Sumsets in the Hypercube

Noga Alon * Or Zamir †

Abstract

A subset S of the Boolean hypercube $\mathbb{F}_2^n$ is a *sumset* if $S = A + A = \{a + b \mid a, b \in A\}$ for some $A \subseteq \mathbb{F}_2^n$. We prove that the number of sumsets in $\mathbb{F}_2^n$ is asymptotically $(2^n - 1)2^{2^{n-1}}$. Furthermore, we show that the family of sumsets in $\mathbb{F}_2^n$ is almost identical to the family of all subsets of $\mathbb{F}_2^n$ that contain a complete linear subspace of co-dimension 1.

1 Introduction

A subset S of an Abelian group G is a *sumset* if $S = A + A = \{a + b \mid a, b \in A\}$ for some $A \subseteq G$. Sumsets are among the most fundamental objects studied in *additive combinatorics* (for a comprehensive survey, see the book of Tao and Vu [TV06]). When G is finite, then arguably the simplest question regarding sumsets in G is how many distinct ones there are. A classical work of Green and Ruzsa [GR04] and later refinements [Sar15] study this question for $G = \mathbb{F}_p$. Other well studied counting questions with regards to sumsets include estimating the number of different sums $A + B$ when the sizes of $|A|$ and $|B|$ are both large [AGU10, SS20]; or estimating the number of sets A with a small sumset $|A + A| \leq K|A|$ [Fre73, Gre05a, ABMS14, GM16, Cam18].

Most of the results cited above study the case of $G = \mathbb{F}_p$ for some large prime p. In this work, we focus on the Boolean hypercube $G = \mathbb{F}_2^n$, that is, the vector space of dimension n over F_2. This choice naturally comes up in applications of additive combinatorics to theoretical computer science (see for example the surveys of [BTW07, Tre09, Vio11, Bib13, Lov17]). Furthermore, vectors spaces $\mathbb{F}_p^n$ over finite fields (also called finite field models) are often easier to study due to the availability of tools from linear algebra [Gre05b]. Recent breakthroughs in additive combinatorics also focus on finite field models: Kelley and Meka's bounds for 3-progressions [KM23] are proven starting with an analysis in $\mathbb{F}_p^n$; Gowers, Green, Manners and Tao prove the polynomial Freiman–Ruzsa conjecture in $\mathbb{F}_2^n$ [GGMT23].

Put $N := 2^n$ and denote the family of sumsets in $\mathbb{F}_2^n$ by $\mathcal{S}_n := \{A + A \mid A \subseteq \mathbb{F}_2^n\}$. Our main focus in this work is to understand the cardinality of $\mathcal{S}_n$ and the typical structure of an element in it. It was shown in [Sar15] that $|\mathcal{S}_n| = 2^{N/2+o(N)}$. In a recent work on property testing of sumsets [CNR+24], this bound was improved to show that $|\mathcal{S}_n|$ is between $2^{N/2}$ and $2^{N/2+n^2+O(1)}$. The following theorem provides an asymptotic formula for $|\mathcal{S}_n|$. Here and in what follows the notation $f \sim g$ for two functions f and g denotes that $f = (1 + o(1))g$ where the $o(1)$-term tends to 0 as the parameters of the functions grow to infinity. Equivalently, this means that the limit of the ratio f/g as the parameters grow is 1.

Theorem 1. $|\mathcal{S}_n| \sim (2^n - 1)2^{N/2}$.

*Princeton University, Princeton, NJ, USA and Tel Aviv University, Tel Aviv, Israel. Email: nalon@math.princeton.edu. Research supported in part by NSF grant DMS-2154082.

†Tel Aviv University, Tel Aviv, Israel. Email: orzamir90@gmail.com

Furthermore, we provide a simple characterization of almost all sumsets in $\mathbb{F}_2^n$. Denote the family of subsets of $\mathbb{F}_2^n$ that contain a complete linear subspace of co-dimension 1 by

$$\mathcal{H}_n := \{S \subseteq \mathbb{F}_2^n \mid \exists\, v \in \mathbb{F}_2^n \,.\, v^\perp \subseteq S\}.$$

We show that almost all sets in $\mathcal{S}_n$ are also in $\mathcal{H}_n$ and vice versa.

Theorem 2. $|\mathcal{S}_n \Delta \mathcal{H}_n| = o\left(|\mathcal{S}_n|\right)$.

The containment of large linear subspaces within sumsets was studied beforehand. As a generalization to a question of Bourgain on arithmetic progressions in sums of sets of integers [Bou90], Green asked whether for every large $A \subseteq \mathbb{F}_2^n$, the sumset $A + A$ must contain a large linear subspace [Gre05b]. It is simple to observe that if $|A| > \frac{1}{2}|\mathbb{F}_2^n|$, then $A + A = \mathbb{F}_2^n$. Sanders [San11] showed that if $|A| > \left(1/2 - \frac{1}{2^9\sqrt{n}}\right)|\mathbb{F}_2^n|$ then $A + A$ contains a subspace of co-dimension 1, or with our notation, $A + A \in \mathcal{H}_n$. Other works (e.g., [San12]) continued studying the relation between the density of A and the largest linear subspace contained in $A + A$.

Our lower bound to $|\mathcal{S}_n|$ is rather straightforward. For the upper bound, we need several structural claims on both sumsets in $\mathbb{F}_2^n$ and independent sets in the n-th dimensional hypercube Q_n, that may be of independent interest. As a byproduct to the discussion here, we show that almost all subsets of $\mathbb{F}_2^n$ cannot be expressed as unions of less than $O(N/(n^2 \log n))$ sumsets, and this is tight up to a factor of $(\log n)^2$. This is somewhat counter-intuitive as cardinality-wise, even the family of pairs of sumsets is larger than the power set of the hypercube, that is, $|\mathcal{P}\left(\mathbb{F}_2^n\right)| < |\mathcal{S}_n|^2$. The proof works for any finite abelian group, providing similar estimates.

2 Notation

When the value of n is fixed and not ambiguous, we use N instead of 2^n and A^c instead of $\mathbb{F}_2^n \setminus A$. We write $A + B := \{a + b : a \in A, b \in B\}$. If one of the sets is a singleton, e.g. if $A = \{a\}$, we will sometimes write $a + B := \{a\} + B$ instead. We use $v^\perp$ to denote the set of all vectors in $\mathbb{F}_2^n$ orthogonal to v. As already mentioned, the notation $f \sim g$ means that $f = (1 + o(1))g$, that is, $\lim(f/g) = 1$.

