern)^t (s^*)^{-s^*-q-3(t-q-s^*)} \leq \big(O(r)\big)^t n^{t\{1-(3t-2q-2s^*)/(5t)\}}. \tag{3.28}$$

By our choice of $C_\varepsilon$, we have

$$\frac{3t-2q-2s^*}{5t} = \frac{3}{5} - \frac{2}{5}(C_\varepsilon^{-1/2} + C_\varepsilon^{-1}) < \frac{3-\varepsilon}{5}.$$

Therefore, for $n \geq n_0(\varepsilon)$ (with sufficiently large $n_0$),

$$N \leq \big(O(r)\big)^t n^{t(2+\varepsilon)/5} < n^{t(2/5+\varepsilon)}.$$

Recall that $N$ is the number of $B_3$-sets of cardinality $t$ which are not $(s^*, r)$-bad and that the $B_3$-sets of cardinality $t$ which are $(s^*, r)$-bad are accounted for by the first term in (3.23). Therefore the bound (3.23) follows by the inequality above and the lemma is proved. □

### 3.3.1 Proof of Lemma 3.3.2

For this proof, we will define tailored linear orders for every subset of $V = V(G)$ and use them to bound the number of independent sets. Roughly speaking we first show that under such linear orders, any independent set $I$, $|I| \geq q+1$, admits a unique *q-prefix*, which is a sequence of $q$ distinct elements from $I$ we define below. We later establish an upper bound on the number of independent sets of size $q+m$ that have the same $q$-prefix.

It will be convenient to assume (without loss of generality) that $V \subset [n]$. For any $V' \subset V$, let $\max(V')$ be the vertex of maximum degree in $G[V']$ of largest value as a number in $[n]$. With this definition we may construct a linear order $>_{V'}$ of $V'$ as $v_1, v_2, \ldots, v_m$, $m = |V'|$, by setting $v_1 = \max(V')$ and $v_{i+1} = \max\big(V' \setminus \{v_1, \ldots, v_i\}\big)$ for $1 \leq i < m$.

Given any independent set $I$ in $G$ with $|I| \geq q+1$, the *q-prefix* of $I$ is a

sequence $(v_1, \ldots, v_q)$ constructed as follows. Let $v_1$ denote the first element of $I$ in the linear order of $>_{V_1}$, where $V_1 = V$. For $1 \leq i \leq q$ let

$$V_{i+1} = \{v \in V_i \colon :: v_i >_{V_i} v, v \notin N_G(v_i)\} \tag{3.29}$$

and let $v_{i+1}$ denote the first element of $I \setminus \{v_1, \ldots, v_i\}$ that appears in $V_{i+1}$ in the order $>_{V_{i+1}}$.

To check that the procedure above does not stop before producing the desired sequence of length $q$, we prove by induction that

$$I \setminus \{v_1, \ldots, v_i\} \subset V_{i+1}, \quad \text{for } i = 1, \ldots, q. \tag{3.30}$$

Indeed, $v_1$ is chosen in such a way that $v_1 >_{V_1} w$ for all $v_1 \neq w \in I$, thus $I \setminus \{v_1\} \subset V_2$. If (3.30) holds for $i = j$, $1 \leq j < q$, then for every $w \in I \setminus \{v_1, \ldots, v_j\} \subset V_{j+1}$, $w \neq v_{j+1}$, we have $v_{j+1} >_{V_{j+1}} w$ and $w \notin N_G(v_{j+1})$. Hence $w \in V_{j+2}$ and (3.30) holds for $i = j + 1$.

Let us now estimate the number of independent sets $I$ of size $q + m$ that admit the same $q$-prefix $(v_1, \ldots, v_q)$. By (3.30), it follows that the number of choices for the $m$ elements in $I \setminus \{v_1, \ldots, v_q\}$ is at most $\binom{|V_{q+1}|}{m}$. Lemma 3.3.2 will follow once we prove that

$$|V_{q+1}| < \delta\,|V|. \tag{3.31}$$

Suppose for the sake of a contradiction that $|V_{q+1}| \geq \delta\,|V|$. By construction, for every $1 \leq i \leq q$, we have $D := \deg_{G[V_i]}(v_i) \geq \deg_{G[V_i]}(w)$ for every $w \in V_{i+1} \subset V_i$. Since $G[V_{i+1}] \subset G[V_i]$, it follows that $D \geq \Delta(G[V_{i+1}])$. Since by our assumption $|V_{i+1}| \geq |V_{q+1}| \geq \delta\,|V|$, condition (3.12) yields

$$D \geq \Delta(G[V_{i+1}]) \geq \frac{2e_G(V_{i+1})}{|V_{i+1}|} \geq 2\beta\,|V_{i+1}|.$$

From the definition of $V_{i+1}$ in (3.29) we have $N_{G[V_i]}(v_i) \cap V_{i+1} = \emptyset$ and $N_{G[V_i]}(v_i) \cup$

$V_{i+1} \subset V_i$. Thus, for every $1 \le i \le q$,

$$|V_i| \ge |N_{G[V_i]}(v_i)| + |V_{i+1}| = D + |V_{i+1}| \ge (1+2\beta)|V_{i+1}|.$$

Consequently, $|V| = |V_1| \ge (1+2\beta)^q |V_{q+1}| \ge (1+2\beta)^q \delta\,|V|$ and therefore $(1+2\beta)^q\delta \le 1$, contradicting the hypothesis of the lemma—see (3.11).

### 3.3.2 Proof of Lemma 3.3.4

Let $A \subset [n] \setminus S$, $|A| \ge 16n/s^2$, be an arbitrary subset. Consider the auxiliary labeled bipartite graph $\Gamma$ defined as follows. The vertex classes of $\Gamma$ are $A$ and a disjoint copy of $[3n]$. The edge set of $\Gamma$ is defined as

$$E(\Gamma) = \big\{(x,u) \in A \times [3n] \colon :: u = x + a_1 + a_2 \text{ for some } a_1, a_2 \in S, a_1 \ne a_2\big\}.$$

Note that because $S$ is a $B_3$-set, for fixed $x, u$, there is at most one solution $\{a_1, a_2\}$ for $u = x + a_1 + a_2$ with $a_1, a_2 \in S$. We will now argue that the multiplicity of a pair $\{x, y\} \in \binom{A}{2}$ in the multigraph $\widetilde{\mathrm{CG}}_S$, which by definition is $R_S(y-x)$, is given by the number of two-paths connecting $x$ to $y$ in $\Gamma$. Indeed, there is a bijection between pairs $\big(\{a_1,a_2\},\{b_1,b_2\}\big) \in \binom{S}{2}^2$ satisfying $y - x = a_1 + a_2 - b_1 - b_2$ and paths $xuy$ in $\Gamma$, where

$$u = x + a_1 + a_2 = y + b_1 + b_2.$$

Recalling the definition of $R_S(\cdot)$ in (3.15), it follows that $R_S(y-x)$ equals the number of two-paths connecting $x$ to $y$ in $\Gamma$.

Consequently, $e_{\widetilde{\mathrm{CG}}_S}(A)$ 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 $f(\alpha) = \binom{\alpha}{2} = \alpha(\alpha-1)/2$,

$$e_{\widetilde{\mathrm{CG}}_S}(A) \ge \sum_{u \in [3n]} \binom{\deg_\Gamma(u)}{2} \ge 3n \binom{e(\Gamma)/3n}{2}.$$

On the other hand, since $|A| \geq 16n/s^2$, we may assume that $s \geq 4$ and

$$e(\Gamma) = \sum_{x \in A} \deg_\Gamma(x) = |A| \binom{s}{2} = 8n\Big(1 - \frac{1}{s}\Big) \geq 6n.$$

It follows that $e(\Gamma)^2/36n^2 \geq e(\Gamma)/6n$ and thus,

$$e_{\widetilde{\mathrm{CG}}_S}(A) \geq n\binom{e(\Gamma)/3n}{2} \geq n\left(\frac{e(\Gamma)^2}{18n^2} - \frac{e(\Gamma)}{6n}\right) \geq \frac{e(\Gamma)^2}{36n} > \frac{s^4}{100n}|A|^2.$$

This concludes the proof of Lemma 3.3.4.

### 3.3.3 Proof of Corollary 3.3.10

Suppose that $S$ is an $r$-bounded $B_3$-set and that $A \subset [n] \setminus S$ is an arbitrary set with $|A| \geq 10000rn/s^2$. From the boundedness of $S$ it follows that the multiplicity of an edge $\{x, y\} \in \widetilde{\mathrm{CG}}_S[A]$ with $x - y \notin S - S$ is at most $r$. On the other hand, edges $\{x, y\} \in \widetilde{\mathrm{CG}}_S[A]$ with $x - y \in S - S$ are degenerate (recall Remark 3.3.8) and have large multiplicity. For this reason, we define the following auxiliary subgraph:

$$H = \left\{\{x, y\} \in \binom{A}{2} : \ :: x - y \in S - S\right\}.$$

We now split the proof in the two cases according to whether

$$|H| < \frac{s^3|A|^2}{200rn} \tag{3.32}$$

holds or not.

First let us assume that (3.32) holds. In this case, by Proposition 3.3.5 and equation (3.16), the total multiplicity of edges in $\widetilde{\mathrm{CG}}_S[A]$ that also appear in $H$ is at most

$$|H| \cdot 2s < \frac{s^4}{100rn}|A|^2 < \frac{e_{\widetilde{\mathrm{CG}}_S}(A)}{2},$$

where the last inequality follows from Lemma 3.3.4. Therefore, recalling that the multiplicity of edges of $\widetilde{\mathrm{CG}}_S[A]$ not in $H$ is bounded by $r$, it follows by Lemma 3.3.4 that

$$e_{\mathrm{CG}_S}(A) \geq \frac{e_{\widetilde{\mathrm{CG}}_S}(A)}{2r} \geq \frac{s^4}{200rn}|A|^2,$$

and the corollary follows in this case.

Now we assume that (3.32) does not hold. Define $\mathcal{T}$ as the set of all triples $(u, \{x, y\}) \in A \times \binom{A}{2}$ with $x - y \notin S - S$ and such that there exists distinct $a, b, c, d \in S$ such that $x - u = a - b$ and $y - u = c - d$.

**Claim 3.3.14** *For any $u \in A$ with $\deg_H(u) \geq 8s$ there are at least $\deg_H(u)^2/4$ triples $(u, \{x, y\}) \in \mathcal{T}$.*

We now prove the claim. Fix an arbitrary element $u \in A$ with degree $d = \deg_H(u) \geq 8s$ and let $\{v_1, \ldots, v_d\}$ denote the neighborhood of $x$ in $H$. For each $v_i$, associate a pair $(\alpha_i, \beta_i) \in S^2$ that is the unique solution to $v_i - u = \alpha_i - \beta_i$. For any fixed $v_i$, $1 \leq i \leq d$ there are at least $d - 4s \geq d/2$ elements $v_j$, with $1 \leq j \leq d$, satisfying $\{\alpha_i, \beta_i\} \cap \{\alpha_j, \beta_j\} = \emptyset$ (which implies that $\alpha_i, \beta_i, \alpha_j, \beta_j$ are all distinct).

We will also show that $v_i - v_j \notin S - S$. Assuming otherwise means that there exist $\gamma_1, \gamma_2 \in S$ such that

$$\alpha_j + \beta_i + \gamma_1 = \alpha_i + \beta_j + \gamma_2.$$

Due to the fact that $S$ is a $B_3$-set, we have $\{\alpha_j, \beta_i, \gamma_1\} = \{\alpha_i, \beta_j, \gamma_2\}$. Since by construction, $\alpha_i \neq \beta_i$, $\alpha_j \neq \beta_j$, this forces $\{\alpha_i, \beta_i\} \cap \{\alpha_j, \beta_j\} \neq \emptyset$, a contradiction.

The number of (unordered) pairs $x = v_i, y = v_j$ as above is at least $\frac{1}{2}d(d - 4s) \geq d^2/4$. Note that for all such pairs $\{x, y\}$, the triple $(u, \{x, v\}) \in \mathcal{T}$. The claim is now proved. □

Since we are assuming that (3.32) does not hold, the average degree of $H$ is at least $2\frac{s^3|A|}{200rn} > 32s$. In particular, there exists a subgraph $H' \subset H$ with

minimum degree $\delta(H') \geq 8s$ and $|H'| \geq |H|/2$. Hence, due to Claim 3.3.14,

$$|\mathcal{T}| \geq \sum_{u \in V(H')} \frac{d_{H'}(u)^2}{4} \geq \frac{|H'|^2}{|V(H')|} \geq \frac{|H|^2}{4\,|A|} > \frac{s^6|A|^3}{10^6 r^2 n^2} \geq \frac{s^4}{100rn}|A|^2. \tag{3.33}$$

We will now estimate $e_{\mathrm{CG}_S}(A)$ by counting for every $\{x, y\} \in \mathrm{CG}_S[A]$, $x - y \notin S - S$, how many $u \in A$ form a triple $(u, \{x, y\}) \in \mathcal{T}$. For every $u$ forming a triple $(u, \{x, y\}) \in \mathcal{T}$ we have $x - u, y - u \in S - S$ and if we let $a_1, a_2, b_1, b_2 \in S$ be the unique values for which $x - u = a_1 - b_1$ and $y - u = b_2 - a_2$, then $x - y = a_1 + a_2 - b_1 - b_2$ and $a_1, a_2, b_1, b_2$ are all distinct. Consequently, to each $u$ such that $(u, \{x, y\}) \in \mathcal{T}$ there is a distinct solution to (3.10).

Since $x - y \notin S - S$ and $S$ is $r$-bounded, there can be at most $r$ solutions to (3.10) and thus at most this many elements $u$ with $(u, \{x, y\}) \in \mathcal{T}$. Thus

$$e_{\mathrm{CG}_S}(A) \geq \frac{|\mathcal{T}|}{r} \overset{(3.33)}{\geq} \frac{s^4}{100r^2n}|A|^2.$$

The corollary follows.

### 3.3.4 Proof of Lemma 3.3.12

The strategy of the proof of Lemma 3.3.12 is the following. For an ordered $B_3$-set $S^\sharp$ consider the following procedure that constructs a maximal subset $S \subset S^\sharp$ which is $r$-bounded. Assume that $S^\sharp = (x_1, \ldots, x_m)$. Take $S_0 = \emptyset$ and, for $1 \leq i \leq m$, define the set $S_i$ as $S_{i-1} \cup \{x_i\}$ if adding $x_i$ to $S_{i-1}$ does not cause the set to lose the $r$-boundedness property. The set $S_m$ is, by construction, $r$-bounded and maximal.

If the set $S \subset S^\sharp$ satisfies $|S| \geq s^*$ then clearly $S^\sharp$ is not $(s^*, r)$-bad. Hence, counting the pairs $(S, S^\sharp)$ with $|S| < s^*$ provides an upper bound for the number of ordered $(s^*, r)$-bad sets.

For any integer $\ell \geq 1$, define

$$Q_{S,\ell} = \sum_{z \notin S-S} R_S(z)^\ell.$$

In the definition above the sum is over $z \notin S - S$ because we intend to use $Q_{S,\ell}$ to eventually bound $\mathbf{R}_S$ (defined in eq. (3.22)) and hence we may ignore the degenerate cases $z \in S - S$ (see Remark 3.3.8).

Lemma 3.3.12 will follow from the following claim, which we will prove later.

**Claim 3.3.15** *If $S$ is a $B_3$-set, $s = |S| \leq s^*$, and $Q_{S,\ell} \leq 2^{\ell+2}\binom{s}{4}$ then for all but at most $3^\ell (s^*)^4$ elements $x$ we have $Q_{S\cup\{x\},\ell} \leq 2^{\ell+2}\binom{s+1}{4}$.*

The following observation will be useful to us later.

**Claim 3.3.16** *Suppose that $\ell = \frac{\varepsilon}{2}\log n$ and $r = r(\varepsilon) = e^{2+2/\varepsilon}$. If for a $B_3$-set $S$ of cardinality $s \leq s^*$ we have $Q_{S,\ell} \leq 2^{\ell+2}\binom{s}{4}$, then $\mathbf{R}_S \leq r$.*

**Proof** For any $z \notin S - S$,

$$R_S(z)^\ell \leq Q_{S,\ell} \leq 2^{\ell+2}\binom{s}{4} < 2^{\ell+2} n \leq e^{(\ell+2)+2\ell/\varepsilon}.$$

Therefore $R_S(z) \leq e^{2+2/\varepsilon} = r$. Since $z \notin S - S$ was arbitrary, we conclude that $\mathbf{R}_S \leq r$. □

From now on we set $\ell = \frac{\varepsilon}{2}\log n$ and $r = r(\varepsilon) = e^{2+2/\varepsilon}$ as in Claim 3.3.16.

We now obtain an upper bound on the number of ordered sets $S^\sharp$ of size $m$ which are $(s^*, r)$-bad by sequentially constructing a pair $(S, S^\sharp)$ with

- $S \subset S^\sharp$,
- $s = |S| \leq s^*$,
- $Q_{S,\ell} \leq 2^{\ell+2}\binom{s}{4}$, which, by Claim 3.3.16, implies that $S$ is $r$-bounded,
- $S^\sharp$ is a $B_3$-set.

Start with $S = S^\sharp = \emptyset$, which trivially satisfies all the conditions above. Whenever an element $x$ is added to $S^\sharp$ we check whether the conditions above would be satisfied with $S' = S \cup \{x\}$ in place of $S$. If so, we also add $x$ to $S$ (otherwise, $S$ stays the same).

By Claim 3.3.15, for all but at most $n^{4/5+\varepsilon}$ choices of $x$, if we add $x$ to $S^\sharp$ then we also add $x$ to $S$. Since $S^\sharp$ must be $(s^*, r)$-bad, the number of times we may add an element to $S$ must be less than $s^*$. To find an upper bound on the number of ordered sets $S^\sharp$ of size $m$ which produce a corresponding $S \subset S^\sharp$ of size $s < s^*$ we will do the following:

- Choose the $s$ steps in which an element is added to $S$ in $\binom{m}{s}$ ways.
- In each of the $m - s$ steps in which the new element $x$ is not added to $S$ we extend $S^\sharp$ by picking $x$ among the at most $n^{4/5+\varepsilon}$ elements that are not eligible to be added to $S$.
- In the steps in which $x$ is added to $S$, there are at most $n$ choices for $x$.

The total number of choices (for fixed $s < s^*$) is

$$\binom{m}{s} \left(n^{4/5+\varepsilon}\right)^{m-s} n^s.$$

By summing over all choices of $s < s^*$ and accounting for the $m!$ permutations of the set $S^\sharp$, the total number of $(s^*, r)$-bad sets $S^\sharp$ of cardinality $m$ is at most

$$[*] := \frac{1}{m!} \sum_{s=0}^{s^*-1} \binom{m}{s} \left(n^{4/5+\varepsilon}\right)^{m-s} n^s.$$

Under the assumption that $m \geq c_\varepsilon s^*$ with $c_\varepsilon \geq 2$, we have $m \geq 2s^*$ and thus the binomial coefficients in the sum above are strictly increasing. Moreover, the term $\left(n^{4/5+\varepsilon}\right)^{m-s} n^s$ increases by $n^{1/5-\varepsilon} \gg 2$ for a unit increment of $s$. Hence,

the sum $[*]$ can be bounded by

$$[*] \leq \frac{1}{m!}\binom{m}{s^*}(n^{4/5+\varepsilon})^{m-s^*}n^{s^*} < c^m\left(\frac{n^{4/5+\varepsilon}}{m-s^*}\right)^{m-s^*}\left(\frac{n}{s^*}\right)^{s^*}, \tag{3.34}$$

where $c > 0$ is an absolute constant. Setting $c_\varepsilon$ sufficiently large (say, $c_\varepsilon > 1 + 1/\varepsilon$) yields

$$(n/s^*)^{s^*} < n^{s^*} < n^{s^*\cdot\varepsilon(c_\varepsilon-1)} \leq n^{\varepsilon(m-s^*)}$$

and since $m - s^* \geq m/2$, we can simplify the denominator in (3.34) thus obtaining

$$[*] \leq (2c)^m\left(\frac{n^{4/5+2\varepsilon}}{m}\right)^{m-s^*}.$$

When $m > n^{4/5+2\varepsilon}$ and $n \geq n_0$ is large enough, the R-H-S of the above inequality is $< 1$ and thus $[*] = 0$. On the other hand, when $m \leq n^{4/5+2\varepsilon}$, we have $[*] \leq \left(\frac{2c\,n^{4/5+2\varepsilon}}{m}\right)^m$. In both cases, by setting $C = 2c$ and rescaling $\varepsilon$, Lemma 3.3.12 follows. It only remains to prove Claim 3.3.15, which we do next.

**Proof of Claim 3.3.15** Suppose that $S$ is a $B_3$-set, $s = |S| \leq s^*$, and $Q_{S,\ell} \leq 2^{\ell+2}\binom{s}{4}$. We would like to show that for most choices of $x \notin S$, with $S' = S \cup \{x\}$ a $B_3$-set, we have $Q_{S',\ell} \leq 2^{\ell+2}\binom{s+1}{4}$. To that end, it is enough to show that for most choices of $x$,

$$Q_{S',\ell} - Q_{S,\ell} \leq 2^{\ell+2}\binom{s}{3} \tag{3.35}$$

since $\binom{s}{4} + \binom{s}{3} = \binom{s+1}{4}$. Note that

$$\begin{aligned} Q_{S',\ell} - Q_{S,\ell} &= \sum_{z\notin S'-S'} R_{S'}(z)^\ell - \sum_{z\notin S-S} R_S(z)^\ell \\ &\leq \sum_{z\notin S'-S'} \big(R_{S'}(z)^\ell - R_S(z)^\ell\big), \end{aligned} \tag{3.36}$$

since $(S-S) \subset (S'-S')$.

In view of the above inequality, we are interested in expressing $R_{S'}(z)$ in terms of $R_S(z)$ for $z \in \mathbb{Z} \setminus (S'-S')$. Indeed, denoting by $T_S(w)$ the number of

triples $\{a_1, a_2, b_1\} \in \binom{S}{3}$ with $a_1 + a_2 - b_1 = w$, we will show that

$$R_{S'}(z) = R_S(z) + T_S(z + x) + T_S(x - z). \tag{3.37}$$

By definition, the L-H-S of (3.37) counts solutions to $z = a_1 + a_2 - b_1 - b_2$ with $\big(\{a_1, a_2\}, \{b_1, b_2\}\big) \in \binom{S'}{2}^2$. We will classify those solutions into three types and show that:

a. solutions without $x$, namely, with $\{a_1, a_2, b_1, b_2\} \subset S$ are counted by $R_S(z)$;

b. solutions with $x \in \{b_1, b_2\}$ are counted by $T_S(z + x)$;

c. solutions with $x \in \{a_1, a_2\}$ are counted by $T_S(x - z)$;

It is clear that the solutions of the first type are counted by $R_S(z)$, therefore we only need to count the second and third types. First we note that for $z \notin S' - S'$ any solution to $z = a_1 + a_2 - b_1 - b_2$ satisfies

$$\{a_1, a_2\} \cap \{b_1, b_2\} = \emptyset. \tag{3.38}$$

In particular, a solution cannot be both of second and third type. In other words, the above classification into types is a partition of the set of solutions counted by $R_{S'}(z)$.

Let us consider solutions of second type, say $b_2 = x$. For those we have

$$z + x = a_1 + a_2 - b_1, \quad \text{with } a_1, a_2, b_1 \in S \tag{3.39}$$

and there are $T_S(z + x)$ sets $\{a_1, a_2, b_1\} \in \binom{S}{3}$ for which the equality (3.39) holds. Note that if (3.39) is satisfied then $a_1 + b_1 - a_2$ cannot equal $z + x$ (otherwise $a_2 = b_1$, which is not possible by (3.38)). Similarly, if (3.39) is satisfied then $a_2 + b_1 - a_1$ cannot equal $z + x$. Consequently, each set counted by $T_S(z + x)$ corresponds to a unique solution of second type.

The same argument shows that if $a_2 = x$ then $x - z = b_1 + b_2 - a_1$ with $a_1, b_1, b_2 \in S$ and the number of solutions $\{b_1, b_2, a_1\} \in \binom{S}{3}$ to this equation

is $T_S(x-z)$. Therefore the number of solutions of third type is $T_S(x-z)$ and hence (3.37) holds.

From (3.37), we see that

$$R_{S'}(z)^\ell - R_S(z)^\ell \le \sum_{i,j} \binom{\ell}{i,j} R_S(z)^{\ell-i-j} T_S(z+x)^i T_S(x-z)^j,$$

where the sum is over all $0 \le i,j \le \ell$ with $1 \le i+j \le \ell$, and $\binom{\ell}{i,j} = \frac{\ell!}{i!j!(\ell-i-j)!}$ is a multinomial coefficient. Since $S$ is a $B_3$-set, we have $T_S(w) \in \{0,1\}$ for all $w \in \mathbb{Z}$, and thus

$$T_S(z+x)^i T_S(x-z)^j \le T_S(z+x) + T_S(x-z).$$

Consequently,

$$\begin{aligned} R_{S'}(z)^\ell - R_S(z)^\ell &\le \big(T_S(z+x) + T_S(x-z)\big) \\ &\quad \times \left\{ \sum_{\substack{0\le i,j\le \ell \\ i+j=\ell}} \binom{\ell}{i,j} \cdot 1 + \sum_{\substack{0\le i,j\le \ell \\ 1\le i+j\le \ell-1}} \binom{\ell}{i,j} R_S(z)^{\ell-1} \right\} \qquad (3.40) \\ &\le \big(T_S(z+x) + T_S(x-z)\big)\big(2^\ell + 3^\ell R_S(z)^{\ell-1}\big). \end{aligned}$$

Since

$$\sum_{w\in\mathbb{Z}} T_S(w) = \binom{s}{3}, \qquad (3.41)$$

it follows that

$$\begin{aligned} Q_{S',\ell} - Q_{S,\ell} &\overset{(3.36)}{\le} \sum_{z\notin S'-S'} \big(R_{S'}(z)^\ell - R_S(z)^\ell\big) \\ &\overset{(3.40)}{\le} 2^{\ell+1}\binom{s}{3} + 3^\ell \sum_{z\notin S'-S'} R_S(z)^{\ell-1} \cdot \big(T_S(z+x) + T_S(x-z)\big). \end{aligned} \qquad (3.42)$$

Let $I_{S,\ell}(x)$ denote the last term on the R-H-S of (3.42) so that

$$Q_{S',\ell} - Q_{S,\ell} - 2^{\ell+1}\binom{s}{3} \le I_{S,\ell}(x).$$

We will now estimate how many $x$ are such that $S' = S \cup \{x\}$ is a $B_3$-set and $I_{S,\ell}(x) > 2^{\ell+1}\binom{s}{3}$. Observe that for every $x$ which does not satisfy this inequality, (3.35) holds.

We have

$$\begin{aligned}
\sum_x I_{S,\ell}(x) &= 3^\ell \sum_x \sum_{z \notin S'-S'} R_S(z)^{\ell-1} \cdot \big(T_S(z+x) + T_S(x-z)\big) \\
&= 3^\ell \sum_{z \notin S'-S'} R_S(z)^{\ell-1} \sum_x \big(T_S(z+x) + T_S(x-z)\big) \\
&\overset{(3.41)}{\le} 3^\ell \sum_{z \notin S'-S'} R_S(z)^{\ell-1} \cdot 2\binom{s}{3} \\
&\le 3^\ell Q_{S,\ell-1} \cdot 2\binom{s}{3}.
\end{aligned} \tag{3.43}$$

Since $Q_{S,\ell-1} \le Q_{S,\ell}$ and, by the assumption of this claim, $Q_{S,\ell} \le 2^{\ell+2}\binom{s}{4} < 2^\ell (s^*)^4$,

$$\sum_x I_{S,\ell}(x) \le 3^\ell (s^*)^4 \cdot 2^{\ell+1}\binom{s}{3}.$$

Consequently, at most $3^\ell (s^*)^4$ elements $x$ satisfy $I_{S,\ell}(x) > 2^{\ell+1}\binom{s}{3}$.

On the other hand, since $Q_{S,\ell} \le 2^{\ell+2}\binom{s}{4}$, for any $x$ such that $I_{S,\ell}(x) \le 2^{\ell+1}\binom{s}{3}$,

$$Q_{S',\ell} \le Q_{S,\ell} + 2^{\ell+1}\binom{s}{3} + I_{S,\ell}(x) \le 2^{\ell+2}\binom{s}{4} + 2 \cdot 2^{\ell+1}\binom{s}{3} = 2^{\ell+2}\binom{s+1}{4}.$$

The claim follows. □

## 3.4 Lower bounds of Theorem 3.1.4

The proof of the lower bound is divided into two parts. The first part (given by Lemma 3.4.1 below) deals with the case $0 \le a \le 1/5$ and yields that in this range $b_1(a) = a$ holds. Since $F_3([n]_m)$ is monotone with respect to $m$, for $a > 1/5$ we have $b_1(a) \ge b_1(1/5) = 1/5$.