We write $H(x)$ to denote the binary entropy function $-x \log_2 x - (1-x)\log_2(1-x)$. Stirling's approximation gives the following helpful identity in which $\Theta^*(f)$ denotes, as usual, $f(\log f)^{\Theta(1)}$:

$$\binom{n}{\alpha n} = \Theta^*(2^{H(\alpha)n}); \text{ or, equivalently, } \binom{2^n}{\alpha 2^n} = 2^{H(\alpha)2^n} \cdot 2^{\Theta(n)}. \tag{1}$$

Given a subset D of an Abelian group G, we write $\Gamma_G(D)$ to denote the *Cayley (sum) graph* of G with respect to the generator set D; that is, the graph on the vertex set G that contains the edge (x, y) if and only if $x + y \in D$. When $D = \{x\}$ is a singleton for some $x \in G$, we abuse notation slightly and write $\Gamma_G(x)$ for $\Gamma_G(\{x\})$.

3 Simple Bounds

In this section we cover the bounds of [CNR+24], where it is shown that $|\mathcal{S}_n|$ is between $2^{N/2}$ and $2^{N/2+n^2+O(1)}$.

Lemma 3. *Fix a sumset $A + A = S \subseteq \mathbb{F}_2^n$ and a set $D \subseteq \mathbb{F}_2^n$ with $S \cap D = \emptyset$. Then A is an independent set in the Cayley graph $\Gamma_{\mathbb{F}_2^n}(D)$.*

Proof. Assume for contradiction that A is not an independent set in $\Gamma_{\mathbb{F}_2^n}(D)$. This implies the existence of $x, y \in A$ with $x + y = s$ for some $s \in D$. Hence, $s \in D \cap (A + A) = D \cap S$, which is a contradiction. □

Proposition 4. *The number of sumsets in $\mathbb{F}_2^n$ is at most*

$$2^{2^{n-1}+n^2+O(1)}.$$

Proof. Consider a sumset $S = A + A$; we consider two cases depending on the linear rank of the set $\mathbb{F}_2^n \setminus S$.

Case 1: $\mathbb{F}_2^n \setminus S$ does not have full rank. In other words, there exists a vector $v \in \mathbb{F}_2^n$ such that

$$\langle x, v \rangle = 1 \qquad \text{implies that} \qquad x \in S.$$

In particular, we have that $S = S' \cup \left(\mathbb{F}_2^n \setminus v^\perp\right)$ for some $S' \subseteq v^\perp = \{x \in \mathbb{F}_2^n : \langle x, v \rangle = 0\}$. As there are at most 2^n choices for v, and for each choice of v there are at most $2^{2^{n-1}}$ choices for S', we have that there are at most $2^{2^{n-1}+n}$ many sumsets of this form.

Case 2: $\mathbb{F}_2^n \setminus S$ has full rank. In particular, there are n linearly independent vectors *not* in S. As we have n linearly independent vectors not in S it follows by Lemma 3 that there exists a nonsingular transformation of $\mathbb{F}_2^n$ such that A must be an independent set in the graph of the hypercube Q_n. As the number of independent sets in the hypercube Q_n is at most $2^{2^{n-1}+O(1)}$ (see for example [KS83, Gal19]), and as the number of nonsingular transformations of the hypercube is at most 2^{n^2}, it follows that the total number of sumsets of this form is at most

$$2^{2^{n-1}+n^2+O(1)}.$$

Both cases together complete the proof. □

Proposition 5. *The number of sumsets in $\mathbb{F}_2^n$ is at least $2^{2^{n-1}}$.*

Proof. For any subset $A \subseteq \mathbb{F}_2^{n-1}$ of the $(n-1)$-th dimensional hypercube, we define a subset $A' \subseteq \mathbb{F}_2^n$ as

$$A' := \{\vec{0}\} \cup \{(1, a) \mid a \in A\},$$

where for a $(n-1)$-dimensional vector a, the concatenation $(1, a)$ is defined as the n-dimensional vector where the first coordinate is 1 and the other $(n-1)$ coordinates are equal to a. We observe that $\left(A' + A'\right) \cap \left(\mathbb{F}_2^n \setminus e_1^\perp\right) = \{(1, a) \mid a \in A\}$. That is, in the sumset $(A' + A')$ all vectors in which the first coordinate is 1 exactly correspond to the set A. In particular, for any $A_1 \neq A_2 \in \mathbb{F}_2^{n-1}$, we have $(A_1' + A_1') \neq (A_2' + A_2')$. □

4 Lower Bound

In this section we improve Proposition 5. We prove that almost all subsets in $\mathcal{H}_n$ are also sumsets, and that the size of $\mathcal{H}_n$ is asymptotically $2^{2^{n-1}}(2^n - 1)$.

Theorem 6. $|\mathcal{H}_n| \sim 2^{2^{n-1}}(2^n - 1)$.

Theorem 7. $|\mathcal{H}_n \setminus \mathcal{S}_n| = o\left(|\mathcal{H}_n|\right)$.

Proof of Theorem 6. The set $\mathcal{H}_n$ is a union of the (2^n-1) sets $H(v)$, where for each nonzero vector v in $\mathbb{F}_2^n$, $H(v)$ is the family of all subsets containing $v^\perp$. As the size of each set $H(v)$ is $2^{2^{n-1}}$ this shows that $|\mathcal{H}_n| \leq 2^{2^{n-1}}(2^n-1)$.

Since the intersection of each two distinct sets $H(u), H(v)$ is of cardinalty $2^{2^{n-2}}$ it follows that

$$|\mathcal{H}_n| \geq 2^{2^{n-1}}(2^n-1) - \frac{(2^n-1)(2^n-2)}{2} \cdot 2^{2^{n-2}} \geq 2^{2^{n-1}}(2^n-1) - 2^{2n-1+2^{n-2}} = \left(1-o\left(1\right)\right) 2^{2^{n-1}}(2^n-1).$$

□

To prove Theorem 7, we use the following observation.

Lemma 8. *There are at most $2^n \cdot 3^{2^{n-1}}$ subsets $A \subseteq \mathbb{F}_2^n$ for which $A+A \neq \mathbb{F}_2^n$.*

Proof. By Lemma 3, if $x \notin A+A$ then A is an independent set in the 1-regular Cayley graph $\Gamma_{\mathbb{F}_2^n}(x)$, which is a perfect matching in Q_n. Thus, there are at most $3^{2^{n-1}}$ subsets A such that $x \notin A+A$. We complete the proof by taking a union bound over all 2^n choices for x. □

Proof of Theorem 7. Take $S \in \mathcal{H}_n$, and denote by v the vector such that $v^\perp \subseteq S$. Denote by $A := \{\vec{0}\} \cup \left(S \setminus v^\perp\right)$. Note that $(A+A) \cap \left(\mathbb{F}_2^n \setminus v^\perp\right) = S \setminus v^\perp$. Thus, $A+A = S$ (and hence $S \in \mathcal{S}_n$) if and only if $(A+A) \cap v^\perp = v^\perp$. On the other hand, $(A+A) \cap v^\perp = \{\vec{0}\} \cup \left(\left(S \setminus v^\perp\right) + \left(S \setminus v^\perp\right)\right)$. Denote by $S' := \left(S \setminus v^\perp\right) - \{u\}$, where u is, say, the lexicographically first vector not in $v^\perp$. The shift by u does not change the sumset, and it is now a subset of $\mathbb{F}_2^n \cap v^\perp$ which is isomorphic to the hypercube $\mathbb{F}_2^{n-1}$ of one dimension less. By Lemma 8 there are at most $2^{n-1} \cdot 3^{2^{n-2}}$ such subsets S' for which $S'+S'$ is not complete. We conclude that $|\mathcal{H}_n \setminus \mathcal{S}_n| \leq 2^n \cdot 2^{n-1} \cdot 3^{2^{n-2}} = o\left(2^{2^{n-1}}\right)$. □

5 Structural Tools for the Upper Bound

5.1 Unions of Sumsets

In this section we prove an upper bound for the number of subsets of $\mathbb{F}_2^n$ that can be written as a union of at most k sumsets. In the proof of the upper bound in Section 6, we in fact use this statement only with regards to unions of two sumsets. The proof we describe below works for any finite abelian group of order N, and implies that only a negligible number of subsets of such a group can be expressed as a union of at most $O(N/\log^3 N)$ sumsets. As we describe later, this can be improved to $O(N/(\log^2 N \log\log N))$, which is tight up to a factor of $(\log\log N)^2$. Since for our purpose here we only need the case of unions of two sumsets we first describe a simple self-contained proof of a weaker estimate that suffices. As done throughout the paper here too we make no attempt to optimize the absolute constants.

Theorem 9. *Let G be a finite abelian group of order N. Then for any integer $s \geq 64\log^2 N$, the number of subsets of G that can be expressed as a union of at most $k = \lfloor \frac{N}{2s\ln(eN/s)} \rfloor$ sumsets is at most $2^{N-s/8} + e^{N/2}$.*

Note that taking, say, $s = N/20$ it follows that the number of sets that can be expressed as a union of two sumsets is (much) smaller than $2^{N-N/200}$.

The proof is based on the approach used in the study of the typical independence number of random Cayley (or Cayley sum) graphs of abelian groups, see, for example, [AO95, Alo13, CNR+24]. For completeness we repeat (or slightly paraphrase) those arguments to derive an explicit tail bound.

Definition 10. *We say that a set $A \subseteq G$ has* many sums *if $|A+A| \geq \frac{1}{4}|A|^2$.*

Lemma 11 (Appears in [Alo13, CNR+24])**.** *Let $A \subseteq G$ be a non-empty set, then there exists $A' \subseteq A$ such that $|A'| \geq \sqrt{|A|}$ and A' has many sums.*

Proof. We construct A' iteratively in a greedy manner. Starting from $A' = \{a\}$ for an arbitrary $a \in A$, as long as there is any $x \in A \setminus A'$ such that $\left(A'+x\right) \setminus \left(A'+A'\right) \geq \frac{1}{2}\left(|A'|+1\right)$ then we add x to A'.

Let $a \in A'$. We observe that $|A'+A'| \leq \frac{|A'|\cdot\left(|A'|-1\right)}{2}$, and thus for a uniformly chosen x from A we have

$$\mathbf{Pr}[a+x \in A'+A'] \leq \frac{|A'+A'|}{|A|} \leq \frac{|A'|\cdot\left(|A'|-1\right)}{2|A|}.$$

Hence, for a uniformly chosen x from A the expected number of elements $a \in A'$ for which $a+x \in A'+A'$ is at most $|A'| \cdot \frac{|A'|\cdot\left(|A'|-1\right)}{2|A|} = \frac{|A'|-1}{2} \cdot \frac{|A'|^2}{|A|}$. In particular, there exists such an outcome of $x \in A$. If $|A'| \leq \sqrt{|A|}$ then $x+a \in A'+A'$ for at most $\frac{|A'|-1}{2}$ elements $a \in A'$. Each of the sums $(x+a)$, for all $a \in A'$, is unique as x is fixed. Thus, $\left(A'+x\right) \setminus \left(A'+A'\right) \geq \frac{1}{2}\left(|A'|+1\right)$. We conclude that the greedy process would not halt before A' is of size at least $\sqrt{|A|}$.

Finally, note that by definition of the greedy process we have

$$|A'+A'| \geq \frac{1}{2}\left(1+2+\ldots+|A'|\right) \geq \frac{1}{4}|A'|^2.$$

□

Lemma 12 (Adapted from [Alo13, CNR+24])**.** *Let G be a finite abelian group of order N and let S be a uniformly random subset of G. For any integer $s \geq 64\log^2 N$, the probability that S contains a sumset $A+A$ for a set A of size s is at most $2^{-s/8}$.*

Proof. For any fixed subset A' of size $\sqrt{s}$ that has many sums the probability that $A'+A'$ is contained in S is at most $2^{-s/4}$. Therefore, by a union bound, the probability that there is such an A' is at most

$$\binom{N}{\sqrt{s}} 2^{-s/4} \leq 2^{-s/8},$$

where here we used the fact that $s \geq 64\log^2 N$. By the previous lemma, if there is no such A' then there can be no A of size s so that $A+A \subset S$, completing the proof. □

Proof of Theorem 9. Let S be a uniform random subset of G. By the last lemma, the probability that there is a subset of size s whose sumset is in S is at most $2^{-s/8}$. Therefore, the number of such subsets is at most $2^{N-s/8}$. For any other choice of the set S, if S can be expressed as a union of sumsets, then each such sumset $A+A$ has $|A| < s$. The number of choices of such a sumset is at most $\sum_{i<s}\binom{N}{i} \leq (eN/s)^s$. The number of possible unions of k such sumsets is thus at most $(eN/s)^{sk} \leq e^{N/2}$, by the choice of k. This completes the proof. □

5.2 Parity Balance of Independent Sets in the Hypercube

Korshunov and Sapozhenko [KS83] proved the following tight bound on the number of independent sets in the hypercube. See also Galvin's exposition of their proof [Gal19].

Theorem 13 ([KS83])**.** $i(Q_n) \sim 2\sqrt{e} \cdot 2^{N/2}$.

In almost all such independent sets, all but very few of the vertices have the same Hamming weight parity. Let $\mathcal{E}$ denote the set of vertices of even weight of Q_n and let $\mathcal{O}$ denote the set of vertices with odd weight. As both $\mathcal{E}$ and $\mathcal{O}$ are independent sets in Q_n of size $N/2$, it follows that there are $2 \cdot 2^{N/2} - 1$ independent sets in which all vertices have the same parity. This turns out to not be too far off from the total number of independent sets. For a small k, assume that $|I \cap \mathcal{E}| = k$. There are $\binom{N/2}{k} \approx \frac{2^{k(n-1)}}{k!}$ choices for $|I \cap \mathcal{E}|$, each of those is an independent set. As Q_n is n-regular, there are at most nk vertices of $\mathcal{O}$ that neighbor any vertex of $I \cap \mathcal{E}$.[1] Hence, there are at least $N/2 - nk$ vertices in $\mathcal{O}$ among which we may choose any subset to complete the independent set I. Therefore, the number of independent sets I with $|I \cap \mathcal{E}| = k$ is at least $\binom{N/2}{k} \cdot 2^{N/2-nk} \approx \frac{2^{k(n-1)}}{k!} \cdot 2^{N/2-nk} = \frac{2^{-k}}{k!} \cdot 2^{N/2}$. By symmetry, we conclude that the number of independent sets in Q_n in which $\min\{|I \cap \mathcal{E}|,\ |I \cap \mathcal{O}|\} \leq k$ is at least

$$\left(1 - o\left(1\right)\right) \cdot 2 \cdot \sum_{i=0}^{k} \left(\frac{2^{-i}}{i!} \cdot 2^{N/2} \right) = 2 \cdot \left(\sqrt{e} - o_k\left(1\right)\right) \cdot 2^{N/2}.$$

As this simple lower bound is already matching the upper bound in Theorem 13, we conclude the following.