The second part (given by Lemma 3.4.2) yields that $b_1(a) \ge a/3$ for all $0 \le a \le 1$. Note that in the range $1/5 \le a \le 3/5$, this is superseded by the bound obtained in the first part, that is, $b_1(a) \ge 1/5$. Combining this two bounds we obtain (3.4).

**Lemma 3.4.1** *For $1 \ll m \le o(n^{1/5})$, almost surely we have $m \ge F_3([n]_m) \ge (1 - o(1))m$.*

Let $S$ be an arbitrary $B_3$-set of cardinality $s = o(n^{1/5})$. Note that for any element $x \in [n] \setminus (3S - 2S)$, the set $S' = S \cup \{x\}$ is a $B_3$-set. Since $|3S - 2S| \le \binom{s}{3}\binom{s}{2} \le s^5 = o(n)$, at least $(1 - o(1))n$ elements $x$ can be added to $S$ to form a larger $B_3$-set. We will use this observation to prove the lemma.

Let $1 \ll m \le o(n^{1/5})$ be fixed and $X_1, \dots, X_m$ be a sequence of uniform and independent random variables over $[n]$. Let us construct a $B_3$-set $S$ from the sequence $X_1, \dots, X_m$ as follows: start with $S_1 = \{X_1\}$ and, for each $2 \le j \le m$, set $S_j = S_{j-1} \cup \{X_j\}$ if $S_{j-1} \cup \{X_j\}$ is a $B_3$-set ($X_j \notin S_{j-1}$), otherwise, set $S_j = S_{j-1}$. Let $S = S_m$ and consider the random variable $|S| = |S_m|$.

Note that for every $2 \le j \le m$,

$$\mathbf{P}\big[S_{j-1} \cup \{X_j\} \text{ is not a } B_3\text{-set}\big] = \mathbf{P}\big[X_j \in 3S_{j-1} - 2S_{j-1}\big] \le \frac{j^5}{n} \le \frac{m^5}{n}.$$

The inequality above holds regardless of the history of $X_1, \dots, X_{j-1}$. Since

$$|S| = |S_m| = 1 + \sum_{j=2}^{m} \mathbf{1}\big[X_j \in [n] \setminus (3S_{j-1} - 2S_{j-1})\big],$$

the probability that $|S| \le m-k$, with $k \ge 1$, is at most

$$\binom{m}{k}\Big(\frac{m^5}{n}\Big)^k \le \Big(\frac{em^6}{kn}\Big)^k.$$

Since $\frac{2em^6}{n} = o(m)$, we can pick $k = k(m) \gg 1$ with $\frac{2em^6}{n} \ll k = o(m)$. For such a choice of $k$, the probability $|S| < m-k$ is at most $2^{-k} = o(1)$.

To complete the proof, we now observe that almost surely, $|\{X_1, \dots, X_m\}| = m$. Indeed, the expected number of pairs $i \ne j$ with $X_i = X_j$ is $\binom{m}{2}\frac{1}{n} = o(1)$. Given our choice of $k = k(m) = o(m)$,

$$\begin{aligned}\mathbf{P}\big[F_3([n]_m) \le m-k\big] &\le \mathbf{P}\big[|S| \le m-k \;\big|\; |\{X_1, \dots, X_m\}| = m\big],\\ &\le \mathbf{P}\big[|S| \le m-k\big]/\mathbf{P}\big[|\{X_1, \dots, X_m\}| = m\big]\\ &= \frac{o(1)}{1-o(1)} = o(1).\end{aligned} \tag{3.44}$$

Summarizing, almost surely, $F_3([n]_m) = (1-o(1))m$, and Lemma 3.4.1 follows.

**Lemma 3.4.2** *For any $1 \ll m \le n$, almost surely we have $F_3([n]_m) = \Omega(m^{1/3})$.*

For this proof, it will be convenient to use the model $[n]_p$ with $p = m/n$ rather than $[n]_m$ (recall Remark 3.1.6). Without loss of generality we assume that $1/p, pn, pn/3 \in \mathbb{N}$. Our strategy here follows that of [31]. In order to show the existence of a Sidon set of order $(pn)^{1/3}$ in a typical instance of a random set $[n]_p$ we will use the following theorem of Bose and Chowla [10] (with the statement adapted for our purposes).

**Theorem 3.4.3** *There exists $m_0$ such that for all $m \ge m_0$, there exists a $B_3$-set $X \subset \mathbb{Z}_m$ with $|X| = \Omega\big(m^{1/3}\big)$.* □

We will apply Theorem 3.4.3 with $m = pn$ to produce a $B_3$-set $X \subset \mathbb{Z}_{pn}$ with $|X| = \Omega\big((pn)^{1/3}\big)$. Then, we will show that there is a projection $\pi\colon U \subset [n] \to \mathbb{Z}_{pn}$ such that

(*a*) any set $S \subset \pi^{-1}(X)$ with $|S \cap \pi^{-1}(x)| \le 1$ for all $x \in X$ is a $B_3$-set;

($b$) almost surely there are $\Omega(|X|)$ elements $x \in X$ for which $[n]_p \cap \pi^{-1}(x) \neq \emptyset$.

Note that the first condition is purely deterministic.

Define a set $S$ by selecting the smallest element from $[n]_p \cap \pi^{-1}(x)$ for each $x \in X$. Combining ($a$) and ($b$) yields that the (random) set $S$ is a $B_3$-set and almost surely $|S| = \Omega(|X|) = \Omega\big((pn)^{1/3}\big)$. Consequently, the lower bound of this lemma will be established after we prove ($a$) and ($b$).

In order to define the projection $\pi$ and its domain $U \subset [n]$ we first partition $[3n]$ into intervals

$$I_j = \left[\frac{j}{p} + 1, \frac{j+1}{p}\right], \quad j = 0, \ldots, 3pn - 1. \tag{3.45}$$

Furthermore, we subdivide each of the intervals above in three equal length intervals, namely,

$$I_{j,k} = \left[\frac{j}{p} + 1 + \frac{k}{3p}, \frac{j}{p} + \frac{k+1}{3p}\right], \quad j = 0, \ldots, 3pn - 1 \text{ and } k = 0, 1, 2. \tag{3.46}$$

The domain of $\pi$ is defined as

$$U = \bigcup_{j=0}^{pn-1} I_{j,0}. \tag{3.47}$$

Note that $U \subset [n]$ since $j < pn$ in the union above. The projection is then set as $\pi\colon x \mapsto j \in \mathbb{Z}_{pn}$, whenever $x \in I_{j,0}$.

Let us now prove ($a$). Let $S \subset \pi^{-1}(X)$ be a set satisfying the condition on ($a$), namely, $|S \cap \pi^{-1}(x)| \leq 1$ for all $x \in X$. This condition ensures that $\pi|_S$ is a one-to-one map. Moreover, $\pi(S) \subset X$ is a $B_3$-set. Let $\{a_1, a_2, a_3\} \in \binom{S}{3}$ be arbitrary and $0 \leq \ell \leq 3pn - 1$ be such that $a_1 + a_2 + a_3 \in I_\ell$. We claim that $\pi(a_1) + \pi(a_2) + \pi(a_3) = \ell \bmod pn$. Indeed, let $j_i$ be such that $a_i \in I_{j_i} = I_{j_i,0}$,

for $i = 1, 2, 3$, and observe that by (3.46), $a_i \in \left[\frac{j_i}{p} + 1, \frac{j_i}{p} + \frac{1}{3p}\right]$. Therefore,

$$a_1 + a_2 + a_3 \in \left[\frac{j_1 + j_2 + j_3}{p} + 3, \frac{j_1 + j_2 + j_3}{p} + 3 \times \frac{1}{3p}\right] \subset I_{j_1+j_2+j_3}.$$

Hence $\ell = j_1 + j_2 + j_3$ and since $\pi(a_i) = j_i \bmod pn$, it follows that $\pi(a_1) + \pi(a_2) + \pi(a_3) = \ell \bmod pn$. Since $\pi(S)$ is a $B_3$-set and $\pi|_S$ is one-to-one, it follows that no other triple $\{b_1, b_2, b_3\} \in \binom{S}{3}$ can satisfy $\pi(b_1) + \pi(b_2) + \pi(b_3) = \ell \bmod pn$. In other words, no other triple $\{b_1, b_2, b_3\}$ satisfies $b_1 + b_2 + b_3 \in I_\ell$ and hence $S$ must be a $B_3$-set.

It remains to prove $(b)$. By construction, $\{\pi^{-1}(x) \colon :: x \in X\}$ is a family of pairwise disjoint intervals (of the form $\pi^{-1}(x) = I_{j_x,0}$, $0 \le j_x \le pn - 1$). For any $x \in X$, the probability that $[n]_p \cap \pi^{-1}(x) = \emptyset$ is

$$q = (1-p)^{|\pi^{-1}(x)|} = (1-p)^{1/3p} \le e^{-p/3p} = e^{-1/3} < 3/4.$$

It follows from the fact that the sets $\pi^{-1}(x)$, $x \in X$, are disjoint, that the number of elements $x \in X$ for which $[n]_p \cap \pi^{-1}(x) = \emptyset$ is a random variable following a Binomial distribution with parameters $|X|$ and $q < 3/4$. Consequently, by the additive form of Chernoff's bound,

$$\mathbf{P}\left[\left|\left\{x \in X \colon :: [n]_p \cap \pi^{-1}(x) = \emptyset\right\}\right| > q\,|X| + \frac{1}{8}|X|\right] \le \exp\{-c\,|X|\},$$

for some absolute constant $c > 0$. Therefore, almost surely there are at least $(1 - q - 1/8)|X| \ge \frac{1}{8}\,|X|$ elements $x \in X$ which satisfy $[n]_p \cap \pi^{-1}(x) \neq \emptyset$, thus proving $(b)$.

# Chapter 4

# Infinite Sidon sets

## 4.1 Introduction

Let $\mathbb{N}$ be the set of positive integers. We study Sidon sets contained in a sparse, random subset of $\mathbb{N}$. First we introduce the random model of interest. In the definition below, we will use the letter $m$ to denote an arbitrary integer and $p_m$ to be the probability associated it. We use the letter $n$ when considering the set of elements of $S$ on an initial segment, that is, $S[n] := S \cap [n]$.

**Definition 4.1.1** (**Random set $R$ and the probability spaces $(\Omega, \mathcal{S}, \mathbb{P})$ and $(\Omega_A, \mathcal{S}_A, \mathbb{P}_A)$**). *Fix $0 \leq p_m \leq 1$ for each $m \in \mathbb{N}$. We generate a random set $R \subset \mathbb{N}$ by adding $m$ to $R$ with probability $p_m$, independently for each $m$. We let $(\Omega, \mathcal{S}, \mathbb{P})$ be the probability space of the random sets $R$. More generally, for $A \subset \mathbb{N}$, let $(\Omega_A, \mathcal{S}_A, \mathbb{P}_A)$ be the probability space of the random sets $R \cap A$.*

In general, we shall fix absolute constants $\alpha > 0$ and $0 < \delta \leq 1$, and let $p_m = \min\{1, \alpha m^{-1+\delta}\}$ for all positive integers $m$. Note that we restrict our probabilities only to the above probabilities, ignoring the case when, say, $p_m = m^{-1/2} \log m$. Covering the remaining cases would not require a new proof technique, but it would be a bit more cumbersome.

Readers interested in the details of the construction of the spaces $(\Omega, \mathcal{S}, \mathbb{P})$

and $(\Omega_A, \mathcal{S}_A, \mathbb{P}_A)$ are encouraged to consult, for example, Halberstam and Roth [21, Theorem 13, page 142]. Using the natural correspondence between subsets of $\mathbb{N}$ and 0–1 vectors indexed by $\mathbb{N}$, we may identify $(\Omega, \mathcal{S}, \mathbb{P})$ with the product of the two-point spaces $(\Omega_m, \mathcal{S}_m, \mathbb{P}_m)$ $(m \in \mathbb{N})$, where $\Omega_m = \{0, 1\}$, $\mathcal{S}_m = 2^{\Omega_m}$, and $\mathbb{P}_m(\{1\}) = p_m$ and $\mathbb{P}_m(\{0\}) = 1 - p_m$. Thus, $\mathcal{S}$ is the $\sigma$-algebra generated by the sets

$$C(m) = \{R \subset \mathbb{N} \colon m \in R\} \quad (m \in \mathbb{N}), \tag{4.1}$$

that is, the smallest family of subsets of $\mathbb{N}$ that is closed under complementation, finite intersections, and countable unions that contains the sets in (4.1). Furthermore, $\mathbb{P}(C(m)) = \mathbb{P}(m \in R) = p_m$ for all $m$, and this suffices to define $\mathbb{P}$ on every member of $\mathcal{S}$ uniquely.

Similarly, $(\Omega_A, \mathcal{S}_A, \mathbb{P}_A)$ may be identified with the product of the two-point spaces $(\Omega_m, \mathcal{S}_m, \mathbb{P}_m)$ $(m \in A)$ above. In what follows, we shall often write $\mathbb{P}$ instead of $\mathbb{P}_A$, as this will not cause any confusion.

We will study how dense Sidon sets are contained in $R$. We introduce the notion of the *growth* of a set $S \subset \mathbb{N}$. We say that $S$ has *lower growth at least* $h(n)$ if $S[n] \geq h(n)$ for every sufficiently large $n$. We also say that $S$ has *upper growth at most* $h(n)$ if $S[n] \leq h(n)$ for every sufficiently large $n$. Let $R$ be a set of $\mathbb{N}$. We will abbreviate the fact that there exists a Sidon subset $S \subset R$ with lower growth at least $h(n)$, by writing

$$lgr^{\exists} S(R) \geq h(n).$$

Similarly,

$$ugr^{\forall} S(R) \leq h(n)$$

will mean that all Sidon subsets $S \subset R$ have upper growth at most $h(n)$.

An abridged version of our results of this paper is the following.