Corollary 14. *Let $k(n) = \omega_n(1)$ be any super-constant function in n, then the number of independent sets I in Q_n in which $\min\{|I \cap \mathcal{E}|,\ |I \cap \mathcal{O}|\} > k(n)$ is $o\left(2^{N/2}\right)$.*

Corollary 14 also follows formally from several works on the hardcore distributional model of the hypercube [Kah01, Gal11, JP20] which with parameter $\lambda = 1$ coincides with drawing a uniform independent set of Q_n. For example, Theorem 1.4 clause 2 in [Gal11] proves that the distribution of $\min\{|I \cap \mathcal{E}|,\ |I \cap \mathcal{O}|\}$ when I is a uniformly chosen independent set in Q_n, converges to a Poisson distribution with parameter $\frac{1}{2}$. In this section we derive a simple *tail bound* on the probability that both parities of a random independent set are of non-negligible size. We note that a stronger estimate follows from the results of Jenssen and Perkins in [JP20], and a nearly tight estimate of $2^{N/2-\Omega(N/\sqrt{n})}$ follows from the proof of Park in [Par22] (see equation (10) in her paper and the two lines preceding it). The bound below is weaker, but suffices (with room to spare) for our purpose here, and as its proof is simple and very different from the ones in the papers above we include it.

Theorem 15. *For any constant $\beta > 0$, there are at most $2^{N/2-\sqrt{N}/2^{\Theta(\sqrt{n})}}$ independent sets I in Q_n for which $\min\{|I \cap \mathcal{E}|,\ |I \cap \mathcal{O}|\} > \beta N$.*

The proof is based on the simple estimate in Corollary 14, together with the recursive structure of the hypercube. These are combined with a known result from the theory of VC-dimension, and the FKG Inequality, or its earlier versions due to Harris and Kleitman.

Definition 16. *We say that a subset $J \subseteq [n]$ of the coordinates $[n] = \{1, 2, \ldots, n\}$ is* shattered *by a family $\mathcal{F} \subseteq 2^{[n]}$, if for every $J' \subseteq J$ there exists $F \in \mathcal{F}$ for which $J' = J \cap F$.*

The following lemma is a variant of the Sauer-Perles-Shelah lemma [Sau72, She72], due to Pajor [Paj85], see also [ARS02].

Lemma 17. *Any family $\mathcal{F} \subseteq \mathbb{F}_2^n$ shatters at least $|\mathcal{F}|$ different subsets $J \subseteq [n]$.*

Definition 18. *We say that a family $\mathcal{J} \subseteq \mathcal{P}\left([n]\right)$ is a* down-closed family *if for any $J \in \mathcal{J}$ and any $J' \subseteq J$ we have that also $J' \in \mathcal{J}$.*

[1] We note that when k is small, this is also the typical number of such vertices.

The following is a special case of the Fortuin–Kasteleyn–Ginibre (FKG) inequality [FKG71], first proved by Harris and by Kleitman [Har60, Kle66].

Lemma 19 (FKG)**.** *Let $\mathcal{J},\mathcal{K} \subseteq \mathcal{P}\left([n]\right)$ be two down-closed families. Then, $\frac{1}{N}|\mathcal{J}\cap\mathcal{K}| \geq \frac{1}{N}|\mathcal{J}|\cdot\frac{1}{N}|\mathcal{K}|$.*

We also use the following simple application of the Chernoff–Hoeffding inequality [Hoe94].

Lemma 20. *Let $\mathcal{J} \subseteq \mathcal{P}\left([n]\right)$ be a family of size $|\mathcal{J}| > \gamma N$, then there exists $J \in \mathcal{J}$ of size $|J| > \frac{1}{2}n - \sqrt{\frac{1}{2}\ln\left(1/\gamma\right)}\cdot\sqrt{n}$.*

Proof. Let x be random vector drawn uniformly from $\mathbb{F}_2^n$. By the Chernoff–Hoeffding inequality, the probability that the Hamming weight of x is at most $n - x\sqrt{n}$ is smaller than e^{-2x^2}. In particular, there are less than γN different vectors in $\mathbb{F}_2^n$ of weight at most $\frac{1}{2}n - \sqrt{\frac{1}{2}\ln\left(1/\gamma\right)}\cdot\sqrt{n}$. The desired result follows by identifying each subset of $[n]$ with its corresponding characteristic vector. □

We are now ready to prove Theorem 15.

Proof of Theorem 15. Let I be an independent set in Q_n for which $\min\{|I\cap\mathcal{E}|,\ |I\cap\mathcal{O}|\} > \beta N$. For any vertex $x \in \mathbb{F}_2^n$ we define $S_x := \{i \in [n] \mid x_i = 1\} \subseteq [n]$ to be the set of coordinates i in which $x_i = 1$. Denote by $\mathcal{F}_\mathcal{E} := \{S_x \mid x \in I\cap\mathcal{E}\}$, and respectively $\mathcal{F}_\mathcal{O} := \{S_x \mid x \in I\cap\mathcal{O}\}$. By our assumption, we have $|\mathcal{F}_\mathcal{E}|, |\mathcal{F}_\mathcal{O}| > \beta N$. Denote by $\mathcal{J}_\mathcal{E} \subseteq \mathcal{P}\left([n]\right)$ (respectively, $\mathcal{J}_\mathcal{O}$) the family of all subsets $J \subseteq [n]$ of coordinates such that $\mathcal{F}_\mathcal{E}$ (respectively, $\mathcal{F}_\mathcal{O}$) shatters J. By Lemma 17, $|\mathcal{J}_\mathcal{E}|, |\mathcal{J}_\mathcal{O}| > \beta N$. We note that if a family shatters J then it also shatters any $J' \subseteq J$, and hence $\mathcal{J}_\mathcal{E}, \mathcal{J}_\mathcal{O}$ are down-closed families. Denote by $\mathcal{J} := \mathcal{J}_\mathcal{E}\cap\mathcal{J}_\mathcal{O}$ the family of all subsets of coordinates shattered by both $\mathcal{F}_\mathcal{E}$ and $\mathcal{F}_\mathcal{O}$, by Lemma 19 we thus have $|\mathcal{J}| \geq \beta^2 N$. Using Lemma 20, we conclude there exists a set of coordinates $J \in \mathcal{J}$ that is shattered by both $\mathcal{F}_\mathcal{E}$ and $\mathcal{F}_\mathcal{O}$, of size $|J| > \frac{1}{2}n - \sqrt{\ln\left(1/\beta\right)}\cdot\sqrt{n}$. Let $K \subseteq J$ be an arbitrary subset of J of size, say, $|K| = \lceil\frac{1}{2}n - 2\sqrt{\ln\left(1/\beta\right)}\cdot\sqrt{n}\rceil$. For any $K' \subseteq K$, denote by $I^{(K')}$ the subset of I containing only the vertices in which the coordinates of K exactly correspond to K', that is $I^{(K')} := \{x \in I \mid S_x \cap K = K'\}$. We observe that $I^{(K')}$ is an independent set in the $\left(n - |K|\right)$-th dimensional hypercube defined by the same restriction $\{x \in \mathbb{F}_2^n \mid S_x \cap K = K'\} \cong \mathbb{F}_2^{n-|K|}$. As $J \supset K$ is shattered by both $\mathcal{F}_\mathcal{E}$ and $\mathcal{F}_\mathcal{O}$, we conclude that there are at least $2^{|J|-|K|}$ vertices in $I^{(K')}$ of each parity. Indeed, for every J' satisfying $K \subset J' \subset J$ there is $F_1 \in \mathcal{F}_\mathcal{E}$ so that $F_1 \cap J = J'$, and similarly there is $F_2 \in \mathcal{F}_\mathcal{O}$ so that $F_2 \cap J = J'$. In particular, $I^{(K')}$ is an independent set in an $\left(n - |K|\right)$-th dimensional hypercube with at least $2^{|J|-|K|} = 2^{\Omega\left(\sqrt{\ln\left(1/\beta\right)}\cdot\sqrt{n}\right)} = \omega_n(1)$ vertices of each parity. By Corollary 14, for a large enough n there are at most $\frac{1}{2}\cdot 2^{\left(2^{n-|K|}\right)/2}$ possible such independent sets $I^{(K')}$. Note that this family of possible independent sets only depends on n, β, K, K' and not on anything else (such as I itself or even J). As $I = \bigcup_{K'\subseteq K} I^{(K')}$, we may repeat the argument above for every K' and conclude that there are at most $\left(\frac{1}{2}\cdot 2^{\left(2^{n-|K|}\right)/2}\right)^{2^{|K|}} = 2^{-2^{|K|}}\cdot 2^{N/2}$ possible such independent sets I, and that family of possible sets only depends on n, β, K. By the pigeonhole principle, at least a 2^{-n} fraction of the independent sets I in Q_n for which $\min\{|I\cap\mathcal{E}|,\ |I\cap\mathcal{O}|\} > \beta N$ will result in the same subset $K \subseteq [N]$ in the above argument. Thus, the number of such independent sets is no more than $2^n \cdot 2^{-2^{|K|}} \cdot 2^{N/2}$. We finally note that for any constant $\beta > 0$,

$$2^n \cdot 2^{-2^{|K|}} < 2^n \cdot 2^{-2^{n/2-\sqrt{\ln(1/\beta)}\cdot\sqrt{n}/2}} = 2^{-\sqrt{N}/2^{\Theta(\sqrt{n})}}.$$

□

As mentioned before Theorem 15, Park [Par22] proved recently that there are asymptotically $2^{\left(1-\Theta\left(1/\sqrt{n}\right)\right)N/2}$ independent sets I of Q_n in which $|I\cap\mathcal{E}| = |I\cap\mathcal{O}|$. Therefore, even when $\beta > 0$ is a constant, the bound in Theorem 15 cannot be improved to anything below $2^{N/2-N/\Theta\left(\sqrt{n}\right)}$, and is thus inherently of the form $2^{N/2-o(N)}$.

6 Upper Bound

In this section, we are finally ready to prove the upper bound on $|\mathcal{S}_n|$.

Theorem 21. $|\mathcal{S}_n \setminus \mathcal{H}_n| = 2^{-\sqrt{N}/2^{\Theta(\sqrt{n})}} \cdot |\mathcal{S}_n| = o\left(|\mathcal{S}_n|\right)$.

We first use Section 5.1 to prove the following lemma.

Definition 22. *Denote by* $\mathcal{H}'_n := \{S \subseteq \mathbb{F}_2^n \mid \exists\, v \in \mathbb{F}_2^n \,.\, \left(\mathbb{F}_2^n \setminus v^\perp\right) \subseteq S\}$ *the family of all subsets of* $\mathbb{F}_2^n$ *that contain the full complement of co-dimension* 1 *linear subspace.*

Lemma 23. $|\mathcal{S}_n \cap \mathcal{H}'_n| = 2^{N/2-\Omega(N)}$.

Proof. Let $S \in \mathcal{S}_n \cap \mathcal{H}'_n$. As $S \in \mathcal{S}_n$, there exists some $A \subseteq \mathbb{F}_2^n$ for which $S = A + A$. As $S \in \mathcal{H}'_n$, there exists some $v \in \mathbb{F}_2^n$ such that $\left(\mathbb{F}_2^n \setminus v^\perp\right) \subseteq S$. Denote by $A_0 = A\cap v^\perp$, and by $A_1 = A\setminus v^\perp$. We have $A = A_0 \cup A_1$ and moreover, $(A+A)\cap v^\perp = (A_0+A_0)\cup(A_1+A_1)$. We note that $\mathbb{F}_2^n \cap v^\perp \cong \mathbb{F}_2^{n-1}$ is isomorphic to a $(n-1)$-th dimensional hypercube. By Theorem 9 the number of possible unions of two sumsets in $\mathbb{F}_2^{n-1}$ is $2^{2^{n-1}-\Omega\left(2^{n-1}\right)}$. As S is fully described by v and by such a union, we conlcude that the number of possible sets S is at most

$$2^n \cdot 2^{2^{n-1}-\Omega(2^{n-1})} = 2^{N/2-\Omega(N)}.$$

□

We next use Section 5.2 to prove the following lemma.

Lemma 24. *Let* $v \in \mathbb{F}_2^n$ *be a vector of even Hamming weight. The number of independent sets in* $Q_n \cup \Gamma_{\mathbb{F}_2^n}(v)$ *is at most* $2^{N/2-\sqrt{N}/2^{\Theta(\sqrt{n})}}$.