**Theorem 4.1.2** (Kohayakawa, Lee, and Rödl [30]) *For every $\varepsilon > 0$, there exist positive constants $c_1$, $c_1'$, $c_2 = c_2(\delta)$, $c_3$, and $c_4 = c_4(\alpha)$ such that with probability 1, the following holds:*

*a.* $(1-\varepsilon)\frac{\alpha}{\delta}n^{\delta} \le lgr^{\exists}S(R),\ \ ugr^{\forall}S(R) \le \frac{\alpha}{\delta}n^{\delta}$ *if* $0<\delta<1/3$

*b.* $(1-\varepsilon)(1-18\alpha^3)3\alpha n^{1/3} \le lgr^{\exists}S(R),\ \ ugr^{\forall}S(R) \le 3\alpha n^{1/3}$ *if* $\delta = 1/3$ *and* $0<\alpha\le 0.1$

*c.* $c_1' n^{1/3} \le lgr^{\exists}S(R),\ \ ugr^{\forall}S(R) \le c_1\big(\log[\alpha+1]\big)^{1/3}n^{1/3}$ *if* $\delta = 1/3$ *and* $\alpha \ge 0.1$

*d.* $c_1' n^{1/3} \le lgr^{\exists}S(R),\ \ ugr^{\forall}S(R) \le c_2(n\log n)^{1/3}$ *if* $1/3<\delta<2/3$

*e.* $c_1' n^{1/3} \le lgr^{\exists}S(R),\ \ ugr^{\forall}S(R) \le c_3 n^{1/3}(\log n)^{4/3}$ *if* $\delta = 2/3$

*f.* $c_1' n^{1/3} \le lgr^{\exists}S(R),\ \ ugr^{\forall}S(R) \le c_4 n^{\delta/2}$ *if* $2/3<\delta\le 1$

## 4.2 Main results

Theorem 4.1.2 will be proved by showing the following two lemmas – one is about $lgr^{\exists}S(R)$ and the other is about $ugr^{\forall}S(R)$. The result of Theorem 4.1.2 can be rewritten using the notation 'lim sup' and 'lim inf' because of the following:
A set $S \subset \mathbb{N}$ satisfies that

$$|S[n]| \ge h(n)(1+o(1)) \qquad \Big(\text{or } |S[n]| \le h(n)(1+o(1))\Big)$$

for every sufficiently large $n$ if and only if $S$ satisfies that

$$\liminf \frac{|S[n]|}{h(n)} \ge 1 \qquad \Big(\text{ or } \limsup \frac{|S[n]|}{h(n)} \le 1\Big).$$

Therefore, the lower and upper bounds in Theorem 4.1.2 can be restated as follows.

**Lemma 4.2.1** *There exist positive constants $c_1$, $c_2 = c_2(\delta)$, $c_3$, and $c_4 = c_4(\alpha)$ such that with probability 1, all Sidon subsets $S$ of a random set $R$ satisfy the following:*

*a.* $\limsup \dfrac{\big|S[n]\big|}{n^{\delta}} \le \dfrac{\alpha}{\delta}$ *if* $0 < \delta < 1/3$

*b.* $\limsup \dfrac{\big|S[n]\big|}{n^{1/3}} \le 3\alpha$ *if* $\delta = 1/3$ *and* $0 < \alpha \le 0.1$

*c.* $\limsup \dfrac{\big|S[n]\big|}{n^{1/3}} \le c_1\big(\log[\alpha+1]\big)^{1/3}$ *if* $\delta = 1/3$ *and* $\alpha \ge 0.1$

*d.* $\limsup \dfrac{\big|S[n]\big|}{(n\log n)^{1/3}} \le c_2$ *if* $1/3 < \delta < 2/3$

*e.* $\limsup \dfrac{\big|S[n]\big|}{n^{1/3}(\log n)^{4/3}} \le c_3$ *if* $\delta = 2/3$

*f.* $\limsup \dfrac{\big|S[n]\big|}{n^{\delta/2}} \le c_4$ *if* $2/3 < \delta \le 1$

**Lemma 4.2.2** *For every $\varepsilon > 0$, a random set $R$ contains a Sidon subset $S$ for which the following holds with probability 1:*

*a.* $\liminf \dfrac{\big|S[n]\big|}{n^{\delta}} \ge (1-\varepsilon)\dfrac{\alpha}{\delta}$ *if* $0 < \delta < 1/3$

*b.* $\liminf \dfrac{\big|S[n]\big|}{n^{1/3}} \ge (1-\varepsilon)(1-18\alpha^3)3\alpha$ *if* $\delta = 1/3$ *and* $0 < \alpha \le 0.1$

Remark that if either $\delta = 1/3$, $\alpha \ge 0.1$ or $\delta > 1/3$, then by the monotonicity, Lemma 4.2.2 (b) implies that with probability 1, a random set $R$ contains a Sidon subset $S$ such that

$$\liminf \frac{\big|S[n]\big|}{n^{1/3}} \ge c,$$

where $c$ is an absolute constant. This yields the lower bounds of (c), (d), (e), and (f) in Theorem 4.1.2.

## 4.3 Preliminaries

### 4.3.1 Sidon quadruples

If a set $A \subset \mathbb{N}$ is not a Sidon set, then there exist $a_1, a_2, a_3, a_4 \in A$ (necessarily distinct but not all equal) such that $a_1+a_4 = a_2+a_3$. Without loss of generality, assume that $a_1 < a_2 \le a_3 < a_4$. We call a set $\{a_1, a_2, a_3, a_4\}$ a *Sidon quadruple* if both $a_1 + a_4 = a_2 + a_3$ and $a_1 < a_2 \le a_3 < a_4$ hold. Note that the size of a Sidon quadruple is either 3 or 4.

### 4.3.2 A maximum Sidon subset of a random set in $[n]$.

We will use the following result on the maximum size of a Sidon subset of a random set in $[n] := \{1, 2, \cdots, n\}$, which was stated in Section 2.2.1. In order to make this chapter self-contained, we recall definitions and results in Section 2.2.1.

Let $[n]_p$ be a random subset of $[n]$ obtained by choosing each element of $[n]$ independently with (uniform) probability $p = p(n)$. Recall that $F([n]_p)$ denotes the maximum size of a Sidon subset of $[n]_p$. The following theorem is a consequence of the result in Section 2.2.1 of this dissertation or in [31].

**Theorem 4.3.1** (Kohayakawa, Lee, Rödl, and Samotij [31]) *Fix absolute constants $\alpha > 0$ and $0 < \delta \le 1$, and suppose $p = \alpha n^{-1+\delta}$. There exist positive absolute constants $c_1$, $c_2$, $c_3$, $c_5$, $c_6$, $c_7$, $c_8$, and a constant $c_4 = c_4(\delta)$, only depending on $\delta$, such that the following holds with probability $1 - O(1/n^2)$.*

*a.* $F([n]_p) = \alpha n^{\delta}\big(1 + o(1)\big)$ *if* $0 < \delta < 1/3$

*b.* $c_1\big(\log[\alpha+1]\big)^{1/3} n^{1/3} \le F([n]_p) \le c_2\big(\log[\alpha+1]\big)^{1/3} n^{1/3}$ *if* $\delta = 1/3$ *and* $\alpha \ge 0.1$

*c.* $c_3\big(n \log n\big)^{1/3} \le F([n]_p) \le c_4\big(n \log n\big)^{1/3}$ *if* $1/3 < \delta < 2/3$

*d.* $c_5 (n \log n)^{1/3} \le F([n]_p) \le c_6 n^{1/3} (\log n)^{4/3}$ *if* $\delta = 2/3$

*e.* $c_7\sqrt{\alpha}\cdot n^{\delta/2} \le F([n]_p) \le c_8\sqrt{\alpha}\cdot n^{\delta/2}$ *if* $2/3 < \delta \le 1$

Remark that in [31], a similar statement of (b) was proved only for a bit narrower range $\alpha \ge 2$ as follows: $c_1\big(\log\alpha\big)^{1/3}n^{1/3} \le F([n]_p) \le c_2\big(\log\alpha\big)^{1/3}n^{1/3}$. However, in fact, by replacing $\log\alpha$ with $\log[\alpha+1]$, one can show (b) with the same argument.

### 4.3.3 Borel–Cantelli lemma

We introduce the Borel–Cantelli lemma which translates a result in $[n]$ into a result in $\mathbb{N}$.

**Lemma 4.3.2** (**Borel–Cantelli Lemma**) *Let $\{E_n\}$, where $n \in \mathbb{N}$, be an infinite sequence of measurable events in a probability space. If $\sum_{n=1}^{\infty}\mathbb{P}[E_n] < \infty$, then with probability 1, only finitely many events of the $E_n$ occur, that is*

$$\mathbb{P}\Big[\ \bigcap_{i=1}^{\infty}\bigcup_{n=i}^{\infty} E_n\ \Big] = 0.$$

The Borel–Cantelli lemma implies the following corollary. Recall that $\Omega$ is the probability space of infinite random sets $R \subset \mathbb{N}$.

**Corollary 4.3.3** *Let $\{F_n\}$ be an infinite sequence of events in $\Omega$. If $F_n$ holds with probability $1 - O(1/n^2)$, then with probability 1, a random set $R$ satisfies that there exists a positive integer $n_0 = n_0(R)$ such that for every $n \ge n_0$, the event $F_n$ holds.*

**Proof** Let $\overline{F_n}$ be the complement of $F_n$. Since $\mathbb{P}[F_n] = 1 - O(1/n^2)$, we have that $\mathbb{P}[\overline{F_n}] = O(1/n^2)$, and hence

$$\sum_n \mathbb{P}[\overline{F_n}] = \sum_n O(1/n^2) < \infty.$$

The Borel–Cantelli lemma with $E_n = \overline{F_n}$ implies that with probability 1, only finitely many events of the $\overline{F_n}$ occur. Hence, with probability 1, a random set

$R$ satisfies that there exists a positive integer $n_0 = n_0(R)$ such that for every $n \geq n_0$, the event $F_n$ holds. □

### 4.3.4 The size of a random set $R$ in an interval

First, we estimate the expected size of $R[an+1, bn]$ where $0 < a < b$.

**Fact 4.3.1** *Let $0 \leq a < b$ and $\beta > 0$ be real numbers. We have*

$$\frac{\alpha}{\delta}(b^\delta - a^\delta)n^\delta - O(1) \leq \mathbb{E}(|R[an+1, bn]|) \leq \frac{\alpha}{\delta}(b^\delta - a^\delta)n^\delta. \tag{4.2}$$

*In particular, we have*

$$\big[1 - (1-\delta)\beta/2\big]\alpha\beta n^\delta - O(1) \leq \mathbb{E}(|R[n+1, (1+\beta)n]|) \leq \alpha\beta n^\delta. \tag{4.3}$$

**Proof** First, we show (4.2). Since only finitely many $m$ satisfy $\alpha m^{-1+\delta} > 1$, the probability $p_m = \min\{1, \alpha m^{-1+\delta}\} = \alpha m^{-1+\delta}$ for all but finitely many $m$. Thus we infer

$$\sum_{m=an+1}^{bn} \alpha m^{-1+\delta} - O(1) \leq \mathbb{E}(|R[an+1, bn]|) \leq \sum_{m=an+1}^{bn} \alpha m^{-1+\delta}. \tag{4.4}$$

Since both

$$\int_{an+1}^{bn} \alpha x^{-1+\delta} dx \leq \sum_{m=an+1}^{bn} \alpha m^{-1+\delta} \leq \int_{an}^{bn} \alpha x^{-1+\delta} dx,$$

and

$$\int_{an+1}^{bn} \alpha x^{-1+\delta} dx = \int_{an}^{bn} \alpha x^{-1+\delta} dx - O(1)$$

hold, inequality (4.4) implies that

$$\int_{an}^{bn} \alpha x^{-1+\delta} dx - O(1) \leq \mathbb{E}(|R[an+1, bn]|) \leq \int_{an}^{bn} \alpha x^{-1+\delta} dx. \tag{4.5}$$

Since $\int_{an}^{bn} \alpha x^{-1+\delta} dx = \alpha(b^\delta - a^\delta)n^\delta/\delta$, inequality (4.5) implies inequality (4.2).

Next, we show that inequality (4.3) follows from (4.2). Indeed, inequality (4.2) with $a = 1$ and $b = 1 + \beta$ implies that

$$\frac{\alpha}{\delta}\big[(1+\beta)^\delta - 1\big]n^\delta - O(1) \le \mathbb{E}(|R[n+1,(1+\beta)n]|) \le \frac{\alpha}{\delta}\big[(1+\beta)^\delta - 1\big]n^\delta. \quad (4.6)$$

Since

$$(1+\beta)^\delta = 1 + \delta\beta + \frac{\delta(\delta-1)}{2}\beta^2 + \cdots,$$

we infer $\delta\beta + \frac{\delta(\delta-1)}{2}\beta^2 \le (1+\beta)^\delta - 1 \le \delta\beta$, and hence

$$\delta\beta\big[1 - (1-\delta)\beta/2\big] \le (1+\beta)^\delta - 1 \le \delta\beta. \quad (4.7)$$

In view of (4.7), inequality (4.6) yields (4.3). □

Next we consider the concentration of $\big|R[an+1,bn]\big|$. We will apply the following version of the Chernoff bound:

**Lemma 4.3.4** (**The Chernoff bound**) *Let $X_i$ be independent random variables such that $\mathbb{P}[X_i = 1] = p_i$ and $\mathbb{P}[X_i = 0] = 1 - p_i$, and let $X = \sum_{i=1}^n X_i$. Then*

$$\mathbb{P}\Big[|X - \mathbb{E}(X)| \ge \lambda\mathbb{E}(X)\Big] \le 2\exp^{-\frac{\lambda^2}{3}\mathbb{E}(X)}.$$

Based on Fact 4.3.1, the Chernoff bound above implies the following:

**Lemma 4.3.5** *Let $0 \le a < b$ and $\beta > 0$ be real numbers. We have that with probability $1 - O(1/n^2)$,*

$$\Big(1 - \frac{1}{\log n}\Big)\frac{\alpha}{\delta}(b^\delta - a^\delta)n^\delta \le |R[an+1,bn]| \le \Big(1 + \frac{1}{\log n}\Big)\frac{\alpha}{\delta}(b^\delta - a^\delta)n^\delta. \quad (4.8)$$

*In particular, we have that with probability $1 - O(1/n^2)$,*

$$\Big(1 - \frac{1}{\log n}\Big)\big[1 - (1-\delta)\beta/2\big]\alpha\beta n^\delta \le |R[n+1,(1+\beta)n]| \le \Big(1 + \frac{1}{\log n}\Big)\alpha\beta n^\delta. \quad (4.9)$$

### 4.3.5 The Kim–Vu polynomial concentration result

The Kim–Vu polynomial concentration result introduced in Section 2.5.2 is also one of important lemmas for the infinite Sidon subsets contained in $\mathbb{N}$. In order to make this chapter self-contained, we recall it again.