Proof. By Theorem 15 there are at most $2^{N/2-\sqrt{N}/2^{\Theta(\sqrt{n})}}$ independent sets I in Q_n for which $\min\{|I\cap \mathcal{E}|,\ |I\cap\mathcal{O}|\} > \frac{1}{200}N$. On the other hand, the induced subgraph of $\Gamma_{\mathbb{F}_2^n}(v)$ on $\mathcal{E}$ (or $\mathcal{O}$) is a perfect matching, and hence it contains only $3^{N/4}$ independent sets. Hence, the number of independent sets I of $Q_n \cup \Gamma_{\mathbb{F}_2^n}(v)$ in which $\min\{|I\cap\mathcal{E}|,\ |I\cap\mathcal{O}|\} \le \frac{1}{200}N$ is at most

$$2 \cdot 3^{N/4} \cdot 2^{H(1/100)\cdot N/2+\Theta(n)} = O\left(1.95^{N/2}\right).$$

□

Remark 25. *A bound of the form* $2^{N/2-o(N)}$ *is the best possible in Lemma 24. Assume* n *is even and consider the even-weight edge* $v := \vec{1} = (1,1,\ldots,1)$. *The set of all even-weight vectors of weight* $< n/2$ *and all odd-weight vectors of weight* $> n/2$ *is of size* $\left(1-\Theta\left(\frac{1}{\sqrt{n}}\right)\right)\frac{N}{2}$ *and is an independent set in* $Q_n \cup \Gamma_{\mathbb{F}_2^n}(v)$ *(hence so is any subset of it).*

Proof of Theorem 21. Let $S \in \mathcal{S}_n \setminus \mathcal{H}_n$. If S^c is not of full linear rank, then $S \in \mathcal{H}'_n$. By Lemma 23 the number of such sumsets S is small enough and thus we may assume that there exists a linear basis $v_1, \ldots, v_n$ in S^c. We apply the linear transformation mapping this basis to the standard one, and then observe using Lemma 3 that the set A for which $S = A + A$ must be an independent set in Q_n (after the linear transformation, which we from now on work under the application of). For any even-weight edge v, we know from Lemma 24 that there are at most $2^{N/2-\sqrt{N}/2^{\Theta(\sqrt{n})}}$ sets A that are both an independent set of Q_n and also have $v \notin (A+A)$ (equivalently, A is also an independent set in $\Gamma_{\mathbb{F}_2^n}(v)$). By a union bound, there are only $2^{n-1} \cdot 2^{N/2-\sqrt{N}/2^{\Theta(\sqrt{n})}}$ sets A that are both an independent set in Q_n and also have any even-weight edge v not in $A + A$. If A is not of this form, then the linear subspace $\left(\vec{1}\right)^{\perp}$ of all even-weight vectors is contained in $A+A = S$, and thus $S \in \mathcal{H}_n$ which is a contradiction. We conclude that S can be described by a linear transformation, of which there are at most 2^{n^2}, and either a set in $\mathcal{H}'_n$ or one of the $2^{n-1} \cdot 2^{N/2-\sqrt{N}/2^{\Theta(\sqrt{n})}}$ sets described above. The overall number of possible sets S is thus $2^{N/2-\sqrt{N}/2^{\Theta(\sqrt{n})}}$. □

The lower bound in Theorem 7 and the upper bound in Theorem 21 together conclude the proof of Theorem 2. In combination with Theorem 6, we also deduce Theorem 1.

7 Expressing random sets as unions of sumsets

Theorem 9 provides a lower bound on the minimum number of sumsets required to express a random subset of a finite abelian group as their union. We can in fact improve the statement of this theorem and prove a nearly tight result, as we show in this section. Throughout the section we consider general finite abelian groups. The arguments can be easily extended to the non-abelian case, but to simplify the presentation we restrict the discussion to abelian groups. It is convenient to adopt here the convention that in a sumset $A + A$ we only include the sums of distinct elements of A.

Theorem 26. *Let G be an abelian group of order N and let S be a random subset chosen uniformly among all subsets of G. Then*

- *With high probability S cannot be expressed as a union of less than $\Omega(N/(\log^2 N \log\log N))$ sumsets.*
- *With high probability S is a union of at most $O(N \log\log N/(\log^2 N))$ sumsets.*

Any abelian group of order N provides a properly edge-colored complete graph K_N in which the vertices correspond to the group elements and every edge ab is colored by the sum $a + b$. A sumset $A + A$ is thus simply the set of all colors that appear in the induced subgraph on A. The results described here apply to general properly edge-colored graphs (which are far more general than the colorings corresponding to groups). The following result is proved in a recent paper of Conlon, Fox, Pham and Yepremyan. Since the paper has not appeared yet, we note that the result for the special case of Cayley (sum-) graphs of the abelian group $\mathbb{F}_2^n$ is proved already in the earlier work of Green [Gre05a] (which also contains a proof of a slightly stronger result for Cayley sum-graphs of cyclic groups).

Theorem 27 ([CFPY24])**.** *Let K be a properly edge-colored complete graph on N vertices, and let S be a random subset of the colors obtained by picking each color, randomly and independently, with probability $1/2$. Then with high probability, the maximum number of vertices of a clique in the graph consisting of all edges colored by colors in S is at most $O(\log N \log\log N)$.*

The first part of Theorem 26 follows quickly from the result above. Indeed, by this theorem, with high probability no sumset of a subset of size larger than $q = O(\log N \log\log N)$ is contained in a random subset S. There are at most $\sum_{i=0}^{q} \binom{N}{i} \leq N^q$ such subsets, so the number of ways to choose at most k of them is not larger than N^{kq}. If this number is smaller than, say, $2^{0.9N}$, then most subsets of the group cannot be expressed as a union of (at most) k sumsets, providing the assertion of the first part of theorem 26.

In order to prove the second part we also consider the more general setting of arbitrary proper edge colorings of complete graphs. Let K be a properly edge-colored complete graph on N vertices. Suppose further that each color appears $(1+o(1))N/2$ times. Let $\varepsilon > 0$ be a fixed small positive real, and put $k = (2-\varepsilon)\log_2 N$ (where as before we assume that N is sufficiently large as a function of ε and where we omit all floor or ceiling signs when these are not crucial). Call a subset A of vertices of K, $|A| = k$, a *rainbow clique* if all edges of it have distinct colors. Let R be a random subset of the set of colors obtained by picking each color, randomly and independently, with probability $1/2$. Call a rainbow clique A *available* if all the $\binom{k}{2}$ colors of its edges belong to R, and let $C(A)$ denote the set of these colors. For convenience call also every single edge e in R an available clique and denote its color by $C(e)$.

Theorem 28. *In the above notation, with high probability, there is a collection $\mathcal{C}$ of at most $N\log\log N/(\log^2 N)$ available cliques so that $R = \cup_{A\in\mathcal{C}} C(A)$.*

Proof. The main part of the proof is a second moment argument that shows that with high probability the total number of available rainbow cliques is close to its expectation and that with high probability almost every edge whose color lies in R belongs to roughly the same number of such available cliques. The required collection $\mathcal{C}$ can then be chosen greedily among the available rainbow cliques by repeatedly adding such a clique that covers a maximum number of yet uncovered colors, together with a smaller additional number of edges. We proceed with the details.

The second moment argument resembles the one used in proof of the typical behaviour of the maximum clique in the random graph $G(n, 1/2)$ as described, for example, in [AS16] section 4.5. However, the dependence between edges of the same color leads to several complications and requires some additional arguments.

Note, first, that almost every set of k vertices spans a rainbow clique in K. Indeed the fraction of k-cliques that contain two edges of the same color is at most $O(k^4/N) = N^{-1+o(1)}$. For each rainbow k-clique A, let X_A denote the indicator random variable with value 1 iff A is available, that is, all the colors of its edges lie in R. Let $X = \sum X_A$ be the number of available k cliques. Since any rainbow k-clique is available with probability $2^{-\binom{k}{2}}$, by linearity of expectation, the expected number of available rainbow k-cliques is

$$\mu = (1-o(1))\binom{N}{k}2^{-\binom{k}{2}} > \left(\frac{N2^{-(k-1)/2}}{k}\right)^k > N^{\varepsilon k/5} > N^{100}.$$

Let Var denote the variance of X. This is the sum of the variances of the indicator variables X_A (which is smaller than μ) plus the sum over all ordered pairs A, A' of the covariances $Cov(X_A, X_{A'})$. If A and A' share no common color this covariance is 0. Otherwise, it is at most the probability that both cliques A, A' are available. For each j, $1 \leq j \leq k-1$, let Δ_j denote the sum of probabilities that $X_A = X_{A'} = 1$ where (A, A') range over all ordered pairs of rainbow k-cliques in which the largest forest in the set of all edges of A' that share colors with the edges in A contains j edges. By the discussion above Var $\leq \mu + \Delta$, where $\Delta = \sum_{j=1}^{k-1} \Delta_j$. The following upper bound for Δ_j (which can be improved) suffices for our purpose here.

Claim: For every $1 \leq j \leq k-1$

$$\Delta_j \leq \binom{N}{k}\binom{\binom{k}{2}}{j}\binom{k}{2}^j N^{k-j}\frac{1}{(k-2j)!}2^{-2\binom{k}{2}+\binom{j+1}{2}},$$

where for $j > k/2$ the $(k-2j)!$ term should be replaced by 1.

Proof of claim: There are at most $\binom{N}{k}$ ways to choose the first rainbow clique. Then there are $\binom{\binom{k}{2}}{j}$ ways to choose j colors that appear in it. There are less than $\binom{k}{2}^j$ ways to place these colored edges as a forest on k labelled vertices. Once this forest is chosen, we can select the second clique by selecting its first vertex in each connected component of the forest (including the ones of size 1). This corresponds to the N^{k-j} factor. When $2j < k$ at least $k-2j$ of these first vertices are in components of size 1, hence we can divide by $(k-2j)!$. The crucial point here is that since the edge coloring is proper, this information suffices to reconstruct all vertices of the second rainbow clique. Finally, the total number of common colors between the two cliques is at most $\binom{j+1}{2}$ (obtained only when the forest is a tree and the set of edges of the second clique that share colors with the first forms a clique on $j+1$ vertices). Even in this case, the probability that all required colors are in R is at most $2^{-2\binom{k}{2}+\binom{j+1}{2}}$. This completes the proof of the claim. □

Returning to the proof of the theorem we next show that the variance Var is much smaller than μ^2. Indeed,

$$\frac{\Delta_j}{\mu^2} \leq \left(\frac{k^6 2^{(j+1)/2}}{N}\right)^j.$$

For small j, say $j \leq \log\log N$, this ratio is less than $N^{-(1+o(1))j}$. For $\log\log N < j \leq (2-\varepsilon)\log_2 N$ this ratio is less than $N^{-\varepsilon j/3}$ which is much smaller than, say, N^{-100}. This shows that the variance Var $\leq \mu + \Delta$ is at most $\mu^2/N^{1-o(1)}$. Therefore, by Chebyshev's Inequality, with high probability the total number of available rainbow k-cliques is $(1+o(1))\mu$ where μ is the expectation of this quantity (which is much larger than N^{100}).

We next show that for every fixed edge of K, if the random set of colors R contains its color, then with high probability it lies in roughly the expected number of available rainbow k-cliques that contain it. The computation here is similar to the previous one, and here too the variance is at most the square of the expectation divided by $N^{1-o(1)}$. As the computation is very similar, we omit it. By Chebyshev's Inequality and Markov'v Inequality, with high probability every color that appears in R besides $N^{o(1)}$ of them appears in roughly the same number of available rainbow k-cliques, which is a $(1+o(1))(k^2/N)$ fraction of the total number of such cliques. We can now choose greedily $3N\log\log N/k^2$ available rainbow k cliques one by one, where in each step we choose such a clique that covers the maximum number of yet uncovered colors in R. As in each step we cover at least a $(1+o(1))k^2/N$ fraction of the remaining colors, this will cover all colors of R besides $o(N/\log^2 N)$ of them. These can be covered by edges, completing the proof and implying also the assertion of the second part of Theorem 26. □

8 Concluding Remarks and Open Problems

Natural extensions of the main question studied in this paper include studying the cardinality of the family of sumsets in $\mathbb{F}_p^n$ for $p > 2$, and studying the cardinality of the family of higher-orders of iterated sums (e.g., sets of the form $A+A+A$ for some $A \subseteq \mathbb{F}_2^n$, or generally kA for any $k > 2$).

The lower bound in Theorem 26 can be improved to $\Omega(N/\log^2 N)$ for some groups, like the cyclic group Z_N. Indeed, Green showed in [Gre05a] that the largest A so that the sumset $A+A$

lies in a random subset of Z_N is, with high probability, of size $O(\log N)$. This, together with the counting argument described in the proof of the first part of theorem 26 here, supplies the improved bound. It seems plausible that the $\log \log N$ factor can be removed in both the upper and the lower bounds in Theorem 26 for every abelian group of order N. A somewhat related known result of Frieze and Reed ([FR95]), that holds in the simpler case of the usual random graph $G = G(N, 1/2)$, is that the minimum number of cliques required to cover all edges of G is, with high probability, $\Theta(N^2/\log^2 N)$.

It is worthwhile to add that much fewer sumsets suffice to express a random subset of $\mathbb{F}_2^n$ as their intersection. Indeed, for a random subset S (containing 0), define, for every $1 \leq i \leq n$, $A_i = S \cap (\mathbb{F}_2^n \setminus e_i^{\perp}) \cup \{0\}$, where e_i is the vector of Hamming weight 1 with its unique 1 coordinate in the i-th place. Then $(A_i + A_i) \cap (\mathbb{F}_2^n \setminus e_i\perp) = (S \cap (\mathbb{F}_2^n \setminus e_i\perp))$, and with high probability (that is, with probability tending to 1 as n tends to infinity) $(A_i + A_i) \cap e_i^{\perp} = e_i^{\perp}$ for every i. It is easy to check that S is the intersection of these n sumsets $A_i + A_i$.

Acknowledgments: We thank Matthew Jenssen for helpful comments and references, and thank an anonymous referee for helpful remarks.