Let $\mathcal{H} = (V, E)$ be a hypergraph with $n$ vertices. Let $R$ be a random subset of $V$ obtained by choosing each element $i \in V$ independently with probability $q_i$. Let $\mathcal{H}[R]$ be the sub-hypergraph of $\mathcal{H}$ induced on $R$ and set $Y = |\mathcal{H}[R]|$. Kim–Vu [25] obtained a result which, although the hyperedges are not chosen to $R$ independently, ensures that $Y$ is concentrated around its mean $\mathbb{E}(Y)$ similarly as in the Chernoff bound. (See also Theorem 7.8.1 in Alon–Spencer [3]). First, we introduce basic definitions.

**Definition 4.3.6.** *Let $k$ be the maximum size of hyperedges in $\mathcal{H}$, and let $A \subset [n]$ be with $|A| \leq k$.*

- $Y_A := |\{e \in \mathcal{H}[R] \mid A \subset e\}|$
- $\mathbb{E}_A := \mathbb{E}(Y_A \,|\, A \subset R)$
- $\mathbb{E}_i := \max \mathbb{E}_A \quad$ *over all $A \subset [n]$ with $|A| = i$.*
- $$\mathbb{E}' := \max_{1 \leq i \leq k} \mathbb{E}_i \qquad \text{and} \qquad \mathbb{E}^* := \max\{\mathbb{E}', \mathbb{E}(Y)\}. \tag{4.10}$$

Now we are ready to state the result by Kim–Vu [25].

**Theorem 4.3.7** (**Kim–Vu polynomial concentration inequality**) *Assume the above notation. For every $\lambda > 1$,*

$$\mathbb{P}\Big[|Y - \mathbb{E}(Y)| > a_k (\mathbb{E}^* \mathbb{E}')^{1/2} \lambda^k\Big] < 2e^2 e^{-\lambda} n^{k-1},$$

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

Based on Theorem 4.3.7, we give a corollary which will be frequently applied.

Let $f(n)$ be a function satisfying that

$$\mathbb{E}(Y) \leq f(n) \quad \text{and} \quad \mathbb{E}' \leq f(n) \cdot n^{-\varepsilon},$$

for some positive constant $\varepsilon$. We infer that $(\mathbb{E}^*\mathbb{E}')^{1/2} \leq f(n) \cdot n^{-\varepsilon/2}$. Setting $\lambda = n^{\varepsilon/4k}$, the following holds:

- $(\mathbb{E}^*\mathbb{E}')^{1/2}\lambda^k \leq f(n) \cdot n^{-\varepsilon/2} \cdot n^{\varepsilon/4} = f(n) \cdot n^{-\varepsilon/4} = o(f(n))$.
- $e^{-\lambda}n^{k-1} \leq n^{-2}$ for every sufficiently large $n$.

Hence, we obtain the following from Theorem 4.3.7:

**Corollary 4.3.8** *Assume the above notation, and let $f(n)$ be a function satisfying that*

$$\mathbb{E}(Y) \leq f(n) \quad \textit{and} \quad \mathbb{E}' \leq f(n) \cdot n^{-\varepsilon},$$

*for some positive constant $\varepsilon$. Then, $Y \leq f(n)(1 + o(1))$ with probability $1 - O(1/n^2)$.*

## 4.4 Proof of Lemma 4.2.1

First, we prove (a) and (b) of Lemma 4.2.1. Fix $0 < \delta \leq 1/3$ and $0 < \alpha \leq 1$. Inequality (4.8), with $a = 0$ and $b = 1$, gives that $|R[n]| \leq (1 + \frac{1}{\log n})\frac{\alpha}{\delta}n^{\delta}$, with probability $1 - O(1/n^2)$, and hence, with probability $1 - O(1/n^2)$,

$$\frac{|R[n]|}{n^{\delta}} \leq \Big(1 + \frac{1}{\log n}\Big)\frac{\alpha}{\delta}. \tag{4.11}$$

In order to obtain a result about an infinite random set, we apply Corollary 4.3.3. We define an event $F_n$ of $\Omega$ as

$$F_n := \Big\{R \subset \mathbb{N} \mid \frac{|R[n]|}{n^{\delta}} \leq \Big(1 + \frac{1}{\log n}\Big)\frac{\alpha}{\delta}\Big\}.$$

Note that $F_n$ holds with probability $1 - O(1/n^2)$. Corollary 4.3.3 implies that with probability 1, a random set $R \subset \mathbb{N}$ satisfies that there exists a positive integer $n_0 = n_0(R)$ such that for every $n \geq n_0$, the event $F_n$ holds, that is inequality (4.11) holds. For every Sidon subset $S \subset R$, we have that for every $n \geq n_0$,

$$\frac{|S[n]|}{n^\delta} \leq \frac{|R[n]|}{n^\delta} \leq \Big(1 + \frac{1}{\log n}\Big)\frac{\alpha}{\delta},$$

and hence,

$$\limsup \frac{|S[n]|}{n^\delta} \leq \lim \Big(1 + \frac{1}{\log n}\Big)\frac{\alpha}{\delta} = \frac{\alpha}{\delta},$$

which completes the proof of (a) and (b) of Lemma 4.2.1.

Next, we show (c)–(f) of Lemma 4.2.1. We give a lemma which translates a result on an upper bound on $F([n]_p)$ in Theorem 4.3.1 into a result on the lim sup of $|S[n]|$ where $S$ is a Sidon subset of an infinite random set in $\mathbb{N}$.

**Lemma 4.4.1** *For $i \in \mathbb{N}$, let $p_i$ be a decreasing function of $i$. Let $b > 0$, $1/3 \leq \nu \leq 1$ and $\rho \geq 0$ be constants such that $F([n]_{p_n}) \leq bn^\nu(\log n)^\rho$ holds with probability $1 - O(1/n^2)$. Then the following holds with probability 1:*
*There exists a positive constant $c = c(b)$, only depending on $b$, such that for any Sidon subset $S \subset R$,*

$$\limsup \frac{\big|S[n]\big|}{n^\nu(\log n)^\rho} \leq c.$$

Observe that combining Theorem 4.3.1 and Lemma 4.4.1 implies (c)–(f) of Lemma 4.2.1.

**Proof of Lemma 4.4.1** It suffices to show that there exists a positive constant $c = c(b)$ such that for every sufficiently large $n$,

$$\big|S[n]\big| \leq cn^\nu(\log n)^\rho. \tag{4.12}$$

To this end, we will consider the following steps:

- **STEP1:** We estimate $\big|S[2^i+1, 2^{i+1}]\big|$ for every sufficiently large integer $i$.

- **STEP2:** We estimate $\left|S[2^j]\right|$ for every sufficiently large integer $j$.
- **STEP3:** We estimate $\left|S[n]\right|$ for every sufficiently large integer $n$.
- **STEP1:** First we estimate an upper bound on $\left|S[n+1,2n]\right|$. Since $p_i$ is a decreasing function of $i$, each element $i$ in $[n+1,2n]$ is chosen to $R$ with probability $p_i$ which is at most $p_n$. The random set $[n+1,2n]_{p_n}$ can be viewed as $R[n+1,2n] \cup R^*[n+1,2n]$, where a random set $R^*$ is obtained by choosing each element $i$ with probability $q_i$ such that $p_i + (1-p_i)q_i = p_n$. Therefore we infer $R[n+1,2n] \subset [n+1,2n]_{p_n}$, and hence,

$$F(R[n+1,2n]) \leq F([n+1,2n]_{p_n}),$$

where $F(A)$ denotes the maximum size of a Sidon subset of $A \subset \mathbb{N}$. Since a Sidon set is invariant by translation, we infer that

$$F(R[n+1,2n]) \leq F([n+1,2n]_{p_n}) = F([n]_{p_n}).$$

Under the assumption that $F([n]_{p_n}) \leq bn^{\nu}(\log n)^{\rho}$ holds with probability $1 - O(1/n^2)$, we have that with probability $1 - O(1/n^2)$,

$$F(R[n+1,2n]) \leq bn^{\nu}(\log n)^{\rho}. \tag{4.13}$$

In order to obtain a result about an infinite random set, we apply Corollary 4.3.3. We define an event $F_n$ of $\Omega$ as

$$F_n := \Big\{ R \subset \mathbb{N} \mid F(R[n+1,2n]) \leq bn^{\nu}(\log n)^{\rho} \Big\}.$$

Note that $F_n$ holds with probability $1 - O(1/n^2)$. Corollary 4.3.3 implies that with probability 1, a random set $R \subset \mathbb{N}$ satisfies that there exists an integer $n_0 = n_0(R) > 0$ such that for every $n \geq n_0$,

$$F(R[n+1,2n]) \leq bn^{\nu}(\log n)^{\rho}. \tag{4.14}$$

Fix an integer $j_0$ such that $2^{j_0} \geq n_0$, and let $S$ be an arbitrary Sidon subset of $R$. Inequality (4.14) yields that for every $i \geq j_0$,

$$\left|S[2^i+1, 2^{i+1}]\right| \leq F(R[2^i+1, 2^{i+1}]) \leq b(2^i)^\nu (\log 2^i)^\rho = b(2^\nu)^i (\log 2^i)^\rho. \quad (4.15)$$

- **STEP2:** We estimate an upper bound on $\left|S[2^j]\right|$ for every sufficiently large $j$. Inequality (4.15) implies that for every $j > j_0$,

$$\left|S[2^j]\right| = \left|S[2^{j_0}]\right| + \sum_{i=j_0}^{j-1} \left|S[2^i+1, 2^{i+1}]\right| \leq \left|R[2^{j_0}]\right| + \sum_{i=j_0}^{j-1} b(2^\nu)^i (\log 2^i)^\rho.$$

Since $R[2^{j_0}]$ is finite, by taking a sufficiently large $j$, we have

$$\left|S[2^j]\right| \leq 2b \sum_{i=j_0}^{j-1} (2^\nu)^i (\log 2^i)^\rho.$$

Due to the fact that $(\log n)^\rho$, with $\rho \geq 0$, is a non-decreasing function of $n$, we infer

$$\begin{aligned} \left|S[2^j]\right| &\leq 2b(\log 2^{j-1})^\rho \sum_{i=j_0}^{j-1} (2^\nu)^i \leq 2b(\log 2^{j-1})^\rho \cdot (2^\nu)^{j_0} \frac{(2^\nu)^{j-j_0}-1}{2^\nu - 1} \\ &\leq \frac{2b}{2^\nu - 1} (\log 2^{j-1})^\rho (2^\nu)^j \leq c(\log 2^{j-1})^\rho 2^{\nu j} \\ &= c(2^j)^\nu (\log 2^{j-1})^\rho, \end{aligned} \quad (4.16)$$

where $c := 8b \geq 2b/(2^\nu - 1)$ for $1/3 \leq \nu \leq 1$.

- **STEP3:** We estimate an upper bound on $\left|S[n]\right|$ where $2^{j-1} < n \leq 2^j$. Inequality (4.16) implies that

$$\left|S[n]\right| \leq \left|S[2^j]\right| \leq c \cdot (2^j)^\nu (\log 2^{j-1})^\rho.$$

Since $2^j < 2n$ holds and $(\log n)^\rho$ with $\rho \geq 0$ is a non-decreasing function of $n$,

we infer that

$$\big|S[n]\big| \le c \cdot (2n)^{\nu} (\log n)^{\rho} \le c \cdot 2^{\nu} n^{\nu} (\log n)^{\rho} \le c' \cdot n^{\nu} (\log n)^{\rho},$$

where $c' := 2c = 16b$ is a constant only depending on $b$. This completes the proof of (4.12). □

It remains to show Lemma 4.2.2 (a) and (b).

## 4.5 Proof of Lemma 4.2.2 (a) and (b)

### 4.5.1 Proof of Lemma 4.2.2 (a) and (b)

The proof of (a) and (b) is almost identical, and therefore, we will prove these cases simultaneously. In case (a) we consider $\delta$, $0 < \delta < 1/3$, and $\alpha$ arbitrary. In case (b) we consider $\delta = 1/3$ and $\alpha \le 0.1$. A random set $R$ is obtained by choosing each element $m$ independently with probability $p_m = \alpha m^{-1+\delta}$. Our proof of (a) and (b) of Lemma 4.2.2 is based on the following key lemma.

**Lemma 4.5.1** *For every $\varepsilon > 0$, a random set $R$ has the following property with probability 1: There exist an integer $n_0 = n_0(\varepsilon, R) > 0$ and a Sidon set $S \subset R$ such that for every $n \ge n_0$, the following holds:*

$$|S[n]|/|R[n]| \ge 1 - \varepsilon \qquad \text{if} \quad 0 < \delta < 1/3 \quad \text{or} \tag{4.17}$$

$$|S[n]|/|R[n]| \ge (1-\varepsilon)(1 - 18\alpha^3) \qquad \text{if} \quad \delta = 1/3. \tag{4.18}$$

We shall prove Lemma 4.5.1 after the proof of (a) and (b) of Lemma 4.2.2 below.

**Proof of (a) and (b) of Lemma 4.2.2** First we prove (a), which is the case when $0 < \delta < 1/3$. Lemma 4.3.5 implies that $|R[n]| \ge \Big(1 - \dfrac{1}{\log n}\Big)\dfrac{\alpha}{\delta} n^{\delta}$ holds with probability $1 - O(1/n^2)$, and hence, Corollary 4.3.3 implies that with

probability 1, a random set $R$ satisfies that there is an integer $n_1 = n_1(R) > 0$ such that for every $n \geq n_1$,

$$|R[n]| \geq \Big(1 - \frac{1}{\log n}\Big)\frac{\alpha}{\delta} n^{\delta}. \tag{4.19}$$

Combining (4.17) and (4.19) implies that for every $n \geq \max\{n_0, n_1\}$,

$$|S[n]| \geq (1-\varepsilon)|R[n]| \geq (1-\varepsilon)\Big(1 - \frac{1}{\log n}\Big)\frac{\alpha}{\delta} n^{\delta}$$

and hence,

$$\liminf \frac{|S[n]|}{n^{\delta}} \geq \liminf (1-\varepsilon)\Big(1 - \frac{1}{\log n}\Big)\frac{\alpha}{\delta} = (1-\varepsilon)\frac{\alpha}{\delta},$$

which completes the proof of (a) of Lemma 4.2.2.