References

[ABMS14] Noga Alon, József Balogh, Robert Morris, and Wojciech Samotij. A refinement of the Cameron–Erdős conjecture. *Proceedings of the London Mathematical Society*, 108(1):44–72, 2014.

[AGU10] Noga Alon, Andrew Granville, and Adrián Ubis. The number of sumsets in a finite field. *Bull. London Math. Soc*, 42(5):784–794, 2010.

[Alo13] Noga Alon. The chromatic number of random Cayley graphs. *European Journal of Combinatorics*, 34(8):1232–1243, 2013. Special Issue in memory of Yahya Ould Hamidoune.

[AO95] Noga Alon and Alon Orlitsky. Repeated communication and Ramsey graphs. *IEEE Transactions on Information Theory*, 41:1276–1289, 1995.

[AS16] Noga Alon and Joel H Spencer. *The Probabilistic Method*, Fourth Edition. John Wiley & Sons, 2016.

[ARS02] Richard P Anstee, Lajos Rónyai, and Attila Sali. Shattering news. *Graphs and Combinatorics*, 18:59–73, 2002.

[Bib13] Khodakhast Bibak. Additive combinatorics: With a view towards computer science and cryptography—an exposition. In Jonathan M. Borwein, Igor Shparlinski, and Wadim Zudilin, editors, *Number Theory and Related Fields*, pages 99–128, New York, NY, 2013. Springer New York.

[Bou90] Jean Bourgain. On arithmetic progressions in sums of sets of integers. in: *A tribute to Paul Erdős*, pages 105–109, 1990.

[BTW07] Boaz Barak, Luca Trevisan, and Avi Wigderson. A mini-course on additive combinatorics, 2007. Available at: https://www.math.cmu.edu/ af1p/Teaching/AdditiveCombinatorics/allnotes.pdf.

[Cam18] Marcelo Soares Campos. On the number of sets with a given doubling constant. *arXiv preprint arXiv:1811.05793*, 2018.

[CNR+24] Xi Chen, Shivam Nadimpalli, Tim Randolph, Rocco A Servedio, and Or Zamir. Testing sumsets is hard. *arXiv preprint arXiv:2401.07242*, 2024.

[CFPY24] David Conlon, Jacob Fox, Huy Tuan Pham, and Liana Yepremyan. In preparation. 2024.

[FKG71] Cees M Fortuin, Pieter W Kasteleyn, and Jean Ginibre. Correlation inequalities on some partially ordered sets. *Communications in Mathematical Physics*, 22:89–103, 1971.

[Fre73] G.A. Freiman. *Foundations of a Structural Theory of Set Addition*. Translations of mathematical monographs. American Mathematical Society, 1973.

[FR95] Alan Frieze and Bruce Reed. Covering the edges of a random graph by cliques. *Combinatorica*, 15(4): 489–497, 1995.

[Gal11] David Galvin. A threshold phenomenon for random independent sets in the discrete hypercube. *Combinatorics, Probability and Computing*, 20(1):27–51, 2011.

[Gal19] David Galvin. Independent sets in the discrete hypercube, 2019.

[GGMT23] William Timothy Gowers, Ben Green, Freddie Manners, and Terence Tao. On a conjecture of Marton. *arXiv preprint arXiv:2311.05762*, 2023.

[GM16] Ben Green and Robert Morris. Counting sets with small sumset and applications. *Combinatorica*, 36:129–159, 2016.

[GR04] Ben Green and Imre Ruzsa. Counting sumsets and sum-free sets modulo a prime. *Studia Scientiarum Mathematicarum Hungarica*, 41(3):285–293, 2004.

[Gre05a] Ben Green. Counting sets with small sumset, and the clique number of random Cayley graphs. *Combinatorica*, 25:307–326, 2005.

[Gre05b] Ben J. Green. Finite field models in additive combinatorics. In Bridget S. Webb, editor, *Surveys in combinatorics*, pages 1–27. Cambridge Univ. Press, 2005.

[Har60] Theodore E Harris. A lower bound for the critical probability in a certain percolation process. In *Mathematical Proceedings of the Cambridge Philosophical Society*, volume 56, pages 13–20. Cambridge University Press, 1960.

[Hoe94] Wassily Hoeffding. Probability inequalities for sums of bounded random variables. *The collected works of Wassily Hoeffding*, pages 409–426, 1994.

[JP20] Matthew Jenssen and Will Perkins. Independent sets in the hypercube revisited. *Journal of the London Mathematical Society*, 102(2):645–669, 2020.

[Kah01] Jeff Kahn. An entropy approach to the hard-core model on bipartite graphs. *Combinatorics, Probability and Computing*, 10(3):219–237, 2001.

[Kle66] Daniel J Kleitman. Families of non-disjoint subsets. *Journal of Combinatorial Theory*, 1(1):153–155, 1966.

[KM23] Zander Kelley and Raghu Meka. Strong bounds for 3-progressions. *arXiv preprint arXiv:2302.05537*, 2023.

[KS83] A. Korshunovand and A. Sapozhenko. The number of binary codes with distance 2. *Problemy Kibernet*, 40:111–130, 1983.

[Lov17] Shachar Lovett. *Additive Combinatorics and its Applications in Theoretical Computer Science*. Number 8 in Graduate Surveys. Theory of Computing Library, 2017.

[Paj85] Alain Pajor. Sous-espaces ℓ_1^n des espaces de Banach. *Hermann, Paris, Collection Travaux en cours*, 1985.

[Par22] Jinyoung Park. Note on the number of balanced independent sets in the Hamming cube. *The Electronic Journal of Combinatorics*, Paper no. 2-34, 9 pp., 2022.

[San11] Tom Sanders. Green's sumset problem at density one half. *Acta Arithmetica*, 146(1):91–101, 2011.

[San12] Tom Sanders. On the Bogolyubov–Ruzsa lemma. *Analysis & PDE*, 5(3):627–655, 2012.

[Sar15] Vahe Gnelovich Sargsyan. Counting sumsets and differences in an abelian group. *Journal of Applied and Industrial Mathematics*, 9:275–282, 2015.

[Sau72] Norbert Sauer. On the density of families of sets. *Journal of Combinatorial Theory, Series A*, 13(1):145–147, 1972.

[She72] Saharon Shelah. A combinatorial problem; stability and order for models and theories in infinitary languages. *Pacific Journal of Mathematics*, 41(1):247–261, 1972.

[SS20] Aleksandr A Sapozhenko and Vahe G Sargsyan. The number of sumsets in abelian group. *Discrete Mathematics and Applications*, 30(5):339–345, 2020.

[Tre09] Luca Trevisan. Additive combinatorics and theoretical computer science. *ACM SIGACT News*, 40(2):50–66, 2009.

[TV06] Terence Tao and Van H Vu. *Additive Combinatorics*, volume 105. Cambridge University Press, 2006.

[Vio11] Emanuele Viola. *Selected Results in Additive Combinatorics: An Exposition*. Number 3 in Graduate Surveys. Theory of Computing Library, 2011.