Next we prove (b), which is the case when $\delta = 1/3$. Our proof of (b) is similar to the proof of (a). Combining (4.18) and (4.19) yields that for every $n \geq \max\{n_0, n_1\}$,

$$|S[n]| \geq (1-\varepsilon)(1-18\alpha^3)|R[n]| \geq (1-\varepsilon)(1-18\alpha^3)\Big(1 - \frac{1}{\log n}\Big)3\alpha n^{1/3}$$

and hence,

$$\liminf \frac{|S[n]|}{n^{1/3}} \geq \liminf (1-\varepsilon)(1-18\alpha^3)\Big(1 - \frac{1}{\log n}\Big)3\alpha = (1-\varepsilon)(1-18\alpha^3)3\alpha,$$

which completes the proof of (b) of Lemma 4.2.2. □

It still remains to prove Lemma 4.5.1. We use Corollary 4.3.3 to prove Lemma 4.5.1. For a suitable sequence of events in Corollary 4.3.3, we introduce a sequence of events $A_n$.

**Definition 4.5.2** (Event $A_n$)**.** *For $\varepsilon > 0$ and $0 < \beta \leq 1$, let $A_n = A_n(\varepsilon, \beta)$ denote the event that $R[n+1, (1+\beta)n]$ contains a Sidon subset $S[n+1, (1+\beta)n]$*

*satisfying the following two properties:*

*(i) for any Sidon set $S[n] \subset R[n]$, the set $S[n] \cup S[n+1,(1+\beta)n]$ is a Sidon set,*

*(ii)*
$$\frac{|S[n+1,(1+\beta)n]|}{|R[n+1,(1+\beta)n]|} \geq 1-\varepsilon \qquad \text{if} \quad 0<\delta<1/3 \quad \text{or} \tag{4.20}$$

$$\frac{|S[n+1,(1+\beta)n]|}{|R[n+1,(1+\beta)n]|} \geq (1-\varepsilon)(1-18\alpha^3) \qquad \text{if} \quad \delta = 1/3. \tag{4.21}$$

Let us remark that Lemma 4.5.1 would follow if one could prove that

$$\mathbb{P}[\overline{A_n}] = O(1/n^2) \tag{4.22}$$

holds for every sufficiently large $n$. Indeed, observe that one could construct an infinite Sidon set $S = \bigcup_{i=0}^{\infty} S[n_i+1, n_{i+1}]$ in $R$ by concatenating Sidon sets $S[n_i+1,n_{i+1}] \subset R[n_i+1,n_{i+1}]$, $i \geq 0$, where $n_{i+1} = \lfloor (1+\beta)n_i \rfloor$. Unfortunately, (4.22) fails to be true without an additional condition on a set $R_{(1+\beta)n}$. For the additional condition we define a sequence of events $B_n$ below. Note that for a fixed $\varepsilon > 0$ we will use $\beta$ with $\beta \ll \varepsilon$.

**Definition 4.5.3** (Event $B_n$)**.** *For $0 < \beta \leq 1$, let $B_n = B_n(\beta)$ be the event that $R[(1+\beta)n]$ is* well distributed, *that is, for all integers $k \in [n]$,*

$$\big|R[k+1,k+\beta n]\big| = (1\pm\beta)\cdot\mathbb{E}\Big(\big|R[k+1,k+\beta n]\big|\Big). \tag{4.23}$$

Note that it is easy to check that if $R[(1+\beta)n]$ is well distributed, then

$$\big|R[n]\big| = (1\pm 3\beta)\mathbb{E}\Big(\big|R[n]\big|\Big) = (1\pm 3\beta)(\alpha/\delta)n^{\delta} - O(1), \tag{4.24}$$

where the last identity follows from Fact 4.3.1 with $a = 0$ and $b = 1$. The Chernoff bound easily implies the following.

**Fact 4.5.1** *For every $\beta > 0$, we have $\mathbb{P}[\overline{B_n}] = O(1/n^2)$, where $\overline{B_n}$ is the complement of $B_n$.*

We shall prove the following lemma about event $A_n$ in the next Section.

**Lemma 4.5.4** *For every $\varepsilon > 0$, there exists a constant $\beta_0 = \beta_0(\varepsilon, \delta)$ such that for every $0 < \beta < \beta_0$, we have $\mathbb{P}[\overline{A_n}|B_n] = O(1/n^2)$.*

We deduce the following from Fact 4.5.1 and Lemma 4.5.4.

**Corollary 4.5.5** *For every $\varepsilon > 0$, there exists a constant $\beta_0 = \beta_0(\varepsilon, \delta)$ such that for every $0 < \beta < \beta_0$, we have $\mathbb{P}[\overline{A_n \cap B_n}] = O(1/n^2)$.*

**Proof** Note that

$$\begin{aligned} \mathbb{P}[\overline{A_n \cap B_n}] &= \mathbb{P}[\overline{B_n} \cup \overline{A_n}] \leq \mathbb{P}[\overline{B_n}] + \mathbb{P}[B_n \cap \overline{A_n}] \\ &= \mathbb{P}[\overline{B_n}] + \mathbb{P}[B_n]\mathbb{P}[\overline{A_n}|B_n] \leq \mathbb{P}[\overline{B_n}] + \mathbb{P}[\overline{A_n}|B_n]. \end{aligned}$$

Consequently, Fact 4.5.1 and Lemma 4.5.4 imply that $\mathbb{P}[\overline{A_n \cap B_n}] = O(1/n^2)$. □

Now we are ready to apply Corollary 4.3.3 for our proof of Lemma 4.5.1.

**Proof of Lemma 4.5.1** Fix $\varepsilon$ as an arbitrary small positive real number. Let $\beta$ be a constant satisfying

$$(1 - 3\beta)/(1 + 3\beta)^2 \geq 1 - \varepsilon \tag{4.25}$$

and $0 < \beta < \beta_0$, where $\beta_0$ is given by Corollary 4.5.5. Recall that $A_n = A_n(\varepsilon, \beta)$ and $B_n = B_n(\beta)$ are the events introduced in Definitions 4.5.2 and 4.5.3. By Corollary 4.5.5, it follows from Corollary 4.3.3 with $F_n = A_n \cap B_n$ that with probability 1, a random set $R$ satisfies that there exists an integer $n_0 = n_0(\varepsilon, R) > 0$ such that for all $n \geq n_0$, both $A_n$ and $B_n$ simultaneously hold.

Now we generate a Sidon subset $S$ of a random set $R \subset \mathbb{N}$ which satisfies the condition of Lemma 4.5.1. First, we consider an infinite sequence of integers

$n_0 < n_1 < n_2 < \cdots$ such that $n_{i+1} = \lfloor (1+\beta)n_i \rfloor$. Since $A_{n_i}$ holds for every $i \geq 0$, there is a subset $S[n_i+1, n_{i+1}] \subset R[n_i+1, n_{i+1}]$ such that

(i) for any Sidon set $S[n_i] \subset R[n_i]$, the set $S[n_i] \cup S[n_i+1, n_{i+1}]$ is a Sidon set,

(ii)
$$\frac{|S[n_i+1, n_{i+1}]|}{|R[n_i+1, n_{i+1}]|} \geq 1-\varepsilon \qquad \text{if} \quad 0 < \delta < 1/3 \quad \text{or} \tag{4.26}$$

$$\frac{|S[n_i+1, n_{i+1}]|}{|R[n_i+1, n_{i+1}]|} \geq (1-\varepsilon)(1-18\alpha^3) \quad \text{if} \quad \delta = 1/3. \tag{4.27}$$

Set $S = \cup_{i=0}^{\infty} S[n_i+1, n_{i+1}]$ and consider it as our desired Sidon set. Note that property (i) clearly guarantees that $S$ is a Sidon subset of $R$.

It remains to show that $S$ satisfies condition (4.17) and (4.18). First, we consider the case where $n = n_i$ for some $i$. Since $|R[n_0]|$ is finite but $|R[n_0+1, \infty]|$ is infinite, it follows from inequality (4.26) and (4.27) that there exists an integer $i_0 > 0$ such that for all $i \geq i_0$, the following holds:

$$|S[n_i]|/|R[n_i]| \geq 1-2\varepsilon \qquad \text{if} \quad 0 < \delta < 1/3 \quad \text{or} \tag{4.28}$$

$$|S[n_i]|/|R[n_i]| \geq (1-2\varepsilon)(1-18\alpha^3) \quad \text{if} \quad \delta = 1/3. \tag{4.29}$$

Now we consider the case when $n$ is an arbitrary integer between $n_i$ and $n_{i+1}$, where $i \geq i_0$, and estimate the ratio $|S[n]|/|R[n]|$. Clearly, we have

$$\frac{|S[n]|}{|R[n]|} \geq \frac{|S[n_i]|}{|R[n_{i+1}]|} = \frac{|S[n_i]|}{|R[n_i]|} \cdot \frac{|R[n_i]|}{|R[n_{i+1}]|}. \tag{4.30}$$

Since both $R[n_i]$ and $R[n_{i+1}]$ are well distributed, by inequality (4.24), we have

$$\begin{aligned} \frac{|R[n_i]|}{|R[n_{i+1}]|} \geq \frac{(1-3\beta)\mathbb{E}(R[n_i])}{(1+3\beta)\mathbb{E}(R[n_{i+1}])} \geq \frac{(1-3\beta)}{(1+3\beta)} \frac{(\alpha/\delta)n_i^{\delta} - O(1)}{(\alpha/\delta)\{(1+\beta)n_i\}^{\delta}} \\ \geq \frac{1-3\beta}{(1+3\beta)^2} \geq 1-\varepsilon, \end{aligned} \tag{4.31}$$

where the second inequality follows from (4.2) and the last inequality follows from (4.25). Combining (4.30), (4.28), (4.29), and (4.31) implies the following:

$$|S[n]|/|R[n]| \geq (1-2\varepsilon)(1-\varepsilon) > 1-3\varepsilon \qquad \text{if} \quad 0<\delta<1/3 \quad \text{or}$$

$$|S[n]|/|R[n]| \geq (1-3\varepsilon)(1-18\alpha^3) \qquad \text{if} \quad \delta = 1/3.$$

Consequently, for all $n \geq n_{i_0}$, we have $|S[n]|/|R[n]| \geq 1-3\varepsilon$ (if $0<\delta<1/3$) or $|S[n]|/|R[n]| \geq (1-3\varepsilon)(1-18\alpha^3)$ (if $\delta = 1/3$). By rescaling $\varepsilon$, this implies condition (4.17) and (4.18), which completes the proof of Lemma 4.5.1. □

It remains to prove Lemma 4.5.4.

### 4.5.2 Proof of Lemma 4.5.4

We need to show that for every $\varepsilon > 0$ there exists a constant $\beta_0 = \beta_0(\varepsilon, \delta) > 0$ such that for every $0 < \beta < \beta_0$, we have $\mathbb{P}[A_n|B_n] = 1 - O(1/n^2)$. In other words, under the assumption of $B_n$, that is, that $R[(1+\beta)n]$ is well distributed, we need to show that with probability $1 - O(1/n^2)$, event $A_n$ holds, that is, a random set $R[n+1,(1+\beta)n]$ contains a subset $S[n+1,(1+\beta)n]$ satisfying the following:

(i) for any Sidon set $S[n] \subset R[n]$, the set $S[n] \cup S[n+1,(1+\beta)n]$ is a Sidon set,

(ii)
$$\frac{|S[n+1,(1+\beta)n]|}{|R[n+1,(1+\beta)n]|} \geq 1-\varepsilon \qquad \text{if} \quad 0<\delta<1/3 \quad \text{or}$$
$$\frac{|S[n+1,(1+\beta)n]|}{|R[n+1,(1+\beta)n]|} \geq (1-\varepsilon)(1-18\alpha^3) \qquad \text{if} \quad \delta = 1/3.$$

In order to obtain $S[n+1,(1+\beta)n]$ from $R[n+1,(1+\beta)n]$, we use a deletion method. For $i = 1,2,3,4$, let $Q_i$ be the set of all Sidon quadruples $\{a,b,c,d\}$ in $R[(1+\beta)n]$ such that $\big|\{a,b,c,d\} \cap R[n]\big| = 4-i$. Note that each Sidon

quadruple in $\cup_{i=1}^{4} Q_i$ intersects $R[n+1,(1+\beta)n]$. Hence, by deleting an element of each Sidon quadruple in $\cup_{i=1}^{4} Q_i$ from $R[n+1,(1+\beta)n]$, we can destroy all Sidon quadruples in $\cup_{i=1}^{4} Q_i$. Let $D_i$ $(i = 1,2,3,4)$ be a set of elements of $R[n+1,(1+\beta)n]$ which removal destroys all Sidon quadruples of $Q_i$, that is, say

$$D_i = \big\{ d \in R[n+1,(1+\beta)n] : \{a,b,c,d\} \in Q_i, a < b \le c < d \big\}. \tag{4.32}$$

Set $S[n+1,(1+\beta)n] := R[n+1,(1+\beta)n] \setminus \cup_{i=1}^{4} D_i$, and hence

$$|S[n+1,(1+\beta)n]| \ge |R[n+1,(1+\beta)n]| - \sum_{i=1}^{4} |D_i|.$$

Thus we infer

$$\frac{|S[n+1,(1+\beta)n]|}{|R[n+1,(1+\beta)n]|} \ge 1 - \frac{\sum_{i=1}^{4} |D_i|}{|R[n+1,(1+\beta)n]|}. \tag{4.33}$$

We now show that the set $S[n+1,(1+\beta)n]$ satisfies properties (i) and (ii) above. First, observe that property (i) is ensured by the construction of $S[n+1,(1+\beta)n]$. In order to prove property (ii), we are going to estimate $\sum_{i=1}^{4} |D_i|$ and $|R[n+1,(1+\beta)n]|$ in (4.33).

**Claim 4.5.6** *If event $B_n(\beta)$ holds, then there exist constants $c_1 = c_1(\delta)$ and $c_2 = c_2(\delta)$ such that the following holds with probability $1 - O(1/n^2)$:*

*(I)* $|D_1| \le (1+3\beta)^4 (2/\delta^2) \beta \alpha^4 \max\{n^{4\delta-1}, n^{\delta/2}\}$,

*(II)* $|D_2| \le c_1 \beta^{1+\delta} \alpha^4 \max\{n^{4\delta-1}, n^{\delta/2}\}$,

*(III)* $|D_3| \le c_2 \beta^2 \alpha^4 \max\{n^{4\delta-1}, n^{\delta/2}\}$,

*(IV)* $|D_4| \le 2\beta^3 \alpha^4 \max\{n^{4\delta-1}, n^{\delta/2}\}$.

The proof of the claim above will be given in the next subsection. Note that Claim 4.5.6 with a sufficiently small $\beta = \beta(\varepsilon,\delta)$ implies that for $i = 1,2,3,4$,

$|D_i|/\big[(2/\delta^2)\beta\alpha^4 \max\{n^{4\delta-1}, n^{\delta/2}\}\big]$ can be made smaller than $(1+\varepsilon/4)$, $\varepsilon/4$, $\varepsilon/4$, and $\varepsilon/4$, respectively. Consequently, Claim 4.5.6 yields that with probability $1 - O(1/n^2)$,

$$\sum_{i=1}^{4} |D_i| \le (1+\varepsilon)(2/\delta^2)\beta\alpha^4 \max\{n^{4\delta-1}, n^{\delta/2}\}. \tag{4.34}$$

Note that $\max\{4\delta-1, \delta/2\} < \delta$ for $0 < \delta < 1/3$, and we have $\max\{4\delta-1, \delta/2\} = \delta = 1/3$ and $\alpha \le 0.1$ when $\delta = 1/3$. Hence inequality (4.34) implies that the following holds with probability $1 - O(1/n^2)$:

$$\sum_{i=1}^{4} |D_i| \le \varepsilon\beta\alpha n^{\delta} \qquad \text{if} \quad 0 < \delta < 1/3 \quad \text{or} \tag{4.35}$$

$$\sum_{i=1}^{4} |D_i| \le (1+\varepsilon)18\beta\alpha^4 n^{1/3} \qquad \text{if} \quad \delta = 1/3. \tag{4.36}$$

On the other hand, since $R[(1+\beta)n]$ is well distributed and the first inequality in (4.3) holds, we infer that

$$\big|R[n+1,(1+\beta)n]\big| \ge (1-\beta)\mathbb{E}\big(\big|R[n+1,(1+\beta)n]\big|\big) \ge (1-\beta)^2\beta\alpha n^{\delta}. \tag{4.37}$$

Now we are ready to prove property (ii). Combining (4.33), (4.35), (4.36), and (4.37) yields that the following holds with probability $1 - O(1/n^2)$.

- If $0 < \delta < 1/3$, then

$$\frac{|S[n+1,(1+\beta)n]|}{|R[n+1,(1+\beta)n]|} \ge 1 - \frac{\varepsilon\beta\alpha n^{\delta}}{(1-\beta)^2\beta\alpha n^{\delta}} = 1 - \frac{\varepsilon}{(1-\beta)^2} \ge 1 - 2\varepsilon,$$

  where the last inequality follows with a sufficiently small $\beta = \beta(\varepsilon)$.

- If $\delta = 1/3$, then

$$\begin{aligned}\frac{|S[n+1,(1+\beta)n]|}{|R[n+1,(1+\beta)n]|} &\geq 1 - \frac{(1+\varepsilon)18\beta\alpha^4 n^{1/3}}{(1-\beta)^2\beta\alpha n^{1/3}} \geq 1 - \frac{1+\varepsilon}{(1-\beta)^2}18\alpha^3 \\ &\geq 1 - (1+2\varepsilon)18\alpha^3 \geq (1-2\varepsilon)(1-18\alpha^3),\end{aligned}$$

  where the third inequality follows with a sufficiently small $\beta = \beta(\varepsilon)$ and the last inequality follows from $\alpha \leq 0.1$.

By rescaling $\varepsilon$, this implies property (ii), which completes the proof of Lemma 4.5.4.

It still remains to prove Claim 4.5.6.

### 4.5.3 Proof of Claim 4.5.6

We prove Claim 4.5.6 by estimating $|D_i|$ $(i = 1, 2, 3, 4)$ separately. We start with a definition which is related to the definition of $D_1$ in (4.32).

**Definition 4.5.7.**

- $D := \{d \mid \{a,b,c,d\}$ *is a Sidon quadruple,* $a,b,c \in R[n],\ d \in [n+1,(1+\beta)n]\}$
- $T := \{\{a,b,c\} \mid \{a,b,c,d\}$ *is a Sidon quadruple,* $a,b,c \in R[n],\ d \in [n+1,(1+\beta)n]\}$

Since more triples $\{a,b,c\} \in T$ $(a \leq b \leq c)$ can correspond to a number $d = -a+b+c \in D$, we have

$$|D| \leq |T|. \tag{4.38}$$

**Proof of (I) in Claim 4.5.6** Recall that by (4.32)

$$D_1 := \{d \mid \{a,b,c,d\} \text{ is a Sidon quadruple, } a,b,c \in R[n],\ d \in R[n+1,(1+\beta)n]\}. \tag{4.39}$$

Note that $D_1 = D \cap R[n+1,(1+\beta)n]$ from Definition 4.5.7. Note that each element in $[n+1,(1+\beta)n]$ is chosen for $R$ independently with probability less than $p_n := \alpha n^{-1+\delta}$, and hence we have

$$\mathbb{E}(|D_1|) \leq |D|\alpha n^{-1+\delta} \overset{(4.38)}{\leq} |T|\alpha n^{-1+\delta}. \tag{4.40}$$

We will show that

$$|T| \leq (1+3\beta)^3(2/\delta^2)\beta\alpha^3 n^{3\delta}. \tag{4.41}$$

Consequently we infer $\mathbb{E}(|D_1|) \leq (1+3\beta)^3(2/\delta^2)\beta\alpha^4 n^{4\delta-1}$. Since each element $d \in D \subset [n+1,(1+\beta)n]$ is chosen for $D_1 \subset R[n+1,(1+\beta)n]$ independently, by the Chernoff bound, we infer that

$$|D_1| \leq (1+3\beta)^4(2/\delta^2)\beta\alpha^4 \max\{n^{4\delta-1}, n^{\delta/2}\}$$

with probability $1 - O(1/n^2)$, which completes the proof of (I) in Claim 4.5.6. It still remains to show (4.41).

Now we show (4.41). Let $\{a,b,c\} \in T$, where $a \leq b \leq c$. First we claim that, for any given $a$ and $b$, the number $c$ is in an interval of length $\beta n$ contained in $[n/2+1,n]$. Since $\{a,b,c,d\}$ $(a < b \leq c < d)$ is a Sidon quadruple, we have $a+d = b+c$. In other words, the distance between $a$ and $b$ is the same as the distance between $c$ and $d$, that is $b-a = d-c$. Since $d \in [n+1,(1+\beta)n]$ is in an interval of length $\beta n$, we have that, for any given $a$ and $b$, the number $c$ is also in an interval of length $\beta n$. In addition, if $c < n/2+1$, then $b-a \leq c-1 < n/2 \leq d-c$, which is impossible, and hence we infer $c \geq n/2+1$.

Since, for any given $a$ and $b$, the number $c$ is in an interval of length $\beta n$ contained in $[n/2+1,n]$, we have that for any given $a$ and $b$, there exists a

non-negative integer $k = n/2 + a - b$ such that

$$\begin{aligned}&\big|\{c \,|\, \{a,b,c\} \in T \text{ for given } a,b \in R[n]\}\big| = \big|R[n/2+k+1, n/2+k+\beta n]\big| \\ &\overset{(4.23)}{\leq} (1+\beta)\mathbb{E}(\big|R[n/2+k+1, n/2+k+\beta n]\big|) \\ &\leq (1+\beta)\mathbb{E}\Big(\big|R[n/2+1, n/2+\beta n]\big|\Big),\end{aligned}$$

where the last inequality follows from the fact that the probability $p_m$ to chose $m \in \mathbb{N}$ to $R$ is decreasing. Inequality (4.3) with $n/2$ and $2\beta$ instead of $n$ and $\beta$ yields that

$$\big|\{c \,|\, \{a,b,c\} \in T \text{ for given } a,b \in R[n]\}\big| \leq (1+\beta)\alpha 2\beta(n/2)^\delta \leq (1+\beta)2\beta\alpha n^\delta.$$

Thus we infer

$$\begin{aligned}|T| &\leq |R[n]|^2 \cdot (1+\beta)2\beta\alpha n^\delta \overset{(4.24)}{\leq} (1+3\beta)^2[(\alpha/\delta)n^\delta]^2 \cdot (1+\beta)2\beta\alpha n^\delta \\ &\leq (1+3\beta)^3(2/\delta^2)\beta\alpha^3 n^{3\delta},\end{aligned}$$

which implies (4.41), and hence this completes the proof of (I) in Claim 4.5.6. □

Now we are going to estimate $|D_2|$ where $D_2$ is defined in (4.32). We start with the definition of the following auxiliary graphs.

**Definition 4.5.8.** *Let $R[n] \subset [n]$ be given.*

- *Let $G = (V, E)$ be the graph with vertex set $V = [n+1, (1+\beta)n]$ and edge set*

$$E = \{\{c,d\} \,|\, \{a,b,c,d\} \text{ is a Sidon quadruple, } a < b \leq c < d, \\ a,b \in R[n],\ c,d \in [n+1,(1+\beta)n]\}. \quad (4.42)$$

- *Let $G^R = (V^R, E^R)$ be the subgraph of $G$ induced on $V^R = R[n+1, (1+$*

$\beta)n]$.

**Proof of (II) in Claim 4.5.6** Recall that by (4.32)

$$D_2 := \{d \mid \{a,b,c,d\} \text{ is a Sidon quadruple, } a < b \le c < d, \\ a,b \in R[n],\ c,d \in R[n+1,(1+\beta)n]\}.$$

Since several edges of $E^R$ in Definition 4.5.8 can correspond to an integer $d \in D_2$, observe that

$$|D_2| \le |E^R|. \tag{4.43}$$

Thus, in order to show (II) in Claim 4.5.6, which estimates $|D_2|$, it suffices to estimate $|E^R|$.

First we consider $\mathbb{E}(|E^R|)$. To this end we estimate $|E|$ in Definition 4.5.8. For fixed $a \in R[n]$, let

$$\Delta_a = \{b-a \mid 0 < b-a \le \beta n,\ b \in R[n]\} \quad \text{and} \quad \Delta = \bigcup_{a \in R[n]} \Delta_a. \tag{4.44}$$

Since $R[n]$ is well distributed, we have that for each $a \in R[n]$

$$\begin{aligned} |\Delta_a| &= |R[a+1,a+\beta n]| \le (1+\beta)\mathbb{E}[|R[a+1,a+\beta n]|] \le 2\mathbb{E}[|R[\beta n]|] \\ &\le 2(\alpha/\delta)\beta^\delta n^\delta = (2/\delta)\beta^\delta \alpha n^\delta, \end{aligned} \tag{4.45}$$

where the second inequality follows from assumption $\beta \le 1$ and the fact that the probability $p_m$ for choosing $m$ to $R$ is decreasing and the last inequality follows from the second inequality of (4.2) with $a = 0$ and $b = \beta$. We infer that

$$\begin{aligned} |\Delta| \overset{(4.44)}{\le} \sum_{a \in R[n]} |\Delta_a| \overset{(4.45)}{\le} |R[n]|(2/\delta)\beta^\delta \alpha n^\delta \\ \overset{(4.24)}{\le} 4(\alpha/\delta)n^\delta \cdot (2/\delta)\beta^\delta \alpha n^\delta \le (8/\delta^2)\beta^\delta \alpha^2 n^{2\delta}. \end{aligned} \tag{4.46}$$

Consequently, we have

$$|E| \leq (\text{choice of } c)(\text{choice of } d) = (\beta n)|\Delta| \leq (8/\delta^2)\beta^{1+\delta}\alpha^2 n^{2\delta+1}.$$

Note that each edge of $E$ is chosen to $E^R$ if the two end vertices of the edge are chosen to $R[n+1,(1+\beta)n]$, which happens with probability at most $p_n^2$. Thus

$$\begin{aligned}\mathbb{E}(|E^R|) \leq |E| \cdot p_n^2 &\leq (8/\delta^2)\beta^{1+\delta}\alpha^2 n^{2\delta+1}(\alpha n^{-1+\delta})^2 \\ &\leq (8/\delta^2)\beta^{1+\delta}\alpha^4 n^{4\delta-1}. \quad (4.47)\end{aligned}$$

Next, in order to consider the concentration of $|E^R|$, we apply Corollary 4.3.8 with $\mathcal{H} = G$ and $\mathcal{H}[R] = G^R$, where $G = (V,E)$ and $G^R = (V^R, E^R)$ are introduced in Definition 4.5.8. First, we estimate $\mathbb{E}_i$ which are introduced in Definition 4.3.6.

- **Estimating $\mathbb{E}_1$:** Fix $u \in V = [n+1,(1+\beta)n]$. We consider $\mathbb{E}_{\{u\}}$, that is, the expected number of edges in $G^R$ containing vertex $u$ under the condition that $u \in V^R = R[n+1,(1+\beta)n]$. Note that the number of edges of $E$ containing vertex $u$ is at most $2|\Delta|$. Under the condition that vertex $u$ is in $R[n+1,(1+\beta)n]$, such an edge is chosen to $E^R$ if the other end vertex of the edge is chosen to $R[n+1,(1+\beta)n]$, which happens with probability at most $p_n$. Hence we infer that

$$\begin{aligned}\mathbb{E}_1 &:= \max \mathbb{E}_{\{u\}} \leq 2|\Delta|p_n \overset{(4.46)}{\leq} (16/\delta^2)\beta^\delta\alpha^2 n^{2\delta} \cdot (\alpha n^{-1+\delta}) \\ &= (16/\delta^2)\beta^\delta\alpha^3 n^{3\delta-1} = O_\delta(\beta^\delta\alpha^3),\end{aligned}$$

  where the last "=" follows from $\delta \leq 1/3$, that is, $3\delta - 1 \leq 0$ and the notation $f(n) = O_\delta(g(n))$ means that $f(n) \leq c_\delta \cdot g(n)$ with some constant $c_\delta$ which only depends on $\delta$. Under assumption $\beta \leq 1$, we have $\mathbb{E}_1 = O_{\alpha,\delta}(1)$.

- **Estimating $\mathbb{E}_2$:** Clearly, we have $\mathbb{E}_2 \leq 1$.

Recalling the assumption $0 < \delta \leq 1/3$, we set

$$f(n) = (8/\delta^2)\beta^{1+\delta}\alpha^4 \max\{n^{4\delta-1}, n^{\delta/2}\}.$$

Comparing (4.47) with the bounds on $\mathbb{E}_1$ and $\mathbb{E}_2$, we infer that

$$\mathbb{E}(|E^R|) \leq f(n) \quad \text{and} \quad \mathbb{E}' := \max\{\mathbb{E}_1, \mathbb{E}_2\} = O_{\alpha,\delta}(1) \leq f(n) \cdot n^{-\delta/4}$$

for every sufficiently large $n$. Corollary 4.3.8 with $\mathcal{H} = G$ and $\mathcal{H}[R] = G^R$ implies that with probability $1 - O(1/n^2)$,

$$|E^R| \leq 2f(n) = (16/\delta^2)\beta^{1+\delta}\alpha^4 \max\{n^{4\delta-1}, n^{\delta/2}\}.$$

Thus, by (4.43), we have that

$$|D_2| \leq |E^R| \leq (16/\delta^2)\beta^{1+\delta}\alpha^4 \max\{n^{4\delta-1}, n^{\delta/2}\}$$

holds with probability $1 - O(1/n^2)$, which completes the proof of (II) in Claim 4.5.6. □

Now we are going to estimate $|D_3|$ where $D_3$ is defined in (4.32).

**Proof of (III) in Claim 4.5.6** Since our proof of (III) is similar to the proof of (II) above, we only give a sketch. Let $H_3^R = (V^R, E_3^R)$ be the hypergraph with vertex set $V^R = R[n+1, (1+\beta)n]$ and edge set

$$E_3^R = \{\{b,c,d\} \mid \{a,b,c,d\} \text{ is a Sidon quadruple}, a < b \leq c < d, \\ a \in R[n],\ b,c,d \in R[n+1,(1+\beta)n]\}.$$

Recall that by (4.32)

$$D_3 := \{d \mid \{a,b,c,d\} \text{ is a Sidon quadruple}, a < b \leq c < d, \\ a \in R[n],\ b,c,d \in R[n+1,(1+\beta)n]\}.$$

Since several hyperedges of $E_3^R$ can correspond to an integer $d \in D_3$, observe that

$$|D_3| \le |E_3^R|. \tag{4.48}$$

Hence, in order to estimate $|D_3|$, it suffices to estimate $|E_3^R|$. By (4.24) and assumption $0 < \beta \le 1$, we have $|R[n]| \le (4/\delta)\alpha n^\delta$, and hence, one can show the following:

$\mathbb{E}(|E_3^R|) \le |R[n]|(\beta n)^2 p_n^3 \le (4/\delta)\alpha n^\delta (\beta n)^2 (\alpha n^{-1+\delta})^3 \le (4/\delta)\beta^2\alpha^4 n^{4\delta-1}$,

$\mathbb{E}_1(|E_3^R|) \le |R[n]|(\beta n)O(1)p_n^2 = O\Big((4/\delta)\alpha n^\delta(\beta n)(\alpha n^{-1+\delta})^2\Big) = O_\delta\Big(\beta\alpha^3 n^{3\delta-1}\Big)$
$= O_{\alpha,\delta}(1)$,

where the last "=" follows from assumption $\delta \le 1/3$ and $\beta \le 1$,

$\mathbb{E}_2(|E_3^R|) \le |R[n]|O(1)p_n = O\Big((4/\delta)\alpha n^\delta(\alpha n^{-1+\delta})\Big) = O_\delta\Big(\alpha^2 n^{2\delta-1}\Big) = O_{\alpha,\delta}(1)$,

$\mathbb{E}_3(|E_3^R|) \le 1$.

Recalling the assumption $0 < \delta \le 1/3$, we set

$$f(n) = (4/\delta)\beta^2\alpha^4 \max\{n^{4\delta-1}, n^{\delta/2}\}.$$

By the above computation, we infer that

$$\mathbb{E}(|E_3^R|) \le f(n) \quad \text{and} \quad \mathbb{E}' := \max\{\mathbb{E}_1, \mathbb{E}_2, \mathbb{E}_3\} = O_{\alpha,\delta}(1) \le f(n)\cdot n^{-\delta/4}$$

for every sufficiently large $n$. Corollary 4.3.8 implies that with probability $1 - O(1/n^2)$,

$$|E_3^R| \le 2f(n) = (8/\delta)\beta^2\alpha^4 \max\{n^{4\delta-1}, n^{\delta/2}\}.$$

Thus, by (4.48), we have that

$$|D_3| \le |E_3^R| \le (8/\delta)\beta^2\alpha^4 \max\{n^{4\delta-1}, n^{\delta/2}\}$$

holds with probability $1 - O(1/n^2)$, which completes the proof of (III) in Claim 4.5.6. □

Finally we estimate $|D_4|$ where $D_4$ is defined in (4.32).

**Proof of (IV) in Claim 4.5.6** Since our proof of (IV) is similar to the proof of (II) and (III) above, we only give a sketch. Let $H_4^R = (V^R, E_4^R)$ be the hypergraph with vertex set $V^R = R[n+1,(1+\beta)n]$ and edge set

$$E_4^R = \{\{a,b,c,d\} \subset R[n+1,(1+\beta)n] \,|\, \{a,b,c,d\} \text{ is a Sidon quadruple, } a<b\le c<d\}.$$

Recall that by (4.32)

$$D_4 := \{d \mid \{a,b,c,d\} \subset R[n+1,(1+\beta)n] \text{ is a Sidon quadruple, } a<b\le c<d\}.$$

Since several hyperedges of $E_4^R$ can correspond to an integer $d \in D_4$, observe that

$$|D_4| \le |E_4^R|. \tag{4.49}$$

Hence, similarly as before, in order to estimate $|D_4|$, it suffices to estimate $|E_4^R|$. One can show the following:

$\mathbb{E}(|E_4^R|) \le (\beta n)^3 p_n^4 \le (\beta n)^3(\alpha n^{-1+\delta})^4 \le \beta^3\alpha^4 n^{4\delta-1}$,

$\mathbb{E}_1(|E_4^R|) \le (\beta n)^2 O(1) p_n^3 = O\Big((\beta n)^2(\alpha n^{-1+\delta})^3\Big) = O\Big(\beta^2\alpha^3 n^{3\delta-1}\Big) = O_\alpha(1)$,
where the last "=" follows from assumption $\beta \le 1$ and $\delta \le 1/3$,

$\mathbb{E}_2(|E_4^R|) \le (\beta n) O(1) p_n^2 = O\Big(\beta n(\alpha n^{-1+\delta})^2\Big) = O(\beta\alpha^2 n^{2\delta-1}) = O_\alpha(1)$,

$\mathbb{E}_3(|E_4^R|) \le O(1) p_n = O(\alpha n^{\delta-1}) = O_\alpha(1)$,

$\mathbb{E}_4(|E_4^R|) \le 1$.

Recalling the assumption $0 < \delta \le 1/3$, we set $f(n) = \beta^3\alpha^4 \max\{n^{4\delta-1}, n^{\delta/2}\}$. By the above computation, we infer that

$$\mathbb{E}(|E_4^R|) \le f(n) \quad \text{and} \quad \mathbb{E}' := \max\{\mathbb{E}_1,\mathbb{E}_2,\mathbb{E}_3,\mathbb{E}_4\} = O_\alpha(1) \le f(n)\cdot n^{-\delta/4}$$

for every sufficiently large $n$. Corollary 4.3.8 implies that with probability $1 - O(1/n^2)$,

$$|E_4^R| \le 2f(n) = 2\beta^3\alpha^4 \max\{n^{4\delta-1}, n^{\delta/2}\}.$$

Thus, by (4.49), we have that

$$|D_4| \leq |E_4^R| \leq 2\beta^3\alpha^4 \max\{n^{4\delta-1}, n^{\delta/2}\}$$

holds with probability $1 - O(1/n^2)$, which completes the proof of (IV) in Claim 4.5.6. □

# Bibliography


[1] M. Ajtai, J. Komlós, J. Pintz, J. Spencer, and E. Szemerédi, *Extremal uncrowded hypergraphs*, J. Combin. Theory Ser. A **32** (1982), no. 3, 321–335. (Cited on page 34.)

[2] N. Alon, J. Balogh, R. Morris, and W. Samotij, *Counting sum-free sets in Abelian groups*, submitted. (Cited on page 13.)

[3] N. Alon and J. H. Spencer, *The probabilistic method*, second ed., Wiley-Interscience Series in Discrete Mathematics and Optimization, Wiley-Interscience [John Wiley & Sons], New York, 2000, With an appendix on the life and work of Paul Erdős. (Cited on pages 28 and 81.)

[4] J. Balogh and W. Samotij, *The number of $K_{m,m}$-free graphs*, Combinatorica **31** (2011), no. 2, 131–150. (Cited on page 13.)

[5] ———, *The number of $K_{s,t}$-free graphs*, J. Lond. Math. Soc. (2) **83** (2011), no. 2, 368–388. (Cited on page 13.)

[6] B. Bollobás, *Random graphs*, Academic Press Inc. [Harcourt Brace Jovanovich Publishers], London, 1985. (Cited on pages 11 and 12.)

[7] R. C. Bose and S. Chowla, *Theorems in the additive theory of numbers*, Comment. Math. Helv. **37** (1962/1963), 141–147. (Cited on page 41.)

[8] 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. (Cited on pages 6 and 42.)

[9] S. Chen, *On the size of finite Sidon sequences*, Proc. Amer. Math. Soc. **121** (1994), no. 2, 353–356. (Cited on page 41.)

[10] 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. (Cited on pages 5, 8, 36, 41, and 70.)

[11] J. Cilleruelo, *New upper bounds for finite $B_h$ sequences*, Adv. Math. **159** (2001), no. 1, 1–17. (Cited on page 41.)

[12] D. Conlon and W. T. Gowers, *Combinatorial theorems in sparse random sets*, submitted, 70pp, 2010. (Cited on pages 2, 7, 8, 43, and 44.)

[13] D. Dellamonica, Jr., Y. Kohayakawa, S. J. Lee, V. Rödl, and W. Samotij, *The number of $B_3$-sets and the maximum size of $B_3$-sets contained in a sparse random set of integers*, In Preparation. (Cited on pages 42 and 43.)

[14] R. A. Duke, H. Lefmann, and V. Rödl, *On uncrowded hypergraphs*, Proceedings of the Sixth International Seminar on Random Graphs and Probabilistic Methods in Combinatorics and Computer Science, "Random Graphs '93" (Poznań, 1993), vol. 6, 1995, pp. 209–212. (Cited on page 34.)

[15] A. G. D'yachkov and V. V. Rykov, *$B_s$-sequences*, Mat. Zametki **36** (1984), no. 4, 593–601. (Cited on page 41.)

[16] P. Erdős, *On a problem of Sidon in additive number theory and on some related problems. Addendum*, J. London Math. Soc. **19** (1944), 208. (Cited on pages 5, 8, 36, and 41.)

[17] 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. (Cited on pages 5, 8, and 41.)

[18] Z. Füredi, *Random Ramsey graphs for the four-cycle*, Discrete Math. **126** (1994), no. 1-3, 407–410. (Cited on page 13.)

[19] B. Green, *The number of squares and $B_h[g]$ sets*, Acta Arith. **100** (2001), no. 4, 365–390. (Cited on page 41.)

[20] B. Green and T. Tao, *The primes contain arbitrarily long arithmetic progressions*, Ann. of Math. (2) **167** (2008), no. 2, 481–547. (Cited on page 2.)

[21] H. Halberstam and K. F. Roth, *Sequences*, second ed., Springer-Verlag, New York, 1983. (Cited on pages 5, 41, and 74.)

[22] S. Janson, T. Łuczak, and A. Ruciński, *An exponential bound for the probability of nonexistence of a specified subgraph in a random graph*, Random graphs '87 (Poznań, 1987), Wiley, Chichester, 1990, pp. 73–87. (Cited on page 2.)

[23] ———, *Random graphs*, Wiley-Interscience, New York, 2000. (Cited on pages 11, 12, 34, and 45.)

[24] X. D. Jia, *On finite Sidon sequences*, J. Number Theory **44** (1993), no. 1, 84–92. (Cited on page 41.)

[25] J. H. Kim and V. H. Vu, *Concentration of multivariate polynomials and its applications*, Combinatorica **20** (2000), no. 3, 417–434. (Cited on pages 27, 28, and 81.)

[26] D. J. Kleitman and K. J. Winston, *On the number of graphs without* 4*-cycles*, Discrete Math. **41** (1982), no. 2, 167–172. (Cited on page 13.)

[27] Y. Kohayakawa, *Szemerédi's regularity lemma for sparse graphs*, Foundations of computational mathematics (Rio de Janeiro, 1997), Springer, Berlin, 1997, pp. 216–230. (Cited on page 2.)

[28] 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), no. 1, 101–120. (Cited on page 13.)

[29] 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. (Cited on page 2.)

[30] Y. Kohayakawa, S. J. Lee, and V. Rödl, *Infinite Sidon sets contained in a sparse random set of integers*, In Preparation. (Cited on page 74.)

[31] 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*, Submitted. (Cited on pages 6, 7, 13, 70, 77, and 78.)

[32] 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. (Cited on page 41.)

[33] J. Komlós, M. Sulyok, and E. Szemeredi, *Linear problems in combinatorial number theory*, Acta Math. Acad. Sci. Hungar. **26** (1975), 113–121. (Cited on page 35.)

[34] F. Krückeberg, *$B_2$-Folgen und verwandte Zahlenfolgen*, J. Reine Angew. Math. **206** (1961), 53–60. (Cited on page 41.)

[35] B. Lindström, *A remark on $B_4$-sequences*, J. Combinatorial Theory **7** (1969), 276–277. (Cited on page 41.)

[36] K. O'Bryant, *A complete annotated bibliography of work related to Sidon sequences*, Electron. J. Combin. (2004), Dynamic surveys 11, 39 pp. (electronic). (Cited on pages 5 and 41.)

[37] O. Reingold, L. Trevisan, M. Tulsiani, and S. P. Vadhan, *Dense subsets of pseudorandom sets*, FOCS, 2008, pp. 76–85. (Cited on page 2.)

[38] ———, *Dense subsets of pseudorandom sets*, Electronic Colloquium on Computational Complexity (ECCC) **15** (2008), Report No.45. (Cited on page 2.)

[39] K. F. Roth, *On certain sets of integers*, J. London Math. Soc. **28** (1953), 104–109. (Cited on page 2.)

[40] M. Schacht, *Extremal results for random discrete structures*, submitted, 27pp, 2009. (Cited on pages 2, 7, 8, 43, and 44.)

[41] I. E. Shparlinskiĭ, *On $B_s$-sequences*, Combinatorial analysis, No. 7 (Russian), Moskov. Gos. Univ., Moscow, 1986, pp. 42–45, 163. (Cited on page 41.)

[42] J. Singer, *A theorem in finite projective geometry and some applications to number theory*, Transactions of the American Mathematical Society **43** (1938), 377–385. (Cited on pages 5 and 41.)

[43] 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. (Cited on page 2.)

[44] T. Tao and V. Vu, *Additive combinatorics*, Cambridge Studies in Advanced Mathematics, vol. 105, Cambridge University Press, Cambridge, 2006. (Cited on pages 2 and 5.)

[45] L. Trevisan, *Guest column: additive combinatorics and theoretical computer science*, SIGACT News **40** (2009), no. 2, 50–66. (Cited on page 5.)